Developing a Theory of Diagnosticity for Online Reviews

Alton Y. K. Chua, and Snehasish Banerjee

Abstract—**Diagnosticity of a given online review is defined as the extent to which it helps users make informed purchase decisions. Users' perception of review diagnosticity can be associated with five factors, namely, review rating, review depth, review readability, reviewer profile and product type. Review rating refers to the numerical valence of reviews on a scale of one to five. Review depth refers to the quantity of textual arguments provided in reviews. Review readability measures the extent to which the textual arguments are comprehensible. Reviewer profile indicates the past track record of users who contribute reviews. Product type includes experience products and search products. Few studies hitherto have analyzed review diagnosticity taking into account all these factors concurrently. Hence, this paper attempts to augment prior research by developing a theory of diagnosticity for online reviews. The theory posits that review diagnosticity is shaped by the interplay among review rating, review depth, review readability and reviewer profile albeit differently between experience products and search products.**

Index Terms—**online reviews, review diagnosticity, theoretical model, moderated multiple regression**

I. Introduction

THE ever growing popularity of online review websites in recent years has led to what is known as information overload [1]. In consequence, users are overwhelmed by the smorgasbord of online reviews (henceforth, simply known as reviews) available on a single product or service. Furthermore, it is not trivial for them to differentiate reviews that are generally perceived as diagnostic from those that are either frivolous or biased. For the purpose of this study, diagnosticity of a given review is defined as the extent to which it helps users to make informed purchase decisions.

The problem of identifying diagnostic reviews has been mitigated in part by the concept of social navigation, whereby votes cast by users on perceived diagnosticity of reviews are used to prioritize the entries [2]. For example, the popular review website Amazon presents the question "*Was this review helpful to you?*" at the end of each submission to seek users' opinions on diagnosticity of the review. Users respond with either a "*Yes*" or a "*No*", which is utilized to rank order the reviews on a given product. This in turn, is used to place the most diagnostic reviews more conspicuously on the product's information page. Given that social navigation allows for the diagnostic entries to be located easily, it is no wonder that votes cast to evaluate diagnosticity of reviews have the potential to shape users' purchase decisions [3], [4]. Hence, developing a theory of diagnosticity for reviews represents an important endeavor not only with research significance but also with business implications.

Review diagnosticity can be associated with the interplay among five factors, namely, review rating, review depth, review readability, reviewer profile and product type. Review rating refers to the numerical valence of reviews and generally ranges from one star to five stars, the former indicating maximal criticism and the latter revealing maximal appreciation. Review depth refers to the quantity of textual arguments that reviewers provide to justify ratings. Review readability is a measure of the extent to which the textual arguments in reviews are comprehensible. Reviewer profile indicates the past track record of users who contribute reviews. Product type suggests the extent to which the products that are reviewed make users dependent on experiences of their peers.

Extant literature has shed interesting insights on ways the first three factors could affect users' perception of review diagnosticity. For review rating, extreme positive or negative reviews that help users confirm or eliminate options can be perceived as being diagnostic [5]. On the other hand, moderate reviews that highlight both the pros and cons of products or services can also be deemed diagnostic [6]. In terms of review depth, reviews that are lengthy and are supported by robust explanations can inspire confidence, and hence perceived to be diagnostic [7]. On the other hand, users at times can also be deterred to read reviews that are overly detailed [2]. With respect to readability, reviews that are easily readable tend to be voted as being diagnostic [3]. On the other hand, too lucid reviews may suggest a lack of competence on the part of reviewers, thereby rendering them not as diagnostic as sophisticated ones [8].

In contrast, review diagnosticity has been consistently affected by the profile of reviewers who contribute them [9], [10]. Reviews submitted by reviewers with positive track record are generally deemed as being more diagnostic compared to those contributed by newbies. Furthermore, diagnosticity of reviews has been shown to vary consistently across the experiential nature of the products reviewed [4], [11]. Typically, products are classified as either experience or search such that users can be more dependent on the post-

Manuscript received December 4, 2013; revised January 7, 2014.
Alton Y. K. Chua is with Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore 637718 (phone: +65-6790-5810; e-mail: altonchua@ntu.edu.sg).
Snehasish Banerjee is with Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore 637718 (e-mail: snehasis002@e.ntu.edu.sg).

purchase experience of others for the former than for the latter [12], [13]. However, few scholarly studies have looked into review diagnosticity taking into account all the five factors concurrently. Hence, this paper attempts to develop a theory of diagnosticity accounting for the interplay among review rating, review depth, review readability, reviewer profile and product type.

