The Effect of Information Depth and Completeness on Review Diagnosticity

Abstract

Online reviews are an important source of information for individuals and organizations alike. They provide an assessment of disparate entities (e.g., a product, a person, a service encounter) in detail, thus guiding and influencing decision-making. It is the textual content of reviews that provides this rich data. However, extant research has largely ignored it in favor of quantitative elements.

In this research we introduce online review diagnosticity, the extent to which the information presented in a review is perceived helpful in evaluating its target. We posit that review diagnosticity depends on the depth of information provided and its completeness. We analyze the textual contents of 77,864 hotel reviews using probabilistic topic modeling, an approach that does not rely on human raters to extract the thematic structures of a corpus of documents. Our findings indicate that both the length of a review (i.e., depth) and the comprehensiveness with which the review discusses various topics of the service experience (i.e., completeness), as well as their interaction significantly impact review diagnosticity. These results provide a conceptual, empirical, and methodological contribution.

Keywords: *online review, diagnosticity, helpfulness, depth, completeness*

Introduction

The ability to create and publish content on the Internet allows users to widely share their experiences with others. While much of the literature refers to online reviews as "electronic word-of-mouth" (Chevalier and Mayzlin, 2006; Zhang, Craciun and Shin, 2010; Pan and Zhang, 2011; Li and Zhan, 2011; Schindler and Bickart, 2012), we contend that information from user-generated content forums differs greatly in scale, diffusion speed, accessibility, persistence, contextual leanness and visibility. The popularity of online reviews is a byproduct of these unique characteristics and it underscores why individuals no longer rely exclusively on providers' communications, or a few friends' word-of-mouth, but increasingly employ, and contribute to, online forums (Huang, Lurie and Mitra, 2009). Online reviews therefore represent a prominent example of the emerging phenomenon of "digital mediation of everyday experiences" (Yoo, 2010). They constitute a crowdsourced pool of opinions, provided by individuals who voluntarily leave feedback about their experiences through computing devices. Thus, online reviews reflect the heterogeneous taste of consumers—yet, they are highly subjective as they are based on the personal judgment of an individual. Despite their subjectivity, consumers regard them as more trustworthy, credible and interesting than seller-provided content (Tsang and Prendergast, 2009; Pavlou and Dimoka, 2006).

We define *online review diagnosticity* as the readers' perception of the extent to which an online opinion's textual content aids their evaluative judgment of the review's target (e.g., a person, a product, a service encounter) (Aboulnasr, 2006; Menon, Raghubir and Schwarz, 1995; Jiang and Benbasat, 2004; 2007; Pavlou, Liang and Xue, 2007; Dimoka, Hong and Pavlou, 2012). It measures the perception of a review's helpfulness in making a subjective judgment. Understanding review diagnosticity is important given its influence on the decision-making process (e.g., Chevalier and Mayzlin, 2006; Aboulnasr, 2006; Clemons, Gao and Hitt, 2006; Duan, Gu and Whinston, 2008; Ghose and Ipeirotis, 2009). Its theoretical and practical value is the reason for an increased attention to perceived review helpfulness in the IS and related literatures (Forman, Ghose and Wiesenfeld, 2008; Ghose and Ipeirotis, 2009; Mudambi and Schuff, 2010; Cao, Duan and Gan, 2011; Li and Zhan, 2011; Pan and Zhang, 2011; Korfiatis, Baek, Ahn and Choi, 2012; García-Bariocanal and Sánchez-Alonso, 2012; Schindler and Bickart, 2012; Scholz and Dorner, 2013). However, with few exceptions (Ghose and Ipeirotis, 2009; Archak, Ghose and Ipeirotis, 2011; Baek et al., 2012; Ghose, Ipeirotis and Li, 2012), previous research has not incorporated thematic elements of opinions as antecedents of review diagnosticity.

Using a dataset of 77,864 hotel reviews, we extract the thematic structure of the reviews through probabilistic topic modeling and compute review depth and review completeness. We find that both the variety of topics a review reflects upon (i.e., completeness) and the detail with which the review discusses features of the service experience (i.e., depth) have an impact on its diagnosticity.

The contribution of our research is threefold: conceptual, empirical, and methodological. Building on a consolidated tradition, we propose review diagnosticity as a new concept and provide the first empirical analysis of its antecedents. We also propose a new construct, review completeness, influencing a review's diagnosticity; and we measure it through the first application of probabilistic topic modeling in information systems. Thus, our work responds to recent calls for focusing on review content as an essential part of computer-mediated communication in e-commerce (Dimoka et al., 2012).

The paper is structured as follows. We first introduce the notion of review diagnosticity, review depth and completeness along with our research model. We then provide a detailed description of the methodological approach used for text analysis, followed by our empirical findings and a discussion of the results.

Theoretical Framework

Diagnostic Information

Diagnosticity was first introduced to describe the ability of information to aid an individual's evaluative judgment (Menon et al., 1995; Aboulnasr, 2006). Information is considered diagnostic for a judgment or decision if the individual believes "that the decision implied by that input alone would accomplish their decision goals (e.g., maximize utility, choose a justifiable alternative, and so on)" (Lynch, Marmorstein and Weigold, 1988, p. 171). Put differently, it refers to an individual's assessment of whether the information provided is helpful for evaluative purposes. Diagnosticity is therefore a subjective construct, stemming from individuals' perceptions (Lynch et al., 1988). The inverse, non-diagnostic information is evidence that information is either irrelevant or subject to multiple interpretations and thus not able to further an individual's decision goals.

The concept has found wide applicability in information systems research as product diagnosticity, Web site diagnosticity and product description diagnosticity. Product diagnosticity mitigates a buyer's uncertainty when purchasing online (Pavlou et al., 2007), Web site diagnosticity measures the extent to which a Web site contributes to an individual's understanding of the product in question (Jiang and Benbasat, 2007) and product description diagnosticity describes "the extent to which a consumer believes that a retailer (or seller) offers helpful textual information to describe a product" (Dimoka et al., 2012). Since online reviews also represent a form of textual description, we define review diagnosticity as the extent to which an individual believes that an online opinion offers helpful information to evaluate the target of the review. It captures the review's ability to provide relevant information for the reader to assess the target of the opinion (e.g., a product, a service, a person). Review diagnosticity is reflected in responses to the question: "Was this review helpful?" posed by opinion platforms, such as Epinons or Yelp. As such, a vote of helpfulness signifies a subjective endorsement of the opinion's diagnosticity. Individuals typically issue votes of helpfulness after reading a review and before taking action (e.g., making a purchase, or hiring a caretaker). Such endorsements represent the weight readers place on the review text as part of their decision-making process and the degree to which the review provides superior information to others (Weiss, Lurie and MacInnis, 2008). A non-diagnostic review, in turn, is one that does not allow an individual to form an impression and is therefore perceived as unhelpful in furthering his or her understanding of the target of interest.

