Title	Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth
Author(s)	Chua, Alton Yeow Kuan; Banerjee, Snehasish
Citation	Chua, A. Y. K., & Banerjee, S. (2014). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. Journal of the Association for Information Science and Technology, 66(2), 354-362.
Date	2014
URL	http://hdl.handle.net/10220/38693
Rights	© 2014 ASIS&T. This is the author created version of a work that has been peer reviewed and accepted for publication by Journal of the Association for Information Science and Technology, ASIS&T. It incorporates referee’s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [http://dx.doi.org/10.1002/asi.23180].

Chua, A. Y. K, & Banerjee, S. (2015). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. Journal of the Association for Information Science and Technology, 66(2), 354-362

Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth

Alton Y.K. Chua and Snehasish Banerjee

Abstract

This paper examines review helpfulness as a function of reviewer reputation, review rating and review depth. Drawing data from the popular review platform Amazon, results indicate that review helpfulness is positively related to reviewer profile and review depth, but is negatively related to review rating. Users seem to have proclivity for reviews contributed by reviewers with positive track record. They also appreciate reviews with lambasting comments and those with adequate depth. By highlighting its implications for theory and practice, the paper concludes with a few limitations and areas for further research.

Introduction

The ubiquity of reviews for products and services on the Internet has become one of the most valuable sources of information to assist users in making purchase decisions (Dellarocas, 2003; Miao, Li & Zeng, 2010). Commonly found in websites such as Amazon, Tripadvisor and Epinions (henceforth, known as review platforms), reviews are a means for users to share their post-purchase experiences with other potential buyers. These are often perceived as being more authentic and credible than marketer-generated information in the printed press (Jiang & Benbasat, 2007; Senecal & Nantel, 2004). However, given that a single product or service can attract a large volume of reviews, it remains a challenge for users to distinguish helpful reviews from those that were either frivolous or biased.

To ease the task of sieving the grain from the chaff, most review platforms incorporate features to help users locate superior reviews, which in turn, can aid more informed decision making. In particular, Amazon employs the feature of review helpfulness to encourage peer-evaluation of reviews. It presents the question “Was this review helpful to you?” at the end of each submitted review. Users can respond to the question with either a “Yes” or a “No”. They may even choose to report abuse, or leave behind comments. Placing the most helpful reviews conspicuously on the products’ information page, Amazon provides helpfulness information alongside each review in the form of annotations such as “x of y people found the following review helpful”. Recent literature suggests that votes cast on helpfulness of reviews play a significant role in influencing users’ purchase decisions (Ghose & Ipeirotis, 2011; Korfiatis, García-Bariocanal & Sánchez-Alonso, 2012; Mudambi and Schuff 2010). Hence, a better understanding of factors that shape review helpfulness can be a significant research endeavor.

An interesting research perspective is to determine the extent to which helpfulness of a given review varies with reputation of the reviewer who contributed that review. Prior studies have shown that contents submitted by reviewers with positive track record tend to be viewed as being helpful (Wathen & Burkell, 2002). Besides, users’ perceptions of review helpfulness can be affected by factors such as review rating and review depth. Review rating is the numerical valence of a review assigned by the reviewer while review depth refers to the length of a review. Specifically, reviews with extreme positive or negative ratings are often considered more informative, and thus perceived as being helpful (Forman, Ghose & Wiesenfeld, 2008). Furthermore, reviews that are lengthy and laden with robust explanations could inspire confidence, and hence perceived to be helpful (Metzger, 2007). Moreover, review depth could moderate the ways review helpfulness varies across reviewer reputation and review rating (Stringam & Gerdes, 2010). This paper thus seeks to investigate the interplay among reviewer reputation, review rating and review depth to shed light on review helpfulness. The dataset was drawn from Amazon, a popular review platform which allows for the investigation of review helpfulness.

This paper has implications for both theory and practice. On the theoretical front, it provides a better understanding of factors that can contribute to users' perceptions of helpfulness in the context of reviews. It extends prior research by relating review helpfulness with reviewer reputation, review rating and review depth. On the practical front, users may lean on the findings to better understand the characteristics of helpful reviews. This in turn may help users contribute reviews of better quality as well as locate those that could be generally superior. Businesses may also tap into such reviews to keep a pulse on users' preferences and complaints towards specific products or services.

The remainder of the paper is organized as follows. The following section gives an overview of the literature on technology-mediated social participation and introduces the concept of review helpfulness. Next, related work on review helpfulness with respect to reviewer reputation, review rating and review depth are discussed. Then, the Methodology section explains the data collection and analysis procedures. Following that, statistical analyses are presented and explained in the Results section. Four major insights drawn from the results are highlighted in the Discussion section. Finally, the paper concludes with implications for theory and practice.

