HARVARD | BUSINESS | SCHOOL

Online Word of Mouth and Product Review Disagreement

Frank Nagle
Christoph Riedl

Working Paper

13-091

May 5, 2015

Online Word of Mouth and Product Review Disagreement*

Frank Nagle† and Christoph Riedl‡§¶‖

This draft: April 22, 2015

Abstract

Studies of online word of mouth have frequently posited—but never systematically conceptualized and explored—that the level of disagreement between existing product reviews can impact the volume and the valence of future reviews. In this study we develop a theoretical framework of disagreement in online WOM and test our predictions in a dataset of nearly 300,000 online reviews for 425 movies over three years. This framework highlights that rather than thinking of disagreement as dispersion of opinions around a mean, high levels of disagreement can be better conceptualized as opposing opinion poles. Such a conceptualization has important implications for how disagreement can be measured and how results can be interpreted. We theoretically develop, validate, and apply a novel statistical measure of disagreement that can be used alongside existing alternative approaches such as standard deviation. We find that only high levels of disagreement—with opposing opinion poles—influence future reviews while simple dispersion does not. We show that high levels of disagreement among previously posted reviews lead to more future product reviews, a relationship that is moderated by informational content such that higher informational content amplifies the effect. Further, we show that increased disagreement leads to future reviews of lower valence. Our findings highlight that an important role for research on big data in information systems is to examine how existing measurement approaches and interpretations can be improved by fully leveraging the richness that digital trace data offers.

Keywords: Online word of mouth, online communities, consumer behavior, online product reviews, viral marketing.

*Acknowledgments. We wish to thank Xiang Ao, Megan Bernard, Ivo Blohm, Anil Doshi, Gary King, Michal Kolesar, Michael Luca, Andrew Marder, Jennifer Pan, Sarah Reynolds, Molly Roberts, Bill Simpson, and Mike Toffel. We are also grateful to the comments made by seminar participants at AOM, HBS, Harvard, MIT, WIN, and WISE. CR acknowledges support from the German Research Foundation under grant code RI 2185/1-1. All mistakes remain the authors' own.

†Harvard Business School, Wyss 100, Soldiers Field, Boston MA 02163; fnagle@hbs.edu

‡D'Amore-McKim School of Business & College of Computer and Information Science, Northeastern University, 360 Huntington Ave, Boston MA 02115; c.riedl@neu.edu

§Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge MA 02138

¶Authors listed alphabetically

‖Corresponding author

1 Introduction

With the advent of the digital age, consumer word of mouth (WOM), which had traditionally consisted of individuals proffering their opinions of a product to other individuals in personal interactions, has shifted to online mediums where individuals now make their opinions known for the world to see. In addition to shifting the intended audience, digitization of WOM has also drastically increased the amount of digital trace data that can be used by businesses and researchers to better understand consumer opinions. While these data provide unprecedented access to consumer opinions and thus important opportunities for business intelligence as well as potential insights into human behavior and communication patterns, such data often contain complex relationships (Chen et al., 2012; George et al., 2014). Analysis of such data can benefit from finer categorization and improved measurement of the observed phenomena, which can then help resolve contradictions and enable progress (Shmueli and Koppius, 2011). The relationships that exist in online WOM are particularly complex since WOM is dynamic and prior online reviews are likely to affect the accumulation of future reviews. Specifically, disagreement between existing product reviews has frequently been posited as a major influencing factor and many studies of online WOM include some measure intended to capture the heterogeneity of consumer opinions (e.g., Dellarocas and Narayan, 2006; Zhang and Dellarocas, 2006; Moe and Schweidel, 2012; Sun, 2012). However, disagreement can be conceptualized in various ways, as extensive research on public deliberation and political science have demonstrated (Klofstad et al., 2013): is disagreement simply a matter of diverging opinions or diametrically opposing opinions? None of the prior studies on online WOM precisely defines what disagreement is, which aspects of disagreement a given measure of disagreement actually captures, or how subsequent results should be interpreted. Without insightful theory grounded in aspects of human behavior to guide the interpretation of this digital trace data, any measurement of disagreement may provide incorrect conclusions. Therefore, this paper investigates the role of disagreement in online WOM.

There is a growing body of literature on online word of mouth that spans diverse fields of research including management, marketing, and information systems. Research in this area has found that increased WOM leads to increased sales (Clemons et al., 2006; Liu, 2006; Duan et al., 2008; Luca, 2011; Gopinath et al., 2013), although this effect can sometimes be negative (Dewan and Ramaprasad, 2014), and when positive has been found to be short-lived and fade over time (Moe and Trusov, 2011). Further, WOM has been found to lead to information cascades that influence buyer adoption behavior (Duan et al., 2009) and can lead to better predictions of product success in the market (Dellarocas et al., 2007). Regarding review valence, higher average ratings have been shown to lead to higher sales (Zhang and Dellarocas, 2006; Chevalier and Mayzlin, 2006; Luca, 2011; Sun, 2012), although product and consumer characteristics have been found to moderate these effects (Zhu and Zhang, 2010). These findings led to an interest in the dynamics of WOM itself as studies explored what affected the valence of reviews (Moon et al., 2010; Godes and Silva, 2011; Muchnik et al., 2013), the propensity to post a review (Dellarocas and Narayan, 2006; Dellarocas et al., 2010), or both (Moe and Schweidel, 2012).[1] One important aspect of social influence that has been theorized—but never systematically conceptualized and explored—is the heterogeneity of consumer opinions, i.e., the level of disagreement among consumer reviews (Dellarocas et al., 2010; Moe and Schweidel, 2012; Sun, 2012). Existing research in this area has either measured consumer disagreement via a proxy, such as professional reviews (Dellarocas et al., 2010), decomposed it into numerical categories (Moe and Trusov, 2011), or has found mixed results for the impact of disagreement on propensity to review and valence of reviews (Moe and Schweidel, 2012). It is not difficult to imagine that measures that are conceptually distinct and tap into different dimensions of disagreement lead to varied interpretations of observed behavior. Without proper theoretical conceptualization and diverging measures, results are hard to interpret. By bringing such conceptual and measurement differences to the forefront, we aim to contribute a theoretical understanding of disagreement in online WOM.

The goal of this paper is to develop a better understanding of disagreement in online WOM. To do this, we first develop a theory of disagreement in online WOM. Then, we hypothesize how disagreement in prior reviews may shape the accumulation of future WOM. Specifically, we develop hypotheses related to the impact of disagreement on the volume of future reviews and the valence

[1] Some recent studies have started exploring additional avenues such as the effect of WOM on product returns (Sahoo et al., 2013), how to employ WOM to infer product types (Hong et al., 2013), and what characteristics of online reviews are most helpful to others (Yin et al., 2014). This literature has also expanded beyond the management world and is a subject of inquiry in the technical literature on machine learning and text mining (e.g., Zhang et al., 2012). In particular, there is also a large stream of business-related text mining research that aims to extract additional information such as helpfulness, readability, or market structure from online WOM and thus goes beyond the use of discrete numeric ratings provided by consumers (e.g., Netzer et al., 2012; Ghose et al., 2012). Our research focuses on numeric ratings and we do not extract valence from textual reviews.

of the posted reviews. Building on theory of informational content and persuasive arguments, we develop a set of hypotheses to investigate how the informational content of online WOM can alter the perception of disagreement and thus moderate the propensity to write a review. Before presenting an empirical analysis of online WOM about movies, we propose and systematically evaluate a novel measure to capture disagreement among online product reviews that can be used alongside existing alternative measures but which offers interesting characteristics and facilitates a finer grained interpretation of the results. Specifically, we demonstrate how different, but equivalent, levels of disagreement are captured in a single statistic and thus facilitate insightful interpretation of social influence in online WOM. For estimation, we use a 5-week panel of movie reviews from Yahoo Movies for 425 movies (almost 300,000 reviews). To properly account for the panel structure and autocorrelation in our data set, we apply generalized method of moments (GMM)-based dynamic panel data estimators in our main analyses.

We find that prior disagreement leads to an increased propensity to post a review, which is counter to some prior findings that disagreement has no effect at the aggregate population level (Moe and Schweidel, 2012). However, we find that it is not heterogeneity in consumer opinions *per se* that drives these effects, but rather it is the existence of strong opposing opinions that drives the results. General disagreement has little to no effect, but extremely high levels of disagreement—instances in which opinions are clearly bi-polar—do have an effect, a nuanced finding which would not be possible using traditional dispersion-based measures, such as standard deviation, as they cannot discern between uni-polar and bi-polar distributions. Further, we find that the positive overall effect of disagreement is amplified by higher informational content in prior reviews: either because the reviews themselves are longer or because of more outside information due to higher availability of some products in the marketplace. With respect to the impact of disagreement on the valence of posted reviews, we find that higher disagreement among prior reviews leads to lower product ratings, which holds important implications for those who wish to encourage online WOM by stimulating disagreement.

Our work makes three primary contributions to the information systems literature on data analytics and online word of mouth. First, we take on the important task of conceptualizing disagreement in online WOM. By bringing conceptual and measurement differences to the forefront, we hope to add a theoretical understanding of disagreement in online WOM to this literature. We hope to show potential avenues in which data driven research can advance our understanding of human behavior by making explicit use of novel data sources. This methodological advance allows future researchers to better understand the importance of disagreement as they explore what drives the posting and valence of online consumer reviews. Second, we show results of the effects of disagreement in online WOM using two different measures that allow us to provide a nuanced interpretation of our empirical findings. We are able to tie our results to high levels of disagreement, levels at which opinions become opposing rather than merely diverging. Our finding that the impact of disagreement is amplified by the average length of prior reviews and is increased by the product's availability in the market, further contributes to a more nuanced understanding of social influence within online WOM. Third, we contribute to the emerging field of big data analytics by demonstrating the importance of theoretical models for guiding measurement and interpretation. We show how advances in the analysis of large datasets, combined with theory, can provided opportunities for finer categorization of complex relationships in human communication. We argue that an important role for research on big data analytics in general, and online WOM in particular, is to examine how naïve measurement approaches and interpretations can be improved by fully leveraging the richness that digital trace data offers. Thus, we provide what we hope will become a blueprint for data analytics measure development and evaluation.

The remainder of this paper is structured as follows. The next section develops a theory of disagreement in online WOM and proposes hypotheses regarding its effect on the volume and valence of future reviews. Section 3 lays out our dataset and empirical strategy as well as a proposed alternative measure of disagreement in online WOM. Section 4 presents our analysis and results. Finally, Section 5 discusses our findings and concludes the paper.

2 Theory and Hypothesis Development

In this section, we review relevant theory on disagreement and extend it to the field of online WOM. We develop a definition of disagreement that is suitable to the context of online WOM and subsequently characterize conceptually different levels of disagreement. Specifically, we develop the separate notions of diverging and opposing opinions. From this discussion, we are able to deduce several desirable characteristics that a measure suitable for capturing disagreement in online WOM should have. We return to these desirable characteristics in a later section of the paper when we discuss an alternative measure of disagreement which builds on the notion of capturing opposing opinion poles. Finally, we develop four testable hypotheses regarding expected effects that disagreement is likely to have on the accumulation of future reviews and the valence of those reviews.