Review Depth and Review Diagnosticity

We define *review depth* as the quantity of textual information provided in an online opinion – measured by the number of words in the free-form field of the review. Considerable evidence shows that review depth is positively correlated with review diagnosticity (e.g., Mudambi and Schuff, 2010; Pan and Zhang, 2011; Baek et al., 2012). Longer reviews, on average, convey more information, thus reducing uncertainty (Daft, Lengel and Trevino, 1987) and better aiding readers in forming evaluating judgments. The link is particularly strong for experiential products (Nelson, 1970;, 1974), where consumer uncertainty is greater due to the subjective features of the product (Mudambi and Schuff, 2010). Experiential products—or products that have to be used before an evaluation can take place (Nelson, 1970; 1974), such as hotel stays or restaurant visits—are “inherently subjective, characterized by high levels of uncertainty and equivocality and difficult to evaluate” (Huang et al., 2009, p. 57). It is therefore not surprising that individuals spend a disproportionate amount of time on the information search phase when making purchase decisions pertaining to experiential products (Huang et al., 2009).[1] A similar dynamic is at play in online dating, where individuals are forming judgments about other individuals (Ellison, Heino and Gibbs, 2006). Since deeper reviews provide the opportunity of more input for a proper judgment, individuals are also more likely to turn to online opinions when researching experiential rather than search products[2] (Baek et al., 2012). Both effects are driven by individuals’ need for information (Ha and Hoch, 1989) and the fact that more information boosts confidence in the decision made (Tversky and Kahneman, 1974). We therefore propose that, on average and all else being equal, the more text a reviewer creates when posting an opinion, the more likely it is that the review will contain valuable information that increases its diagnosticity.

H1a: Review depth positively influences review diagnosticity.

While the main effect of depth on helpfulness is a stable research result, recent work suggests that the relationship may not be strictly linear because the incremental quantity of information provided tends to diminish with more text in the opinion (e.g., Schindler and Bickart, 2012). In other words, the contribution of review depth to diagnosticity is incrementally diminishing as review depth increases. While depth is always valuable, a review might decrease the reader’s ability to identify useful incremental information as the opinion becomes longer and more detailed. A study analyzing conversations in

[1] A typical decision-making process in the context of purchasing behavior consists of five phases (Dewey, 1910): problem recognition, information search, evaluation of alternatives, purchase decision, and post-purchase behavior.

[2] Typical search products used in the literature include for example: digital camera, cell phone, laser printer (Mudambi and Schuff, 2010).

newsgroups, for example, found that members intuitively curtail the depth of their posts and that those individuals are often side-lined (Atfi, Mandelcwaig and Marcoccia, 2011). Recent research analyzing word repetitions of online reviews shows that the effect of reiteration is detrimental to a review's perceived value (Schindler and Bickart, 2012). Previous work also shows that a reader's cognitive load non-linearly increases with an increase in information received (Schroeder, Driver and Streufert, 1967; Malhotra, Jain and Lagakos, 1982; Keller and Staelin, 1987). As review depth increases, individuals tend to apply simpler heuristics and to process information only selectively (Bettman, Luce and Payne, 1998). Thus, while lengthy reviews are more diagnostic on average, increasing depth may progressively make useful evidence harder to discern, thus hampering decision making and reducing the effect of depth on diagnosticity.

H1b: Review depth has a diminishing marginal effect on review diagnosticity.

Review Completeness and Review Diagnosticity

We define *review completeness* as the degree to which an online review describes all the elements of the review target. Thus, review completeness measures the extent to which the opinion provides information about the various aspects of the product (or person, or service) that are relevant to a reader's judgment. Review completeness relates to the more general concept of information completeness—a facet of information quality (Knight and Burn, 2005; Nelson, Todd and Wixom, 2005; Setia, Venkatesh and Joglekar, 2013). In the context of commercial reviews, where the objective is to provide a crowdsourced description of consumer experiences for others prior to purchase, this variation in topics is primarily dependent upon the type of product or service described (Huang et al., 2009). Information provided about experiential products, for example, is typically multi-faceted; it consists of sets of attributes that are harder to pinpoint and quantify when compared to search products (Baek et al., 2012; Huang et al., 2009). Previous work shows that providing additional descriptors (or aspects) in order to capture a product's make-up increases an individual's perceptions of the product's capability and thus results in improved evaluations and sales (Mukherjee and Hoyer, 2001)—an effect that is particularly accentuated for experiential products (Sela and Berger, 2012).

Research on consumer feedback indicates that the more product-descriptive information is provided in a review, the higher its perceived value (Schindler and Bickart, 2012) and that textual content is an important determinant of consumer choice—over and above valence (i.e., rating) and volume (Archak et al., 2011). Since experiential products are, on average, more difficult to describe than search products, covering the numerous facets that make up the entirety of a product experience is vital. The more aspects of an experience are addressed in a review, the more likely is it perceived as complete and able to reduce the reader's uncertainty in decision-making (Daft et al., 1987). A similar dynamic is at play in online dating where the target of evaluation is a person (e.g., Brand, Bonatsos, D'Orazio and DeShong, 2012; Toma and Hancock, 2012). We posit that complete reviews are, on average and all else being equal, more diagnostic when compared to incomplete ones.

H2a: Review completeness positively influences review diagnosticity.

Just as the quantity of information might have a diminishing marginal effect on diagnosticity, we propose that information completeness follows the same dynamic. In other words, we expect that greater degrees of information completeness contribute to an increase in review diagnosticity, however, this increase flattens in its impact.

While very little theoretical guidance exists in this regard our proposition is supported by studies that examine individuals' information-processing strategies when selecting product alternatives. As the number of attributes for a product increases, people necessitate greater cognitive effort to process them (Payne, Bettman and Johnson, 1988; Malhotra et al., 1982). As a result, they apply simpler heuristics and process information only selectively in order to arrive at a decision (Bettman et al., 1998), which also tends to be of lesser quality (Keller and Staelin, 1987; 1989). Providing more aspects, and thus a broader coverage, of a product (or service, or person) may add to a review's diagnosticity in general, but it may do so to a lesser extent as the number of aspects (or topics) addressed in a review increase. Accordingly, we hypothesize:

H2b: Review completeness has a diminishing marginal effect on review diagnosticity.

The Joint Effect of Review Depth and Completeness on Review Diagnosticity

While review depth and review completeness are conceptually distinct, they are also complementary. Whereas one is reflective of the quantity of textual information provided in a review, the other is indicative of the breadth of content. Together, we expected them to interact in their relationship with review diagnosticity. In fact, the interplay between review depth and completeness can be viewed as an expression of conciseness.

Conciseness describes the extent to which information is compactly represented without being overwhelming – brief in presentation, yet complete and to the point (Knight and Burn, 2005, p. 162). Some studies have categorized conciseness as part of representational elements of information quality, i.e. whether the information is portrayed in a suitable and digestible format (Wang and Strong, 2006; Dedeke, 2000; Naumann and Rolker, 2000). Yet, others have labeled conciseness as a pragmatic element at the semiotic level, whether it is useful (or not) for readers (Price and Shanks, 2005); and others as sound (versus useful and relevant) information regarding a product (Kahn, Strong and Wang, 2002) or content (Eppler and Muenzenmayer, 2002). In this study, we propose that the interplay between review depth and completeness (i.e., conciseness) measures the appropriate balance of depth and completeness, or the ability of a review to maximize feature coverage efficiently. We posit that, given the ephemeral nature of online transactions (Herring, Stein and Virtanen, 2013), all else being equal, the level of review completeness moderates the influence of review depth on the diagnosticity of information provided in a review. We accordingly state:

H3: Review completeness negatively moderates the relationship between review depth and review diagnosticity.