Literature Review

The emergence of Web 2.0 has shifted the role of Internet from a mere medium for information transfer to an amalgamation of platforms for technology-mediated social participation. For example, the video sharing platform YouTube, which boasts 10 billion video views per month, has become the default choice to post personal videos, product advertisements or political messages (Preece & Shneiderman, 2009). Social networking platforms such as Facebook has over 600 million users who upload 2.5 billion photos per month along with comments, photos, videos and status updates (DuBois, Golbeck & Srinivasan, 2011). Similarly, other Web 2.0 applications such as recommendation systems, social tagging systems, micro-blogging sites and review platforms have immensely grown in popularity (Preece & Shneiderman, 2009; Trant, Bearman & Chun, 2007).

Conceivably, proliferation of technology-mediated social participation has also piqued substantial scholarly interest. This gives rise to a number of research themes including usability, sentiment analysis and trust. Usability refers to the simplicity of a system, which affects the ease, flexibility and speed with which it can be used (Flavián, Guinalíu & Gurrea, 2006; Nielsen, 1994). Usability of platforms that support technology mediated social participation is regarded as an important predictor of users' satisfaction and loyalty (Casaló, Flavián & Guinalíu, 2008). Sentiment analysis deals with the inference of polarity in opinions about specific entities as reflected through technology-mediated social participation (Abbasi, Chen & Salem, 2008; Feldman, 2013). Its significance stems from the fact that opinions of online communities can often play a vital role in shaping individual user's decision making (Zhou, He & Wang, 2008). Furthermore, given the abundance of content created through technology-mediated social participation, trust in Web 2.0 has become an interesting area of research (Chen & Dhillion, 2003; Gefen, Benbasat & Pavlou, 2008; Jeacle & Carter, 2011). Trustworthiness of content allows users to sort and filter information, receive recommendations and make better informed decisions (DuBois, Golbeck & Srinivasan, 2011).

Another emerging research theme focuses on social navigation. In gist, social navigation deals with ways the problem of information overload can be mitigated in part by allowing users to evaluate the quality of entries, which in turn, can be used to rank order the submissions (Otterbacher, 2009). In particular, a single product can often attract a huge number of reviews in highly popular review platforms. Hence, social navigation feature such as review helpfulness which draws on the wisdom of crowds can be used to accentuate the more helpful reviews. Given that helpfulness of a review is taken as a proxy of the perceived value it offers to users in decision making (Korfiatis, García-Bariocanal & Sánchez-Alonso, 2012), a better understanding of the factors affecting review helpfulness not only represents a burgeoning research agenda but also holds far-reaching commercial implications.

Related Work

Related work suggests that review helpfulness could be shaped by the collective interplay among reviewer reputation, review rating and review depth. The ways in which review helpfulness can vary across the three constructs are discussed as follows.

Reviewer Reputation

Reviewer reputation refers to the identity-descriptive information displayed on review platforms for users who have contributed reviews. Such information typically includes user names, summaries of past contributions, ranks and special badges such as top-50 reviewer or top-100 reviewer (Ghose & Ipeirotis, 2011). Though reviewers contribute substantial time and energy to write reviews (Wang, 2010), they may differ widely in their motivation. While some reviewers can be inclined to show their bravado in evaluating products for peer recognition, others may have the proclivity to write untruthful reviews with malicious intentions of hoodwinking others.

In social psychology literature, message source characteristics have long been found to influence readers' judgment, behaviour and perception (Chaiken, 1980). In fact, authority of the source has been shown to largely affect the perceived credibility of messages (Wathen & Burkell, 2002). Likewise, prior research on reviews argues that users' responses towards a given review could be shaped in part by its contributor's reputation (Pavlou & Dimoka, 2006). In fact, reviews contributed by influential reviewers can have a significant impact on the sales of products or services (Forman, Ghose & Wiesenfeld, 2008). To gauge the past track record of a reviewer, users may choose to examine the proportion of helpful votes the reviewer has hitherto garnered. More eclectic users may even venture to read all past reviews submitted to better appraise the reviewer's ability to write helpful reviews (Ghose & Ipeirotis, 2011). Thus, the following is hypothesized:

H1: There exists a positive relationship between reviewer reputation and review helpfulness.

Review Rating

In most review platforms, reviewers are required to rate their experiences of products or services using a single indicator to reflect the overall valence of their reviews (Wu, Heijden & Korfiatis, 2011). Such review ratings are meant to numerically summarize the entire content of review texts (Chevalier & Mayzlin, 2006). They usually range from one-star which denotes extreme disapproval to five-star which indicates utmost appreciation. Ratings tend to capture immediate attention because they are displayed conspicuously at the beginning of entries in most review platforms.