The remainder of the paper is structured as follows. The next section reviews related literature on review diagnosticity, culminating into a theoretical model. Next, the methods for data collection and analysis used to empirically test the model are described. This is followed by the results of the analysis. The main findings gleaned from the results are discussed next to offer insights into the theory of review diagnosticity. Finally, the paper concludes highlighting its implications, limitations and directions for future research.

II. Theoretical Model

Most review websites display reviews that consist of two parts, namely, review ratings and review texts. The former refers to numerical valence of reviews, generally ranging from one star to five stars, while the latter denotes the open-ended textual arguments. Before making a purchase decision, users can browse reviews of various ratings. They may choose to read selected reviews that offer adequate depth and readability. In particular, users may want to look for reviews contributed by reviewers with laudable profiles, especially for products that are more experiential in nature. It is thus conceivable that the interplay among review rating, review depth, review readability, reviewer profile and product type is crucial to pave the way towards understanding the diagnosticity of reviews.

Review ratings are typically captured on a five-point star rating scale. Users assign review ratings as a way to summarize their post-purchase sentiments. For example, they may award four or five stars ratings to express their delight. They may conversely show their disapproval through extreme negative reviews with ratings of one or two stars. Yet others with middle ground attitude may choose to express their neutrality through moderate reviews with ratings of three stars. Prior studies investigating the relationship between review ratings and review diagnosticity have yielded largely inconsistent results. Extreme reviews may be considered diagnostic because they apparently help users confirm or eliminate options [5]. However, given that they are often deemed too good or bad to be true, extreme reviews may not be perceived as credible [14]. Likewise, moderate reviews may appear diagnostic as they highlight both pros and cons [6], which in turn, help paint a more realistic picture of products and services. However, they are often too ambiguous to effectively aid decision making [15]. Hence, it appears that review ratings can have both linear and curvilinear relationship with review diagnosticity.

Review depth reflects the extent to which the reviews are lengthy enough to offer robust explanation to justify the assigned review ratings. In general, greater review depth tends to enhance the perceived value a review offers to users in decision making [7]. However, overly lengthy reviews may be intimidating for users to read [2]. In fact, users may favor terse extreme reviews and lengthy moderate reviews. This is because reviews with extreme ratings need to include only either positive or negative comments, while those with moderate ratings need to highlight both [3]. Hence, an extreme review that elaborately rants and raves about a product or a service may be perceived as being ambiguous. On the other hand, a sketchy moderate review is unlikely to be deemed as diagnostic due to its inadequacy in highlighting both pros and cons.

Review readability is a measure of the effort and expertise required by users to comprehend the meaning of a given review [16]. Simple reviews that are better readable tend to enhance comprehension, retention, and reading speed [8]. By attracting a wider audience, such reviews can be deemed as being more diagnostic compared to the sophisticated ones that may not always be easily readable. However, when users browse reviews, they not only study their content but also try to gauge the competence of the corresponding reviewers [6]. Too simplistic reviews containing overly lucid language may reflect reviewers' incompetence in writing sophisticated reviews [8]. This in turn can render the more readable reviews less diagnostic among users.

With respect to reviewer profile, most review websites display identity-descriptive information about reviewers alongside every review. Such information typically includes user names, summaries of past contributions and special badges such as top-50 reviewer or top-100 reviewer [8]. Prior research has consistently demonstrated the positive influence of information source characteristics on readers' perceptions [10], [17]. In fact, profile of sources has been largely shown to affect the perceived credibility and diagnosticity of information [9]. Likewise, perceived diagnosticity of a given review can be shaped in part by the profile of the respective reviewer [15]. Furthermore, reviews submitted by reviewers with laudable profiles have been shown to have a significant impact on the sales of products or services [5].

Product type can be another determinant of review diagnosticity. Typically, there are two types of products, namely, experience and search [12]. Those for which user satisfaction cannot be easily appraised prior to purchase are termed as experience products. Music albums are examples of experience products [4]. On the other hand, search products are those for which users can gauge their satisfaction more easily prior to purchase. Digital cameras, for example, qualify as search products as user satisfaction can be approximated on the basis of product specifications [18]. Even though the advent of review websites may have somewhat blurred the line between experience and search products, users' perception of review diagnosticity has been consistently shown to differ between the two types [12], [13].