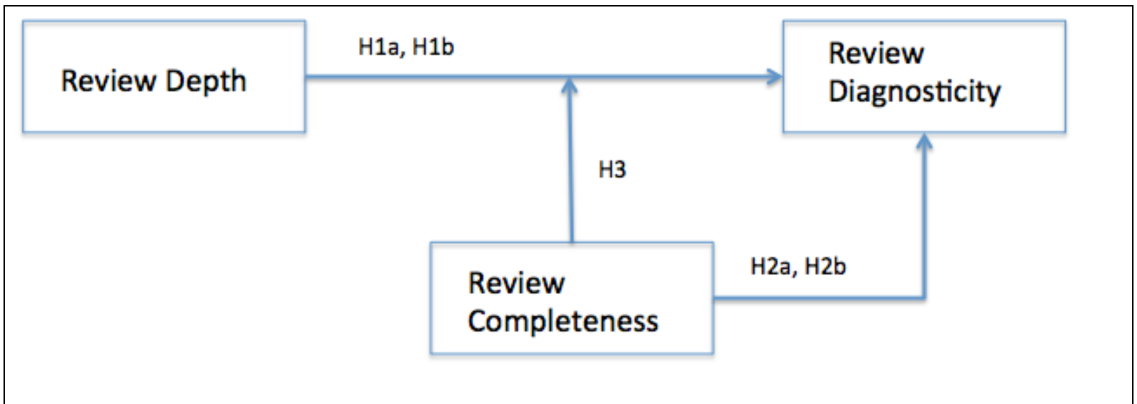

Figure 1: Research Model

Research Method

Our investigation is based on an archival research methodology, using a set of reviews from a major opinion platform. The original dataset comprised 293,295 reviews posted between January 1st and December 20th, 2012, pertaining to the 25 most populous US cities—a total of 3,686 hotels. From this dataset we extracted a four-months subsample of 77,967 reviews. We chose the subsample to exclude any major holidays (e.g., Christmas, new years eve) and to include enough time after the posting of the review for the recording of helpfulness votes to occur. We therefore selected the timeframe from February 1st, 2012 to May 30th, 2012. For each review, we downloaded its title and text, as well as the numeric ratings for the various elements of the hotel experience: value, location, service, room. We also collected identifying information about the hotel and the customers who wrote the review.

Variables and Measures

The measure of review diagnosticity is the total number of helpfulness votes a review received in the specified time frame, ranging from zero to 78 in our sample. Since the aggregator platform does not publish the number of individuals who have read the review, the total number of votes can be misleading as the probability that a review receives a vote is dependent upon its chance of exposure to readers on the site. The chance of exposure depends upon the number of individuals visiting as well as the number of reviews displayed on each page. We control for the exposure of each review by estimating its chance of receiving a helpfulness vote within a given timeframe. We are not aware of any research that has explicitly evaluated the relationship between review position and exposure. However, work in related areas suggests that the chances of achieving popularity decreases dramatically when a product is not on the first page (Salganik and Watts, 2008). We estimated two control variables to capture exposure: *nreviews* and *firstpage*.

nreviews computes, for each review, the total number of subsequent reviews the hotel has received since the posting of the focal review. It therefore can be viewed as a proxy of the number of individuals who have seen the review, assuming that the number of reviews a hotel receives in the opinion platform is a fixed percentage of the number of consumers who researched it. *firstpage* computes the number of days a given review was visible on the first page of the hotel on the review site – each page displays 10 reviews. More specifically, it captures the time difference between the posting of the focal review and the posting of the 10^{th} review following it. For reviews that remained on the front page until the end of the sampling frame (n = 1,703, 2.2%), we used the last date available in our dataset (i.e., December 20^{th}, 2012) for the computation. Both control variables, *nreviews* and *firstpage*, are correlated, but capture two different aspects of exposure. While *nreviews* focuses on the inflow of new reviews as a measure of the number of potential readers, *firstpage* estimates the length of time a review was presented to all visitors. We control for each as well as for their interaction effect in our model.

As independent variables, we measured review depth and completeness. Consistent with the literature, we measure depth as the total number of words in each review (Mudambi and Schuff, 2010). Completeness, on the other hand, required an innovative approach to measuring how comprehensively an opinion references the elements of a hotel experience: service, value, location, room and food. In the next section, we briefly describe probabilistic topic modeling, the technique we used to extract the thematic structures of each review.

Topic Modeling

Topic models are generative probabilistic approaches with a wide array of applications in machine learning (Blei, 2012). The LDA model (Blei, Ng, Jordan and Lafferty, 2003), which we adopt for this analysis, operates on the bag-of-words assumption. It models thematic text structures (i.e., topics) through distributions over words and documents with their own distribution over topics; in other words, it generates a probabilistic model that captures (and is able to reproduce) the structure of topics implicitly contained in a set of documents. One advantage of using topic modeling for texts is that topics can be inspected and "understood" in terms of the distribution of terms that comprises them, and—more importantly—documents can be represented in terms of their topic probabilities.

In this study, we used a seeded sentence-level LDA model (Lu, Ott, Cardie and Tsou, 2011) after tokenizing and sentence splitting each review with the Stanford POS Tagger (Toutanova, Klein, Manning and Singer, 2003). Given our specific interest in the five elements of the lodging service experience, we seeded each topic with four terms each (see also the Appendix).

The output of the sentence-level LDA algorithm is a vector of θ weights for each sentence, representing the relevance of the five topics of the hotel experience in a given sentence. θ weights are reflective of the probability that a particular topic is associated with a sentence. The topic with the largest weight for each sentence can be thought of representing the topic a sentence refers to. We chose a threshold of $\theta > 0.7$ for topic assignment.[3] Thus if a sentence had a topic (e.g., service) with a θ weight greater than 0.7 it would be representative of that topic. Sentences with no explicit topic above the threshold were classified as "undefined." Upon tagging each sentence with a topic, we recombined this information over the original reviews. We then computed the extent of completeness by counting the different number of unique mentions of the five topics in each review (Maldberger and Nakayama, 2013). Thus, completeness ranges from 0 to

[3] We ran the analysis with θ thresholds ranging from 0.5 to 0.9. Results did not qualitatively change.

5, with 0 representing a review with only undefined sentences and 5 representing a review with at least one sentence addressing each of the five features of the hotel experience. (A detailed description can also be found in the Appendix.)

Additional Control Variables

Apart from *nreviews* and *firstpage*, we control for numeric experience ratings by measuring the overall rating score for each review (*ratings.overall*).[4] We also control for hotel specific and author specific effects. Posting channel (i.e., Web or mobile) and hotel class were used, but proved no difference in results and were thus dropped.

A descriptive summary of the variables is given in Table 1. For this summary and in the subsequent analysis, reviews containing more than 1,000 words were omitted in order to avoid the undue influence of extreme observations (i.e., 102 reviews, 0.1%). In addition, one review was removed where the overall rating was missing. This led to 77,864 opinions being used for analysis.

---

[4] The overall rating is the only numeric variable required by the opinion platform—ratings per categories (i.e., cleanliness, service, room, etc.) are optional. For comparative purposes, we ran a version of the model with individual, i.e., categorical, numeric rating scores. However, since the overall rating score summarizes the other scores and thus captures their variability, results do not statistically change.