To study disagreement, we must first define what we mean by disagreement. In the context of online WOM, we define disagreement as an interaction among consumers who hold divergent viewpoints and perspectives regarding a product or service. This definition is similar to an accepted definition of disagreement used in the political science literature (Huckfeldt et al., 2004). These viewpoints are often dimensionless and are not explicitly labeled as referring to "product quality" although they often have this connotation. Amazon.com, for example, provides no categorization of their star reviews (i.e., they are not explicitly tied to quality) and instead provides anchor points where the lowest rating is labeled as "I hate it" and the highest as "I love it." Even hotel reviews that often explicitly mention specific review categories like "location," "service," or "cleanliness" typically include an aggregated, dimensionless "overall review" category.

Many studies of online WOM include some measure intended to capture the distribution of consumers' opinions as a measure of disagreement (e.g., Zhang and Dellarocas, 2006; Dellarocas and Narayan, 2006; Sun, 2012; Moe and Schweidel, 2012).[2] However, none of these studies precisely defines what disagreement is and which aspects of disagreement a given measure of disagreement actually captures. While measures of opinion distributions have often been included in prior studies of online WOM, defining the underlying concept of disagreement has not been a main thrust of those studies and consequently disagreement remains under-theorized in the WOM literature. Further, many of these prior studies have pointed out challenges with precisely capturing relevant aspects of disagreement given a set of consumer reviews (Dellarocas and Narayan, 2006; Hu et al., 2009). Consequently, different conceptualizations and different measures could very well explain some of the inconsistent findings about the effect of disagreement observed in the WOM literature. By

[2]Throughout this work we equate "opinion," "individual," and "review," implicitly assuming that any individual can have only one opinion about a given product or service, and expresses that opinion through at most one written product review (i.e., consumers do not post multiple reviews). This is consistent with the approach employed by the vast majority of product review sites, including the site we use in our empirical analysis below.

bringing such conceptual and measurement differences to the forefront, we hope to add a theoretical understanding of disagreement in online WOM to this literature and then make further progress in our understanding of the role that disagreement plays in online WOM.

To start, it is important to clarify who is disagreeing with whom. The view taken in studies of online WOM—although rarely expressed explicitly—is that of disagreement of previously posted reviews *amongst each other* (Zhang and Dellarocas, 2006; Dellarocas et al., 2010; Sun, 2012; Moe and Schweidel, 2012). This "disagreement within the crowd" is then observed by an ego, whose behavior and opinion is potentially influenced by having observed disagreement among previously stated opinions. We call this *crowd-centric disagreement*. This view does not explicitly model (or measure) what ego's own opinion is, compared to the opinions expressed by the crowd. The communications literature takes a different viewpoint and focuses on disagreement when an ego's own opinion diverges from that of another individual or group of individuals (Klofstad et al., 2013). We call such disagreement *ego-centric disagreement*. Further, the communications literature often focuses on disagreement when the question is based on fact rather than opinion, and there is a correct answer (e.g., Asch, 1951). Consistent with prior work in WOM, in this study we focus on the crowd-centric view of disagreement of opinions where there is no correct answer.

How can we then conceptualize disagreement within a crowd? At question is both the presence and degree of disagreement: what constitutes disagreement in a crowd, and how do we quantify the amount—or level—of disagreement? If we conceptualize an outsider observing opinions expressed within a crowd, we could classify the observed distribution of opinions as falling between two possible endpoints: complete agreement or complete disagreement. Thus, we speak of a spectrum of disagreement. In the case of complete agreement, all opinions are identical and we observe a complete absence of disagreement. As disagreement grows, opinions increasingly diverge from each other. While some consumers "love it," other may love it a little bit less. As disagreement increases further, more and more opinions move to become polar opposites and eventually each pole becomes more pronounced. In the case of extreme disagreement, opinions do not just diverge from each other but rather oppose each other and are completely polarized. We move to a case of opposing "I love it" vs. "I hate it" reviews. We provide a conceptual illustration of such a spectrum of disagreement in Figure 1. The figure is overly simplified and used for illustrative purposes only.

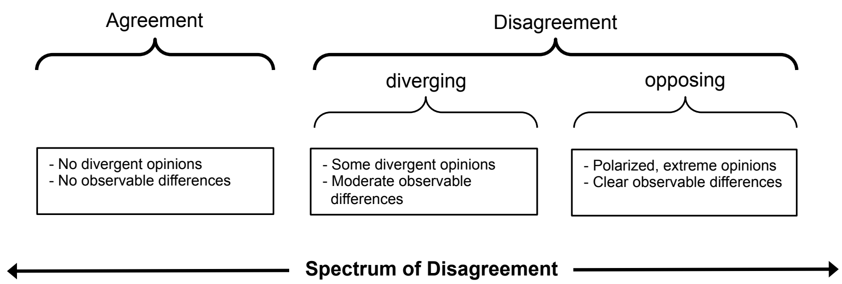

Figure 1: Conceptual illustration of a spectrum of disagreement in online WOM.

Following this conceptualization of disagreement in online WOM, the basic theoretical questions are: at what level does the observation of disagreement become obvious, and at what level does disagreement start to affect consumer behavior (if at all) and in what way?

In the context of online WOM, extreme levels of disagreement are very likely to occur. Prior research has theorized that due to reporting bias (Anderson, 1998; Hu et al., 2009) disagreement can be very high as consumers with extreme opinions over-report while those with less extreme opinions under-report. Furthermore, it has been shown that purchasing bias—an effect by which consumers who are more likely to like a product are also more likely to buy it—shifts the rating distribution to be, on average, slightly positive (Chevalier and Mayzlin, 2006; Hu et al., 2009). This has been empirically observed in numerous online WOM settings in categories such as movies (Dellarocas and Narayan, 2006; Dellarocas et al., 2010), books (Chevalier and Mayzlin, 2006; Sun, 2012), and music (Hu et al., 2009). Further, in the empirical dataset of online movie reviews used in this study, described in more detail in the next section, we observe that 63% of movie-week pairs exhibit a bi-modal rating distribution pattern. The fact that extreme levels of disagreement are prevalent in online WOM underlines the importance of precisely formulating the concept of disagreement in online WOM and measuring it appropriately.

Because online product reviews are public, the valence of stated opinions and their distribution are salient to consumers engaging in online WOM and consequently disagreement is salient, especially in these more extreme forms. In offline settings, disagreement has been shown to be an important social influencer with regard to opinion and behavior (McPhee et al., 1963; Klofstad et al., 2013). Therefore, it is likely that disagreement may influence a consumer's decision to post a review or not (thus affecting volume of reviews that get posted) as well as the valence of any posted review (thus affecting mean valence as well as level of disagreement). We consider the direction of these effects, as well as possible moderators, in the following sections.

2.1 Disagreement and Propensity to Review

Before theorizing any specific effects that disagreement might have on a consumer's behavior when writing reviews, we must consider the overall incentives for writing a review. The goal of a review writer is to offer their opinion of a good or service that they have consumed in an effort to inform the purchase decisions of future consumers. Such reviews are useful for any type of good, but are particularly useful for experience goods as the quality of experience goods is generally more difficult to observe in advance (Nelson, 1970). However, writing a review is costly in terms of both time and cognitive effort. Therefore, motivation-based theory indicates that the review writer must be incentivized to incur those costs (Wu and Huberman, 2008; Godes and Silva, 2011). Early research on offline WOM identified self-enhancement and other-involvement (to help others, altruism) as motivating factors to engage in WOM (Dichter, 1966). Especially in the context of experience goods, consumers wish to help others make better purchasing decisions. Dichter (1966) finds a number of motivating factors that compel consumers to engage in WOM. These include the need to share their positive consumption experiences through WOM communication in an effort to enhance their image among others by projecting themselves as intelligent shoppers, the desire to help others in making purchasing decisions (helping others to make a satisfying purchase decision or helping others to prevent negative experiences), and the belief that the impact of their review will be high. This is consistent with the general idea that people are concerned about the impact of their contribution as they derive not only purely altruistic benefits from their contribution but also private benefits such as moral satisfaction, joy of contribution, or self image (Zhang and Zhu, 2011).

How then is this consumer behavior affected by observing disagreement in previously posted reviews? Studies of contribution behavior in general find that social effects encourage contribution (Zhang and Zhu, 2011). If consumers engaging in writing costly reviews care about the impact of their review, we would expect them to contribute more if they consider their contributed opinion to be more valuable to others—if their opinion is less redundant (Ling et al., 2005). In cases of strong agreement, consumer opinions are very similar and adding more of the same has relatively little impact. In cases of strong disagreement, however, the presumed impact of an additional contribution is higher as the contributed opinion is less redundant. Consequently, we would expect consumers' propensity to review to be higher in cases of dissenting prior reviews. Controlling for a given volume of prior reviews, an arbitrary consumer opinion is less redundant and adds higher informational value if those prior reviews are more diverse.

In addition to the general motivational effect affecting contribution propensity, there is likely also a more direct effect of disagreement. Work on disagreement in political science by Huckfeldt et al. (2004) demonstrated that ambivalence leads to lower rates of reporting and participation. Ambivalence is highest when most existing opinions are in agreement: when either the positive component or the negative component is very high. The perception that the crowd has already decided increases people's feeling of ambivalence which leads to lower participation (Jackson, 1983). Conversely, disagreement polarizes consumers' opinions which leads to more over-reporting since these opinions are more extreme. Furthermore, disagreement is known to evoke high levels of arousal, which has been found to drive sharing of content and opinions (Berger and Milkman, 2012). Accordingly, we expect to see a higher likelihood of consumers to contributing to online WOM if previously posted reviews have been more disparate in their ratings, reflecting a higher level of disagreement about a product by the consuming population. We formalize this as follows:

> HYPOTHESIS 1. *A higher level of disagreement amongst previously posted reviews about a product leads to a higher propensity to review the product online post-consumption.*

2.2 Perception of Disagreement

A great deal of research has been devoted to studying how individuals and groups process relevant information and the effects of this information processing on group polarization (e.g., Isenberg, 1986; El-Shinnawy and Vinze, 1998; Sia et al., 2002). Theory on informational influence, and related persuasive argument theory, has found effects not only of shared information itself but also that the amount of information and the persuasiveness of the arguments affect perceived levels of disagreement (Hinsz and Davis, 1984). Consequently, it is likely that the information contained in the textual reviews qualitatively modifies the perception of disagreement among the numeric reviews. If the information contained in the textual reviews is more persuasive leading to a higher level of perceived disagreement, we would expect the effect of disagreement on propensity to review to be higher. Consequently, informational content can qualitatively modify, and thus moderate, the effects of disagreement. Here we investigate two mechanisms through which informational content may vary. First, the information contained in online WOM itself may vary with longer and shorter reviews. Second, the information available about a product outside of WOM may vary, thus giving more or less informational value to the information contained within WOM.