On a five-point star scale, reviewers may express their neutrality through three-star ratings. However, reviews with such moderate ratings are often perceived less helpful than those with extreme ratings (Forman, Ghose & Wiesenfeld, 2008; Pavlou & Dimoka, 2006). After all, reviews that put forth strong arguments either in favor of or against a product or service help users confirm or eliminate alternatives. Extreme reviews are therefore more helpful for users in making purchase decisions (Korfiatis, García-Bariocanal & Sánchez-Alonso, 2012). However, there exists another school of thought which considers reviews with moderate ratings as being more helpful than those with extreme ratings (Crowley & Hoyer, 1994; Eisend, 2006). This is because moderate reviews tend to present both the pros and the cons of products or services (Connors, Mudambi & Schuff, 2011). In this way, users can exercise their own judgement in making informed decisions. Hence, extreme reviews with one-star or five-star ratings tend to influence review helpfulness differently vis-à-vis moderate reviews. Thus, the following is hypothesized:

H2: There exists a curvilinear relationship between review rating and review helpfulness.

Review Depth

Review depth is a measure of the amount of open-ended textual contents that reviewers provide to justify the review ratings (Mudambi & Schuff, 2010). When the length of the information expressed in the text matches users' expectations and information processing strategies, a cognitive fit occurs (Vessey & Galletta, 1991). Such a fit in turn promotes users' perceptions of helpfulness, resulting in more informed decision making (Nah et al., 2010). In general, greater review depth tends to enhance the perceived value a review offers to users in decision making (Jiang & Benbasat, 2007; Metzger, 2007). This is especially applicable if users obtain the information free of charge (Johnson & Payne, 1985). Thus, the following is hypothesized:

H3: There exists a positive relationship between review depth and review helpfulness.

Undoubtedly, reviews are generally considered to be helpful if they are of sufficient depth. Concurrently, reviews contributed by reviewers with positive track record are perceived as helpful (Pavlou & Dimoka 2006). However, a given review that lacks depth is unlikely to be perceived as helpful even if it has been contributed by a reputed reviewer. Review depth thus appears to have an impact on the way reviewer reputation affects review helpfulness.

Furthermore, it is conceivable that reviews with extreme ratings need not be as lengthy as those with moderate ratings (Korfiatis, García-Bariocanal & Sánchez-Alonso, 2012). This is because the former only needs to contain either positive or negative arguments, while the latter is expected to shed light on both sides of the coin (Chevalier & Mayzlin, 2006; Stringam & Gerdes, 2010). A given moderate review that lacks depth is unlikely to be perceived as helpful due to its inadequacy in explicating both pros and cons. Review depth thus also appears to have an impact on the way review rating affects review helpfulness. Thus, the following are hypothesized:

H4a: Review depth moderates the relationship between reviewer reputation and review helpfulness.

H4b: Review depth moderates the relationship between review rating and review helpfulness.

Methodology

Choice of Dataset

While review platforms are aplenty on the Internet, review helpfulness has yet to become a commonly-adopted feature in these sites. Even where available, its form and format have been largely unstandardized. For example, review helpfulness in Epinions is expressed using annotations such as "Rated a Very Helpful Review by the Epinions community" and "Rated a Somewhat Helpful Review by the Epinions community". However, in the absence of any other numerical value, it is hard to assess the extent to which the review helpfulness is reflective of the online community. In review platforms such as Tripadvisor and Yelp, review helpfulness is conceived as the number of users who voted a review as being helpful or useful.

Nonetheless, without any indication of the number of users who would have otherwise voted the review as being unhelpful or not useful, the significance of review helpfulness in these review platforms is limited. In contrast, other review platforms such as IMDb and Amazon not only report the number of users who found a review helpful but also the total number of users who evaluated the review. Conceivably, such platforms appear to offer a more meaningful measure of review helpfulness.

Among the various review platforms, Amazon was deemed as the appropriate choice because it is not only a pioneering site that supports peer-evaluation of reviews (Forman, Ghose & Wiesenfeld, 2008), but also thrives well hitherto. Deemed almost as a de facto standard of review platforms, its longevity and popularity allow for a substantial scope of data collection. Moreover, review helpfulness in Amazon has attracted considerable scholarly attention in recent years (eg. Ghose & Ipeirotis, 2011; Korfiatis, García-Bariocanal & Sánchez-Alonso, 2012; Mudambi and Schuff, 2010). This study thus dovetails past efforts to better understand the helpfulness of reviews contributed in Amazon.

Specifically, reviews on books were drawn from Amazon for analysis. This is because books represent a classic case of experience goods, the utility of which can be judged only after consumption (Nelson, 1970). The typical nature of experience goods makes users more reliant on experiences of their cohorts (Wu, Van der Heijden & Korfiatis, 2011). Thus, reviews on books offer an appropriate context to investigate users' perceptions of review helpfulness.