Table 1: Descriptive Statistics				
Variable	Mean	Std. dev.	Min	Max
diagnosticity	0.84	1.36	0	78
nreviews	181.80	161.03	1	1059
firstpage	39.04	50.95	0	323
ratings.overall	4.07	1.08	1	5
depth	142.67	116.98	3	999
completeness	3.00	1.30	0	5

Analytical Procedure

We fit a generalized linear mixed model (GLMM) using the R package lme4 (R Core Team, 2014; Bates, Maechler, Bolker and Walker, 2014). Given the nature of helpfulness votes as a count variable, we modeled review diagnosticity with a Poisson distribution and a log link function. The logarithm of the expected number of votes is modeled by a linear predictor, including random effects for hotel and author variation:

$$
\begin{aligned}
\log(\text{diagnosticity}_i) = \beta_0 &+ \beta_1 \log(\text{nreviews})_i \\
&+ \beta_2 \log(\text{firstpage} + 1)_i \\
&+ \beta_3 \text{depth}_i \\
&+ \beta_4 \text{depth}^2{}_i \\
&+ \beta_5 \text{completeness}_i \\
&+ \beta_6 \text{completeness}^2{}_i \\
&+ \beta_7 \text{ratings.overall}_i \\
&+ \beta_8 \log(\text{nreviews})_i * \log(\text{firstpage} + 1)_i \\
&+ \beta_9 \text{depth}_i * \text{completeness}_i \\
&+ \gamma_{0\text{hotel } i} + \gamma_{0\text{author\_id } i},
\end{aligned}
$$

where $\gamma_{0\text{hotel } i} \sim N(0, \sigma^2_{\text{hotel}})$ and $\gamma_{0\text{author\_id } i} \sim N(0, \sigma^2_{\text{author\_id}})$ and i = [0, 77,864]

The above model enabled us to compute both marginal and conditional R^2 (Nakagawa and Schielzeth, 2013). Marginal R^2 measures the variance explained by fixed effects in relation to the total variance, while the conditional R^2 value reports the total variance explained by fixed and random effects. The latter therefore includes random effects for hotel and author variation in explaining diagnosticity.

Findings

After controlling for hotel and author specific effects, the number of reviews posted after the focal one (*nreviews*), the number of days the review was available on the front of the hotel's review page (*firstpage*), and the numeric evaluations of the overall service experience (*ratings.overall*), our analysis finds that review depth is a strong predictor of review diagnosticity (H1a). We also detect a significant negative quadratic effect of review depth, supporting our prediction that additional text has a positive but diminishing effect on the diagnosticity of the review (H1b). Our hypotheses about review completeness are also supported with both a significant positive impact on diagnosticity (H2a) and a decreasing marginal effect (H2b). Finally, we find a significant negative interaction between depth and completeness. This result confirms our expectation about the diagnostic value of conciseness (H3). It shows that completeness has a stronger positive impact on the diagnosticity of a short versus a long review. The model accounts for 46.4% of the total variability in review diagnosticity, including 8.9% that can be ascribed to the fixed effects (Table 2).[5]

[5] We evaluated multiple models with different control variables. We also tested a "full interaction" model estimating the interaction of depth and completeness as well as their squared values. The results of all these robustness checks are qualitatively the same with slight changes in the parameters.

Table 2: Results		
	Null Model	Full Model
(Intercept)	-1.9888 *** (0.2501)	-2.6728 *** (0.2477)
log(nreviews)	0.1441 ** (0.0484)	0.1291 ** (0.0478)
log(firstpage + 1)	0.2164 *** (0.0482)	0.2285 *** (0.0475)
log(nreviews) * log(firstpage + 1)	0.0549 *** (0.0095)	0.0529 *** (0.0094)
ratings.overall	-0.1981 *** (0.0044)	-0.1401 *** (0.0045)
depth		0.0038 *** (0.0001)
$depth^2$		-0.0000 *** (0.0000)
completeness		0.0564 *** (0.0151)
$completeness^2$		-0.0061 * (0.0029)
depth * completeness		-0.0002 *** (0.0000)
Conditional R^2	0.4633	0.4640
Marginal R^2	0.0539	0.0894

Fixed effect coefficients (standard errors in parentheses) (N = 77,864)
*p < 0.05; **p < 0.01; ***p < 0.001

Moreover, the model predicts that for a review addressing only one aspect of the service experience, an increase in words from 50 to 100 leads to an increase in diagnosticity of 18%. This effect is slightly less pronounced for reviews that cover three topics. Here an increase from 50 to 100 words leads to an increase in diagnosticity of 16%; and for reviews that cover all five topics this increase accounts for 15%. Holding review depth constant while altering completeness also shows differential impacts on diagnosticity. For example, for reviews with 50 words, increasing the number of topics from zero to five leads to a 9% increase in diagnosticity. A closer examination shows that a move from one to two topics results in an increase of 3% in diagnosticity while a move from two to five topics results in an increase of 2%. For reviews with 100 words, increasing the number of topics from none to five leads to a 5% increase in diagnosticity (Figure 2).

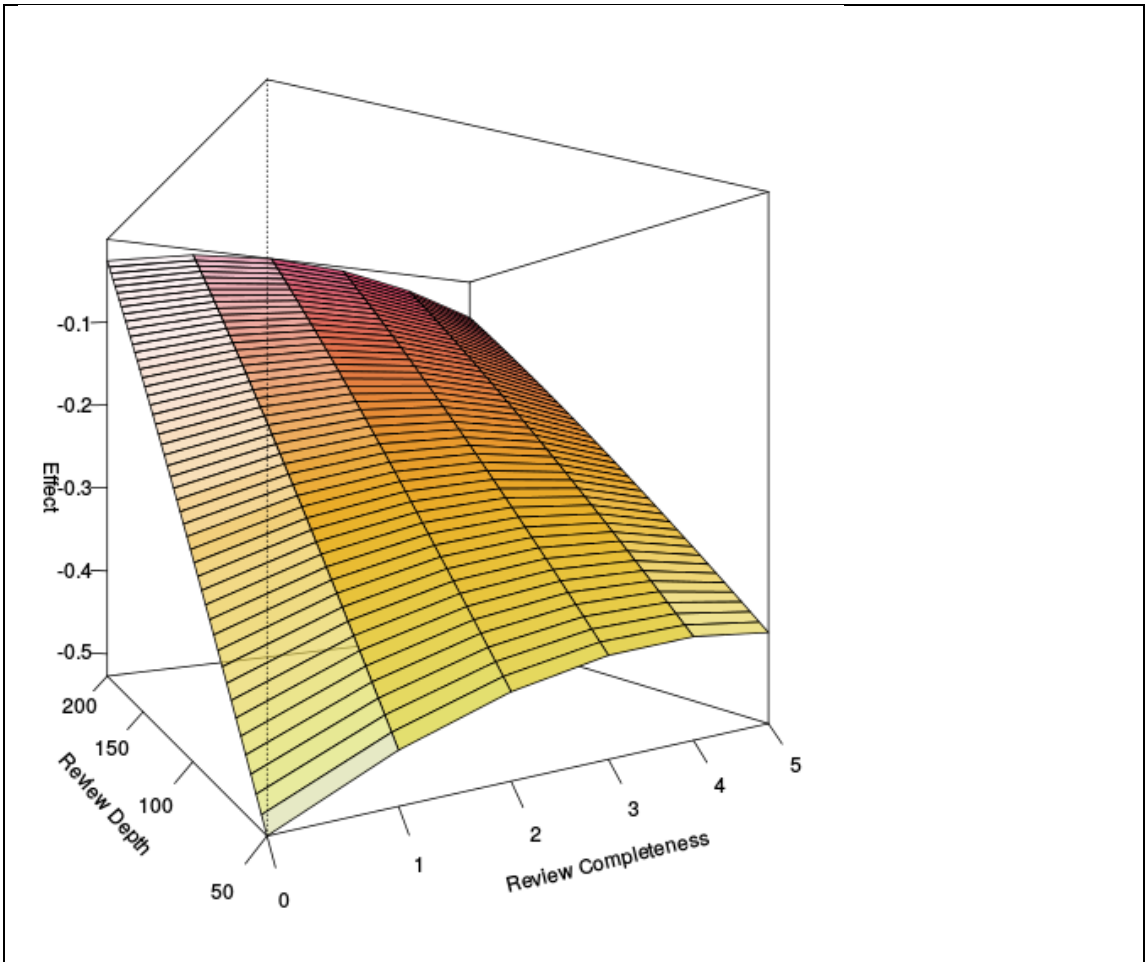

Figure 2: Combined Effect of Review Depth and Review Completeness[6]

[6] Color represents the density of observations. Higher transparency levels indicate less observations.