2.2.1 Disagreement and Length of Reviews

As writing a review is a costly endeavor (Wu and Huberman, 2008), it follows that the cost incurred by the writer increases as the review length increases. In addition to the altruistic motivation discussed above, passion has also been shown to be an important motivator in online settings (Wang et al., 2008). Combined with altruism, passion can compel reviewers to incur the cost of writing longer reviews. Therefore, longer reviews indicate contributors who are likely more passionate about their feelings towards their experience of the product and likely more persuasive than shorter reviews. Furthermore, because longer reviews can contain more informational content, longer reviews and messages have been shown to be more persuasive in both offline (Wright, 1980) and online (Zhang et al., 2010) settings. Therefore, the passionate feelings of the review writer are likely to increase the sense of disagreement to a review reader such that the same level of disagreement in the numeric ratings is perceived more strongly if the average length of the textual reviews is longer.

In addition to the influence of passion, higher informational content in longer reviews can have another effect. When a person with a particular opinion is presented with additional information that is divergent, they tend to become more confident about their own opinion (Kelly, 2008), inducing them to be more likely to post their opinion. Hence, as length is a characteristic of the review that is highly salient to readers (Chevalier and Mayzlin, 2006), we expect longer reviews to moderate the effect of disagreement on future reviewers. We formalize this as follows:

Hypothesis 2. *The positive effect of disagreement on the propensity to review a product online post-consumption is moderated by the average length of previously posted reviews* *such that longer reviews increase the effect of disagreement.*

2.2.2 Disagreement and Product Availability

The previous hypothesis put forward a case by which the informational content of previously posted reviews qualitatively modifies the perception of prior disagreement, thus strengthening its effect. A product's availability in the market place, and consequently consumers' awareness of that product, can have a similar effect on the perception of WOM. Due to larger advertising budgets, consumers are generally more aware of products with large availability in the market place, compared to those that have lower availability in the market place. Since the producers of these widely available products are sponsoring advertising about the product, advertising will be focused on creating a positive perception of the product in an attempt to increase sales (Assmus et al., 1984). Although any advertising for products that are less available in the market will also be positive, there will be much less advertising for these products. Therefore, compared to less available products, widely available products will have a larger amount of positive information (Nelson, 1970) about them in the marketplace before the good is released and consumers have a chance to experience the good and offer review information of their own.

When a product is released, if the consumer reception is generally positive, and therefore in agreement with the pre-release information available through advertising, then consumers feel they were not misled by the advertising (Olson and Dover, 1978; Mizuno and Odagiri, 1990). However, if the consumer reception of the product, as made salient through online WOM, is more mixed, meaning there is more disagreement about feelings toward the product, there is likely to be a different effect. The consumer response is no longer in sync with the pre-release positive information, and consumers experience cognitive dissonance (Festinger, 1957). This cognitive dissonance will lead consumers to seek to reduce the dissonance by altering or adding contributions to the group discourse (Festinger, 1957). This process of cognitive dissonance reduction can lead to overconfidence, which can manifest itself in an assertion of the individual's opinion to the group in an attempt to sway the group's opinion toward that of the individual (Blanton et al., 2001). Since this dissonance only occurs when there is a large degree of positive information available in the market prior to the release of a product, it is likely that consumers of products that are more available in the market, and hence have been exposed to more positive information than consumers of less available products, will experience larger dissonance if the online WOM shows disagreement within consumer experiences. Therefore, the baseline effect of disagreement discussed above is likely to be larger for movies that are more broadly available in the market than those that are not. Our formal hypothesis is as follows:

> HYPOTHESIS 3. *The positive effect of disagreement on the propensity to review a product online post-consumption is moderated by the product's market availability such that higher market availability increases the effect of disagreement.*

2.3 Disagreement and Valence of Reviews

The above sections theorize effects that prior disagreement might have on a population's propensity to review. Specifically, we expect that increased disagreement leads to more reviews, moderated by review length and product availability. However, will disagreement also affect the valence of reviews and, if so, in what direction? If disagreement leads consumers to simply be more likely to express their opinions, then we would not expect valence to change. If, however, disagreement leads some groups of reviewers (say the majority opinion holders) to express their opinions more insistently (more extremely) or participate more than other groups, then we would expect a shift in valence as a consequence.[3] Existing research has shown that prior reviews can indeed influence the valence of future reviews (Godes and Silva, 2011; Moe and Trusov, 2011). Assuming a shift towards the increased reporting of the majority opinion, for a product with high average rating this would imply proportionally more positive reviews in the future. For a product with low ratings, this would imply proportionally more negative ratings in the future. Given that, due to purchasing biases (Hu et al., 2009) the majority opinion usually is positive for most products, the valence of future reviews is likely to go up.

However, an alternative effect is also likely. A robust body of literature in communication and political science has demonstrated strong effects of group and belief polarization—the tendency of people to become more extreme in their thinking following group discussion (Isenberg, 1986).

[3] Given that we cannot observe people's true opinion before they express it in a review, a proportional shift in propensity to report and changes in opinion are observationally equivalent and cannot be separated.

Following this tendency of opinions to become increasingly polarized we expect two effects: within-group divergence on either side of the mean decreases, while across-group dissent becomes larger as each opinion pole becomes increasingly polarized (Kelly, 2008). Especially in computer-mediated and anonymous online settings, group polarization is increased due to facilitated generation of more novel arguments and one-upmanship (Sia et al., 2002; El-Shinnawy and Vinze, 1998). Consequently, initial moderate levels of disagreement can lead to increasingly higher levels of disagreement due to polarization. Given a bounded opinion spectrum, as in most online product review settings, we argue that the negative opinion pole is proportionally more strongly affected by this polarization simply because there is more "room" for the opinion pole to shift downward (remember that due to purchasing bias the average review is positive-leaning). As a consequence, mean valence will decrease as disagreement increases due to group polarization.[4] We argue that polarization dominates the dynamic of social influence and consequently the valence of future reviews will decrease with increasing levels of disagreement. Formally:

> Hypothesis 4. *A higher level of disagreement amongst previously posted reviews about a product leads to a lower valence of future online product reviews about that product.*

3 Data and Empirical Strategy

Our study uses a new dataset we collected from multiple sources to identify the importance of disagreement to both the likelihood to post a review, as well as the valence of reviews that are posted about movies. The following sections construct a new measure for disagreement, outline our variables and present our data and empirical strategy.

3.1 Measuring Disagreement

The natural question that follows from our theoretical discussion of disagreement is how can we accurately describe and capture the existence, and gauge the level, of disagreement in online WOM

[4] An alternative presentation of this argument relies more heavily on mathematical actualities. In a bounded opinion spectrum, mean valence and disagreement are technically linked. High levels of disagreement imply polarized opinions of both the very high and very low valence and consequently mean valence around the center-point of the opinion spectrum. Since this center-point is (on average) lower than typical mean valence (which is positive-leaning due to purchasing bias), mean valence of future reviews will go down as disagreement becomes increasingly polarized.

in a single statistic. An obvious candidate is standard deviation. Indeed, standard deviation has been used as a measure of disagreement in many studies of online WOM (e.g., Dellarocas and Narayan, 2006; Sun, 2012; Moe and Schweidel, 2012). Standard deviation measures the amount of variation or dispersion from the average. A low standard deviation indicates that the data are close to the mean while a high standard deviation indicates that the data are spread out further from the mean. However, standard deviation does not characterize how many data points are closely centered around the mean. Thus, standard deviation may be more appropriately conceived of as measuring the *absence* of disagreement rather than precisely distinguishing between the more extreme levels of disagreement. The concept underlying this measurement approach may lack accuracy in distinguishing between moderate and high levels of disagreement and may therefore not always accurately explain outcomes, especially when those outcomes vary only with high levels of disagreement. In the framework from the previous section, standard deviation is well-suited for measuring diverging disagreement, but may be less precise for measuring opposing disagreement. As a consequence, we argue that measuring disagreement through standard deviation alone potentially overlooks the effects of the more extreme, but very common, levels of disagreement found in online WOM. In empirical settings with low or only moderate levels of disagreement, standard deviation can quite accurately describe differences in the level of disagreement. However, if disagreement is extreme, standard deviation less accurately describes differences, as it is not well suited to distinguish between medium and high levels of disagreement.

From our conceptualization of disagreement in Section 2, we can derive desirable characteristics of a measure to accurately capture the full spectrum of disagreement, including precise characterizations of extreme levels of disagreement comprised of opposing opinions. Such a measure allows us to capture the distribution of opinions such that we can derive the presence of poles if they are present (disagreement is higher if there are two poles), how far the poles are apart from each other (disagreement is higher if the poles are further apart), how defined each pole is—how far data are spread out *within* each pole (disagreement is higher if each pole is more clearly defined), as well as the relative importance of each pole—the proportion of opinions comprised in each pole (disagreement is higher if poles are equally important; disagreement is lower the more lopsided the mixing becomes).

From here it is not a stretch to imagine that conceptually distinct measures that tap into different aspects of disagreement, specifically their ability to accurately capture opposing levels of disagreement, could hold differing implications for behavioral outcomes. In this section, we construct a measure that can capture the full spectrum of disagreement in a single statistic. We acknowledge that ours is but one possible way to construct such a statistic. However, we believe that this measure can serve as an alternative to standard deviation that is particularly useful when levels of disagreement are high. We argue that comparing these two measures of disagreement can provide important insights into behavioral consequences of disagreement in online WOM. Our subsequent empirical analysis focuses on the extent to which these two measures provide us with similar or divergent pictures of how disagreement influences consumer behavior in online WOM. Our goal is not to prescribe either measure as "better," rather to argue that both measures can provide meaningful insights in different research contexts. Further, we seek to demonstrate that clear conceptual understanding and measurement choices hold important consequences for the study of disagreement in online WOM.

