

Journal of the Association for Information Systems

JAIS

Research Article

What Makes a Review Voted? An Empirical Investigation of Review Voting in Online Review Systems

Kevin K.Y. Kuan
The University of Sydney
kevin.kuan@sydney.edu.au

Kai-Lung Hui
Hong Kong University of Science and
Technology
klhui@ust.hk

Pattarawan Prasarnphanich
Sasin Graduate Institute of Business
Administration of Chulalongkorn University
pattarawan.prasarnphanich@sasin.edu

Hok-Yin Lai
Hong Kong Baptist University
jeanlai@comp.hkbu.edu.hk

Abstract

Many online review systems adopt a voluntary voting mechanism to identify helpful reviews to support consumer purchase decisions. While several studies have looked at what makes an online review helpful (review helpfulness), little is known on what makes an online review receive votes (review voting). Drawing on information processing theories and the related literature, we investigated the effects of a select set of review characteristics, including review length and readability, review valence, review extremity, and reviewer credibility on two outcomes—review voting and review helpfulness. We examined and analyzed a large set of review data from Amazon with the sample selection model. Our results indicate that there are systematic differences between voted and non-voted reviews, suggesting that helpful reviews with certain characteristics are more likely to be observed and identified in an online review system than reviews without the characteristics. Furthermore, when review characteristics had opposite effects on the two outcomes (i.e. review voting and review helpfulness), ignoring the selection effects due to review voting would result in the effects on review helpfulness being over-estimated, which increases the risk of committing a type I error. Even when the effects on the two outcomes are in the same direction, ignoring the selection effects due to review voting would increase the risk of committing type II error that cannot be mitigated with a larger sample. We discuss the implications of the findings on research and practice.

Keywords: *Online Review Systems, Review Voting, Review Helpfulness, Review Length, Readability, Review Valence, Review Extremity, Reviewer Credibility, Vividness, Diagnosticity, Sample Selection Bias.*

* Paul Pavlou was the accepting senior editor. This article was submitted on 13th February 2012 and went through two revisions.

Volume 16, Issue 1, pp. 48-71, January 2015

What Makes a Review Voted? An Empirical Investigation of Review Voting in Online Review Systems

1. Introduction

Consumers increasingly rely on online product reviews in guiding purchases. According to eMarketer (2010), 92 percent of consumers read online product reviews before making purchases, and 89 percent said that their purchase decisions were affected (favorably or unfavorably) after reading the reviews. However, online consumer reviews are not limited to online retailers. Traditional retailers can also take advantage of online consumer reviews to add value to in-store shopping experience. In fact, 82 percent of consumers consider online consumer reviews to be better than researching in-store with a sales associate (ZDNet, 2008). As a result, it is not surprising that over 80 percent of retailers planned to use online consumer reviews by the end of 2010 (eMarketer, 2010). For example, Sephora, the leading beauty retailer with presence in 13 countries and over 500 stores in the US, launched a mobile service in 2009 that allows in-store shoppers to read product reviews online.

Among the challenges posed by online consumer reviews is their explosive growth in number. It is unlikely and impossible for consumers to read all reviews in detail¹. To help consumers find helpful reviews among hundreds of reviews on a particular product, online review providers and retailers such as Amazon² have implemented a voting mechanism whereby consumers can rate whether a review is helpful. For example, Figure 1 shows a list of movie DVDs from Amazon. Figure 2 shows that for each single item the number of reviews often exceed several hundreds. Figure 2 shows that, on clicking an item, one can see a list of specific reviews, the ratings given by the reviewers to the item, and the proportion of helpful votes that each review received. Many websites, including the online Apple Store, eBay, IMDB.com, and so on, have implemented a similar voting mechanism in their review systems.

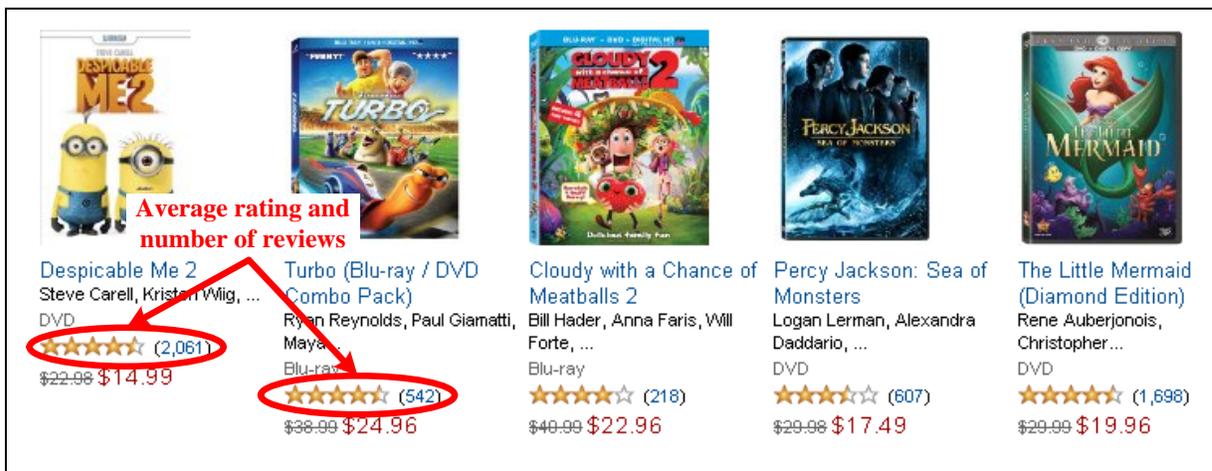


Figure 1. Reviews and Helpfulness—Item List with Number of Reviews

¹ In fact, 64 percent of consumers spent 10 minutes or more reading online reviews, 33 percent spent 30 minutes or more, and 39 percent of consumers read eight or more reviews. Only 12 percent read 16 or more reviews (eMarketer, 2010).

² Amazon is among the most popular source of online consumer reviews, with 26 percent of online consumers reading the reviews on Amazon before making their purchases (ZDNet, 2008).



The voluntary voting mechanism provides a practical way for someone to identify helpful reviews, provided that the reviews are voted on their helpfulness in the first place. Reviews that are supposed to be helpful but hasn't received votes cannot be identified by online review systems. Given the voluntary nature of the voting mechanism, reviews are not equally likely to receive votes, just as they are not considered equally helpful. Some reviews are more or less likely to receive votes because of systematic reasons other than their helpfulness. For example, a review that attracts readers' attention and motivates them to read it will be evaluated on its helpfulness, which may be helpful or unhelpful. However, online review systems will not show reviews with no such votes because they lack a helpfulness rating. Understanding the types of reviews that are more likely to receive votes can provide practical insights on the types of reviews that are inherently helpful but are often overlooked by online review systems because they receive no vote.

While there exists considerable amount of research on how different characteristics of a review affect its helpfulness (Chevalier & Mayzlin, 2006; Forman, Ghose, & Wiesenfeld, 2008; Korfiatis, Rodríguez, & Sicilia, 2008; Otterbacher, 2009; Zhang & Tran, 2009; Mudambi & Schuff, 2010; Zhang, Craciun, & Shin, 2010; Ghose & Ipeiritos, 2011; Baek, Ahn, & Choi, 2012; Korfiatis, Garcia-Bariocanal, & Sanchez-Alonso, 2012; Liu, Jin, Harding, & Fung, 2013), little is known on how these characteristics affect the chance of the review receiving votes in the first place. Understanding how review characteristics affect review voting is important from theoretical and methodological perspectives. Theoretically, what makes a review helpful does not necessarily mean it will receive votes. Drawing on information processing theories and related literature, we argue that review voting and review helpfulness are driven by different theoretical considerations. Methodologically, review helpfulness has been studied based on voted reviews only because one cannot observe the helpfulness of non-voted reviews. Examining the impact of review characteristics on review helpfulness, without considering their impacts on review voting, is subject to sample selection bias (Heckman, 1976, 1979). While past studies have acknowledged this bias (Mudambi & Schuff, 2010), little is known about the bias's influence, such as the magnitude and the direction, on the true impacts of review characteristics on review helpfulness. Hence, assessing the impacts of review characteristics on review voting is necessary to assess the true impacts of the same characteristics on review helpfulness.

Hence, we examine the effects of review characteristics as past studies have suggested on two distinct but related outcomes: review voting and review helpfulness. In line with prior research in online consumer reviews, we argue that the effects of review characteristics on review helpfulness can be explained in terms of review diagnosticity. Our basic premise is that people are more likely to vote a review with characteristics that provide diagnostic value to their product evaluation as helpful. However, for the effects of review characteristics on review voting, we draw on information processing theories and related literature and argue from a different theoretical

perspective: our basic premise is that a vivid review that attracts attention and that motivates one to further process the review is more likely to receive votes. Conversely, a pallid review that fails to attract attention is less likely to receive votes.

We examined the effects of review characteristics on review voting and review helpfulness using reviews that had both received and not received votes for DVD and book titles from Amazon, and analyzed them using the sample selection model. Furthermore, we assessed the non-random sampling under different conditions and its implications to practice and research (e.g., type I and type II errors). This holistic approach provides a more-comprehensive understanding of the unbiased effects of review characteristics on both outcomes (review voting and helpfulness) that otherwise would be overlooked by examining each of the outcomes separately.