Product type thus appears to moderate the interplay among review rating, review depth, review readability and reviewer profile in shaping review diagnosticity. Given the subjective nature of experience products, moderate reviews are more likely to be perceived as diagnostic. For search products however, users tend to find extreme arguments more credible [19]. Since reviews for search products are

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

more likely to contain objective details, additional depth can add more value in reviews for search products than in entries for experience products [4]. Likewise, since objective details can be more easily substantiated, reviews for search products can be more readable than those for experience products. With respect to reviewer profile, users' reliance on prolific reviewers may also differ between experience products and search products [12], [13].

Taking all the five factors together, a theoretical model on review diagnosticity is proposed (Fig. 1). The model is summarized as follows. Users' perception of review diagnosticity appears to be associated with the interplay among review rating, review depth, review readability and reviewer profile albeit differently between experience and search products.

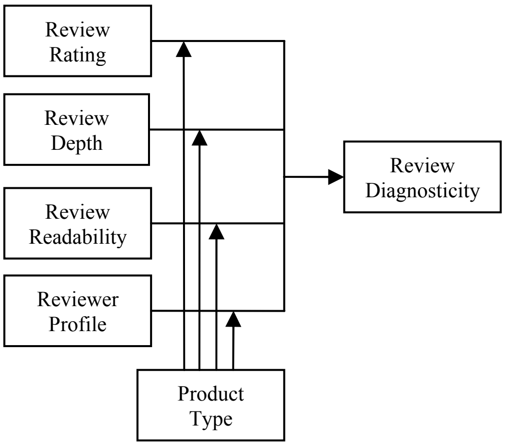

Fig. 1. Theoretical model of review diagnosticity.

III. Methods

A. Data Collection

Data for this paper were drawn from Amazon. It is one of the pioneering review websites that supports peer-evaluation of reviews through social navigation. In fact, its popularity and longevity has made it almost a de facto standard of review websites for scholarly inquiry [3], [4], [8]. Amazon allows for all the components in the theoretical model (Fig. 1) to be captured and analyzed, which further makes it appropriate for this paper.

Data were collected in October, 2012 and involved three steps. The first step was to identify the top 100 best seller product items from the following nine categories in Amazon, namely, books, skin care, music, digital SLR cameras, point-and-shoot digital cameras, contract cell phones, no-contract cell phones, unlocked cell phones, and laser printers. Among these products, books [3], skin care [18] and music [4] represent experience products. On the other hand, digital cameras [13], cell phones [20] and laser printers [18] represent search products. Hence, the first step yielded an initial pool of 900 best seller product items across the six products (books = 100, skin care = 100, music = 100, digital cameras = 200, cell phones = 300, printers = 100).

In the second step, ten product items for each of the six products were identified. Among the 900 best seller items, those that had attracted less than 30 or more than 100 reviews were eliminated. Product items with less than 30 reviews could be either recently launched or comprise those that rarely attract reviews. Such items may not aid a meaningful analysis. Those with more than 100 reviews were avoided as helpfulness of reviews for too popular items could be skewed due to bandwagon effect [4], which in turn may obscure the findings. A filtered pool of 195 product items (books = 24, skin care = 23, music = 20, digital cameras = 43, cell phones = 66, laser printers = 19) were obtained. From this pool, ten product items were randomly selected for each of the six products (10 x 6 = 60 product items).

In the third step, all reviews posted against these 60 product items were collected using a web scraper. For each review, the following data items were obtained: review rating, review date, review content, 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 attracted by the reviewer across all previously contributed reviews were also retrieved. Of the reviews collected, those that had missing data points were eliminated. The remaining 2,307 reviews were admitted for analysis. Specifically, the dataset included 1,113 reviews for experience products (books = 414, skin care = 348, music = 351) and 1,194 reviews for search products (digital cameras = 347, cell phones = 458, printers = 389).

B. Data Analysis

To empirically test the proposed theoretical model, moderated multiple regression was used. Review rating, review depth, review readability and reviewer profile were the independent variables while product type was the moderator. Mean centering was used to avoid problems of multicollinearity. Review diagnosticity was the dependent variable.