Based on the desirable characteristics of a measure of disagreement laid out above, the empirical rating distribution r of a given product can be expressed as a mixture of two normal distributions, one representing the reviews of positive valence and one representing reviews of negative valence:

$$r = \pi\mathcal{N}(\mu_1, \sigma_1) + (1 - \pi)\mathcal{N}(\mu_2, \sigma_2), \tag{1}$$

where π is the mixing proportion of the two distributions (i.e., the proportion of consumers who liked the product and consumers who disliked the product); μ_1 and μ_2 are the means of the two distributions and σ_1 and σ_2 are the respective standard deviations. We can use the various variables that define this mixed distribution to create an index of disagreement. To construct this index, we build on work by Wang et al. (2009) who proposed such a measure for the analysis of gene expressions in cancer research. Here, we generalize their approach by relaxing the assumption of equal variance between the groups and present an adaptation of their approach developed for separating gene expressions in biomedical research to the study of disagreement in online social media. The exposition of the approach closely follows the original work. We define δ, the standardized distance between the two distributions as

$$\delta = \frac{|\mu_1 - \mu_2|}{\frac{\sigma_1 + \sigma_2}{2}}. \tag{2}$$

To demonstrate how the shape of the density of a polar distribution changes as π and δ vary, we show a set of synthetic opinion distributions, holding μ_1 constant at 0 and $\sigma_1 = \sigma_2 = .5$ (thus, δ effectively becomes a function of μ_2). Figure 2 presents the density plots in the (π, δ) plane. Because of symmetry in π from 0.0 to 1.0, we only illustrate the plots using π from $0.5 - 1.0$. The plots indicate that the ability to index significant levels of polarity among post-purchase reviews depends on (a) μ_1 and μ_2: the distance between the two means of the two opinion poles; (b) σ_1 and σ_2: how pronounced each pole is; (c) π: the mixing proportion of the two distributions. This corresponds to the desirable characteristics of a disagreement measure laid out above. Medium levels of disagreement can be characterized when polarity is no longer visually distinguishable when μ_1 and μ_2 reach critical values and the poles become inseparable, either when the poles are overlapping due to large standard deviations or when one pole is very small compared to the other due to lopsided mixing proportion. The plots also suggest a pattern by which high levels of disagreement with opposing opinion poles can be distinguished. Using a curve in the (π, δ) coordinate system we can distinguish distributions such that distributions with a clear polar pattern appear above the curve, while those that are harder to discern as exhibiting a polar pattern appear below the curve. In Figure 2 this curve is indicated by the plots in pink. The degree to which poles are opposing can be made objective by reference to a standard sample size computation as shown by Wang et al. (2009), by defining *Disagreement Index*, DI, as a function of the standardized distance, δ, and the ratio of the mixing proportion π as

$$DI = \delta\sqrt{\pi(1 - \pi)}. \tag{3}$$

In practice, we can attempt to estimate π and δ for a given set of consumer reviews and use the estimated values to compute DI. Combinations of π and δ that give the same values of DI describe rating distributions that are equally separable as belonging to a bi-polar distribution. Constant DI values in Equation 3 define curves in the (π, δ) plane. The curves with a constant disagreement index take on their minimum value at $\pi = 0.5$ (i.e., when the sizes of the two subgroups are identical),

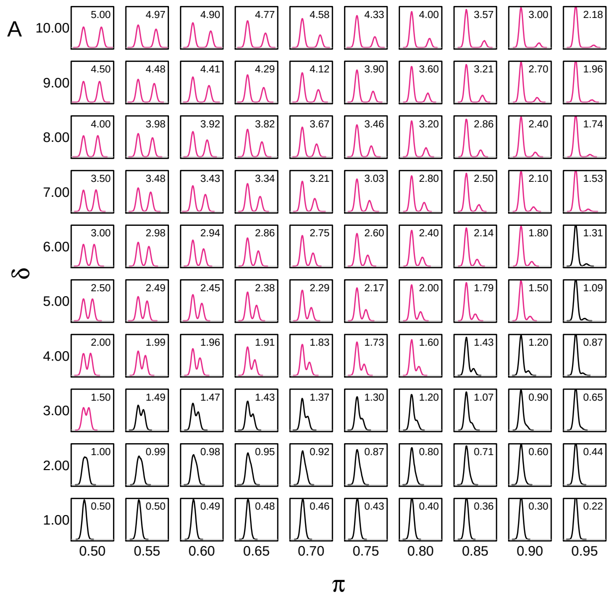

Figure 2: Relationships between polarity, mixing proportion π, and distribution means μ_1 and μ_2. Density plots of simulated rating data as π and δ vary, with resulting value of DI inset in top right corner (μ_1 held constant at 0; $\sigma_1 = \sigma_2 = .5$). These plots indicate that high consumer polarity is evident when mixing proportion π is close to a $50 : 50$ split and when the means of the poles μ_1 and μ_2 are sufficiently separated from each other. Polarity in consumer reviews is harder to discern when the means are closer together or when the mixing proportion is very uneven. The density plots colored pink correspond roughly to those distributions that are "visually" distinguishable as belonging to two distinct consumer populations by a polarity measure of $DI \geq 1.5$. A similar graph is used by Wang et al. (2009) to illustrate separability of gene expressions.

which results in the most power to distinguish a bi-polar pattern for a given total sample size. When the group sizes are very unequal, for example, when π is close to 0.9 (or 0.1), separation of the two distributions becomes harder given identical sample size. In other words, as the proportion of reviews in one of the distributions (π) increases, the standardized distance between the two distributions (δ) must also increase to arrive at the same value of DI. Overall, these analyses using synthetic data demonstrate that equally polar distributions lead to equal DI values, irrespective of the exact combination of μ_1, μ_2, σ_1, σ_2, and π values. Thus, the disagreement index is able to index the full spectrum of disagreement—ranging from the absence of disagreement to high levels of disagreement that result in opposing poles of consumer opinions—in a single continuous measure.

The remaining key issue is the practical estimation of δ and π (and thus DI). To do this, we employ parameterized finite mixture modeling methods using expectation-maximization (EM; McLachlan and Peel, 2000) to estimate δ and π for a set of product reviews and then use these estimated values to compute DI. Mixture models have been shown to be useful for classification when a single class of data (all product reviews) is constructed of multiple subclasses (Witten et al., 2011), and are increasingly being used in information systems research (Bapna et al., 2011). In our setting studying online product reviews, these subclasses represent reviews from consumers who enjoyed the product and consumers who did not enjoy the product. Furthermore, since mixture models are a probability based clustering method, they have the benefit that they do not classify data into disjunct categories but rather assign proportions of class membership (Witten et al., 2011).

3.2 Simulation Studies

To evaluate the performance of the disagreement index for capturing significant levels of polarity in consumer opinions expressed through online product reviews, we performed simulation studies. We used the **mixtools** package (Benaglia et al., 2009) for the R language and environment for statistical computing (R Development Core Team, 2012) to perform mixture model-based classification of consumer reviews belonging to consumers who liked the product and those who did not like the product. Based on the classification we can then obtain the statistical parameters necessary to compute DI for online product reviews for each product. Since we apply the mixture modeling to simulated data, the true underlying distributions from which the data are drawn is known and we can compute an error measure to evaluate the accuracy and precision of our approach. Specifically, we compute the Mean Squared Error as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(DI_{obs} - DI_{sim})^2. \tag{4}$$

The error will depend on the number of observations available for classification and how well the two distributions can be separated. We would expect higher errors for instances with a lower number of available ratings to be classified and for distributions that are harder to separate, i.e., for rating distributions with lower levels of disagreement. However, situations with low separation between the two poles result in small values of DI. We simulate data for different values of δ ranging from 1 to 10 in steps of 1. A value of $\delta \geq 8$ corresponds to high disagreement, $\delta \approx 6$ corresponds to medium disagreement, while $\delta \leq 4$ corresponds to weak or no disagreement. For simplicity, we set $\sigma_1 = \sigma_2 = 1$, and $\mu_1 = 0$ in which case δ becomes a function of μ_2 which we vary from 0.5 to 5 in steps of 0.5, which yields the desired range from $1 - 10$ for δ. Furthermore, we simulate distributions for different mixing proportions π, ranging from 0.5 to 0.95 in steps of 0.05. We then generate datasets for all combinations of δ and π for six different sample sizes $n = 50 - 300$ (in steps of 50). Finally, we repeat each simulation 500 times to achieve more precise error estimates. In sum, we computed error measures for 6 $\textit{sample sizes} \times 10\ \textit{different}\ \delta \times 10\ \textit{different}\ \pi \times 500\ \textit{repetitions} = 300,000$

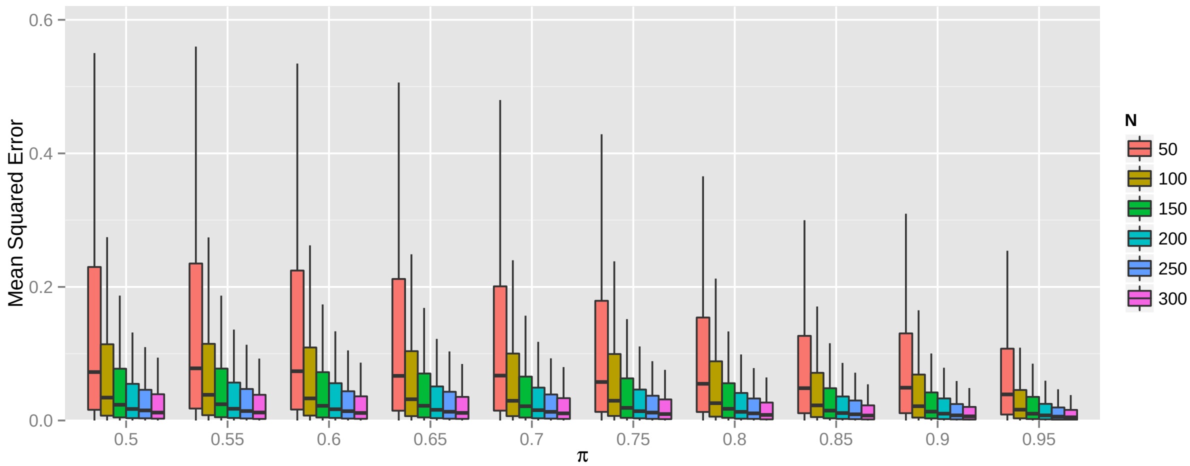

Figure 3: Boxplot showing MSE between estimated and observed DI based on $300,000$ simulated datasets. Error rates are not significantly different between different sample sizes, and combinations δ and π.

distributions. For each of the $300,000$ datasets we apply EM based mixture modeling using the **mixtools** package to estimate the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$, and π from which we compute δ and ultimately DI.

Figure 3 shows box-plots of the resulting MSE of DI grouped by sample size, and π. The simulation study results indicate that, even with small sample sizes ($n = 50$), our approach performs well with an average MSE of 0.152. Increasing the sample size to $n = 300$ improves the MSE to 0.028 (across all combinations of π and δ). Regarding variation in the mixing proportion π, in cases of very balanced mixing proportions of $\pi = 0.5$ we find a MSE of 0.078. In instances with more significantly unbalanced mixing proportions of $\pi = 0.95$ we find an MSE as low as 0.043 (across all sample sizes). Overall, error rates are low (0.063) and are not significantly different between sample sizes or combinations of δ and π (rejecting the test for different means with $p < .001$). The simulation study shows that our measure of disagreement based on the *Disagreement Index* performs well in various settings of disagreement, and is robust across various combinations of mixing proportions π, δ, and sample sizes. In summary, the *Disagreement Index* provides a single statistic that characterizes the full spectrum of disagreement and maps similar levels of disagreement, resulting from a plethora of different combinations of π and δ found in real-world online product reviews, onto similar numerical values.