2. Theoretical Foundation and Model

Online consumer reviews have received considerable interest in research (e.g., Chevalier & Mayzlin, 2006; Dellarocas, Zhang, & Awad, 2007; Forman et al., 2008; Pathak, Garfinkel, Gopal, & Venkatesan, 2010; Zhu & Zhang, 2010). One line of study has focused on the characteristics that make a review considered to be helpful (Pang & Lee 2004; Ghose & Ipeirotis, 2006; Pavlou & Dimoka, 2006; Korfiatis et al., 2008; Mudambi & Schuff, 2010). These studies view the relationships between review characteristics and review helpfulness as driven by the diagnostic value reflected by their characteristics. In other words, reviews with certain characteristics are more likely to be voted helpful because they provide diagnostic value to consumers (Jiang & Benbasat, 2004, 2007; Sen & Lerman, 2007; Wang, Teo, & Wei, 2007; Lee & Youn, 2009; Mudambi & Schuff, 2010; Zhang et al., 2010; Purnawirawan, de Pelsmacker, & Dens, 2012; Qiu, Pang, & Lim, 2012).

One interesting question is whether review characteristics also affect the chance for a review to receive votes. In other words, does the diagnostic value of a review also determine whether it is more likely to receive votes? Drawing on theories and models in the information processing literature (Hovland, Janis, & Kelley, 1953; McGuire, 1968, 1969; Petty & Cacioppo, 1981, 1986; Herr, Kardes, & Kim, 1991), we argue that the effects of review characteristics on review voting and helpfulness are driven by different theoretical considerations.

2.1. Review Voting and Review Helpfulness

Classical information processing theories such the message learning theory (Hovland et al., 1953) and McGuire's information processing model (McGuire, 1968, 1969) suggest that people go through a series of stages in information processing (see Figure 3) (Hamilton & Nowak, 2005). Beginning with message exposure, people go through message reception (attention and comprehension) and message yielding (evaluation, belief change, and attitude change). Message evaluation (the first stage in message yielding) is influenced by both attention and comprehension (the two stages in message reception). In other words, how someone evaluates a message depends not only on what information they pay attention to, but also how they comprehend this information to reach the evaluation. The elaboration likelihood model (ELM) also views message evaluation as being influenced by multiple stages (Petty & Cacioppo, 1981, 1986): the model posits that message evaluation is determined by both an individual's motivation and ability to elaborate on the message. Similarly, the accessibility-diagnostics model (Herr et al., 1991) posits that message evaluation depends on both the information's vividness (accessibility) and diagnosticsity.

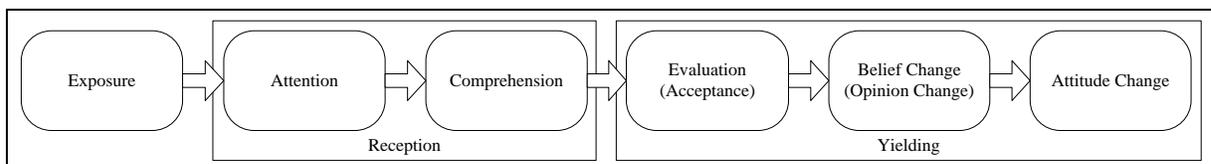


Figure 3. Stages in Information Processing (Adapted from Hovland et al., 1953; McGuire, 1968,

A common theme across the different theories and models in information processing is that the influence of message characteristics on message evaluation depends not only on their effects on comprehension/diagnosticity/ability, but also on their effects on attention/vividness/motivation. Each stage can be influenced by message characteristics in different manners. For example, a vivid message that attracts greater attention (Tversky & Kahneman, 1974; Nisbett & Ross, 1980) and enhances an individual's motivation to process the content (MacInnis, Moorman, & Jaworski, 1991) is not necessarily more diagnostic than a pallid message (Herr et al., 1991).

Online consumer review studies have primarily looked into the effects of review characteristics on review helpfulness from diagnosticity-comprehension perspectives. The basic premise is that reviews' characteristics that enhance their diagnosticity are expected to be evaluated favorably. However, given the large number of online consumer reviews, it is unlikely that every review is comprehended and determined on its diagnosticity. Hence, the diagnosticity-comprehension paradigm, while providing a strong theoretical explanation on why a review is voted helpful or not helpful (i.e., review helpfulness), may not explain why a review is voted or not voted in the first place (i.e., review voting).

Instead, we examine the effects of review characteristics on review voting from the vividness-attention perspective. The basic premise is that a review's vividness, as enhanced by review characteristics, determines whether the review is likely to receive votes in the first place. Hence, given the large number of online consumer reviews, a vivid review, which tends to attract readers' attention (Tversky & Kahneman, 1974; Nisbett & Ross, 1980) and motivate them to read (MacInnis et al., 1991) it, is more likely to be voted. Conversely, a pallid review, which tends to draw less attention (Herr et al., 1991), is more likely to be overlooked or abandoned. An overlooked or abandoned review is unlikely to receive votes.

Past studies have examined several review characteristics related to reviews' text (Pang & Lee, 2004; Ghose & Ipeiritis, 2006; Pavlou & Dimoka, 2006; Korfiatis et al., 2008; Mudambi & Schuff, 2010; Cao, Duan, & Gan, 2011), their numerical rating (Forman et al., 2008; Mudambi & Schuff, 2010), their author (Forman et al., 2008; Ghose & Ipeiritis, 2011), and so on. In this study, we examine the effects of these review characteristics on review helpfulness from the diagnosticity-comprehension perspective, whereas we review their effects on review voting from the vividness-attention perspective.

2.2. Review Content

Review length and readability are two frequently studied review characteristics (Mudambi & Schuff, 2010; Ghose & Ipeiritis, 2011). Review length measures the amount of information in a review, and a longer review, *ceteris paribus*, should be more comprehensive in information and, thus, more diagnostic and more helpful than a shorter review (Mudambi & Schuff, 2010). Readability measures the ease of reading the content (Korfiatis et al., 2008), which affects how one comprehends the review (Ghose & Ipeiritis, 2011). Consistent with past studies, we argue that review length and readability reflect the diagnosticity of a comprehensive and comprehensible review, which, in turn, affects its helpfulness.

Yet, in the context of online consumer review, not every review is fully attended to, comprehended, and evaluated on its helpfulness. In many cases, most reviews are simply ignored without being attended to at all. For a review to be possibly evaluated on its helpfulness, it has to be attended to in the first place, or else it has no chance to be comprehended to determine its diagnosticity. In this regard, the review's vividness plays a prominent role in influencing the amount of attention it receives (Tversky & Kahneman 1974; Nisbett & Ross, 1980; Herr et al., 1991) and readers' motivation to further process the review further (MacInnis et al., 1991).

Review length reflects the size of a review's text on the screen. Without having to comprehend the content, a longer review is visually more salient and less likely to be overlooked on the screen than a shorter review. Readability, which reflects the linguistic sophistication of a review (Ghose & Ipeiritis, 2011), affects the attention required to skim through the review. Furthermore, researchers have suggested that readability can impact a message's perceived concreteness and interestingness

(Sadoski, Goetz, & Fritz, 1993; Sadoski, 1999), both of which affect its vividness (Hamilton & Hunter, 1998; Hamilton & Nowak, 2005). Hence, we argue that review length and readability affect the salience of and the attention drawn to a review, which have a direct influence on review voting.

H1a: *A longer and readable review is more likely to receive votes.*

H1b: *A longer and readable review is more likely to be rated as helpful.*

2.3. Review Valence

Positive and negative reviews have been shown to exert different degrees of influence in product evaluation (Sen & Lerman, 2007; Forman et al., 2008; Lee & Youn, 2009; Zhang et al., 2010; Khare, Labrecque, & Asare, 2011; Cui, Lui, & Guo, 2012). Research on negativity bias suggests that negative information is generally considered more diagnostic than positive information (Skowronski & Carlston, 1987; Herr et al., 1991). Furthermore, apart from its diagnosticity, negative information is also more vivid than positive information (Fiske, 1980; Herr et al., 1991; Vonk, 1993). With people being mildly positive most of the time, negative information becomes more salient in contrast and attracts greater attention than positive information (Fiske, 1980; Skowronski & Carlston, 1987; Smith, Cacioppo, Larsen, & Chartrand, 2003).

Hence, the literature suggests that review valence may have similar effects on review voting and helpfulness, but for different theoretical explanations. The diagnosticity explanation for review valence on review helpfulness is based on the premise that negative information is more probative or less ambiguous than positive information and, thus, more helpful to someone making a judgment. This is in line with the purchasing bias identified in online consumer reviews, which suggests that positive reviews are more likely to be reported than negative reviews (Hu, Pavlou, & Zhang, 2009). The purchasing bias results in review valence skewed toward the positive on average. Positive reviews inflate the positively skewed bias further, whereas negative reviews reduce it. Hence, negative reviews, being more diagnostic to unbiased judgment, are more likely to be voted as helpful compared to positive reviews.

The vividness explanation for review valence on review voting is based on the premise that negative information is more salient than positive information and, thus, more likely to be attended to. Given the overwhelming number of online consumer reviews being positive, negative reviews are expected to be visually more salient in contrast and attract greater attention than positive reviews. Hence, negative reviews are more likely to receive votes than positive reviews because they are more vivid. Negative reviews are more likely to be considered helpful than positive reviews because they are more diagnostic.