Review rating was measured as the star value of a review indicated by the reviewer [3]. To account for the curvilinear effect, square of review rating was computed into the model. Review depth was calculated as the length of reviews in words [4]. For readability, three metrics were used, namely, Gunning-Fog Index (FOG), Coleman-Liau Index (CLI), and Automated-Readability Index (ARI) [3], [8]. Lower values in these metrics suggest more readable reviews. Review readability of a given review was calculated as the arithmetic mean of these three indicators. Reviewer profile was operationalized using the summary of reviewers' past contributions [8]. It was computed as the ratio of the number of diagnostic votes to the total votes attracted by all the reviews contributed by a reviewer. Given the moderating nature of product type, it was dummy-coded with 1 and 0 indicating reviews for experience and search products respectively.

The dependent variable review diagnosticity was operationalized as the proportion of users who voted "*Yes*" to the question "*Was this review helpful to you?*" for a given review [4]. However, such proportions are often prone to biases. For example, reviews for which Amazon reports "*5 of 10 people found the following review helpful*" will have

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

numerically identical review diagnosticity with those for which Amazon indicates "*50 of 100 people found the following review helpful*". In order to mitigate such a confounding effect, the total number of votes provided in evaluation of a review's diagnosticity was taken as a control variable for the analysis [3], [4].

IV. Results

Table I shows the descriptive statistics of all reviews in the dataset as well as those for experience products (Exp) and search products (Srch) separately. It appears that reviews for experience products garnered higher ratings (RAT) than those for search products in spite of having relatively lesser depth (DEP). Reviews contributed for experience products seem to have less readability (REA) than those for search products. However, there is hardly any difference between the two product types in terms of reviewer profile (PRO). Moreover, reviews for experience products generally attracted more favorable votes on review diagnosticity (DIA) and more total votes (TOT) than those for search products.

TABLE I
DESCRIPTIVE STATISTICS

	All (N=2307) Mean ± SD	Exp (N_e=1113) Mean ± SD	Srch (N_s=1194) Mean ± SD
RAT	3.91 ± 1.44	4.21 ± 1.24	3.63 ± 1.56
DEP	186.74 ± 257.46	170.70 ± 209.57	201.70 ± 294.50
REA	8.07 ± 3.83	8.19 ± 4.50	7.96 ± 3.08
PRO	0.79 ± 0.21	0.81 ± 0.19	0.78 ± 0.22
DIA	0.80 ± 0.29	0.84 ± 0.25	0.76 ± 0.92
TOT	20.77 ± 47.56	25.42 ± 39.17	16.43 ± 53.88

As shown in Table II, results of the multiple regression analysis moderated by product type (TYP) suggest a good fit of the data with the proposed theoretical model ($p < 0.001$; $R^2 = 0.34$). Based on the results, four observations can be made. First, even though review rating has neither linear nor curvilinear relationship with review diagnosticity, both the effects are significantly moderated by product type. Experience products demonstrate significant positive linear relationship ($\beta = 0.15$, $p < 0.001$) and negative curvilinear relationship ($\beta = -0.04$, $p < 0.05$). This suggests that though perceived diagnosticity of reviews improves with increase in rating, moderate reviews were generally deemed more diagnostic than extreme ones. No statistically significant relationship was however found between rating and diagnosticity for search products.

Second, there was significant positive relationship between review depth and review diagnosticity ($\beta = 0.07$, $p < 0.05$). Reviews with substantial depth were considered more diagnostic than shorter ones. Product type did not significantly moderate the relationship between review depth and review diagnosticity. The relationship was slightly stronger for search products ($\beta = 0.10$, $p < 0.05$) than that for experience products ($\beta = 0.03$, $p < 0.05$).

Third, there was significant positive relationship between review readability and review diagnosticity ($\beta = 0.09$, $p < 0.05$). Sophisticated reviews were generally perceived as being more diagnostic than lucid reviews. Product type significantly moderated the effect of review readability on review diagnosticity. In particular, there was no significant relationship between readability and diagnosticity for reviews of experience products. For reviews of search products however, users' preference for sophisticated reviews with lower readability was statistically significant ($\beta = 0.07$, $p < 0.001$).