3.3 Variables

3.3.1 Dependent Variables

To investigate the effect of disagreement on the accumulation of future reviews, we use the log-transformed volume of reviews a movie i received in week t as our dependent variable. We control for the logged box office revenue of the same movie in the same week (effectively the number of tickets sold for that movie). Given the natural interpretation of weekly cycles and to be consistent with prior work on the movie industry, this data is aggregated at the weekly level.

To investigate the effect of disagreement on the valence of posted reviews, our second dependent variable, $STARS_{j,i} \in 1, 2, 3, 4, 5$, is the rating assigned by reviewer j to movie i. Five stars being the best rating, one star being the worst. We perform this analysis of valence at the individual level, rather than the aggregate weekly level, due to the more granular observation level. Specifically, our dataset contains the exact time when individual reviews were posted, but box office revenue is only available in weekly aggregates. Given this exact timing data, we can reconstruct exactly what reviews were on the review site just before a consumer posts a review, allowing us to perform the valence estimation at the individual consumer level.

3.3.2 Variables of Interest

The primary variable of interest for predicting propensity to post a review is cumulative disagreement of previously posted reviews, $DI_{i,1:t-1}$. To calculate this measure for movie i in week t, we consider all reviews posted prior to week t for movie i to obtain a measure of disagreement, as outlined in Section 3.1.[5] For comparative purposes, we also consider the standard deviation of all reviews posted in prior weeks about a specific movie, $SD_{i,1:t-1}$. When using disagreement to predict review valence, we alter the disagreement variable to be $DI_{i,1:j-1}$ to represent the cumulative level of disagreement about movie i before reviewer j posts her review.

We measure the average length of prior WOM by averaging the word count per review over all previously posted reviews for a given movie i prior to week t ($AVGLEN_{i,1:t-1}$). Similarly, when predicting review valence, we consider the average length of WOM about movie j posted prior

[5]Since we are approximating discrete rating distribution through mixtures of normal distributions, we apply a continuity correction of $\frac{1}{2}$ to all ratings (Cox, 1970). All results reported are independent of this continuity correction and do not change substantively when not applying the correction.

Variable	Specification
Dependent Variables	
$VOL_{j,t}$	Volume of reviews posted for movie j during week t of its theatrical release
$STARS_{i,j}$	The star-rating posted by reviewer i for movie j
Independent and Control Variables	
$BOX_{j,t}$	Box office revenues of movie i during week t of its theatrical release (in million USD)
$DI_{j,1:t-1}$	Disagreement between all reviews posted for movie j prior to week t ($DI_{j,1:i-1}$ for valence regression)
$SD_{j,1:t-1}$	Standard deviation of all reviews posted for movie j prior to week t
$AVGLEN_{j,1:t-1}$	Average length of reviews posted for movie j prior to week t
$THEATERS_{j,t}$	Count of theaters where movie j is screened during week t (in thousands)
$AVG_{i,1:j}$	Average valence (star-rating) of reviews posted for movie j prior to week t (or prior to review i)
$TIME_{i,j}$	Time (in hours) from when movie j was released until review i was posted
$ORDER_{i,j}$	A counter for which number review for movie j review i is (e.g. the first review has an $ORDER$ of 1, the second is 2, etc.)
$INDIE\ MOVIE_j$	Dummy coded variable indicating if movie j was released in below-average number of theaters in its opening week
$GOOGLE\ TRENDS_{j,t}$	The volume of Google searches for movie j in week t
$Calendar\ Controls_{j,t}$	Fixed calendar effects controlling for year-week

Table 1: Summary of week-level and individual-level variables.

to the review of person j ($AVGLEN_{i,1:j-1}$). We consider one additional explanatory variable to measure the impact of product effects when interacted with disagreement. Specifically, we measure a movie's market availability in its opening week by the number of theaters in which a given movie i is shown. We convert this into a time-invariant indicator variable $INDIE\ MOVIE_i$ which is 1 for a movie that was shown in a below average number of theaters in its opening week and 0 if it was shown in an above average number of theaters in its opening week. We interpret this as a measure of blockbuster or independent movie status which serves as a proxy for the availability of outside information about a movie. We show a summary of our key variables in Table 1.

3.4 Data Collection and Summary Statistics

Our dataset includes all movies released nationwide in the US (wide release) between 2007-2009. We collected weekly box office results (BOX), number of screens ($THEATERS$) from Box Office Mojo.[6] We collected review ratings and text from Yahoo! Movies.[7] Box Office Mojo and Yahoo!

[6] http://www.boxofficemojo.com/
[7] http://www.movies.yahoo.com/

Movies are commonly used sources for obtaining such movie and review characteristics (e.g. Moon et al., 2010). For the time frame covered in our analysis, Yahoo! Movies was the dominant website for movie reviews. Similarly to Dellarocas et al. (2010) we collected all of these measures for the first five weeks the movie was in theaters and dropped any movies for which the Yahoo! Movies website was not available or no box-office revenue data was available, indicating that the movie was in theaters for less than five weeks. We collected data for product search volume from Google Trends.[8] Our final dataset includes three years: 2007, 2008, and 2009, with a total of 425 movies and 298,007 reviews within the first five weeks after the original release.[9] On average, movies received 140 reviews per week. Movies that were widely available (one standard deviation above mean availability) received an average of 406 reviews per week while moves that were less widely available received only 17 reviews. Table 2 and Table 3 show descriptive statistics and correlation patterns of the data. Figure 4 shows raw data of movie-week rating distributions using smoothing splines. The polarized pattern of many movie-week rating distributions is clearly visible with major peaks at 5-star and 1-star ratings. In fact, reviews in 63% of movie-week pairs exhibit a bi-modal rating distribution pattern in our dataset. The importance of this fact is further emphasized in the correlations shown in Table 3.The correlation between $DI_{i,1:j-1}$ and $SD_{i,1:j-1}$, not including week one when no prior reviews exist, is $\rho = 0.63$ ($p < .001$) indicating that the two measures do indeed capture substantially different aspects of disagreement.

[8] http://www.google.com/trends/

[9] We also collected data for the years 2010 and 2011 but exclude them from the analysis as Yahoo! Movies had massively lost popularity by that time: While the average number of reviews per movie posted within five weeks after release was 840 in 2007, it was only 215 in 2011. Consequently, this latter data might be biased as consumers still writing reviews on Yahoo! Movies might be systematically different from those who left the community.

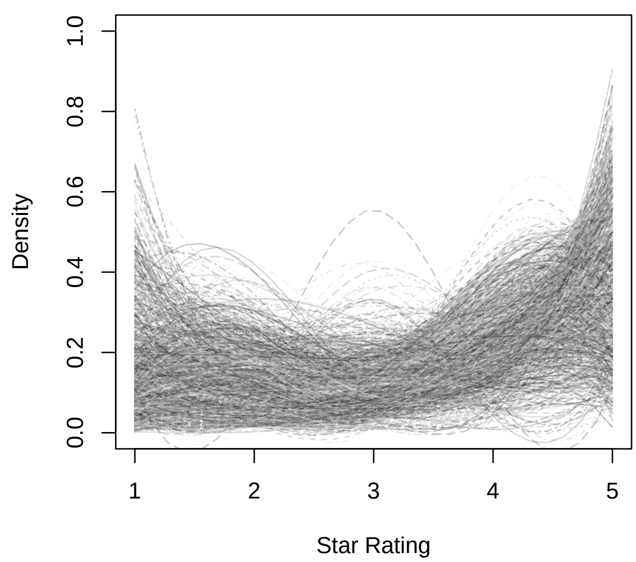

Figure 4: Distribution of raw rating data on movie-week-level illustrated through smoothing splines.

	2007	2008	2009	All Years
Movies	147	137	141	425
Total Reviews	134,211	102,754	61,042	298,007
Mean Reviews per Week	183	150	87	140
Mean Reviews per Week for Wide Availability Movies	536	478	245	406
Mean Reviews per Week for Limited Availability Movies	18	16	17	17

Table 2: Summary of movies and reviews by year. Mean distributions for wide/limited availability movies are shown as one SD above/below average availability.

	Mean	SD	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VOL (1)	140.24	359.06	0.00	6389.00								
BOX (2)	11.50	18.90	0.00	238.62	0.74							
DIS (3)	1.76	0.45	0.00	6.94	0.08	-0.02						
SD (4)	1.26	0.28	0.00	1.85	-0.03	-0.12	0.63					
AVG (5)	3.66	0.67	1.42	5.00	0.06	0.17	-0.32	-0.74				
AVGLEN (6)	75.50	17.47	41.44	268.17	0.13	0.10	-0.02	0.02	-0.09			
THEATERS (7)	1.96	1.17	0.00	4.46	0.36	0.60	-0.05	-0.02	0.10	-0.04		
INDIE MOVIE (8)	0.38	0.49	0.00	1.00	-0.17	-0.29	0.04	-0.11	0.10	0.09	-0.58	
GOOGLE TRENDS (9)	0.49	1.21	0.00	16.00	0.46	0.50	0.00	-0.08	0.09	0.13	0.30	-0.17

Table 3: Summary statistics. Number of observations: 2,125 movie-week pairs.

3.5 Empirical Model

We model the impact of past product reviews on the accumulation of future reviews by directly incorporating past product review information in a linear equation while controlling for product sales. Our estimation equation is given by

$$log(VOL)_{j,t} = \theta log(VOL_{j,t-1}) + \gamma log(BOX_{j,t}) + \beta X_{j,1:t-1} + \delta Z_{j,t} + \lambda Z_j + d_j + \epsilon_{j,t} \quad (5)$$

where $VOL_{j,t}$ is the volume of reviews for movie j in week t, $BOX_{j,t}$ is the box-office revenue of movie j in week t (effectively, the number of people who saw movie j in week t), $X_{j,1:t-1}$ is the vector of cumulative review measures for reviews that have been posted for movie j prior to week t, Z_j and $Z_{j,t}$ are vectors of time-variant and time-invariant control variables, d_j is the product-specific fixed effect, and $\epsilon_{j,t}$ is product and time specific error term. To estimate our moderator hypotheses we estimate additional models that include interaction terms between review measures (H2) and time-invariant movie availability characteristics (H3).