Overall, our findings suggest that for review depth “more” equates to “more helpful,” though its effect flattens out. Individuals tend to appreciate in-depth reviews, but their enthusiasm wanes as the review grows overly expansive. The same applies to review completeness. Providing more aspects in a review leads to more helpfulness votes in general, indicating that individuals appreciate a discussion of various facets of a product experience, but once again the effect lessens; individuals are less likely to appreciate an encompassing portrayal of aspects as their number increases. Review depth and review completeness seem to stagnate in their effect on diagnosticity and become increasingly marginal. Their interaction effect also shows that the coverage of aspects carries only little importance for long reviews while it is important for shorter ones.

Discussion and Implications

While available data continues to grow at exponential rates (IDC, 2014), human beings remain limited in their information processing capabilities. It is therefore imperative to better understand how to design the systems that will enable humans to benefit from the increasing “digital mediation of everyday experiences” (Yoo, 2010, p. 215). Seeking information as well as leaving information behind for others to read are fundamental aspects of the human experience, and they are increasingly shaped by information systems. Thus, review diagnosticity, or the extent to which an individual believes that an online opinion offers helpful information to evaluate the target of the review, should be a key concern in the design of online opinion platforms. While people write online reviews for a multitude of reasons (Cheung and Lee, 2012), obtaining helpful decision-making information is the principal driver (Hennig-Thurau and Walsh, 2003). Their design should therefore make review diagnosticity the primary metric of success and understanding its drivers should be a key concern for IS scholars.

Numerous studies have shown that online reviews have an impact on sales (Ghose and Ipeirotis, 2009; Sonnier, McAlister and Rutz, 2011; Lu, Ba, Huang and Feng, 2013), sales ranks (Archak et al., 2001; Chevalier and Mayzlin, 2006; Ghose et al., 2012), sale growth rates (Clemons et al., 2006), box office revenues (Dellarocas, Zhang and Awad, 2007; Duan et al., 2008; Liu, 2006), conversion rates (Ludwig, de Ruyter, Friedmann, Brüggen, Wetzels and Pfann, 2013) and price premiums (Pavlou and Dimoka, 2006). Moreover, recent research has addressed the extent to which online reviews help individuals make decisions (e.g., Baek et al., 2012; Korfiatis et al., 2012; Forman et al., 2008; Li and Zhan, 2011; Pan and Zhang, 2011; Mudambi and Schuff, 2010; Pavlou and Dimoka, 2006). However, no research to date has leveraged the thematic structures of the textual content of online opinions antecedents to explain review diagnosticity.

Thus, our work contributes to the information systems literature in multiple ways. First, we are able to extract the features of experiential product reviews using probabilistic topic modeling. We complement previous work that has analyzed textual review content (e.g., Schindler and Bickart, 2012; Li and Zhan, 2011; Pavlou and Dimoka, 2006) by automating topic extraction rather than relying on human raters. Further, we algorithmically categorize all 77,864 reviews in our sample, focusing on key features of the service experience—the actual subject matter of the reviews. Thus our work complements previous research that uses quantifiable, but arguably tangential, elements of text, including spelling errors (Ghose and Ipeirotis, 2009; Schindler and Bickart, 2012; Scholz and Dorner, 2013) or average word length (Cao et al., 2011). In short, we respond to recent calls for focusing on the essential elements of computer-mediated communication in e-commerce (Dimoka et al., 2012). We do so in the context of experiential products, an important but under-researched target of online opinions.

With respect to the contribution of our results, we replicate the findings of previous research and show that review depth is a significant determinant of review diagnosticity for online reviews (H1a). We extend previous work (e.g., Mudambi and Schuff, 2010; Baek et al., 2012; Korfiatis et al., 2012) to show that the positive impact of review depth is tampered by a decreasing marginal effect (H1b). These results are important because they indicate that the more an individual writes about his or her experiences, the higher the likelihood that other individuals, despite their idiosyncratic expectations and information needs, will find the opinion valuable for decision-making. However, longer reviews are more cognitively taxing and time consuming to process. As humans seek to minimize the cognitive efforts involved in forming judgments (Bettman et al., 1998), the informational value of longer reviews is dampened by the very driver of diagnosticity: quantity of information.

Furthermore, our study is the first to consider the effect of information completeness on review diagnosticity and showcasing its positive (H2a), yet marginally diminishing (H2b) effect. While information quality has long been theorized as a key element of IS success (DeLone and McLean, 1992), its formative elements—among them, information completeness—remain mostly ill-defined (Knight and Burn, 2005; Petter, DeLone and McLean, 2013). Currency, format and accuracy for example—other, frequently cited formative elements of information quality (Nelson et al., 2005; Setia et al., 2013)—are already pre-defined in the context of online reviews, simply by the choice of the medium. For example, a review's currency is captured via a timestamp; a review's format follows a pre-set layout and structure, typically including star rating scales and open text boxes; and a review's accuracy is assumed (rightfully or not) and sometimes even enhanced through the disclosure of a reviewer's identity (Forman et al., 2008). In contrast, a review's completeness is entirely at the discretion of the individual, and thus it is the most variable element of information quality. By looking at completeness in the online review context, we were able to look at one aspect of information quality in isolation, thus precisely evaluating its discriminatory effect in relation to perceptions of diagnosticity.

Depth and completeness, in combination, also shed light on the notion of information conciseness—a concept that has not been clearly defined by IS research (Petter et al., 2013). The study proposes a new conceptualization of information conciseness based upon the interplay of depth and completeness in relation to diagnosticity (H3). Online reviews are a means of computer-mediated communication that is subject to some fundamental principles ensuring that information can be shared effectively among individuals (Grice, 1975; 2008; Wänke, 2007; Koch et al., 2013). In this context, information conciseness is a fundamental principle, suggesting a delicate balance between depth and completeness by providing "just enough" and "just the right" information. Apart from introducing information conciseness as the interaction of depth and completeness, our study suggests that a review, on average, is more likely to offer helpful information when its content provides neither too much nor too little, neither redundant nor irrelevant, information in order to evaluate an experiential product. Furthermore, our study suggests that IS researchers, in order to ensure effective computer-mediated communication, have to pay closer attention to factors beyond mere information quality. Since the traditional perspective of information quality (Nelson et al., 2005; Setia et al., 2013) does not examine the quantity of textual information provided, studying conciseness as a new informational attribute is valuable. Over time, IS researcher might even be able to infer a set of pragmatic rules to guide contributors in providing highly diagnostic online reviews or to define design elements of online platforms that provide diagnosticity maximizing structure to online reviews. Since contributors draw a high level of personal fulfillment from establishing a credible online persona by volunteering time and writing skills (Forman et al., 2008), providing guidelines that reduce these efforts would be beneficial.

Our findings also help to extend the notion of diagnosticity beyond online retailers (Jiang and Benbasat, 2007; Pavlou et al., 2007; Dimoka et al., 2012; Xu, Benbasat and Cenfetelli, 2014). Diagnosticity applies to any computer-mediated communication that aids the formation of judgment irrespective of its source—in fact, upon replication and corroboration, our findings may be generalizable to any experiential encounter between individuals. For example, online dating studies have shown that digital profile descriptions contain diagnostic information (apart from visuals) that individuals use to formulate perceptions about a candidate's attractiveness (e.g., Brand et al., 2012; Toma and Jeffrey, 2010). Likewise, online recruitment systems capture information that is diagnostic in nature, for both applicant and recruiting organization (Dineen, Ash and Noe, 2002). Whenever human experiences are described, diagnostic information is important in order to assist in evaluations and decision-making. As experiential computing is increasingly gaining the attention of IS researchers (Yoo, 2010), being able to capture the varying levels of diagnosticity of these experiences might prove essential.