H2a: *A negative review is more likely to receive votes than a positive review.*

H2b: *A negative review is more likely to be rated as helpful than a positive review.*

2.4. Review Extremity

Online consumer-review studies have found review extremity as another review characteristic that affects review helpfulness (Forman et al., 2008; Mudambi & Schuff, 2010; Cao et al., 2011). Review extremity is commonly measured by the rating given to a review, and reviews with extreme ratings (e.g., one star or five stars) are often associated with “brag or moan” views (Hu et al., 2009). The single-sided and biased views associated with extreme reviews make them less diagnostic than reviews with moderate ratings (Mudambi & Schuff, 2010). Hence, reviews with extreme ratings are less likely to be considered helpful.

However, extreme information has been shown to be more vivid and attract greater attention (Skowronski & Carlston, 1987; Herr et al., 1991). In the context of online consumer reviews, with most reviews being four-star or five-star reviews (Hu et al., 2009), a one-star review is more visually distinct and more likely to be noticed. The reverse holds for a five-star review among mostly one-star or two-star

reviews. In fact, extreme reviews, positive or negative, are more likely to receive a higher number of helpfulness votes (Cao et al., 2011), which suggests that they generally attract greater attention. Furthermore, the seemingly discrepant view suggested by an extreme review creates cognitive dissonance, a psychological tension that induces motivation to attend to the review (Festinger, 1957; Strong, 1968; Greenwald & Ronis, 1978; Akerlof & Dickens, 1982; Harmon-Jones & Mills, 1999; Cooper, 2007). Hence, while being less diagnostic, extreme reviews are more vivid. As such, they are more likely to be attended to, but they are less likely to be considered helpful.

H3a: *An extreme review is more likely to receive votes.*

H3b: *An extreme review is less likely to be rated as helpful.*

2.5. Reviewer Credibility

Reviewers' characteristics have also been found to play an important role in online consumer reviews (Pavlou & Dimoka, 2006; Brown, Broderick, & Lee, 2007; Cheung, Lee, & Rabjohn, 2008; Hu, Liu, & Zhang, 2008; Cheung, Luo, Sia, & Chen, 2009; Ghose & Ipeirotis, 2011; Pan & Zhang, 2011; Baek et al., 2012; Cheung, Sia, & Kuan, 2012; Dou, Walden, Lee, & Lee, 2012). In particular, review characteristics that suggest the reviewer's credibility have a positive impact on review helpfulness (Forman et al., 2008; Ghose & Ipeirotis, 2011). The influence of source credibility on message evaluation has been well documented in the literature (Chaiken & Maheswaran, 1994; Jones, Sinclair, & Courneya, 2003; Sussman & Siegal, 2003; Eisend, 2004; Pornpitakpan, 2004; Tormala & Petty, 2004; Eisend, 2010). In general, information from a credible source is more diagnostic and given more thoughtful consideration (Chaiken, 1980; Petty & Cacioppo, 1981, 1986; Herr et al., 1991; Chaiken & Maheswaran, 1994; Van Hove & Lievens, 2009). It follows that reviews from credible reviewers are more likely to be considered helpful.

In Amazon and many online review systems, reviews by reputable reviewers are visually tagged with a "top reviewer" badge, which serves a visual cue that helps consumers to skim through the overwhelming number of reviews and heuristically determine which reviews to consider further (Bikhchandani, Hirshleifer, & Welch, 1998; Hansen & Haas, 2001; Forman et al., 2008). The "top reviewer" badge makes the review tagged with the badge visually more prominent on the screen regardless of the content. Hence, reviews by top reviewers are more likely to be attended to.

H4a: *A review from a top reviewer is more likely to receive votes.*

H4b: *A review from a top reviewer is more likely to be rated as helpful.*

Figure 4 depicts the research model. The model contains two dependent variables: review voting and review helpfulness. Review helpfulness is only observed for voted reviews. Hence, we examined review helpfulness by considering review voting to account for the sampling bias (Heckman, 1976, 1979). Consistent with past literature, we argue that the effects of review characteristics on review helpfulness are theoretically driven by the review's diagnosticity. However, we argue that the effects of review characteristics on review voting are theoretically driven by the review's vividness. Given the different theoretical considerations, review characteristics may impact review voting and review helpfulness in different manners.

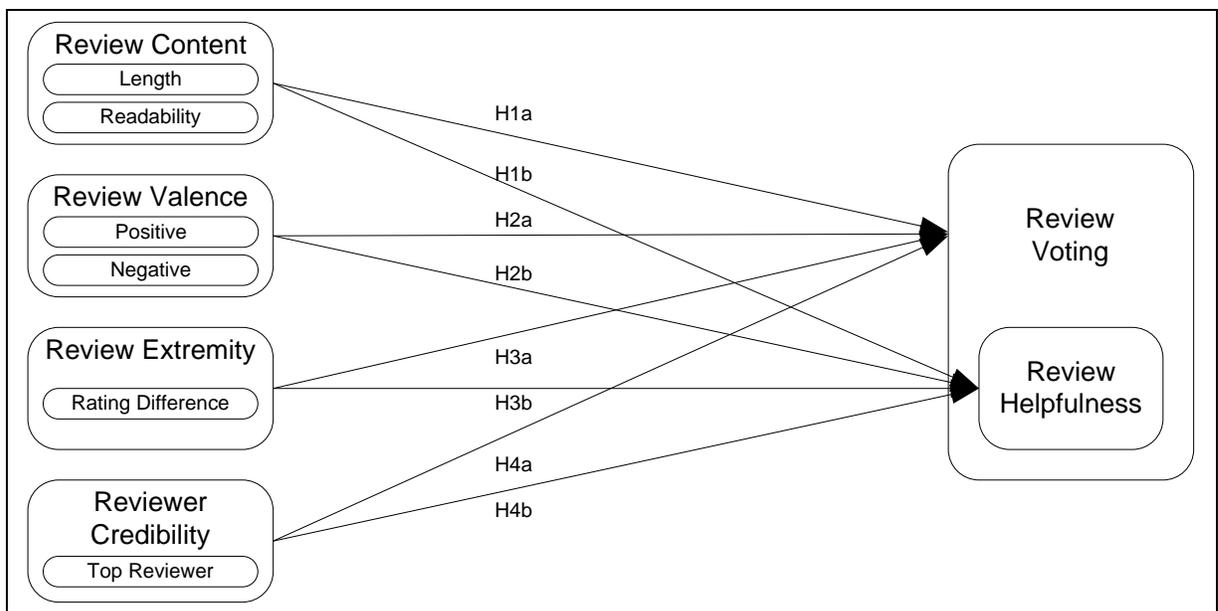


Figure 4. Research Model

3. Methodology

3.1. Data

In this paper, the units of analysis are reviews about specific items. Our data contains a random sample of 37,007 reviews on 629 DVD titles and 89,362 reviews on 1,003 books from Amazon³. To ensure that all the reviews were sufficiently exposed to consumers, we discarded reviews that were published for less than three months. On average, each DVD and book received 59 and 89 reviews, respectively. The average ratings for DVD and book were 4.28 and 4.36 on a five-point scale from one star (lowest) to five stars (highest).

3.2. Measures

We used two dependent variables: review voting and review helpfulness. We measured review voting with a binary variable that indicated whether a review received votes for being helpful (voting): “0” denotes that the review received no votes, and “1” denotes that the review received at least one vote. We measured review helpfulness, as adapted from past studies (Mudambi & Schuff, 2010; Ghose & Ipeirotis, 2011), by the proportion of helpful votes received (helpfulness), or the ratio of helpful votes to total votes received by a review.

Following past studies, we measured the reviews for their length and readability (Korfiatis et al., 2008; Mudambi & Schuff, 2010; Ghose & Ipeirotis, 2011; Korfiatis et al., 2012). We measured length by the number of words in the review (Words). We measured readability with the Flesch Reading Ease Score (FRES) of the review, a widely used measure that indicates how easily a text is comprehended (Flesch, 1948, 1951, 1974). FRES has also been used in online consumer review studies to measure reviews’ readability in online consumer review studies (Korfiatis et al., 2008; Ghose & Ipeirotis, 2011; Korfiatis et al., 2012). FRES ranges from 0 (very difficult to read) to 100 (very easy to read). In general, a text with a FRES between 0 and 30 would be best understood by college graduates, a text with a FRES between 60 and 70 would be easily understood by 13- to 15- year-old students, and a text with a FRES between 90 and 100 would be easily understood, an average, by 11-year-old students.

³ To ensure the data reflected review helpfulness and review voting at approximately the same time, we limited the data collection to a two-week period. Data associated with 629 (targeted 500) DVD titles (e.g., review rating, review text, number of votes, etc.) during a two-week period in March 2009. Data associated with 1003 (targeted 1000) books were downloaded during a two-week period in April 2010.

We derived review valence based on the positive and negative words in a review, which we identified by using the General Inquirer content analysis software word list (Buvac & Stone, 2001). We measured positive review valence by the proportion of positive words, or the ratio of positive words to total words (positive valence). We measured negative review valence by the proportion of negative words, or the ratio of negative words to total words (negative valence). A review that is high in positive or negative valence suggests that the review consists of mainly positive or negative information, respectively (e.g., a one-sided review). A review that is high in both positive and negative valence suggests that the review includes both positive and negative information (e.g., a two-sided review). A review that is low in both positive and negative valence suggests that the review lacks positive or negative information (e.g., an equivocal review).