We recognize that our study is based on observational data and ordinary least squares may overestimate the causal effect of prior WOM. Therefore, we follow established best practices from similar WOM research to rule out as many confounding factors as possible (e.g., Aral, 2011; Iyengar et al., 2011; Ghose et al., 2012; Wu and Brynjolfsson, 2013). We include temporal controls (calendar fixed effects) for each weekly period to hold constant cross-temporal variation, which could confound our results (Aral, 2011). We follow practices employed in similar work studying online WOM (e.g., Archak et al., 2011; Ghose et al., 2012) and estimate our model using a generalized method of moments (GMM) system estimator (Blundell and Bond, 1998). GMM estimators are designed for panels with few time periods and many individuals (products in our case), fixed effects, heteroskedasticity, and possible endogeneity due to autocorrelation. The system GMM estimator uses the original estimation equation to obtain a system of two equations: one in differences and one in levels. To account for time-series specific effects such as autocorrelation we include a one-period lag of the dependent variable in the model and use additional lags as instruments. Furthermore, to control for possible endogeneity in product popularity due to unobservable exogenous shocks that may stimulate reviewing behavior, we use product search volume (Google Trends) as additional instruments. For each movie, we retrieved the search volume from the Google Trends website using the title of the movie as the search term (in cases where the movie title consists of only a single word, we added the commonly used "the movie" to avoid overly ambiguous search terms; e.g., for the movie "300" we used "300 the movie" as search term). The use of search volume from Google Trends as a measure of product publicity acts as a suitable control for any unobserved factors driving both sales and word of mouth and is commonly used in this capacity (Archak et al., 2011; Ghose et al., 2012; Wu and Brynjolfsson, 2013). Our control for trends using Google search volume data should therefore alleviate most, if not all, endogenous popularity concerns.

For inference we use the robust covariance matrix proposed by Windmeijer (2005) which applies a finite-sample correction for the two-step covariance matrix and increases the efficiency of the GMM estimator. Further, we take into account possible problems with using too many lags as instruments (Roodman, 2009). We perform all estimations using the `pgmm` implementation of the **plm** package (Croissant and Millo, 2008) in R (R Development Core Team, 2012).

In a second set of analyses, we are interested in modeling the valence of future reviews. We estimate the valence of rating i that an individual consumer provides for movie j ($STARS_{i,j}$) as follows:

$$STARS_{i,j} = \beta X_{i,j} + \gamma Y_{i,j} + d_j + \epsilon_{i,j} \tag{6}$$

where $X_{i,j}$ is our key explanatory variable measuring disagreement (H4); $Y_{i,j}$ is a vector of individual-level controls; and d_j are movie-level fixed effects. We use the same approach as Godes and Silva (2011) and estimate ordered logistic regression with robust standard errors to account for potential heteroskedasticity and serial correlations in the error terms. To control for prior WOM and temporal dynamics, which vary for each review posted However, we include several control variables. Specifically, we control for sequential and temporal effects which have been shown to affect WOM (Godes and Silva, 2011) by controlling for the total number of hours elapsed since movie j was released before review i was posted ($TIME_{i,j}$), and the arrival order in which the review was posted ($ORDER_{i,j}$). Finally, we control for $AVG_{i,1:j-1}$, the average valence of all reviews posted prior to review i for movie j. For comparison and easier interpretation, we estimate OLS as well.

4 Analysis and Results

In this section, we discuss the estimation results for the accumulation of reviews and valence of reviews. We start by discussing results from the models estimating if disagreement leads to increased future reviews and then proceed to discuss results from the models estimating the effect of disagreement on the valence of future reviews.

4.1 Disagreement and Review Volume

Table 4 presents our main regression results. We can make several inferences from the regression coefficients. Note that the signs of the coefficients of the control variables are in accordance with what one would expect. The coefficient of the lagged dependent variable is significant and positive. The coefficient for the box-office revenue is significant and positive indicating that we can expect more reviews if the movie sold more tickets. Furthermore, we find no significant effect of the number of theaters that show a movie, whether the movie opened in below-average number of theaters, or the average rating valence of prior reviews. Finally, we find a significant and positive effect of the average length of prior reviews implying that review volume increases if prior reviews were longer. The Sargan (1958) test of the moment conditions does not indicate overidentification. We also report results for first-order and second-order serial correlation in the first-differenced residuals which are asymptotically distributed as $N(0,1)$ under the null of no serial correlation. We reject the null of no first-order serial correlation but do not reject the null of no second-order serial correlation.

Model 1 shows a positive and significant effect of prior disagreement, using the *Disagreement Index* measure developed above. This provides support for H1 that opposing disagreement in prior reviews predicts higher volumes of future reviews. In Model 2, we repeat the same analysis but substitute $SD_{j,1:t-1}$ for $DI_{j,1:t-1}$. This model finds no statistically significant effect of disagreement. This is consistent with our conceptualization of disagreement in which we argued that the effect is driven by consumers perception of polarization rather than simple divergence. Furthermore, the result is consistent with prior work that finds no significant effect of review divergence measured by standard deviation on its own (Moe and Schweidel, 2012). In summary, this indicates that the ability to detect significant effects of disagreement may well depend on how disagreement is conceptualized and measured.

Dependent variable:	$log(VOL_{j,t})$				
	(1)	(2)	(3)	(4)	(5)
$log(VOL_{j,t-1})$	0.425***	0.393***	0.421***	0.415***	0.415***
	(0.038)	(0.041)	(0.037)	(0.039)	(0.038)
$log(BOX_{j,t})$	0.439***	0.461***	0.440***	0.444***	0.437***
	(0.048)	(0.049)	(0.046)	(0.049)	(0.046)
$THEATERS_{j,t}$	−0.021	−0.018	−0.021	−0.017	−0.008
	(0.045)	(0.045)	(0.041)	(0.045)	(0.042)
$INDIE\ MOVIE_j$	0.064	0.069	0.059	0.111*	0.133**
	(0.042)	(0.044)	(0.040)	(0.057)	(0.054)
$AVG_{j,1:t-1}$	0.024	−0.118	0.030	0.043	0.058
	(0.041)	(0.081)	(0.037)	(0.045)	(0.043)
$AVGLEN_{j,1:t-1}$	0.006***	0.006***	0.006***	0.006***	0.006***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
$DI_{j,1:t-1}$	0.215***		0.136**	0.325***	0.286***
	(0.060)		(0.059)	(0.074)	(0.070)
$SD_{j,1:t-1}$		−0.215			
		(0.172)			
$DI_{j,1:t-1} \times AVGLEN_{j,1:t-1}$			0.006***		0.006***
			(0.001)		(0.001)
$DI_{j,1:t-1} \times INDIE\ MOVIE_j$				−0.160*	−0.230***
				(0.096)	(0.087)
Calendar Controls	Fixed	Fixed	Fixed	Fixed	Fixed
Movie effects	Fixed	Fixed	Fixed	Fixed	Fixed
Sargan Test	171.295	160.931	165.520	165.440	162.864
Autocorrelation test (1)	−3.613***	−3.598***	−3.824***	−3.459***	−4.049***
Autocorrelation test (2)	0.008	0.109	−0.077	0.073	−0.028
Wald test for coefficients	248958.322***	203236.834***	266600.042***	268696.990***	284282.420***
Note:					*p<0.10; **p<0.05; ***p<0.01

Table 4: Twostep dynamic panel model analyses using all available lags of VOL as internal instruments and $GOOGLE\ TRENDS$ as external instruments. Balanced Panel with $n = 425$ movies, $T = 5$ time periods, for a total of $N = 2,125$ observations. Model 1: Main effect of disagreement using *Disagreement Index* measure (H1); Model 2: Alternative measure of disagreement using standard deviation; Model 3: Interaction of DI and average length of prior reviews (H2); Model 4: Interaction of DI and product availability in opening week ($INDIE\ MOVIE$; H3); Model 5: Full model using both interactions.

Model 3 introduces the interaction effect between disagreement and the length of prior reviews. We find a statistically significant and positive coefficient. This supports H2, that the effect of disagreement on the review volume is amplified by the average length of previously posted reviews such that longer reviews increase the effect of disagreement. Model 4 probes the second interaction of interest, between disagreement and product availability. We find a statistically significant and negative coefficient for the interaction of disagreement and product availability. This provides support for H3, indicating that movies that have low initial marketplace availability do not benefit from disagreement as much as movies that are more widely available, and therefore had a large amount of positive information available pre-release due to advertising.

We probe the interaction effects visually by computing simple slopes based on the regression coefficients reported in Model 5 (Aiken and West, 1991). Panel A in Figure 5 shows the disagreement

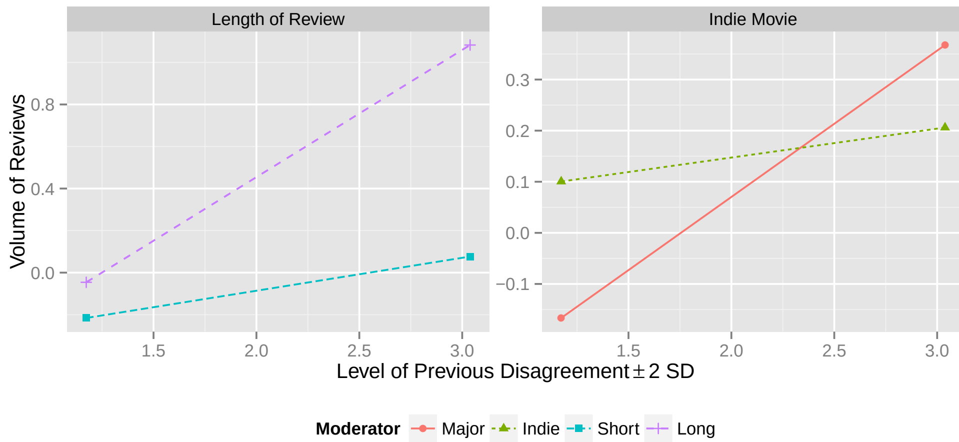

Figure 5: Interaction plots. Panel A shows interaction of prior disagreement and length of prior reviews. The positive effect of prior disagreement is amplified by longer average length of prior reviews. Panel B shows interaction of of a movie's availability with prior disagreement. The positive effect of disagreement is amplified for products with limited availability and decreases slightly for products with wide availability.

and length of review relationship using the mean length of reviews to split the data into movies that had short average review lengths and long average review lengths. Panel B in Figure 5 shows the disagreement and product availability relationship using the mean number of theaters to split the data into movies that had wide availability and those that had limited availability.

To investigate the effect of polarity of prior reviews on the volume of future reviews, we performed additional analyses using a flexible semi-parametric approach to model the partial relationship between volume of reviews to review and polarity (Figure 6). We re-estimate Model 1 using generative additive models (GAM) with random effects (Wood, 2011). The degree of smoothness of model terms is estimated as part of fitting using restricted maximum likelihood (REML). The results suggest a slight departure from the linear model. The semi-parametric model indicates the relationship between volume of reviews and polarity is slightly convex: for polarity values between 0 and about 1.5 the effect is around zero but it increases to values above zero with higher values of disagreement index (values above about 1.5). DI values above 1.5 correspond to rating distributions that are noticeably bi-modal (see simulated distributions in Figure 2).