Data Collection

The data collection was a three-step process. First, a set of 1,000 best seller books indicated in Amazon as of October, 2012 were identified. Specifically, books were selected to include 10 popular genres such as science fiction, computers and technology, and education and reference. Amazon offers a list of 100 best seller books for each genre. Hence, the first step of data collection yielded an initial pool of 1,000 best seller books.

Next, from this initial pool of books, those that had attracted less than 30 or more than 100 reviews were eliminated. Books that did not meet the lower threshold meant that they could either be recently published or comprise those that rarely attract reviews. The small volume of reviews for such books might not afford meaningful analyses. Conversely, books that exceeded the upper threshold were also avoided because helpfulness of reviews for overly popular books could be skewed by bandwagon effect (Mudambi & Schuff, 2011), which in turn might confound the findings. Furthermore, books offered at special offers or discounts were excluded to control for price effects (Pavlou & Dimoka, 2006). Out of the filtered set of some 250 best seller books, a total of 150 were selected randomly. Each selected book could be identified by its unique ISBN number. Care was also taken to ensure that there were no duplicates in the dataset such as instances of hard cover and paperback for the same book (Wu, van der Heijden & Korfiatis, 2011).

In the third step, all reviews posted against these 150 selected books were collected using a web scraper. Of the 7,897 reviews collected, 1,036 did not attract any votes. Another 1,272 reviews were posted by reviewers who were either anonymous or had not received any votes for their contributions. These reviews, which represent incomplete data points, were therefore eliminated from the dataset. The resultant 5,589 reviews were admitted for analysis.

For each review, the following data items were obtained: review rating, review title, review date, review text, number of helpful votes, and number of total votes attracted by the review. In addition, information about the reviewer, including reviewer ID, number of helpful votes, and number of total votes received by the reviewer from all previously contributed reviews, were also retrieved. Figure 1 shows a sample review posted in Amazon. The Reviewer ID has been concealed for the sake of privacy.

Figure 1: A sample review posted in Amazon

Operationalization

For the purpose of this paper, reviewer reputation (REP) is operationalized using the summary of reviewers' past contributions (Ghose & Ipeirotis, 2011). It is quantified as the ratio of the number of helpful votes to the number of total votes received by the reviews contributed by a reviewer. Review rating (RAT) is defined as the star rating of a review provided by the reviewer on a 1-5 scale (Korfiatis, García-Bariocanal & Sánchez-Alonso, 2012) while review depth (DEP) is derived from the number of words in the review text (Mudambi and Schuff, 2010).

The dependent variable review helpfulness (HEL) is operationalized as the proportion of users who found a review to be helpful (Forman, Ghose & Wiesenfeld, 2008; Korfiatis, García-Bariocanal & Sánchez-Alonso, 2012; Mudambi and Schuff, 2010). It is computed as the ratio of the number of helpful votes to the number of total votes that the review attracted. However, such a ratio can be easily skewed for smaller sample sizes. For example, reviews for which Amazon reports "5 of 10 people found the following review helpful" will become numerically identical with those for which Amazon indicates "50 of 100 people found the following review helpful". To mitigate such a confounding effect, the total number of votes provided in evaluation of a review (TOT) is taken as a control variable for the analysis (Mudambi

& Schuff, 2010). The hypotheses and the operationalization of variables are shown in Figure 2.

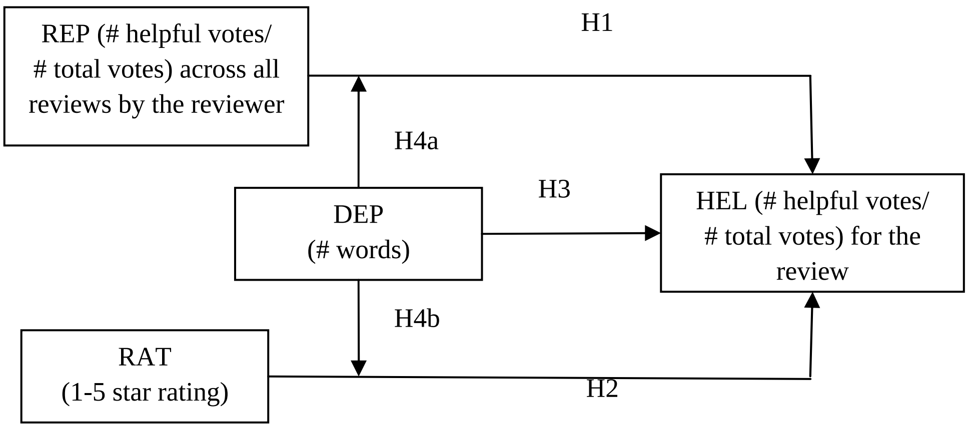

Figure 2: Hypotheses and operationalization of variables

Data Analysis

Multiple regression was used to test the hypotheses. Specifically, Tobit estimation was preferred over ordinary least squares (OLS) estimation due to two reasons. First, the former is suitable when the dataset exhibits censored nature. For example, the dependent variable review helpfulness is bounded in its range. Users can only provide dichotomous evaluations as helpful or unhelpful. They cannot vote reviews as better than helpful or worse than unhelpful (Mudambi & Schuff, 2010).