We measured review extremity by the difference between the rating given by a review and the average rating given by all reviews (rating difference), and in a similar manner for positively and negatively extreme reviews (i.e., we consider the absolute differences) (Cao et al., 2011). For example, we would denote a one-star review on a product with an average rating of 4.5 stars with a rating difference of 3.5. High extremity suggests that the review has a drastically different view, which makes it more likely to attract attention (Cao et al., 2011).

We measured reviewer credibility by a binary variable that indicates whether a top reviewer wrote the review (top reviewer). Amazon ranks all reviewers based on the quantity and quality of all the reviews each reviewer writes, and reviews written by top reviewers are tagged with a “top reviewer” badge. Hence, “1” denotes that a top reviewer wrote the review with a “top reviewer” badge, and “0” denotes otherwise.

Tables 1 and 2 present descriptive statistics and correlations of the variables. As previous studies on online consumer reviews (e.g., Forman et al., 2008; Ghose & Ipeirotis, 2011) suggest, we included control variables, including average review ratings for the reviewed items (average rating)⁴, sales ranks of the items, number of days elapsed since the reviews were published (review age), and number of reviews received by the items (total reviews), in the analyses.

Table 1. Descriptive Statistics

Variable	DVD				Book			
	Mean	St. dev.	Max	Min	Mean	Std. dev.	Max	Min
Voting	0.762	0.425	1	0	0.820	0.384	1	0
Helpfulness	0.716	0.335	1	0	0.667	0.337	1	0
Words	91.37	102.09	2,061	1	141.46	160.06	5,563	0
FRES	68.59	16.94	100	0	63.31	17.98	100	0
Positive valence	0.072	0.075	1	0	0.067	0.053	1	0
Negative valence	0.028	0.052	1	0	0.029	0.028	1	0
Rating difference	0.824	0.778	4	0	0.816	0.738	4	0
Top reviewer	0.057	0.231	1	0	0.032	0.176	1	0
Average rating	4.276	0.480	5	1	4.301	0.419	5	2.5
Sales rank	35,285	276,271	7,449,881	1	2,846	5,865	202,972	127
Review age	938	835	4684	98	1273	1104	5229	91
Total reviews	687	628	2,170	1	588	695	2,712	1

⁴ Since review extremity concerns about the difference in rating given by a review relative to other reviews, we included the average rating as a control variable in the estimation.

Table 2. Correlations

	1	2	3	4	5	6	7	8	9	10	11	12
1. Voting	1											
2. Helpfulness	0.598	1										
3. Words	0.134	0.164	1									
4. FRES	-0.096	-0.088	-0.225	1								
5. Positive valence	-0.052	-0.034	-0.159	0.039	1							
6. Negative valence	0.043	-0.008	0.060	-0.009	0.144	1						
7. Rating difference	0.183	-0.129	0.053	-0.016	-0.088	0.111	1					
8. Top reviewer	0.057	0.082	0.190	-0.114	-0.040	0.019	-0.041	1				
9. Average rating	-0.137	0.053	-0.092	0.027	0.067	-0.095	-0.374	-0.025	1			
10. Sales rank	0.000	0.024	0.004	-0.005	0.010	-0.008	-0.007	-0.001	0.001	1		
11. Review age	0.012	0.014	0.046	-0.016	-0.042	0.003	-0.040	-0.001	0.102	0.007	1	
12. Total reviews	-0.050	-0.073	-0.001	0.013	-0.039	0.014	0.035	-0.037	-0.033	-0.060	0.178	1

3.3. Empirical Model

The primary model we used to analyze the data was as the sample selection model (Heckman, 1976, 1979). We used the same model because some reviews were voted because of systematic reasons, such as their length (Mudambi & Schuff, 2010; Ghose & Ipeirotis, 2011) or readability (Korfiatis et al., 2008; Korfiatis et al., 2012). Hence, a sample that contains only voted reviews is likely to be non-random. Least squares estimation of such a non-random sample would produce biased estimates (Greene, 2007)⁵.

The sample selection model of voting comprises two equations:

Equation 1: Voting: $s_i^* = w_i' \gamma + u_i$, $s_i = 1$ if $s_i^* > 0$, and 0 otherwise.

Equation 2: Helpfulness: $y_i = x_i' \beta + \varepsilon_i$, which is observable only if $s_i = 1$.

The conditional expectation of review helpfulness is:

Equation 3: $E[y_i | s_i, x_i, w_i] = x_i' \beta + \rho \sigma_\varepsilon \lambda(w_i' \gamma)$,

where ρ is the correlation between the two error terms, u_i and ε_i , σ_ε is the standard deviation of ε_i , and $\lambda(\cdot)$ is the inverse Mills ratio. Equation 1 is a Probit "selection" equation that allows us to explore what determine whether a review gets voted. Equation 2 is the key equation of interest – it relates review helpfulness to the review characteristics.

Ordinary least squares regression of Equation 2 using observations for which $s_i = 1$ (i.e., only voted reviews) would produce biased estimates because of the second term in Equation 3. The problem is akin to an omitted variable bias. The parameters in Equation 2, however, can be consistently and efficiently estimated by using the information (specifically, the predicted $w_i' \gamma$) obtained from Equation 1 in a two-step estimation procedure (Heckman 1976, 1979).

4. Results

We first estimated a basic model that contained only the control variables. We report the results in column 1 of Table 3 (DVD dataset) and Table 4 (book dataset). Review age was positively correlated, and sales rank and total reviews were negatively correlated, with voting. These results are well

⁵ Mudambi and Schuff (2010) used a Tobit specification to study a similar problem. However, they included only voted reviews in their estimation. Tobit specification does not apply in our setting because the problem is one of non-random sampling, not censoring. Hence, it did not resolve any bias arising from non-random sampling.

expected because the longer a review has been published, the more likely it would have received at least one vote. Reviews of popular items with lower sales ranks were more likely to receive votes because customers are more likely to buy them (hence their popularity). The negative influence of total reviews could be due to limited attention span. When there were more reviews per item, it was less likely for readers to go through and vote for each of them. There were no unequivocal influences from these variables on helpfulness.

Table 3. DVD Results

	(1)		(2)		(3)		(4)	
	Basic		Full model		With squared length		With syllables/words	
Variables	Voting	Helpfulness	Voting	Helpfulness	Voting	Helpfulness	Voting	Helpfulness
Words [†]			1.048*** (0.103)	0.374*** (0.026)	1.959*** (0.141)	0.869*** (0.049)	1.891*** (0.145)	0.882*** (0.048)
Words squared [†]					-2.692*** (0.278)	-1.264*** (0.087)	-2.597*** (0.281)	-1.265*** (0.084)
FRES [†]			-2.729*** (0.467)	-1.169*** (0.141)	-2.344*** (0.468)	-0.950*** (0.137)		
Syllables per Word [†]							95.167* (45.489)	51.775*** (9.582)
Words per sentence [†]							3.479*** (0.765)	-0.178 (0.172)
Positive valence			-0.938*** (0.107)	-0.219*** (0.041)	-0.828*** (0.108)	-0.161*** (0.040)	-0.806*** (0.108)	-0.144*** (0.039)
Negative valence			1.888*** (0.221)	0.211*** (0.051)	1.758*** (0.220)	0.175*** (0.050)	1.707*** (0.222)	0.140* (0.062)
Rating difference			0.436*** (0.014)	-0.098*** (0.007)	0.437*** (0.014)	-0.099*** (0.007)	0.437*** (0.014)	-0.107*** (0.007)
Top reviewer			0.509*** (0.039)	0.076*** (0.012)	0.494*** (0.039)	0.072*** (0.012)	0.500*** (0.039)	0.067*** (0.012)
Average rating			-0.188*** (0.019)	0.019** (0.006)	-0.184*** (0.019)	0.021*** (0.006)	-0.184*** (0.019)	0.026*** (0.006)
Log (sales rank)	-0.115*** (0.006)	0.001 (0.016)	-0.098*** (0.006)	0.012*** (0.002)	-0.098*** (0.006)	0.012*** (0.002)	-0.099*** (0.006)	0.014*** (0.002)
Log (review Age)	0.204*** (0.008)	0.040 (0.028)	0.204*** (0.009)	0.008*** (0.004)	0.197*** (0.009)	0.003 (0.004)	0.197*** (0.009)	-0.001 (0.004)
Log (total reviews)	-0.046*** (0.007)	-0.042*** (0.006)	-0.056*** (0.008)	-0.027*** (0.002)	-0.052*** (0.008)	-0.024*** (0.002)	-0.050*** (0.008)	-0.023*** (0.002)
Constant	0.247*** (0.063)	0.504 (0.195)	0.916*** (0.112)	0.689*** (0.031)	0.918*** (0.112)	0.685*** (0.030)	0.554*** (0.125)	0.569*** (0.034)
σ^2		0.404		0.362		0.361		0.347
ρ		0.746		0.675		0.681		0.540
Inverse Mills Ratio		0.302 (0.314)		0.244*** (0.049)		0.246*** (0.048)		0.187*** (0.048)
R ² Adjusted R ²		0.015 0.015		0.139 0.139		0.146 0.146		0.146 0.146
N	37,007	28,230	37,007	28,230	37,007	28,230	37,007	28,230