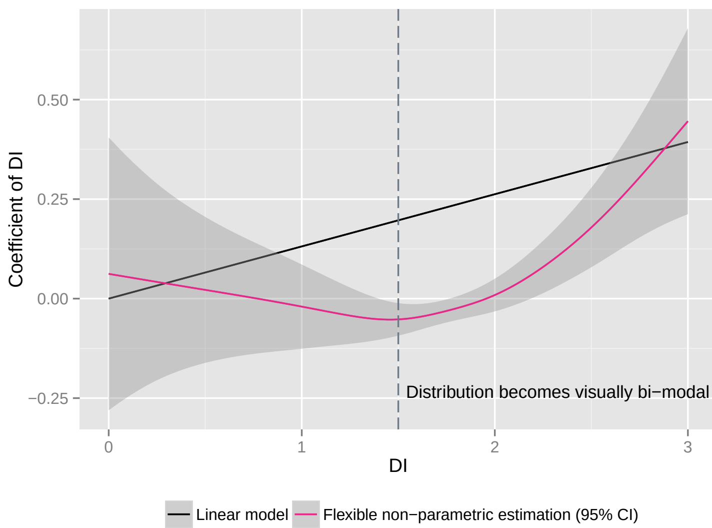

Figure 6: Linear versus semi-parametric specifications of the relationship between volume of reviews and polarity of prior reviews with 95% confidence bands estimated using GAM. The effect of DI is around zero for levels of low or medium disagreement and increases markedly for medium and high levels of disagreement.

4.2 Disagreement and Review Valence

The analyses presented so far investigate the effect of disagreement on the volume of future reviews. Overall, we find that increased disagreement predicts a higher volume of reviews. How does this effect of increased reviews affect the valence of posted reviews? Are the additional reviews garnered by increased disagreement more or less positive than the average reviews posted? To investigate these questions, we perform our second set of analyses, moving to an ordered logistic regression framework with movie fixed effects and employing data on the individual review level. Results are shown in Table 5 using individual level rating valence as the dependent variable ($STARS_{i,j}$).

Prior research (Muchnik et al., 2013) has shown positive herding effects such that products with reviews of higher prior valence receive higher valence ratings in the future (controlling for product effects, including quality). Prior research has also established that both over time and over sequence, reviews tend to exhibit a negative trajectory: reviews posted later are of lower valence and the $n+1$'th review is more negative than the n'th review (Godes and Silva, 2011). Consequently, we begin with a baseline in Model 1 by regressing valence of the next review on valence of previous reviews, the average length of prior reviews, timing and sequence of the new review, as well as movie fixed effects. We find a statistically significant and positive effect of prior

Dependent variable:	*Valence*		*Valence*
	ordered logistic		*OLS*
	(1)	(2)	(3)
$DI_{j,1:i-1}$		-0.173^{***}	-0.076^{***}
		(0.006)	(0.006)
Controls			
$AVG_{j,1:i-1}$	0.947^{***}	0.845^{***}	0.658^{***}
	(0.003)	(0.003)	(0.018)
$AVGLEN_{j,1:i-1}$	-0.003^{***}	-0.003^{***}	-0.002^{***}
	(0.0001)	(0.0001)	(0.0004)
$TIME_i$	-0.0002^{***}	-0.0002^{***}	-0.0002^{***}
	(0.00002)	(0.00002)	(0.00002)
$ORDER_i$	-0.00004^{***}	-0.00003^{***}	-0.00001^{***}
	(0.00000)	(0.00000)	(0.00000)
Movie effects	Fixed	Fixed	Fixed
Observations	296,899	296,899	296,899
AIC	782,542.42	782,368.05	
Adjusted R^2			0.214
Note:			*p<0.1; **p<0.05; ***p<0.01 Robust standard errors in parentheses.

Table 5: Predicting the valence of individual level ratings using ordered logit regression and OLS (for comparison).

valence, a statistically significant and negative effect of the average length of prior reviews, and statistically significant and negative effects of time and sequence. Coefficients are almost identical to those reported by Godes and Silva (2011). Model 2 introduces the main variable of interest, prior disagreement, into the regression. We find a statistically significant and negative coefficient of prior disagreement. The other coefficients remain largely similar in both size and significance level. This supports H4 that the valence of an online product review is negatively related to the disagreement among previously posted reviews about the same product. This is contrary to the results reported in prior work that finds no significant effect of disagreement by itself on the valence of ratings (Moe and Schweidel, 2012). Finally, Model 3 shows an OLS regression for comparison to the ordered logistic models shown in Model 1 and 2 with similar results.

As an additional robustness test, we investigate if our results could be driven by measurement errors of the mixture model-based classification of consumer reviews. Our simulation study in Section 3.2 indicates measurement errors are somewhat higher for DI values below 0.5. Hence, as an additional robustness test (not shown) we repeat our analysis on a subset of the data for which

$DI > 0.5$. We find results that are substantively similar.

5 Discussion and Conclusions

Increasing digitization of human behavior and the resultant increased access to digital trace data combined with advances in the development of machine learning and quantitative analysis methods have led to widespread popularity of big data analytics (Chen et al., 2012; George et al., 2014). Analyses of consumer-generated content and online WOM have been among the first areas to see widespread applications of these methods. Within this context, numerous studies have touched upon disagreement in consumer opinions (e.g., Dellarocas et al., 2010; Sun, 2012; Hu et al., 2009). However, upon close inspection of prior work, it is apparent that disagreement in online WOM has not been precisely conceptualized and has not been investigated systematically. As a consequence, the important concept of disagreement remains underdeveloped and poorly understood.

Using communication and public deliberation theories, we argued that disagreement in prior WOM shapes both the volume and valence of future reviews. Our empirical analysis shows that (1) prior disagreement leads to an increase in volume of reviews in the future; (2) this effect is amplified if the informational content of prior reviews is higher (using the length of prior reviews and product availability in the marketplace as proxies); and (3) the valence of future reviews declines with increased disagreement. Our statistical results suggest that the effect of disagreement is limited to extremely high—opposing—levels while moderate levels of disagreement in which opinions are merely diverging has little to no effect at all.

A key question we posed in the beginning of this work was how can we conceptualize disagreement in online WOM. In this article, we provide a theory of disagreement in online WOM (see Figure 1) and discuss how this conceptualization has important implications on how we measure disagreement and how we interpret the results. Driven by theoretically derived desirable characteristics, we suggest and evaluate an alternative measure that can capture the full spectrum of disagreement in a single statistic. We show results using two different measures and are able to provide nuanced interpretation of our empirical findings that allows us to tie results to high levels of disagreement—levels at which opinions become opposing rather than merely converging. Our results show that only DI values above 1.5 exhibit a strong effect on the propensity to review, which conforms to clearly bi-polar rating distributions across a variety of different means, standard deviations, and mixing proportions of those distributions. This insight would not be possible using standard deviation as a measure of disagreement as it is not able to distinguish bi-polar distributions from those that are uni-polar. This adds important nuances to existing research investigating effects of disagreement in online WOM.

Our study has implications for several issues that are important for theory building and research on online WOM, and specifically the role of disagreement as a mechanism of social influence (Dellarocas and Narayan, 2006; Sun, 2012; Agarwal and Dhar, 2014). If effects of disagreement depend, as we show, on the level of dissent such that diverging disagreement has no effect, but extreme levels of opposing disagreement do have an effect, this can explain differences in prior reported findings as well as differences with regard to the context in which these effects are studied. Conceptually distinct measures that tap into different aspects of disagreement, specifically their ability to accurately capture opposing levels of disagreement, can hold differing implications for our understanding of behavioral outcomes. We argue that the measure we construct can capture the full spectrum of disagreement in a single statistic and can serve as an alternative to standard deviation that is particularly useful when levels of disagreement are high, as is frequently the case in online WOM. Our goal is not to prescribe either measure as "better," rather to argue that both measures can provide meaningful insights in different research contexts. Furthermore, our empirical analysis demonstrates that clear conceptual understanding and measurement choices hold important consequences for the study of disagreement in online WOM and can explain diverging conclusions drawn in prior work.

An additional implication is the relationship of disagreement to informational content contained in online WOM and the information available about a product in general. Not only does the general level of disagreement of prior reviews affect later reviews, but it does so in quite nuanced ways. Building on theories of informational content and persuasive arguments (Vinokur and Burnstein, 1978; Isenberg, 1986; El-Shinnawy and Vinze, 1998), we find that informational content moderates the effect of disagreement. Spefically, we find that the effect of disagreement is amplified for products for which the average length of reviews is longer and thus reviews are likely to contain more information and appear more persuasive. This implication relates to the role of dynamic interactions within the space of online WOM (Aral, 2011; Godes and Silva, 2011). We suggest that an important role for research on these social dynamics is to further examine how prior reviews influence the perception of disagreement. Furthermore, we are able to tie the social dynamics to external informational content which we captured through a product's availability in the marketplace. Our finding that the effect of disagreement on a population's propensity to write reviews is increased by the availability of a product in the market provides additional nuance to prior work on the interaction between product and social effects in online settings (e.g., Dellarocas et al., 2010; Zhu and Zhang, 2010; Verbraken et al., 2014). We suggest that this could provide a first step in gaining additional insights into the formation of long-tail markets in the digital economy (Brynjolfsson et al., 2003), which may be influenced by the type and volume of online WOM that exists for products with lower market availability.

On a broader note, we contribute to the emerging field of big data analytics by demonstrating the importance of theoretical models for guiding measurement and interpretation. As the case of disagreement demonstrates, key effects only materialize in the realm of extremely high—opposing—disagreement. That is, diverging disagreement has little to no effect on consumer WOM, but opposing disagreement does have a significant effect. Consequently, we argue that access to large datasets and data analytic methods puts an increased burden on theorizing in addition to measurement (cf. Bapna et al., 2006). We suggest that an important role for research on big data analytics is to examine how naïve measurement approaches and interpretations could be improved by fully leveraging the richness that many digital trace data offer.

Our work also leads to a number of managerial implications. First, we demonstrate that the level of disagreement encompassed in prior reviews has important social influence effects on both the volume as well as the valence of future reviews. Consequently, the ease (or difficulty) with which consumers can gauge the existing level of disagreement can have important implications on review dynamics. Practitioners could, for example, exploit this finding by attempting to more precisely control the salience of disagreement by selectively displaying, or re-arranging, prior ratings. For example, Amazon.com recently changed the way in which they display product reviews to specifically highlight one very favorable and one very critical review in a side by side comparison. Importantly, these dynamics are different for products with different availability in the marketplace, such that products with lower availability—and thus lower availability of outside information about the product—benefit more from increased disagreement by accumulating a larger volume of future reviews. This could, for example, be incorporated into website designs such that reviews for products of differing availability are displayed differently. Furthermore, we demonstrated that with increased disagreement, the mean valence of future reviews goes down. This could have negative effects on product sales since lower review valence has been linked with lower sales (Clemons et al., 2006; Liu, 2006; Duan et al., 2008; Luca, 2011; Gopinath et al., 2013). Consequently, producers and merchants should be careful in encouraging too much disagreement. However, online WOM is a major content driver itself and some websites exist purely to collect and facilitate online WOM. These websites may be less concerned with the valence of reviews but instead focused on driving WOM itself. Consequently, these websites are likely to employ different optimization strategies.