Second, Tobit estimation can help minimize potential selection bias inherent in the dataset. For example, Amazon indicates only the number of total votes on a review and how many of those were helpful. However, not all users who read the review voted on its helpfulness. This gives rise to a potential selection bias. Thus, Tobit estimation is preferred over OLS estimation which tends to be biased (Kennedy, 1994).

Results

The descriptive statistics for the dataset is summarized in Table 1. The control variable, total number of helpfulness votes provided in evaluation of a review, ranged from 1 to 360. The high mean for review rating is expected as users generally contribute reviews to express satisfaction (Hu, Zhang & Pavlou, 2009). The moderating variable, review depth, had a median of 75 words. A median split was used to label reviews as those with high and low values. Specifically, there were 2,805 (2,784) reviews with 75 words or less (greater than 75 words).

Chua, A. Y. K, & Banerjee, S. (2015). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. Journal of the Association for Information Science and Technology, 66(2), 354-362

Table 1: Descriptive statistics of dataset

	Full Sample (N = 5,589)		Low DEP (N_{low} = 2,805)		High DEP (N_{high} = 2,784)	
Variables	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
TOT	7.39	18.48	3.92	9.66	10.88	23.82
REP	0.75	0.23	0.68	0.32	0.82	0.18
RAT	4.41	1.01	4.53	0.90	4.31	1.10
DEP	119.15	142.14	42.18	16.74	196.70	168.19
HEL	0.59	0.43	0.48	0.45	0.69	0.38

Using Tobit estimation, a hierarchical regression analysis was conducted. The variables were entered into the regression model according to the assumed causal order. The control variable entered in the first block. This was followed by the variables for main effects (H1, H2, H3) in the second block, and those for moderating effects (H4a, H4b) in the third block. This resulted in the three regression models. As shown in Table 2, the first model with the control variable could only explain 2.53 % of the variance in helpfulness. The second model with the main effects could additionally explain more 17.37 % of the variance, while the interaction terms in model 3 could further explain 2.17 % of the variance.

Table 2: Regression results for helpfulness (N = 5,589)

	Model 1		Model 2		Model 3	
	Coeff	**SE**	**Coeff**	**SE**	**Coeff**	**SE**
TOT	0.00***	0.00	0.00*	0.00	0.00	0.00
REP			1.94***	0.07	1.78***	0.07
RAT			-0.05	0.03	-0.13*	0.04
RAT^2			0.00	0.01	-0.01	0.01
DEP			0.01***	0.00	0.01*	0.00
DEP x REP					0.28***	0.05
DEP x RAT					0.16*	0.06
DEP x RAT^2					0.03	0.02
log likelihood	-6066.26 (df: 1)		-5533.24 (df: 5)		-5501.28 (df: 8)	
Δ R^2 (%)	2.53		17.37		2.17	
Total R^2 (%)	2.53		19.90		22.07	

* $p < 0.05$, *** $p < 0.001$

On the basis of the results, three observations could be made. First, there existed significant positive relationship between reviewer reputation and review helpfulness ($\beta = 1.94$, $p < 0.001$), lending support for H1. Reviews contributed by reviewers with strong past track records were generally perceived as more helpful than those posted by newbies.

Second, review rating had no curvilinear relationship with review helpfulness. Hence, H2 was not supported. However, there was a significant negative linear relationship between rating and helpfulness ($\beta = -0.13$, $p < 0.05$), suggesting that negative reviews are generally perceived more helpful.

Third, there was significant positive relationship between review depth and review helpfulness ($\beta = 0.01$, $p < 0.05$), lending support for H3. Reviews with adequate depth were considered more helpful than shorter reviews. Review depth also moderated the relationship between reviewer reputation and review helpfulness (β =

0.28, $p < 0.001$). So, H4a was supported. Review depth also moderated the relationship between review rating and review helpfulness ($\beta = 0.16$, $p < 0.05$), lending support for H4b. However, the moderating effect was significant for the linear relationship and not for the curvilinear relationship.