† We rescaled Words (mean centered), FRES, Syllables per word and Words per sentence by dividing them by 1,000. Words squared = square of rescaled Words.
*** p<0.001, ** p<0.01, * p<0.05

Table 4. Book Results

	(1)		(2)		(3)		(4)	
	Basic		Full model		With squared length		With syllables/words	
Variables	Voting	Helpfulness	Voting	Helpfulness	Voting	Helpfulness	Voting	Helpfulness
Words [†]			-0.047* (0.020)	0.031*** (0.004)	-0.159*** (0.024)	-0.001 (0.006)	-0.177*** (0.024)	-0.102*** (0.005)
Words squared [†]					0.829*** (0.096)	0.224*** (0.022)	0.862*** (0.098)	0.160*** (0.021)
FRES [†]			-0.046*** (0.003)	-0.008*** (0.001)	-0.044*** (0.003)	-0.007*** (0.001)		
Syllables per word [†]							6.384*** (0.347)	1.119*** (0.098)
Words per sentence [†]							8.169*** (0.567)	2.755*** (0.128)
Positive valence			-0.071*** (0.016)	-0.021*** (0.004)	-0.088*** (0.016)	-0.025*** (0.004)	0.007 (0.017)	-0.003 (0.003)
Negative valence			0.124*** (0.036)	0.027*** (0.007)	0.120*** (0.035)	0.025*** (0.007)	0.117** (0.038)	0.014* (0.007)
Rating difference			0.500*** (0.011)	-0.071*** (0.004)	0.500*** (0.011)	-0.075*** (0.004)	0.501*** (0.011)	-0.101*** (0.003)
Top reviewer			0.534*** (0.036)	0.131*** (0.008)	0.541*** (0.036)	0.129*** (0.008)	0.464*** (0.036)	0.075*** (0.007)
Average rating			-0.318*** (0.014)	0.026*** (0.004)	-0.318*** (0.014)	0.029*** (0.004)	-0.308*** (0.014)	0.054*** (0.004)
Log (sales rank)	-0.105*** (0.004)	0.688*** (0.022)	-0.068*** (0.005)	-0.012*** (0.001)	-0.070*** (0.005)	-0.012*** (0.001)	-0.074*** (0.005)	-0.008*** (0.001)
Log (review age)	0.141*** (0.004)	-0.053*** (0.007)	0.299*** (0.006)	0.055*** (0.003)	0.298*** (0.006)	0.052*** (0.003)	0.295*** (0.006)	0.031*** (0.003)
Log (total reviews)	-0.150*** (0.004)	0.091*** (0.010)	-0.166*** (0.005)	-0.033*** (0.002)	-0.166*** (0.005)	-0.032*** (0.002)	-0.166*** (0.005)	-0.021*** (0.001)
Constant	1.559*** (0.044)	-0.088*** (0.011)	1.590*** (0.092)	0.497*** (0.020)	1.558*** (0.092)	0.488*** (0.020)	0.972*** (0.094)	0.382*** (0.021)
σ_e		0.585		0.336		0.332		0.317
ρ		1.294		0.543		0.482		-0.091
Inverse Mills ratio		0.758*** (0.181)		0.183*** (0.029)		0.160*** (0.028)		-0.029*** (0.026)
R ²		0.025		0.092		0.093		0.105
Adjusted R ²		0.025		0.092		0.093		0.104
N	89,362	73,282	89,362	73,282	89,362	73,282	89,362	73,282

† We rescaled words (mean centered), FRES, syllables per word, and words per sentence by dividing them by 1,000. Words squared = square of rescaled words.

*** p<0.001, ** p<0.01, * p<0.05

4.1. Review Content

H1 and H1a state that longer and readable reviews are more likely to receive votes and be rated as helpful, respectively. As Tables 3 and 4 (Column 2) show, the coefficients of words on helpfulness were positive for DVDs and books. However, the coefficients of words on voting were positive for DVDs but negative for books. One possible explanation is that the effects of review length are non-linear: if a review is too long, people may disregard it simply because they do not want to comprehend a long message (Holmqvist, Holsanova, Barthelson, & Lundqvist, 2003).

To test the non-linear effects of review length, we included squared term of words (words squared) in the estimation (Column 3, Tables 3 and 4). The coefficients of words and words squared were significant with opposite signs, which support the non-linear effects of review length and explain the inconsistent linear effects on review voting across DVDs and books. Both DVD reviews and book reviews were very large in quantity (687 DVD reviews and 588 book reviews per item; see Table 1). However, DVD reviews were considerably shorter in length than book reviews (91 words per DVD review and 141 words per book review; see Table 1). Our results suggest that, *ceteris paribus*, a longer DVD review, among many short reviews, is more likely to receive votes and be considered helpful until it becomes too long (about 430 words). However, a book review, among those already long reviews, needs to be sufficiently long enough for it to be increasingly likely voted (at least 230 words, or 90 words longer than average) and considered helpful (at least 140 words). In other words, a slightly longer book review among numerous long book reviews does not necessarily benefit from its greater length unless it goes beyond a certain threshold.

Interestingly, the FRES coefficients on voting and helpfulness were both negative. Most commonly used readability measures take two factors into consideration: sentence length and word complexity. Sentence length is measured by average number of words per sentence. Word complexity is measured by average number of syllables per word. FRES was negatively correlated with number of words per sentence and number of syllables per word⁶. The results suggest that less-readable reviews (i.e., longer sentences with more-complex words) are more likely to receive votes and be rated as helpful. Past literature has shown that reduced readability, while increasing comprehension difficulty, does increase motivation to elaborate (Lowrey, 1998; Bradley & Meeds, 2002). Within a moderate range of readability, the increased motivation to process the review due to reduced readability outweighs the increased comprehension difficulty, which results in an overall enhancement to the review's comprehension and diagnosticity. In our data, approximately 65 percent of reviews had a FRES higher than 60 (easily understood by 13-year old or younger students). Less than five percent of reviews had a FRES lower than 30 (best understood by university graduates). Given the majority reviews were in the easy to moderate range of readability, a decrease in readability may actually enhance the overall comprehension.

To assess the individual effects of sentence length and word complexity on the two outcomes, we re-estimated our model by replacing readability with average sentence length (words per sentence) and word complexity (syllables per word) (see Column 4, Tables 3 and 4). The coefficients of the two replacement variables on the two dependent variables were both positive (cf. negative when FRES was used)⁷. In other words, both sentence length and word complexity were individually important to the two outcomes. Accordingly, provided that a review remains reasonably readable, the use of elaborated sentences and/or sophisticated words do interest consumers to attend to it and read further. A review written in simple language may be considered unprofessional and unhelpful (Petty & Wegener, 1998; Petty & Brinol, 2002).

4.2. Review Valence

H2 and H2a state that negative reviews are more likely to receive votes and be rated as helpful than positive reviews. Our findings (Tables 3 and 4, Column 2) show that the negative valence was positively associated with both voting and helpfulness, and that positive valence was negatively associated with voting and helpfulness. In other words, the results suggest that negative (positive) reviews are more likely to receive votes, and that they are generally considered to be more (less) helpful. The findings are consistent with the literature in negativity research, which suggests that negative information is considered more vivid and diagnostic than positive information (Ahluwalia, 2000; Sen & Lerman, 2007). Given the overwhelming number of online consumer reviews, customers are unlikely to attend to every review, but rather skim through them and selectively attend to some. With the majority of reviews being positive (Chevalier and Mayzlin 2006; Hu et al. 2009)⁸, negative reviews are more likely to attract attention than positive reviews.

⁶ Technically, FRES is computed as $206.876: 1.015 \times (\text{words per sentence}) - 84.6 \times (\text{syllables per word})$ (Flesch 1948). The negative correlations become apparent from this formula.

⁷ The only exception was the effect of words per sentence on helpfulness in the DVD data, which was insignificant.

⁸ In our data, 77 and 82 percent of DVD and book reviews, respectively, were either four- or five-star reviews.

Given that negative reviews are more likely to receive votes, we expect that they will receive higher number of helpfulness votes than positive reviews. To confirm this, we regressed the total number of votes (total votes) obtained by each DVD and book review on positive and negative valence. Table 5 reports the results. We found negative valence (positive valence) to be positively (negatively) associated with the total number of votes. This is consistent with the conjecture that negative reviews are more likely to be attended to and receive votes than positive reviews. To probe further into this phenomenon, we grouped the reviews based on the medians of positive valence and negative valence, and compared the reviews in terms of voting and total votes for reviews with (1) high positive valence and low negative valence, and (2) low positive valence and high negative valence. Table 6 reports the results. The reviews in group (ii) had a significantly higher chance of being voted and received substantially more votes than those in group (i).