Fourth, there was significant positive relationship between reviewer profile and review diagnosticity ($\beta = 0.54$, $p < 0.05$). Reviews contributed by reviewers with strong past track records were generally perceived as being more diagnostic vis-à-vis those posted by newbies. Product type significantly moderated the effect of reviewer profile on review diagnosticity. The relationship was slightly stronger for search products ($\beta = 0.52$, $p < 0.001$) than that for experience products ($\beta = 0.45$, $p < 0.001$).

TABLE II
INFERENTIAL STATISTICS

	All (N=2307) β	Exp (N_e=1113) β	Srch (N_s=1194) β
TOT	-0.02	-0.04	0.03
RAT	0.05	0.15***	0.09
RAT^2	-0.07	-0.04*	-0.02
DEP	0.07*	0.03*	0.10*
REA	0.09*	0.03	0.07***
PRO	0.54***	0.45***	0.52***
RAT x TYP	0.10**		
RAT^2 x TYP	0.06*		
DEP x TYP	-0.01		
PRO x TYP	-0.07*		
REA x TYP	-0.07*		
Model R^2	0.34***	0.30***	0.34***

*** p < 0.001; ** p < 0.01; * p < 0.05

V. Discussion

In light of the results, four major insights into the theory of diagnosticity for reviews could be drawn. First, review rating seems to shape users' perception of review diagnosticity only for experience products. The relationship between rating and diagnosticity of reviews has been generally inconsistent in extant literature. For example, [6] and [21] found a curvilinear relationship with higher diagnosticity for moderate reviews than that for extreme reviews. Other studies such as [15] and [22] found that users perceive positive reviews as being diagnostic. To augment such prior studies, this paper shows that moderate reviews are perceived as being more diagnostic vis-à-vis extreme ones for experience products. No relationship for search products suggest that users are perhaps more keen to look into their objective details instead of making superficial judgment of diagnosticity based on ratings. This finding also serves as a dovetailing effort to studies such as [4], which considered the effect of rating on diagnosticity across product types without comprehensively taking into account other factors such as review readability and reviewer profile.

Second, review depth appears to be a useful antecedent for review diagnosticity for both experience and search products. This is in line with studies such as [4] and [7] which argued that reviews with substantial depth connote a sense of adequacy and hence, perceived to be diagnostic.

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

However, the relationship was slightly stronger for search products than that for experience products. Since reviews for search products are more likely to contain objective details than those for experience products, it is conceivable that additional depth adds more value in reviews for the former than in entries for the latter [4]. As a result, the former appears to show a slightly stronger relationship between review depth and review diagnostcity.

Third, review readability seems to be associated with review diagnosticity only for search products. As indicated earlier, reviews for search products are more likely to contain objective details than those for search products [19]. Furthermore, objective details can be easier to substantiate than subjective impression [4]. Conceivably, reviews for search products are likely to be more lucid than those for experience products [16]. However, it appears that users browsing reviews for search products prefer sophisticated reviews over lucid ones. Perhaps, too simplistic reviews reflect reviewers' incompetence [8], thereby rendering them to be perceived as being less diagnostic. On the other hand, given the difficulty in explicating subjective impression, users perhaps do not consider readability a useful proxy to comprehend diagnosticity of reviews posted in evaluation of experience products.

Fourth, reviewer profile appears to be an important proxy of review diagnosticity for both experience and search products. This is consistent with prior studies such as [9], [15], [17] that advocated the positive influence of source characteristics on information diagnosticity. Modern tech-savvy users appear extremely conscious of whom to rely upon [8]. Interestingly, the relationship was slightly stronger for search products than that for experience products even though users could be more dependent on the post-purchase experience of others for the latter [12], [13]. Users perhaps are aware of the subjective nature of experience and the objective nature of search products. Hence, they put greater emphasis on reviewer profile to assess review diagnosticity for search products than that for experience products. After all, reviews for experience products would remain subjective even if they are contributed by the most prolific reviewers. Hence, users' perhaps do not find the role of reviewer profile as important for experience products as that for search products.