As a robustness check, separate regression analysis was conducted for reviews with low depth (Nlow = 2,805) and those with high depth (Nhigh = 2,784) to investigate the effect of reviewer reputation and review rating on review helpfulness. For this purpose, the control variable was entered in the first block, followed by the variables for main effects in the second block, resulting in two regression models. As shown in Table 3, results indicate that the relationship between reviewer reputation and review helpfulness was statistically significant for reviews with low depth ($\beta = 2.21$, $p < 0.001$)) as well as for those with high depth ($\beta = 1.97$, $p < 0.001$), albeit stronger for the former. The negative linear relationship between rating and helpfulness was statistically significant only for reviews with low depth ($\beta = -0.18$, $p < 0.05$). For both groups of reviews, the curvilinear relationship between review rating and review helpfulness was not statistically significant.

Table 3: Regression results for helpfulness for reviews with low and high depth

	Low DEP (N_{low} = 2,805)				**High DEP (N_{high} = 2,784)**			
	Model 1		**Model 2**		**Model 1**		**Model 2**	
	Coeff	**SE**	**Coeff**	**SE**	**Coeff**	**SE**	**Coeff**	**SE**
TOT	0.01*	0.00	0.02***	0.00	0.00	0.00	0.00	0.00
REP			2.21***	0.12			1.97***	0.09
RAT			-0.18*	0.06			0.02	0.03
RAT^2			-0.04	0.03			0.02	0.01
log likelihood	-3006.23 (df: 1)		-2783.08 (df: 4)		-2818.95 (df: 1)		-2571.45 (df: 4)	
ΔR^2 (%)	1.22		14.91		2.09		15.56	
Total R^2 (%)	1.22		16.13		2.09		17.65	

* p < 0.05, *** p < 0.001

Discussion

Arising from the results, four major insights could be drawn. First, consistent with prior research (Chaiken, 1980; Pavlou & Dimoka, 2006; Wathen & Burkell, 2002), reviewer reputation showed significant positive relationship with review helpfulness. Modern tech-savvy users seem to be conscious of whom to believe. Given that some reviews could be frivolously posted with malicious intentions, users appear to look for reviews contributed by prolific reviewers who apparently devote substantial time and energy to contribute reviews (Wang, 2010). Users could even have thoroughly examined the reputation of reviewers before even looking at their reviews (Ghose & Ipeirotis, 2011). Moreover, being spoilt for choices in terms of review availability, they perhaps only seek reviews that have been contributed by reviewers with positive track records. On the other hand, reviews contributed by newbies remain largely unseen, and hence, predominantly unevaluated. This represents a typical case of online hegemony where the rich seems to get richer, while the poor becomes relatively poorer (Cao, Duan & Gan, 2011; Xiong, Jiang & Wang, 2012).

Second, there was a significant negative relationship between review rating and review helpfulness, especially for reviews with low depth. Prior research on reviews has consistently documented significant relationship between rating and helpfulness. However, the nature of the relationship has not been consistent. For example, some studies found negative relationship suggesting that critical reviews are perceived as helpful (Chatterjee, 2001; Wu, Heijden & Korfiatis, 2011). Some other studies found curvilinear relationship with higher helpfulness for extreme reviews (Chevalier & Mayzlin, 2006; Forman, Ghose & Wiesenfeld, 2008). Yet other studies found curvilinear relationship with higher helpfulness for moderate reviews (Schlosser, 2005; Connors, Mudambi & Schuff, 2011). To augment these studies, this paper shows that for reviews with low depth, users tend to favor negative reviews. For reviews with high depth, no relationship between rating and helpfulness suggests that users are increasingly becoming skeptical. They perhaps do not always treat review rating as a reliable proxy to judge review helpfulness.

Third, consistent with prior research (Jiang & Benbasat, 2007; Metzger, 2007; Mudambi & Schuff, 2010), review depth turned out to be a useful proxy for users' perceptions of review helpfulness. Reviews with substantial depth command a sense of adequacy and competence. They also appear to match users' expectations and information processing strategies (Metzger, 2007; Wang, 2010). This in turn results in cognitive fit (Vessey & Galletta, 1991), proliferating review helpfulness (Nah et al., 2010). Interestingly however, results in Table 3 suggest that the positive relationship between reviewer reputation and review helpfulness was slightly stronger for reviews with low depth than those with high depth. Users seem to be put off by reviews that are too long. Perhaps, there is an optimum length of reviews that users find helpful (Otterbacher, 2009). If a review is overly detailed, users may be reluctant to go through it. On the other hand, a sketchy review can be too trivial and simplistic for users to appreciate.

Fourth, even though statistically significant relationship was found for reviewer reputation, review rating and review depth on review helpfulness, the R2 values (as indicated in Tables 2 and 3) were relatively low. Low R2 values indicate that the regression model with these variables do not result in a good fit with the data. Even so, such low R2 values are common in research on reviews (eg. Chevalier & Mayzlin, 2006; Forman, Ghose & Wiesenfeld, 2008; Otterbacher, 2009; Wu, Heijden & Korfiatis, 2011). Perhaps, this could be vestige of the overall complexity of the online environment in review platforms. Prediction of review helpfulness could be affected by numerous extraneous factors that lie outside the scope of such research. For example, review helpfulness could be affected by the demographics of users who browse the reviews (Ip, Lee & Law, 2012). It could also be dependent on users' motivation and subjective perception to vote for or against reviews. Furthermore, as with reviewers who are not unanimously honest in contributing reviews (Wang, 2010), not all users are veracious in evaluating the helpfulness of reviews (Kornish, 2009).