Table 5. Positive and Negative Valence and Voting Behaviors—Regression Results

Variables	DVD	Book
Positive valence	-2.183*** (0.084)	-0.067*** (0.011)
Negative valence	2.946*** (0.119)	0.195*** (0.019)
Constant	1.361*** (0.007)	1.522*** (0.003)

*** p<0.001, ** p<0.01, * p<0.05

Table 6. Positive and Negative Valence and Voting Behaviors—Mean Comparisons

Review valence	DVD		Book	
	Voting	Total votes	Voting	Total votes
High positive and low negative	0.708	4.497	0.749	7.012
Low positive and high negative	0.820	10.019	0.780	10.594
t-statistic	18.204***	16.600***	3.561***	6.148***

*** p<0.001, ** p<0.01, * p<0.05

4.3. Review Extremity and Reviewer Credibility

H3 and H3a state that review extremity has positive effects on review voting but negative effects on review helpfulness. Our findings (Tables 3 and 4, Column 2) support these hypotheses. The opposite effects suggest an interesting trade-off between review voting and review helpfulness. The positive coefficients of rating difference on voting suggest that reviews with extreme ratings are more likely to receive votes. However, the negative coefficients of rating difference on helpfulness suggest that reviews with extreme ratings, while more likely to receive votes, are considered less helpful. Finally, H4 and H4a state that reviews by top reviewers are more likely to receive votes and be considered helpful. The positive coefficients of top review on voting and helpfulness supported these hypotheses.

4.4. Sample Selection Bias

Our results suggest that sample selection bias is significant in online voting. Reviews with votes were systematically different from those without votes. To assess the impact of such selection bias, we removed all reviews that received no votes and re-estimated the review helpfulness equation using ordinary least squares (OLS)⁹. Table 7 shows the OLS results and the review helpfulness results in the sample selection model.

⁹ The spirit of this regression was similar to previous studies that excluded non-voted review data (Ghose & Iperiotis, 2011; Mudambi & Scuff, 2010). However, the detailed specifications may vary across studies. For example, Mudambi and Scuff (2010) used the Tobit model instead of OLS as in most studies. We also re-estimated the review helpfulness equation using the Tobit model and the results were similar to those using OLS.

The estimated coefficients were remarkably different in terms of magnitude and significance. For variables such as words, FRES, positive and negative valence, top reviewer, and so on, their impacts on voting and helpfulness were in the same direction (i.e., both positive or both negative), but the OLS estimates on helpfulness were biased downwards. For the variables of interest, the underestimation ranged from -19 percent to -60 percent. In the cases of the positive valence and negative valence, the OLS estimates were insignificant in the DVD data (N = 28,230), which is a smaller set than the book data (N = 73,282), though still quite sizeable by most standards. In general, when a review characteristic impacts review voting and review helpfulness in the same direction (i.e., both positive or both negative), ignoring the sample selection effect would result in an increased risk of committing a type II error (false negative).

Conversely, rating difference had opposite effects on voting (positive) and helpfulness (negative), and the OLS estimates on helpfulness were biased upwards by approximately 30 percent (from +28 to +35%). In this case, sample selection poses a greater threat to research because overlooking it would result in an increased risk of committing a type I error (false positive).

Table 7. Heckman vs. OLS

Variables	DVD			Book		
	Heckman	OLS	Diff.	Heckman	OLS	Diff.
Words [†]	0.869*** (0.049)	0.699*** (0.032)	-20%	-0.001 (0.006)	0.008 (0.005)	n.a.
Words squared [†]	-1.264*** (0.087)	-1.011*** (0.065)	-20%	0.224*** (0.022)	0.180*** (0.020)	-20%
FRES [†]	-0.950*** (0.137)	-0.695*** (0.119)	-27%	-0.007*** (0.001)	-0.004*** (0.001)	-34%
Positive valence	-0.161*** (0.040)	-0.050 (0.031)	n.a.	-0.025*** (0.004)	-0.020*** (0.003)	-19%
Negative valence	0.175*** (0.050)	0.053 (0.040)	n.a.	0.025*** (0.007)	0.020** (0.007)	-22%
Rating difference	-0.099*** (0.007)	-0.133*** (0.002)	+35%	-0.075*** (0.004)	-0.096*** (0.001)	+28%
Top reviewer	0.072*** (0.012)	0.028*** (0.007)	-60%	0.129*** (0.008)	0.102*** (0.006)	-20%
Average rating	0.021*** (0.006)	0.042*** (0.004)	+99%	0.029*** (0.004)	0.050*** (0.002)	+70%
Log (sales rank)	0.012*** (0.002)	0.022*** (0.001)	+71%	-0.012*** (0.001)	-0.008*** (0.001)	-32%
Log (review age)	0.003 (0.004)	-0.015*** (0.002)	n.a.	0.052*** (0.003)	0.035*** (0.001)	-32%
Log (total reviews)	-0.024*** (0.002)	-0.019*** (0.002)	-22%	-0.032*** (0.002)	-0.023*** (0.001)	-28%
Constant	0.685*** (0.030)	0.742*** (0.026)	+8%	0.488*** (0.020)	0.493*** (0.019)	+1%
R ²	0.146	0.145		0.093	0.093	
Adjusted R ²	(0.146)	(0.145)		(0.093)	(0.093)	
N	28,230	28,230		73,282	73,282	

† We rescaled Words (mean centered) and FRES by dividing them by 1,000. Words squared = square of rescaled Words.
 *** p<0.001, ** p<0.01, * p<0.05

5. Discussion and Implications

What makes a review helpful do not necessarily make them receive votes. Yet, only helpful reviews that receive votes can be identified by online review systems. We looked at the impacts of review characteristics on both review voting and review helpfulness. Our results indicate that review voting is not a random act. Just as some reviews are generally more helpful, some are more likely to receive votes. By analyzing data on DVD and book reviews from Amazon, the results suggest a few interesting points.

Reviews are not necessarily better being long and easy to read. A longer review attracts attention and motivates reading up to a certain point, and discourages processing when the additional cognitive effort anticipated exceeds the incremental value expected from extra length. This is probably the case for DVD reviews, where the reviews are generally shorter (91.37 words on the average in our data). However, for book reviews where other reviews are generally longer (141.46 words on the average in our data), a longer review among the already long reviews may actually discourage further reading since the additional processing effort anticipated is likely to outweigh the incremental value expected from the extra length. Besides, reviews using elaborated sentences and sophisticated vocabularies, though technically less readable, may appear more professional. Provided that the reviews are sufficiently readable, they are more attractive to read and considered more helpful. Reviews that are too readable (e.g., those composed of short or even broken sentences using simple vocabularies) may appear unprofessional. They are more likely to be disregarded or evaluated unfavorably.

It is not surprising that praising reviews, which constitute the majority of reviews, are less likely to receive votes. Not to mention the fact that they are considered less helpful, it is more difficult for praising reviews to stand out from the crowd and attract people to read. Rather, people are more attracted to, or even purposely screen for, criticizing reviews because they are more helpful for individuals' purchase decisions. This is consistent with the literature that negative information is more vivid and diagnostic than positive information (Herr et al., 1991; Ito, Larsen, Smith, & Cacioppo, 1998; Ahluwalia, 2000; Rozin & Royzman, 2001; Sen & Lerman, 2007). Besides, people also use the "top reviewer" badge as a heuristic cue to screen reviews for further reading (Chaiken, 1980; Petty, Cacioppo, & Schumann, 1983; Petty & Cacioppo, 1986; Chaiken & Maheswaran, 1994). Reviews labeled as written by top reviewers are more likely to be voted, and they are generally also more helpful.

The opposite effects of review extremity on review voting and review helpfulness create an interesting trade-off and demonstrate the different theoretical considerations that drive review voting and review helpfulness. Consistent with past studies, an extreme review is generally considered less helpful. Without considering review voting, one may conclude that a moderate review is preferred because it is more helpful. However, a moderate review with a similar rating as most other reviews appears to offer nothing unique, which makes it less appealing and more likely to be overlooked or disregarded. Instead, the opposite effects of review extremity on review voting and review helpfulness together suggest that it is the reviews that are discrepant enough that appear to offer some different opinions, yet not too extreme to appear as biased, are more likely to be attended to and evaluated favorably.

5.1. Implications for Research

This study has some implications for research. First, by examining the effects of review characteristics on both review voting and review helpfulness together, the study provides some additional insights that would otherwise have been overlooked by studying the two outcomes separately. For example, by simply looking at the negative effect of review extremity on review helpfulness, one might conclude that extreme reviews are bad and moderate reviews are good. However, the positive effect of review extremity on review voting suggests that moderate reviews are more likely to be unnoticed or disregarded since they are perceived to be similar to other reviews and to provide little extra value. In fact, the opposite effects of review extremity suggest that reviews that are discrepant enough to appear different from other reviews and draw people's attention, yet not too discrepant to appear as biased, would motivate people to read and evaluate them favorably.

Second, just as any other voluntary mechanism, review voting is systematically determined by review characteristics. Analyzing review helpfulness based on observed outcomes alone is subject to the sample selection bias that cannot be simply mitigated with a larger sample. While past studies on online consumer reviews acknowledge the bias, this study is among the first attempts that actually address the bias and assess its nature. By comparing the sample selection model estimates and the OLS estimates, we show that, when a review characteristic affects both review voting and review helpfulness in the same direction, ignoring the issue of sample selection would result in a higher risk of committing type II error. However, when the effects are in the opposite direction, as in the case of review extremity, ignoring the issues of sample selection would result in a higher risk of committing type I error.