Implications for Practice

Our findings may prompt IS practitioners to consider two things. First, they need to make sure to capture the different facets of an experience. This could take place, for example, by providing multiple textual entry fields, each soliciting textual input for different topics or, more subtly, by providing real-time feedback while users are typing the review. For example, the system may suggest adding information about missing features (e.g., the location) when a review is particularly short. This algorithmic feedback would change dynamically depending on the missing elements of the review that would maximally improve the review's expected diagnosticity.

Second, practitioners need to ensure that ample space is provided to capture each facet of an experience while keeping both a lower and upper limit in mind. While intuitively simple and easy to do, this suggestion may be difficult to implement as more and more reviews are contributed via mobile platforms. As with review completeness, the optimal review depth is unlikely to be a static number. Thus, educating online reviewers about how to write a diagnostic review is important. By institutionalizing writing guidelines or by editorially identifying high quality reviews, sites, such as Yelp or Tripadvisor, could promote "golden standards" for review completeness and depth. Review diagnosticity may also become an important element in the ranking of opinions for display to users.

Our study is not without limitations. It focuses on one aggregator site and one type of experiential product. While this choice is pragmatic it limits the generalizability of our findings. A comparison of the thematic structure of search versus experiential products and a replication across different review platforms is needed. Future work should also refine our methodology. The aggregator site we used did not track the total number of views for each opinion. While we control for the number of days the review was available on the first page and the overall traffic on the site, future study that can control for actual exposure are needed.

Conclusion

This study was set out to demonstrate the importance of review content. By developing notions of review depth and review completeness along with the notion of conciseness, we were able to show that both the variety of topics a review reflects upon and the detail with which the experience is discussed has an impact on the perceived level of diagnosticity this review portrays. Understanding diagnosticity in more detail will allow researchers and organizations alike to better understand how individuals form judgments in computer-mediated contexts and what informational pieces are considered particularly diagnostic. Thus, our work responds to recent calls for focusing on review content as an essential part of computer-mediated communication in e-commerce and will contribute to helping designers develop online review systems that maximize value for users.

References

Aboulnasr, K. (2006). Country of Origin Effects: The Role of Information Diagnosticity, Information Typicality and Involvement. *Marketing Management Journal,* 16(1), 1-18.

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

Atifi, H., Mandelcwaig, S. and Marcoccia, M. (2011). The Co-operative Principle and Computer-Mediated Communication: The Maxim of Quantity in Newsgroup Discussions. *Language Sciences,* 33, 330-340.

Baek, H., Ahn, J. and Choi, Y. (2012). Helpfulness of Online Consumer Reviews: Readers' Objectives and Review Cues. *International Journal of Electronic Commerce,* 17(2), 99-126.

Bates D., M. Maechler, B. Bolker, and Walker, S. (2014). lme4: Linear Mixed-Effects Models using Eigen and S4. R package version 1.0-6. Retrieved August, 2014, from http://CRAN.R-project.org/package=lme4.

Bettman, J. R., Luce, M. F. and Payne, J. W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research,* 25(3), 187-217.

Blei, D. M. (2012). Probabilistic Topic Models. *Communications of the ACM,* 55(4), 77-84.

Blei, D. M., Ng, A. Y., Jordan, M. I., and Lafferty, J. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research,* 3(4), 993-1022.

Brand, R. J., Bonatsos, A., D'Orazio, R. and DeSong, H. (2012). What is Beautiful is Good, even Online: Correlations between Photo Attractiveness and Text Attractiveness in Men's Online Dating Profiles. *Computers in Human Behavior,* 28(1), 166-170.

Cao, Q., Duan, W., and 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.

Cheung, M. K. and Lee, M. K. O. (2012). What Drives Consumers to Spread Electronic Word of Mouth in Online Consumer-Opinion Platforms. *Decision Support Systems,* 53(1), 218-225.

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

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

Daft, R. L., Lengel, R. H., and Trevino, L. K. (1987). Message Equivocality, Media Selection, and Manager Performance: Implications for Information Systems. *Management Information Systems Quarterly,* 11(3), 354-366.

Dedeke, A. (2000). A Conceptual Framework for Developing Quality Measures for Information Systems. *Proceedings of the 5th International Conference on Information Quality*, 126-128.

Dellarocas, C., Zhang, X., and Awad, N. F. (2007). Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures. *Journal of Interactive Marketing,* 21(4), 23-45.

DeLone, W. H., and McLean, E. R. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research,* 3(1), 60-95.

Dewey, J. (1910). *How We Think*, Boston: Health & Co.

Dimoka, A., Hong,Y., and Pavlou, P. (2012). On Product Uncertainty in Online Markets: Theory and Evidence. *Management Information Systems Quarterly,* 36(2), 395-426.

Dineen, B. R., Ash, S. R. and Noe, R. A. (2002). A Web of Applicant Attraction: Person-Organization Fit in the Context of Web-Based Recruitment. *Journal of Applied Psychology,* 87(4), 723-734.

Duan, W., Gu, B., and Whinston, A. B. (2008). Do Online Reviews Matter? — An Empirical Investigation of Panel Data. *Decision Support Systems,* 45(4), 1007-1016.

Ellison, N., Heino, R. and Gibbs, J. (2006). Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment. *Journal of Computer-Mediated Communication,* 11(2), 415-441.

Eppler, M., and Muenzenmayer, P. (2002). Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology. *Proceedings of 7th International Conference on Information Quality,* 187–196.

Forman, C., Ghose, A., and Wiesenfeld, B. (2008). Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research,* 19(3), 291-313.

Ghose, A., and Ipeirotis, P. G. (2009). 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.

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

Grice, H. P. (1975). Logic and Conversation. In: P. Cole and J. Morgan (eds) *Speech Acts*, New York: Academic Press, 41-58.

Grice, H. P. (2008). Further Notes on Logic and Conversation. In: J. E. Adler and L. J. rips (eds), *Reasoning: Studies of Human Inference and Its Foundations*, New York: Cambridge University Press, 765-773.

IDC. (2014). The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. IDC White Paper. Retrieved May, 2014, from http://idcdocserv.com/1678.

Ha, Y., and Hoch, S. J. (1989). Ambiguity, Processing Strategy, and Advertising-Evidence Interactions. *Journal of Consumer Research,* 16(3), 354-360.

Hennig-Thurau, T. and Walsh, G. (2003). Electronic Word-Of-Mouth: Motives for and Consequences of Reading Customers Articulations on the Internet. *International Journal of Electronic Commerce,* 8(2), 51-74.

Herring, S., Stein, D. and Virtanen, T. (2013). *Pragmatics of Computer-Mediated Communication*, Berlin, Germany: De Gruyter Mouton.

Huang, P., Lurie, N. H., and Mitra, S. (2009). Searching for Experience on the Web: An Empirical Examination of Consumer Behavior for Search and Experience Goods. *Journal of Marketing,* 73(2), 55-69.

Jiang, Z., and 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., and Benbasat, I. (2007). The Effects of Presentation Formats and Task Complexity on Online Consumers' Product Understanding. *Management Information Systems Quarterly,* 31(3), 475-500.