Conclusion

This paper examines review helpfulness as a function of reviewer reputation, review rating and review depth. Drawing data from the popular review platform Amazon, results indicate that review helpfulness is positively related to reviewer reputation and review depth, but is negatively related to review rating. Users seem to have proclivity for reviews contributed by reviewers with positive track record. They also appreciate reviews with lambasting comments and those with adequate depth.

The findings of the paper offer implications for both theory and practice. On the theoretical front, it builds on prior literature by providing a conceptualization of factors that contribute to users' perceptions of helpfulness in the context of reviews. It extends extant literature by relating review helpfulness with reviewer reputation, review rating and review depth. The findings suggest that reviews contributed by those with positive track records are generally perceived helpful. Furthermore, short reviews tend to be favored if they are negative but longer reviews do not exhibit any relationship between review rating and review helpfulness. Also, users appear to appreciate reviews that are neither too short nor too long. The paper also suggests a possible existence of online hegemony in review platforms. Further research along these themes is needed to verify and validate these findings.

On the practical front, the paper provides implications for users, businesses and website developers. First, the findings of this paper may help users write better reviews. Users should strike a balance in terms of review length to ensure that reviews are neither sketchy nor overly detailed. Users could also lean on the findings of this paper to conjecture which reviews are likely to be helpful for making informed decisions. Second, businesses may also tap into helpful reviews to keep a pulse on users' preferences and complaints towards specific products or services. This in turn could lead to a win-win situation where both users and businesses are benefitted from helpful reviews. Third, developers of review platforms might want to include information of users who evaluate the helpfulness of a given review. This will assist users to examine the reputation of those who vote for or against reviews, a feature that is not available in the state-of-the-art websites.

Three limitations inherent in this paper must be acknowledged. First, the results presented in this paper hold true for books only. Caution should be exercised when generalizing the findings to other types of products and services. Second, the variables reviewer reputation, review rating, review depth and review helpfulness were quantitative surrogates and not direct measures. Though it allowed for an objective and data-driven approach, the lack of qualitative approach might have impaired data richness. Finally, the paper measured helpfulness of reviews without taking into account the role of helpful entries in influencing users' purchase decisions.

Going forward, a number of future research directions can be identified. One possible area of investigation would be to sample a different range of products or services from multiple review websites in order to validate if the results from this paper hold. Different brands of the same product category might be used to analyze the relationship between perceptions of brand and helpfulness. Another direction would be to examine review helpfulness using qualitative approaches. A qualitative analysis of the review content with multiple coders could provide a deeper understanding of what makes a review helpful. Moreover, it could be interesting to determine whether users are merely satisfied by reading helpful reviews or are significantly influenced in making purchase decisions by such comments. Yet another possible research direction might involve investigating usability elements and the role of trust in the helpfulness of reviews. With the inclusion of users' perspectives into the study of review helpfulness, the theoretical boundaries of the paper can be extended.

References

Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. ACM Transactions on Information Systems, 26(3), 12-46.

Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. Decision Support Systems, 50(2), 511-521.

Casaló, L. V., Flavián, C., & Guinalíu, M. (2008). The role of satisfaction and website usability in developing customer loyalty and positive word-of-mouth in the e-banking services. International Journal of Bank Marketing, 26(6), 399-417.

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.

Chatterjee, P. (2001). Online Reviews – Do Consumers Use Them? In M. C. Gilly, & J. Myers-Levy (Eds.). Proceedings of Advances in Consumer Research, 129-134.

Chen, S. C. & Dhillon, G. S. (2003). Interpreting dimensions of consumer trust in e-commerce. Information Technology and Management, 4, 303-318.

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.

Connors, L., Mudambi, S.M., & Schuff, D. (2011). Is it the review or the reviewer? A multi-method approach to determine the antecedents of online review helpfulness. Hawaii International Conference on System Sciences (pp. 1-10). IEEE.

Crowley, A.E., & Hoyer, W.D. (1994). An integrative framework for understanding two sided persuasion. Journal of Consumer Research, 20(4), 561-574.

Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., & Lee, L. (2009, April). How opinions are received by online communities: a case study on Amazon.com helpfulness votes. Proceedings of the International Conference on World Wide Web (pp. 141-150). New York: ACM.

Dellarocas, C. (2003). The digitization of word of mouth: promise and challenges of online feedback mechanisms. Management Science, 49(10), 1407-1424.