5.2. Implications for Practice

This study also provides some implications for practice. Our findings suggest some general design guidelines to online consumer review systems. For example, to encourage other consumers to read and evaluate the reviews, online review systems may want to determine the desirable review length for different products and guide or even restrict reviews' length. Automated agents may be deployed to provide support (e.g., spelling and grammar checking, synonym suggestion, readability statistics, etc.) to reviewers when writing reviews. Reviewers may be reminded to take a more-balanced view in the review when it becomes too "extreme", or be more specific in justifying their opinions. Given that consumers are more attracted to reviews with extreme ratings and those written by top reviewers, features may be implemented to allow these reviews to be filtered to help consumers find those reviews more easily.

Our findings also show that reviews are not equally likely to receive votes, and that there are systematic differences between reviews with and without votes in terms of their characteristics. Online review systems should exercise caution when building predictive models using only reviews with votes for providing personalized recommendations. Without considering the fact that some reviews are more likely to receive votes, such models would be biased toward certain types of reviews. In fact, the systems may examine newly published reviews that have not been received votes, identify those that are potentially helpful but likely to be overlooked by consumers, and display them in positions that are more prominent. This would help consumers discover helpful reviews that would otherwise be buried among other reviews.

5.3. Limitations

The findings of this study should be interpreted with caution. First, our data was limited to DVD and book reviews at Amazon. With the growing wealth of literature on the different effects for different product categories, such as search vs. experience products (Bhattacharjee, Gopal, Lertwachara, & Marsden, 2006; Weathers, Sharma, & Wood, 2007; Mudambi & Schuff, 2010), and hedonic vs. utilitarian products (Park & Young, 1986; Okada, 2005; Smith, Menon, & Sivakumar, 2005), the generalizability of our findings to other product categories requires further investigation. For example, by looking at books and DVDs, this study focuses on products with which people tend to have more emotional experiences. It is plausible that the relationship between diagnosticity and attention would be different for goods with more specific, fact-driven characteristics (such as an outdoor grill). Second, we focused on characteristics related to the reviews in our empirical model. Future research could augment our analysis with voters' characteristics, such as their level of expertise and involvement. The literature suggests that these characteristics may moderate the impacts of review characteristics on review voting and review helpfulness (Sussman & Siegal, 2003). Finally, we were limited to observable variables available from Amazon, and we used them as surrogate measures of review characteristics. The use of surrogate variables, together with other uncontrollable variables in secondary data, limits our model's explanatory power. Future studies using controlled experiment could directly examine how different review characteristics affect the processing of review in different stages.

5.4. Conclusions

The success of any online system that involves voluntary use often depends on whether consumers participate in exploring the information presented by the system, and whether they, after exploring the

system, actually prefer to use it. What motivates people to explore a system in the first place is often different from what makes it evaluated positively. The sample selection involved in voting (selection) and helpfulness (response) is clearly not limited to online consumer review. Past studies have mostly addressed either the “selection” (e.g., Chen, Harper, & Konstan, 2010; Zhang & Zhu, 2010) or the “response” (e.g., Mudambi & Schuff, 2010; Ghose & Ipeirotis, 2011) question in isolation. This study suggests that integrating these two outcomes in a holistic analysis may provide richer insights on online consumer behavior.

Acknowledgement

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 144510).

References

- Ahluwalia, R. (2000). Examination of psychological processes underlying resistance to persuasion. *Journal of Consumer Research*, 27(2), 217-232.
- Akerlof, G. A., & Dickens, W. T. (1982). The economic consequences of cognitive dissonance. *The American Economic Review*, 72(3), 307-319.
- Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of online consumer reviews: readers' objectives and review cues. *International Journal of Electronic Commerce*, 17(2), 99-126.
- Bhattacharjee, S., Gopal, R. D., Lertwachara, K., & Marsden, J. R. (2006). Consumer search and retailer strategies in the presence of online music sharing. *Journal of Management Information Systems*, 23(1), 129-159.
- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *The Journal of Economic Perspectives*, 12(3), 151-170.
- Bradley, S. D., & Meeds, R. (2002). Surface-structure transformations and advertising slogans: The case for moderate syntactic complexity. *Psychology & Marketing*, 19(7-8), 595-619.
- Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of Interactive Marketing*, 21(3), 2-20.
- Buvac, V., & Stone, P. (2001). The General Inquirer User's Guide. Retrieved July 5, 2010, from http://www.wjh.harvard.edu/~inquirer/j1_0/manual/manual.html
- Cao, Q., Duan, W. J., & Gan, Q. W. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511-521.
- Chaiken, S. (1980). Heuristic versus systematic information-processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(5), 752-766.
- Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing—effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 66(3), 460-473.
- Chen, Y., Harper, F. M., Konstan, J., & Li, S. X. (2010). Social comparisons and contributions to online communities: A field experiment on MovieLens. *American Economic Review*, 100(4), 1358-1398.
- Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The impact of electronic word-of-mouth—the adoption of online opinions in online customer communities. *Internet Research*, 18(3), 229-247.
- Cheung, C. M. Y., Sia, C. L., & Kuan, K. K. Y. (2012). Is this review believable? A study of factors affecting the credibility of online consumer reviews from an ELM perspective. *Journal of the Association for Information Systems*, 13(8), 618-635.
- Cheung, M. Y., Luo, C., Sia, C. L., & Chen, H. (2009). Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. *International Journal of Electronic Commerce*, 13(4), 9-38.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345-354.
- Cooper, J. (2007). *Cognitive dissonance: Fifty years of a classic theory*. Thousand Oaks, CA: Sage.
- Cui, G., Lui, H. K., & Guo, X. N. (2012). The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce*, 17(1), 39-57.
- Dellarocas, C., Zhang, X. Q., & 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.
- Dou, X., Walden, J. A., Lee, S., & Lee, J. Y. (2012). Does source matter? Examining source effects in online product reviews. *Computers in Human Behavior*, 28(5), 1555-1563.
- Eisend, M. (2004). Is it still worth to be credible? A meta-analysis of temporal patterns of source credibility effects in marketing. *Advances in Consumer Research*, 31(1), 352-357.
- Eisend, M. (2010). Explaining the joint effect of source credibility and negativity of information in two-sided messages. *Psychology & Marketing*, 27(11), 1032-1049.
- eMarketer. (2010). *The role of customer product reviews*. Retrieved from <http://www.emarketer.com/Article/Role-of-Customer-Product-Reviews/1008019>

- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.
- Fiske, S. T. (1980). Attention and weight in person perception—the impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, 38(6), 889-906.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221-233.
- Flesch, R. F. (1951). *How to test readability*. New York: Harper.
- Flesch, R. F. (1974). *The art of readable writing: With the Flesch readability formula*. New York: Harper & Row.
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19(3) 291-313.
- Ghose, A., & Ipeiritos, P. G. (2006). *Designing ranking systems for consumer reviews: The impact of review subjectivity on product sales and review quality*. Paper presented at the Workshop on Information Technology and Systems, Milwaukee, WI.
- Ghose, A., & Ipeiritos, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10), 1498-1512.
- Greene, W. H. (2007). *Econometric analysis*. Upper Saddle River, NJ: Prentice Hall.
- Greenwald, A. G., & Ronis, D. L. (1978). Twenty years of cognitive dissonance: Case study of the evolution of a theory. *Psychological Review*, 85(1), 53-57.
- Hamilton, M. A., & Hunter, J. E. (1998). The effect of language intensity on receiver evaluations of message, source, and topic. In M. Allen & R. Preiss (Eds.), *Persuasion: Advances through Meta-analysis* (pp. 99-138). Cresskill, NJ: Hampton.
- Hamilton, M. A., & Nowak, K. L. (2005). Information systems concepts across two decades: An empirical analysis of trends in theory, methods, process, and research domains. *Journal of Communication*, 55(3), 529-553.
- Hansen, M. T., & Haas, M. R. (2001). Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. *Administrative Science Quarterly*, 46(1), 1-28.
- Harmon-Jones, E. E., & Mills, J. E. (1999). *Cognitive dissonance: Progress on a pivotal theory in social psychology*. Washington, DC: American Psychological Association.
- Heckman, J. J. (1976). Common structure of statistical-models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5(4), 475-492.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153-161.
- Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of word-of-mouth and product-attribute information on persuasion—an accessibility-diagnostics perspective. *Journal of Consumer Research*, 17(4), 454-462.
- Holmqvist, K., Holsanova, J., Barthelson, M., & Lundqvist, D. (2003). Reading or scanning? A study of newspaper and net paper reading. In J. Hyona, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 657-670). Amsterdam: Elsevier Science.
- Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change*. New Haven, CT: Yale University Press.
- Hu, N., Liu, L., & Zhang, J. J. (2008). Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology & Management*, 9(3) 201-214.
- Hu, N., Pavlou, P. A., & Zhang, J. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM*, 52(10), 144-147.
- Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, 75(4), 887-900.
- Jiang, Z. H., & Benbasat, I. (2004). Virtual product experience: Effects of visual and functional control of products on perceived diagnosticity and flow in electronic shopping. *Journal of Management Information Systems*, 21(3), 111-147.
- Jiang, Z. H., & Benbasat, I. (2007). Investigating the influence of the functional mechanisms of online product presentations. *Information Systems Research*, 18(4), 454-470.