Kahn, B. K., Strong, D. M., and Wang, R. Y. (2002). Information Quality Benchmarks: Product and Service Performance. *Communications of the ACM,* 45(5), 84-192.

Keller, K. L., and Staelin, R. (1987). Effects of Quality and Quantity of Information on Decision Effectiveness. *Journal of Consumer Research,* 14(2), 200-213.

Keller, K. L., and Staelin, R. (1989). Assessing Biases in Measuring Decision Effectiveness and Information Overload. *Journal of Consumer Research,* 15(4), 504-508.

Knight, S., and Burn, J. (2005). Developing a Framework for Assessing Information Quality on the World Wide Web. *Informing Science,* 8(01), 159-172.

Koch, A. S., Forgas, J. P., and Matovic, D. (2013). Can Negative Mood Improve Your Conversation? Affective Influences on Conforming to Grice's Communication Norms. *European Journal of Social Psychology,* 43, 326-334.

Korfiatis, N., García-Bariocanal, E., and 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.

Li, J., and Zhan, L. (2011). Online Persuasion: How the Written Word Drives WOM: Evidence from Consumer-Generated Product Reviews. *Journal of Advertising Research,* 51(1), 239-257.

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

Lu, B., Ott, M., Cardie, C., and Tsou, B. K. (2011). Multi-Aspect Sentiment Analysis with Topic Models. *IEEE International Conference on Data Mining Workshops*, 81-88.

Lu, X., Ba, S., Huang, L. and Feng, Y. (2013). Promotional Marketing or Word-of-Mouth? Evidence from Online Restaurant Reviews. *Information Systems Research,* 24(3), 596-612.

Ludwig, S., de Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., and Pfann, G. (2013). More than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing,* 77(1), 87-103.

Lynch Jr., J. G., Marmorstein, H., and Weigold, M. F. (1988). Choices from Sets Including Remembered Brands: Use of Recalled Attributes and Prior overall Evaluations. *Journal of Consumer Research,* 15(2), 169-184.

Madlberger, M., and Nakayama, M. (2013). On Top of The World, Down in the Dumps: Text Mining the Emotionality of Online Consumer Reviews. *21st European Conference on Information Systems*, 1-12.

Malhotra, N. K., Jain, A. K., and Lagakos, S. W. (1982). The Information Overload Controversy: An Alternative Viewpoint. *Journal of Marketing,* 46, 27-37.

Menon, G., Raghubir, P. and Schwarz, N. (1995). Behavioral Frequency Judgments: An Accessibility-Diagnosticity Framework. *Journal of Consumer Research,* 22, 212-228.

Mudambi, S. M., and Schuff, D. (2010). What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.Com. *Management Information Systems Quarterly,* 34(1), 185-200.

Mukherjee, A., and Hoyer, W. D. (2001). The Effect of Novel Attributes on Product Evaluation. *Journal of Consumer Research,* 28(3), 462-472.

Nakagawa, S., and Schielzeth, H. (2013). A General and Simple Method for Obtaining R2 from Generalized Linear Mixed-effects Models. *Methods in Ecology and Evolution,* 4(2), 133-142.

Naumann, F. and Rolker, C. (2000). Assessment Methods for Information Quality Criteria. *Proceedings of 5th International Conference on Information Quality*, 148–162.

Nelson, P. (1970). Information and Consumer Behavior. *Journal of Political Economy,* 78(03), 311-329.

Nelson, P. (1974). Advertising as Information. *Journal of Political Economy,* 82(07), 729-754.

Nelson, R. R., Todd, P A., and Wixom, B. H. (2005). Antecedents of Information and System Quality: An Empirical Examination within the Context of Data Warehousing. *Journal of Management Information Systems,* 21(4), 199-235.

Pan, Y., and Zhang, J. Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing,* 87(4), 598-612.

Pavlou, P. A., and 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.

Pavlou, P. A., Liang, H., and Xue, Y. (2007). Understanding and Mitigating Uncertainty in Online Exchange Relationships: A Principal--Agent Perspective. *Management Information Systems Quarterly,* 31(1), 105-136.

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1988). Adaptive Strategy Selection in Decision Making. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 14(3), 534-552.

Petter, S., DeLone, W., and McLean, E. R. (2013). Information Systems Success: The Quest for the Independent Variables. *Journal of Management Information Systems,* 29(4), 7-62.

Price, R. and Shanks, G. (2005). A Semiotic Information Quality Framework: Development and Comparative Analysis. *Journal of Information Technology,* 20, 88-102.

R Core Team. (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved August, 2014, from http://www.R-project.org.

Salganik, M. J. and Watts, D. J. (2008). Leading the Herd Astray: An Experimental Study of Self-Fulfilling Prophecies in an Artificial Cultural Market. *Social Psychology Quarterly*, 71, 338-355.

Schindler, R. M., and Bickart, B. (2012). Perceived Helpfulness of Online Consumer Reviews: The Role of Message Content and Style. *Journal of Consumer Behaviour,* 11(3), 234-243.

Schroeder, H. M, Driver, M. J., and Streufert, S. (1967). *Human Information Processing*, New York: Holt, Rinehart and Winston.

Scholz, M., and Dorner, V. (2013). The Recipe for the Perfect Review? An Investigation into the Determinants of Review Helpfulness. *Business & Information Systems Engineering,* 5(3), 141-151.

Sela, A., and Berger, J. (2012). How Attribute Quantity Influences Option Choice. *Journal of Marketing Research,* 49(6), 942-953.

Setia, P., Venkatesh, V., and Joglekar, S. (2013). Leveraging Digital Technologies: How Information Quality Leads to Localized Capabilities and Customer Service Performance. *Management Information Systems Quarterly,* 37(2), 565-590.

Sonnier, G. P., McAlister, L., and Rutz, O. J. (2011). A Dynamic Model of the Effect of Online Communications on Firm Sales. *Marketing Science,* 30(4), 702-716.

Toma, C. L., and Hancock, J. T. (2012). What Lies Beneath: The Linguistic Traces of Deception in Online Dating Profiles. *Journal of Communication,* 62(1), 78-97.

Toutanova, K., Klein, D., Manning, C. D., and Singer, Y. (2003). Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. *Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, May 27-June 1, Edmonton, Canada.

Tsang, A. S. L. and Prendergast, G. (2009). Is a "Star" Worth a Thousand Words? *European Journal of Marketing,* 43(11/12), 1269-1280.

Tversky, A., and Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science,* 185(4157), 1124-1131.

Wänke, M. (2007). What is Said and What is Meant: Conversational Implications in Natural Conversations, Research Settings, Media and Advertising. In: K. Fiedler (ed.) *Social Communication*, New York: Psychology Press, 223-256.

Wang, R. Y. and Strong, D. M. (2006). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems,* 12(4), 5-33.

Walther, J. B., Liang, Y., Ganster, T., Wohn, D. Y., and Emington, J. (2012). Online Reviews, Helpfulness Ratings, and Consumer Attitudes: An Extension of Congruity Theory to Multiple Sources in Web 2.0. *Journal of Computer-Mediated Communication,* 18(1), 97-112.

Weiss, A. M., Lurie, N. H., and MacInnis, D. J. (2008). Listening to Strangers: Whose Responses are Valuable, How Valuable are They, and Why? *Journal of Marketing Research,* 45(4), 425-436.

Xu, J., Benbasat, I., and Cenfetelli, R. T. (2014). The Nature and Consequences of Trade-Off Transparency in the Context of Recommendation Agents. *Management Information Systems Quarterly,* 38(2), 379-406.