Thus, the theory of diagnosticity for reviews is summarized as follows. Product type largely appears to moderate the influence of review rating, review depth, review readability and reviewer profile on review diagnosticity. Specifically, for experience products, users tend to prefer moderate reviews irrespective of their readability. Given the subjective nature of such products, extreme positive or negative reviews do not seem to inspire confidence. On the other hand for search products, users seem to have proclivity for sophisticated reviews irrespective of their rating. Given the objective nature of such products, users perhaps are reluctant to rely on overly simplistic reviews that might have been posted by less competent reviewers. Furthermore, users appear to emphasize on review depth and reviewer profile more for search products than for experience products. The differences in the interplay among review rating, review depth, review readability, reviewer profile and review diagnosticity across experience and search products is shown in Fig. 2.

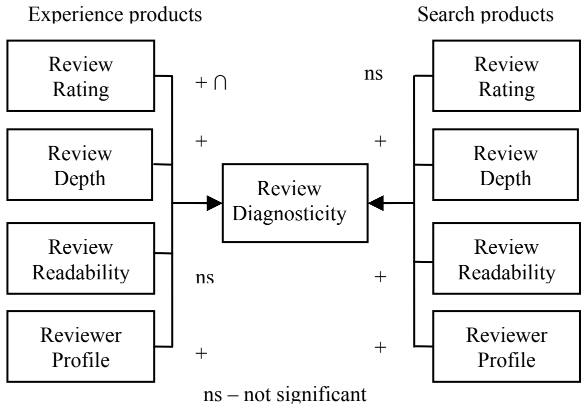

Fig. 2. The interplay among the five factors with respect to review diagnosticity.

VI. Conclusion

This paper attempted to develop a theory of diagnosticity for reviews. In particular, it identified five factors that could be related to users' perception of review diagnosticity. These include review rating, review depth, review readability, reviewer profile and product type. Teasing out ways these five factors may shape review diagnosticity from extant literature, a theoretical model was proposed. It was then tested empirically using data drawn from the popular review website Amazon. Results indicate that review rating has a negative curvilinear relationship with review diagnosticity for experience products but no relationship for search products. Review depth is positively associated with review diagnosticity for both experience products and search products albeit slightly stronger for the latter. Review readability has no relationship with review diagnosticity for experience products but a positive relationship for search products. Reviewer profile is positively related to review diagnosticity for both experience products and search products albeit slightly stronger for the latter. To sum up, the theory of diagnosticity for online reviews specifies four variables in affecting users' perception of review diagnosticity, namely, review rating, review depth, review readability and reviewer profile. More granularly, review diagnosticity of experience products is influenced by review rating, review depth and reviewer profile, while that of search products is influenced by review depth, review readability and reviewer profile.

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 diagnosticity in the context of reviews. It represents a dovetailing effort to extant literature by shedding light on review diagnosticity with respect to review rating, review depth, review readability, reviewer profile and product type concurrently.

On the practical front, the paper provides implications for users as well as businesses. Users could lean on the findings of this paper to conjecture which reviews are likely to be diagnostic in order to make informed purchase decisions.

The findings may offer cues to users to write more diagnostic reviews. In terms of review rating, users should try to include both pros and cons of products. They should strike a balance in terms of review length to ensure that reviews are neither too sketchy nor overly detailed. They should strive to make the reviews readable. Besides, the findings offer insights to businesses to tap into reviews that are apparently more diagnostic than others. This in turn will allow businesses to keep a pulse on users' most dominant preferences and complaints towards specific products or services.

Three limitations inherent in this paper need to be acknowledged. First, the results presented in this paper hold true for three experience products, namely, books, cosmetics and music, and three search products, namely, digital cameras, mobile phones and printers. Caution needs to be exercised while generalizing the results to other products or services. Second, the variables reviewer profile, review rating, review depth and review diagnosticity were quantitative surrogates and not direct measures. Though it allowed for data-driven statistical analysis such as moderated multiple regression, using a qualitative approach might have provided richer data with more scope for triangulation. Third, given the cross-sectional nature of the dataset, causal inference could not be made. For example, even though review depth was found to be positively related to review diagnosticity, it remains unclear if the former causes the latter.

Nonetheless, a number of future research directions can be identified from this paper. 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 diagnosticity. Another direction would be to investigate review diagosticity using qualitative approaches. A qualitative analysis of the review content with multiple coders could further offer a more discerning understanding of what makes a review diagnostic.

REFERENCES

[1] A. Ghose, and P. G. Ipeirotis, "Designing novel review ranking systems: Predicting usefulness and impact of reviews," in *Proc. 2007 ACM International Conf. Electronic Commerce*, pp. 303-310.