It is worth mentioning limitations of our work as well. Although we account for many observable sources of endogeneity through the use of very strict temporal controls and the use of data on product search volume from Google Trends as well as the use of GMM estimation, it is possible that unobserved heterogeneity and simultaneity could still confound the estimates in our data.

In summary, using our disagreement conceptualization, we were able to explore how disagreement affects both the propensity of a consumer to post an online review, as well as the valence of that review. Taken together, our findings help to make the case that disagreement matters in online WOM and underline the importance of theory driven measurement and interpretation in big data analytics.

References

Agarwal, R. and Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for is research. *Information Systems Research*, 25(3):443–448.

Aiken, L. S. and West, S. G. (1991). *Multiple Regression: Testing and interpreting Interactions*. Sage, Thousand Oaks, CA.

Anderson, E. (1998). Customer satisfaction and word of mouth. *Journal of Service Research*, 1(1):5–17.

Aral, S. (2011). Identifying Social Influence: A Comment on Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science*, 30(2):217–223.

Archak, N., Ghose, A., and Ipeirotis, P. G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, 57(8):1485–1509.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. *Groups, leadership, and men*, pages 222–236.

Assmus, G., Farley, J. U., and Lehmann, D. R. (1984). How advertising affects sales: Meta-analysis of econometric results. *Journal of Marketing Research*, pages 65–74.

Bapna, R., Goes, P., Gopal, R., and Marsden, J. R. (2006). Moving from data-constrained to data-enabled research: Experiences and challenges in collecting, validating and analyzing large-scale e-commerce data. *Statistical Science*, 21(2):116–130.

Bapna, R., Goes, P., Wei, K. K., and Zhang, Z. (2011). A finite mixture logit model to segment and predict electronic payments system adoption. *Information Systems Research*, 22(1):118–133.

Benaglia, T., Hunter, D. R., and Young, D. S. (2009). mixtools : An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software*, 32(6).

Berger, J. and Milkman, K. (2012). What Makes Online Content Viral? *Journal of Marketing Research*, 49(2):192–205.

Blanton, H., Pelham, B. W., DeHart, T., and Carvallo, M. (2001). Overconfidence as Dissonance Reduction. *Journal of Experimental Social Psychology*, 37:373–385.

Blundell, R. and Bond, S. (1998). Initital conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

Brynjolfsson, E., Smith, M. D., and Hu, Y. J. (2003). Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers. *Management Science*, 49(11):1580–1596.

Chen, H., Chiang, R. H. L., and Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4):1165–1188.

Chevalier, J. A. and Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing*, 43(3):345–354.

Clemons, E. K., Gao, G. G., and Hitt, L. M. (2006). When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry. *Journal of Management Information Systems*, 23(2):149–171.

Cox, D. (1970). The continuity correction. *Biometrika*, 51(1):217–219.

Croissant, Y. and Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, 27(2).

Dellarocas, C., Gao, G. G., and Narayan, R. (2010). Are Consumers More Likely to Contribute Online Reviews for Hit or Niche Products?. *Journal of Management Information Systems*, 27(2):127–157.

Dellarocas, C. and Narayan, R. (2006). A Statistical Measure of a Populations Propensity to Engage in Post-Purchase Online Word-of-Mouth. *Statistical Science*, 21(2):277–285.

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):23–45.

Dewan, S. and Ramaprasad, J. (2014). Social media, traditional media, and music sales. *MIS Quarterly*, pages 101–121.

Dichter, E. (1966). How Word-of-Mouth Advertising Works. *Harvard Business Review*, 44(6):147–160.

Duan, W., Gu, B., and Whinston, A. (2008). The Dynamics of Online Word-of-Mouth and Product Sales - An Empirical Investigation of the Movie Industry. *Journal of Retailing*, 84(2):233–242.

Duan, W., Gu, B., and Whinston, A. (2009). Informational Cascades and Software Adoption on the Internet: An Empirical Investigation. *MIS Quarterly*, 33(1):23–48.

El-Shinnawy, M. and Vinze, A. S. (1998). Polarization and Persuasive Argumentation: A Study of Decision Making in Group Settings. *MIS Quarterly*, 22(2):165–199.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press, Stanford, CA.

George, G., Haas, M. R., and Pentland, A. (2014). Big Data and Management. *Academy of Management Journal*, 57(2):321–326.

Ghose, a., Ipeirotis, P. G., and Li, B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science*, 31(3):493–520.

Godes, D. and Silva, J. C. (2011). Sequential and Temporal Dynamics of Online Opinion. *Marketing Science*, online ahe:1–26.

Gopinath, S., Chintagunta, P. K., and Venkataraman, S. (2013). Blogs, Advertising, and Local-Market Movie Box Office Performance. *Management Science*, 59(12):2635–2654.

Hinsz, V. and Davis, J. (1984). Persuasive arguments theory, group polarization, and choice shifts. *Personality and Social Psychology Bulletin*, pages 260–268.

Hong, K. Y., Chen, P.-y., and Hitt, L. M. (2013). Measuring Product Type with Sequential Dynamics of Online Product Reviews: Theory and Applications. In *Workshop on Information Systems and Economics*, pages 1–31, Milan, Italy.

Hu, N., Pavlou, P., and Zhang, J. (2009). Overcoming the J-Shaped Distribution of Product Reviews. *Communications of the ACM*, 52(10):144–147.

Huckfeldt, R., Johnson, P. E., and Sprague, J. (2004). *Political Disagreement: The Survival of Diverse Opinions Within Communication Networks*. Cambridge University Press, New York, NY.

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, 50(6):1141–1151.

Iyengar, R., Van den Bulte, C., and Valente, T. W. (2011). Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science*, 30(2):195–212.

Jackson, J. E. (1983). Election night reporting and voter turnout. *American Journal of Political Science*, pages 615–635.

Kelly, T. (2008). Disagreement, dogmatism, and belief polarization. *The Journal of Philosophy*, 105(10):611–633.

Klofstad, C. a., Sokhey, A. E., and McClurg, S. D. (2013). Disagreeing about Disagreement: How Conflict in Social Networks Affects Political Behavior. *American Journal of Political Science*, 57(1):120–134.

Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., li, X., Cosley, D., Frankowski, D., Terveen, l., Rashid, A., Resnick, P., and Kraut, R. (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10.

Liu, Y. (2006). Word of Mouth for Movies : Its Dynamics and Impact on Box Office. *Journal of Marketing*, 70(3):74–89.

Luca, M. (2011). Reviews, Reputation, and Revenue: The Case of Yelp.com. *Harvard Business School Working Paper*, No. 12-016.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*, volume 299. John Wiley & Sons, New York, NY.

McPhee, W. N., Ferguson, J., and Smith, R. B. (1963). A theory of informal social influence. In McPhee, W. N., editor, *Formal Theories of Mass Behavior*, pages 74—-99. Collier-Macmillan, London, UK.

Mizuno, M. and Odagiri, H. (1990). Does advertising mislead consumers to buy low-quality products? *International Journal of Industrial Organization*, 8(4):545–558.

Moe, W. W. and Schweidel, D. A. (2012). Online Product Opinions: Incidence, Evaluation, and Evolution. *Marketing Science*, 31(3):372–386.

Moe, W. W. and Trusov, M. (2011). The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing*, 48(3):444–456.

Moon, S., Bergey, P. K., and Iacobucci, D. (2010). Dynamic Effects Among Movie Ratings, Movie Revenues, and Viewer Satisfaction. *Journal of Marketing*, 74(1):108–121.

Muchnik, L., Aral, S., and Taylor, S. J. (2013). Social Influence Bias: A Randomized Experiment. *Science*, 341(6146):647–651.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2):311–329.

Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3):521–543.

Olson, J. C. and Dover, P. A. (1978). Cognitive effects of deceptive advertising. *Journal of Marketing Research*, pages 29–38.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in stata. *Stata Journal*, 9(1):86–136.

Sahoo, N., Srinivasan, S., and Dellarocas, C. (2013). The Impact of Online Product Reviews on Product Returns and Net Sales Introduction. In *Workshop on Information Systems and Economics*, pages 1–6, Milan, Italy.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, 26(3):393–415.

Shmueli, G. and Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly*, 35(3):553–572.

Sia, C., Tan, B., and Wei, K. (2002). Group polarization and computer-mediated communication: Effects of communication cues, social presence, and anonymity. *Information Systems Research*, 13(1):70–90.

Sun, M. (2012). How Does the Variance of Product Ratings Matter? *Management Science*, 58(4):696–707.

Verbraken, T., Goethals, F., Verbeke, W., and Baesens, B. (2014). Predicting online channel acceptance with social network data. *Decision Support Systems*, 63:104–114.

Vinokur, A. and Burnstein, E. (1978). Novel argumentation and attitude change: The case of polarization following group discussion. *European Journal of Social Psychology*, 8(3):335–348.

Wang, C. K. J., Khoo, A., Liu, W. C., and Divaharan, S. (2008). Passion and intrinsic motivation in digital gaming. *CyberPsychology & Behavior*, 11(1):39–45.

Wang, J., Wen, S., Symmans, W. F., Pusztai, L., and Coombes, K. R. (2009). The bimodality index: a criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data. *Cancer Informatics*, 7:199.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-steps GMM estimators. *Journal of Econometrics*, 126(1):25–51.

Witten, I. H., Frank, E., and Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, Burlington, MA, 3rd edition.

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1):3–36.

Wright, P. (1980). Message-evoked thoughts: Persuasion research using thought verbalizations. *Journal of Consumer Research*, 7(2):151–175.

Wu, F. and Huberman, B. A. (2008). How public opinion forms. In *Internet and Network Economics*, pages 334–341. Springer.

Wu, L. and Brynjolfsson, E. (2013). The future of prediction: How google searches foreshadow housing prices and sales. *Working Paper,* http://ssrn.com/abstract=2022293.

Yin, D., Bond, S. D., and Zhang, H. (2014). Anxious or angry? effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, pages 539–560.

Zhang, J. Q., Craciun, G., and Shin, D. (2010). When does electronic word-of-mouth matter? a study of consumer product reviews. *Journal of Business Research*, 63(12):1336–1341.

Zhang, X. and Dellarocas, C. (2006). The lord of the ratings: Is a movie's fate is influenced by reviews? *ICIS 2006 Proceedings*, page 117.

Zhang, X. and Zhu, F. (2011). Group size and incentives to contribute: A natural experiment at chinese wikipedia. *American Economic Review*, 101(4):1601–1615.

Zhang, Z. H. U., Li, X. I. N., and Chen, Y. (2012). Deciphering Word-of-Mouth in Social Media: Text-Based Metrics of Consumer Reviews. *ACM Transactions on Management Information Systems*, 3(1):Article 5.

Zhu, F. and Zhang, M. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, 74:133–148.