Yoo, Y. (2010). Computing in Everyday Life: A Call for Research on Experiential Computing. *Management Information Systems Quarterly,* 34(2), 213-231.

Zhang, J. Q., Craciun, G., and Shin, D. (2010). When does Electronic Word-of-Mouth Matter? A Study of Consumer Product Reviews. *Journal of Business Research,* 63(12), 1336-1341.

Appendix

In order to illustrate the LDA algorithm used in this study, take the following original review.

> *"My week-long stay at the Chelsea Pines in may marked my second visit to this charming boutique hotel near the meatpacking district, high line park, and Chelsea Market. This is a very personal establishment; every member of the staff team relates to each guest not just as a patron customer but as a friend. This personal touch really pays dividends when you come inside the door, morning noon or night, and you are greeted warmly with welcoming conversation (and the staff has the sense to leave you alone if you appear to be hungover [sic] or otherwise under the weather). The festive atmosphere is contagious such that all the guests feel like included family, and the breakfast room and deck leading outside give travelers a place to come together and share their experiences. The all-day breakfast room, by the way, is a real highlight with the thoughtful inclusions of the complete gamut of what a continental breakfast can be: nespresso machine, a smorgasbord of teas and other beverages, luscious fresh fruits, high-end yogurts, sumptuous breads and pastries, imaginative cereals, even peanut butter (a personal favorite). I also like the fact that one can use the hotel's refrigerator to stash a bottle of wine or nosh. The establishment's silver screen theme is amusing and entertaining (i.e., in which movie star's room are you staying?*
> *Sophia Loren's or Rita Hayworth's or Albert Finney's or ??) all in all, this place is unique, distinctive and memorable, and I look forward to coming back again and again."*

First, we split each review into its sentences (see Table below). Next, LDA computes topic probabilities for the various features discussed in each sentence. We assign a topic to a sentence if the maximum θ weight of at least one feature (e.g., room, location) is greater than 0.7; if no feature is characterized by a θ weight above the threshold, we mark the sentence as "undefined." We compute completeness by counting the number of unique topics covered by the full review (here: 4).

	Service	Value	Location	Room	Food	Feature
my week-long stay at the chelsea pines in may <number> marked my second visit to this charming boutique hotel near the meatpacking district , high line park , and chelsea market .	0.00	0.35	0.65	0.00	0.00	undefined
this is a very personal establishment ; every member of the staff team relates to each guest not just as a patron customer but as a friend .	0.96	0.03	0.00	0.00	0.00	service
this personal touch really pays dividends when you come inside the door , morning noon or night , and you are greeted warmly with welcoming conversation (and the staff has the sense to leave you alone if you appear to be hungover or otherwise under the weather) .	0.95	0.00	0.00	0.01	0.04	service
the festive atmosphere is contagious such that all the guests feel like included family , and the breakfast room and deck leading outside give travelers a place to come together and share their experiences .	0.04	0.56	0.00	0.02	0.38	undefined
the all-day breakfast room , by the way , is a real highlight with the thoughtful inclusions of the complete gamut of what a continental breakfast can be : nespresso machine , a smorgasbord of teas and other beverages , luscious fresh fruits , high-end yogurts , sumptuous breads and pastries , imaginative cereals , even peanut butter (a personal favorite) .	0.00	0.00	0.00	0.00	0.99	food
i also like the fact that one can use the hotel's refrigerator to stash a bottle of wine or nosh .	0.00	0.00	0.00	0.02	0.97	food
the establishment's silver screen theme is amusing and entertaining (i.e. , in which movie star's room are you staying <question>	0.02	0.08	0.05	0.79	0.06	rooms
sophia loren's or rita hayworth's or albert finney's or <question>	0.95	0.02	0.01	0.01	0.01	service
) all in all , this place is unique , distinctive and memorable , and i look forward to coming back again and again .	0.00	0.97	0.02	0.01	0.00	value

We selected seed words to yield topics that were maximally informative, with respect to the characteristics of the experience, and maximally discriminating (i.e., orthogonal). We therefore chose only nouns (no adjectives), referring to the defining elements of the lodging experience: service, value, location, room and food. We did not choose any element that would have to be described in terms of adjectives (e.g., cleanliness). Despite the difficulty in identifying discriminant nouns for the notion of value, we retained such a concept given its importance. To select the specific terms (see Table below), we inspected high-frequency terms and chose the highest frequency terms in our corpus that were coherent with the five elements.

Topic	**Service**	**Value**	**Location**	**Room**	**Food**
Seed 1	Service	Value	Location	Room	Food
Seed 2	Staff	Price	Place	Bed	Breakfast
Seed 3	Desk	Rate	Area	Bathroom	Bar
Seed 4	Reservation	Money	View	Shower	Restaurant

The defining terms for each topic are listed below together with their probability of occurrence in this topic:

Service:	Value:	Location:	Room:	Food:
staff 0.048496	hotel 0.054521	hotel 0.039597	room 0.066664	breakfast 0.034507
friendly 0.022181	stay 0.041433	location 0.037858	rooms 0.027906	free 0.018126
desk 0.021314	stayed 0.030259	<number> 0.025031	clean 0.025743	room 0.017977
room 0.020792	<number> 0.029641	great 0.023377	bed 0.018984	good 0.017421
helpful 0.018493	<exclamation> 0.021368	walk 0.018041	comfortable 0.018038	hotel 0.015699
front 0.018148	great 0.011295	close 0.014786	bathroom 0.015455	bar 0.014993
hotel 0.018007	time 0.010698	walking 0.012498	<number> 0.013594	food 0.014787
service 0.014592	nights 0.010229	restaurants 0.011965	small 0.012306	<exclamation> 0.014668
<number> 0.011723	recommend 0.009888	street 0.011946	nice 0.012041	restaurant 0.014468
<exclamation> 0.010752	hotels 0.009634	area 0.011692	hotel 0.011732	great 0.014451
check 0.010146	place 0.009303	square 0.010673	floor 0.011088	nice 0.013976
time 0.007410	night 0.009259	park 0.009274	beds 0.009382	service 0.013017
told 0.006430	price 0.008986	distance 0.009048	shower 0.009251	pool 0.011271

The Table below offers a variety of sample sentences for each category:

Food	the all-day breakfast room, by the way, is a real highlight with the thoughtful inclusions of the complete gamut of what a continental breakfast can be: nespresso machine, a smorgasbord of teas and other beverages, luscious fresh fruits, high-end yogurts, sumptuous breads and pastries, imaginative cereals, even peanut butter (a personal favorite).
	this hotel offered wine and hot toddies, yum <exclamation>
	i still love the apple pie for room service, but that's it.
Location	not far from the blue and red line.
	in walking distance to space needle and pike place market + many restaurants and shops.
	so landing close to midnight with a domestic flight the following day, this hotel was a 10min walk from terminal e so very well located.
Room	the bathroom is very small and no vent or fan makes it very uncomfortable to shower/get dressed in the bathroom.
	everything about it was so luxurious from the incredibly comfortable beds to the well appointed bathroom.
	it was a spacious room with a large bathroom.
Service	they were very attentive and friendly.
	at front desk the workers were friendly, but it's not enough.
	she prepared the paperwork for us to check in (which we were several hours early for) and told us that she would call my cell phone when the room was ready.
Undefined	parking was <money> a day if you parked your car yourself and <money> if hyatt staff parked it.
	because <question>
	we did not stay at the actual hotel, but did walk through.
Value	i met up with a girlfriend for a weekend in san francisco, and hotel fusion was perfect for what we needed.
	the orchard is very high in tripadvisor ranking and i think it deserve it.
	i enjoyed it and would recommend.