DuBois, T., Golbeck, J., & Srinivasan, A. (2011). Predicting trust and distrust in social networks. IEEE International Conference on Social Computing: Privacy, Security, Risk and Trust, 418–424.

Eisend, M. (2006). Two-sided advertising: a meta-analysis. International Journal of Research in Marketing, 23(2), 187-198.

Feldman, J.M., & Lynch, J.G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behaviour. Journal of Applied Psychology, 73(3), 421-435.

Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82–89.

Flavián, C., Guinalíu, M. and Gurrea, R. (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty. Information & Management, 43 (1), 1-14.

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, 2008, 291–313.

Gefen, D., Benbasat, I. & Pavlou, P. A. (2008). A research agenda for trust in online environments. Journal of Management Information Systems, 24(4), 275-286.

Ghose, A., & Ipeirotis, 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.

Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. Communications of the ACM, 52(10), 144-147.

Ip, C., Lee, H. A., & Law, R. (2012). Profiling the users of travel websites for planning and online experience sharing. Journal of Hospitality & Tourism Research, 36(3), 418-426.

Jeacle, I. & Carter, C. (2011). In TripAdvisor we trust: rankings, calculative regimes and abstract systems. Accounting, Organizations and Society, 36(4-5), 293-309.

Jiang, Z., & Benbasat, I. (2007). Investigating the influence of the functional mechanisms of online product presentations. Information Systems Research, 18(4), 454-470.

Johnson, E., & Payne, J. (1985). Effort and Accuracy in Choice. Management Science, 31(4), 395-415.

Kelman, H.C. (1958). Compliance, identification, and internalization: three processes of attitude change. The Journal of Conflict Resolution, 2(1), 51-60.

Kennedy, P. (1994). A guide to econometrics. Oxford, England: Blackwell Publishers.

Korfiatis, N., García-Bariocanal, E., & Sánchez-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.

Kornish, L. J. (2009). Are user reviews systematically manipulated? Evidence from helpfulness ratings. Working paper, Leeds School of Business, Boulder, Colorado.

Metzger, M.J. (2007). Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. Journal of the American Society for Information Science and Technology, 58(13), 2078-2091.

Miao, Q., Li, Q., & Zeng, D. (2010), Fine-grained opinion mining by integrating multiple review sources. Journal of the American Society for Information Science and Technology, 61(11), 2288–2299.

Mudambi, S.M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon. MIS Quarterly, 34(1), 185-200.

Nah, F. F. H., Hong, W., Chen, L., & Lee, H. H. (2010). Information search patterns in e-commerce product comparison services. Journal of Database Management, 21(2), 26-40.

Nielsen, J. (1994). Usability Engineering. San Francisco: Morgan Kaufmann.

Otterbacher, J. (2009). 'Helpfulness' in online communities: A measure of message quality. Proceedings of the International Conference on Human factors in Computing Systems (pp. 955-964). ACM.

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.

Preece, J., & Shneiderman, B. (2009). The reader-to-leader framework: Motivating technology-mediated social participation. AIS Transactions on Human-Computer Interaction, 1(1), 13–32.

Schlosser, A. (2005). Source perceptions and the persuasiveness of Internet word-of-mouth communication. In G. Menon & A. Rao (Eds.). Proceedings of Advances in Consumer Research, 202-203.

Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. Journal of Retailing, 80(2), 159-169.

Stringam, B.B., & Gerdes, J. (2010). An analysis of word-of-mouse ratings and guest comments of online hotel distribution sites. Journal of Hospitality Marketing & Management, 19(7), 773-796.

Trant, J., Bearman, D., & Chun, S. (2007). The eye of the beholder: steve.museum and social tagging of museum collections. Proceedings of International Cultural Heritage Informatics Meeting, Toronto, Ontario.

Vessey, I., & Galletta, D. (1991). Cognitive fit: an empirical study of information acquisition. Information Systems Research, 2(1), 63-84.

Wang, Z. (2010). Anonymity, social image, and the competition for volunteers: a case study of the online market for reviews. The BE Journal of Economic Analysis & Policy, 10(1), 1-34.

Wathen, C.N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the web. Journal of the American Society for Information Science and Technology, 53(2), 134-144.

Wu, P.F., van der Heijden, H., & Korfiatis, N. (2011, August). The influences of negativity and review quality on the helpfulness of online Reviews. Proceedings of the International Conference on Information Systems, Shanghai, China, 1-10.

Xiong, Z., Jiang, W., & Wang, G. (2012). Evaluating user community influence in online social networks. IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 640-647.

Zhou, L. Z., He, Y. K., & Wang, J. Y. (2008). Survey on research of sentiment analysis. Journal of Computer Applications, 28(11), 2725-2728.