[2] J. Otterbacher, "'Helpfulness' in online communities: A measure of message quality," in *Proc. 2009 ACM SIGCHI Conf. Human Factors in Computing Systems*, pp. 955-964.

[3] N. Korfiatis, E. García-Bariocanal, and S. Sánchez-Alonso, "Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content," *Electronic Commerce Research and Applications*, vol. 11, no. 3, pp. 205-217, May-Jun, 2012.

[4] S. M. Mudambi, and D. Schuff, "What makes a helpful online review? A study of customer reviews on Amazon," *MIS Quarterly*, vol. 34, no. 1, pp. 185-200, Mar. 2010.

[5] C. Forman, A. Ghose, and B. Wiesenfeld, "Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets," *Information Systems Research*, vol. 19, no. 3, pp. 291–313, Sep. 2008.

[6] L. Connors, S. M. Mudambi, and D. Schuff, "Is it the review or the reviewer? A multi-method approach to determine the antecedents of online review helpfulness," in *Proc. 2011 IEEE HICSS*, pp. 1-10.

[7] M. J. Metzger, "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*, vol. 58, no. 13, pp. 2078-2091, Nov. 2007.

[8] A. Ghose, and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer Characteristics," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 10, pp. 1498-1512, Oct. 2011.

[9] C. N. Wathen, and J. Burkell, "Believe it or not: Factors influencing credibility on the web," *Journal of the American Society for Information Science and Technology*, vol. 53, no. 2, pp. 134-144, Dec. 2001.

[10] Y. Liu, X. Huang, A. An, and X. Yu, "Modeling and predicting the helpfulness of online reviews," in *Proc. 2008 IEEE International Conf. Data Mining*, pp. 443-452.

[11] S. Sen, and D. Lerman, "Why are you telling me this? An examination into negative consumer reviews on the web," *Journal of Interactive Marketing*, vol. 21, no. 4, pp. 76-94, 2007.

[12] P. Nelson, "Information and consumer behavior," *The Journal of Political Economy*, vol. 78, no. 2, pp. 311-329, Mar.-Apr., 1970.

[13] L. Xia, and N. N. Bechwati, "Word of mouse: The role of cognitive personalization in online consumer reviews," *Journal of interactive Advertising*, vol. 9, no. 1, pp. 3-13, 2008.

[14] M. J. Metzger, A. J. Flanagin, and R. B. Medders, "Social and heuristic approaches to credibility evaluation online," *Journal of Communication*, vol. 60, no. 3, pp. 413-439, Sep. 2010.

[15] P. A. Pavlou, and A. Dimoka, "The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation," *Information Systems Research*, vol. 17, no. 4, pp. 392-414, Dec. 2006.

[16] B. L. Zakaluk, and S. J. Samuels, *Readability: Its past, present and future*, Newark: International Reading Association, 1988.

[17] S. Chaiken, "Heuristic versus systematic information processing and the use of source versus message cues in persuasion," *Journal of Personality and Social Psychology*, vol. 39, no. 5, pp. 752-766, Nov. 1980.

[18] L. M. Willemsen, P. C. Neijens, F. Bronner, and J. A. de Ridder, "'Highly recommended!' The content characteristics and perceived usefulness of online consumer reviews," *Journal of Computer Mediated Communication*, vol. 17, no. 1, pp. 19-38, Oct. 2011.

[19] G. T. Ford, D. B. Smith, and J. L. Swasy, "Consumer skepticism of advertising claims: Testing hypotheses from economics of information," *Journal of Consumer Research*, vol. 16, no. 4, pp. 433-441, Mar. 1990.

[20] D. Weathers, S. Sharma, and S. L. Wood, "Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods," *Journal of Retailing*, vol. 83, no. 4, pp. 393-401, Dec. 2007.

[21] A. Schlosser, "Source perceptions and the persuasiveness of Internet word-of-mouth communication," in G. Menon and A. Rao, Eds., *Proc. 2005 Advances in Consumer Research*, pp. 202-203.

[22] Y. Pan, and J. Q. Zhang, "Born unequal: A study of the helpfulness of user-generated product reviews," *Journal of Retailing*, vol. 87, no. 4, pp. 598-612, Dec. 2011.