- Jones, L. W., Sinclair, R. C., & Courneya, K. S. (2003). The effects of source credibility and message framing on exercise intentions, behaviors, and attitudes: An integration of the elaboration likelihood model and prospect theory. *Journal of Applied Social Psychology, 33*(1), 179-196.
- Khare, A., Labrecque, L. I., & Asare, A. K. (2011). The assimilative and contrastive effects of word-of-mouth volume: An experimental examination of online consumer ratings. *Journal of Retailing, 87*(1), 111-126.
- Korfiatis, N., Garcia-Bariocanal, E., & Sanchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications, 11*(3), 205-217.
- Korfiatis, N., Rodríguez, D., & Sicilia, M. A. (2008). The impact of readability on the usefulness of online product reviews: A case study on an online bookstore. In M. D. Lytras, J. M. Carroll, E. Damiani, & R. D. Tennyson (Eds.), *Emerging technologies and information systems for the knowledge society* (pp. 423-432). Berlin: Springer-Verlag.
- Lee, M., & Youn, S. (2009). Electronic word of mouth (eWOM): How eWOM platforms influence consumer product judgement. *International Journal of Advertising, 28*(3), 473-499.
- Liu, Y., Jin, J., Ji, P., Harding, J. A., & Fung, R. Y. K. (2013). Identifying helpful online reviews: A product designer's perspective. *Computer-Aided Design, 45*(2), 180-194.
- Lowrey, T. M. (1998). The effects of syntactic complexity on advertising persuasiveness. *Journal of Consumer Psychology, 7*(2), 187-206.
- MacInnis, D. J., Moorman, C., & Jaworski, B. J. (1991). Enhancing and measuring consumers motivation, opportunity, and ability to process brand information from ads. *Journal of Marketing, 55*(4), 32-53.
- McGuire, W. J. (1968). Personality and attitude change: An information-processing theory. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes* (pp. 171-196). San Diego, CA: Academic Press.
- McGuire, W. J. (1969). The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (pp. 136-314). Reading, MA: Addison-Wesley.
- Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly, 34*(1), 185-200.
- Nisbett, R. E., & Ross, L. (1980). *Human Inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Okada, E. M. (2005). Justification effects on consumer choice of hedonic and utilitarian goods. *Journal of Marketing Research, 42*(1), 43-53.
- Otterbacher, J. (2009). "Helpfulness" in online communities: A measure of message quality. *Proceedings of the 27th Annual Conference on Human Factors in Computing Systems, 1-4*, 955-964.
- Pan, Y., & Zhang, J. Q. (2011). Born unequal: A study of the helpfulness of user-generated product reviews. *Journal of Retailing, 87*(4), 598-612.
- Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*.
- Park, C. W., & Young, S. M. (1986). Consumer response to television commercials—the impact of involvement and background music on brand attitude formation. *Journal of Marketing Research, 23*(1), 11-24.
- Pathak, B., Garfinkel, R., Gopal, R. D., Venkatesan, R., & Yin, F. (2010). Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems, 27*(2), 159-188.
- Pavlou, P. A., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research, 17*(4), 392-414.
- Petty, R. E., & Brinol, P. (2002). Attitude change: The elaboration likelihood model of persuasion. In G. Bartels & W. Nelissen (Eds.), *Marketing for sustainability: Towards transactional policy making* (pp. 176-190). Amsterdam: IOS Press.
- Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion—classic and contemporary approaches*. Dubuque, IA: W.C. Brown.

- Petty, R. E., & Cacioppo, J. T. (1986). Communication and persuasion: Central and peripheral routes to attitude change. New York, NY: Springer-Verlag.
- Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and peripheral routes to advertising effectiveness—the moderating role of involvement. *Journal of Consumer Research*, 10(2), 135-146.
- Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 323-390). New York: McGraw-Hill.
- Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, 34(2), 243-281.
- Purnawirawan, N., de Pelsmacker, P., & Dens, N. (2012). Balance and sequence in online reviews: How perceived usefulness affects attitudes and intentions. *Journal of Interactive Marketing*, 26(4), 244-255.
- Qiu, L. Y., Pang, J., & Lim, K. H. (2012). Effects of conflicting aggregated rating on eWOM review credibility and diagnosticity: The moderating role of review valence. *Decision Support Systems*, 54(1), 631-643.
- Rozin, P., & Royzman, E. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4), 296-320.
- Sadoski, M. (1999). Theoretical, empirical, and practical considerations in designing informational text. *Document Design*, 1(1), 25-34.
- Sadoski, M., Goetz, E. T., & Fritz, J. B. (1993). A causal model of sentence recall—effects of familiarity, concreteness, comprehensibility, and interestingness. *Journal of Reading Behavior*, 25(1), 5-16.
- Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the Web. *Journal of Interactive Marketing*, 21(4), 76-94.
- Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory—the role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, 52(4), 689-699.
- Smith, D., Menon, S., & Sivakumar, K. (2005). Online peer and editorial recommendations, trust, and choice in virtual markets. *Journal of Interactive Marketing*, 19(3), 15-37.
- Smith, N. K., Cacioppo, J. T., Larsen, J. T., & Chartrand, T. L. (2003). May I have your attention, please: Electrocortical responses to positive and negative stimuli. *Neuropsychologia*, 41(2), 171-183.
- Strong, S. R. (1968). Counseling: An interpersonal influence process. *Journal of Counseling Psychology*, 15(3), 215-224.
- Sussman, S. W., & Siegal, W. S. (2003). Informational influence in organizations: An integrated approach to knowledge adoption. *Information Systems Research*, 14(1), 47-65.
- Tormala, Z. L., & Petty, R. E. (2004). Source credibility and attitude certainty: A metacognitive analysis of resistance to persuasion. *Journal of Consumer Psychology*, 14(4), 427-442.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty—heuristics and biases. *Science*, 185(4157), 1124-1131.
- Van Hove, G., & Lievens, F. (2009). Tapping the grapevine: A closer look at word-of-mouth as a recruitment source. *Journal of Applied Psychology*, 94(2), 341-352.
- Vonk, R. (1993). The negativity effect in trait ratings and in open-ended descriptions of persons. *Personality and Social Psychology Bulletin*, 19(3), 269-278.
- Wang, X., Teo, H. H., & Wei, K. K. (2007). The acceptance of product recommendation from Web-based word-of-mouth systems: Effects of information, informant, and system characteristics. *Proceedings of the International Conference on Information Systems*.
- Weathers, D., Sharma, S., & Wood, S. L. (2007). Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods. *Journal of Retailing*, 83(4), 393-401.
- ZDNet. (2008). *26% of online customers go to Amazon reviews before buying*. Retrieved from <http://www.zdnet.com/article/26-of-online-customers-go-to-amazon-reviews-before-buying/>
- Zhang, J. Q., Craciun, G., & 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, R. C., & Tran, T. (2009). Helping e-commerce consumers make good purchase decisions: A user reviews-based approach. *E-Technologies-Innovation in an Open World*, 26, 1-11.
- Zhang, X. M., & Zhu, F. (2010). Group size and incentive to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review*, 101(4), 1601-1615.
- Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74(2), 133-148.

About the Authors

Kevin K.Y. KUAN is a senior lecturer in the School of Information Technologies at the University of Sydney. He received his PhD in business administration from the Ross School of Business at the University of Michigan, Ann Arbor. His research interests include online consumer review and electronic word-of-mouth, information presentation and online decision making, and human-computer interaction. His work has appeared in journals such as *Journal of the Association of Information Systems*, *Communications of the Association of Information Systems*, *European Journal of Information Systems*, *IEEE Transactions on Engineering Management, Information & Management*, and *Journal of Management Information Systems*.

Kai-Lung HUI is a professor in the Department of Information Systems, Business Statistics, and Operations Management at the Hong Kong University of Science and Technology. His research interests include information privacy and security, information technology policies, and electronic commerce. His research has been published in scholarly journals including *Management Science*, *MIS Quarterly*, and *Journal of Management Information Systems*, among others. He has provided consulting services to the Ministry of Law of Singapore, the World Intellectual Property Organization (WIPO), WKK Distributions Ltd., and the Intellectual Property Department of the Hong Kong SAR Government. He obtained his BBA and PhD degrees from the Hong Kong University of Science and Technology.

Pattarawan PRASARNPHANICH is a faculty member at Sasin Graduate Institute of Business Administration of Chulalongkorn University and a Certified Information Systems Auditor (CISA). Prior to joining Sasin, she was Assistant Professor of Information Systems at the City University of Hong Kong. She received the PhD degree in Management Information Systems from the University of Memphis, the M.B.A. (Decision Sciences) from Virginia Commonwealth University, and the B.S. in Statistics from Chulalongkorn University. Her research interests include IT Governance, Information System Strategy, Electronic Commerce, Knowledge Management, and Collaboration Technology. She has published research articles in journals such as *Decision Sciences*, *Communications of the ACM*, *IEEE Transactions on Engineering Management*, and *Journal of Computer Information Systems*.

Hok-Yin LAI is a Lecturer in the Department of Computer Science at the Hong Kong Baptist University. Her research interests covers the areas of social computing, decision making in Finance and e-learning system design. She has published articles and book chapters including journals such as *Expert Systems with Applications*, *The International Journal of Intelligent Information Technologies*, and *The Journal of Studies in Education*. She obtained her bachelor degree from the Hong Kong Polytechnic University, her master degrees from the Hong Kong Polytechnic University and the Chinese University of Hong Kong, and her PhD degree from the City University of Hong Kong.