Naive Or Savvy: How Credible Are Online Reviews For Credence Services?

Katherine Stewart, University of Maryland, USA
Rebecca Hamilton, University of Maryland, USA
Shannon Lantzy, University of Maryland, USA

Because consumers cannot assess the quality of credence attributes (e.g., a doctor’s skill in diagnosis), reviews of these attributes should be discounted by other consumers. We examine the claims made in reviews of credence vs. experience services (e.g., doctors vs. hair stylists) and how consumers interpret these claims.

[to cite]:

[url]:
http://www.acrwebsite.org/volumes/1014903/volumes/v41/NA-41

[copyright notice]:
This work is copyrighted by The Association for Consumer Research. For permission to copy or use this work in whole or in part, please contact the Copyright Clearance Center at http://www.copyright.com/.
Let’s Talk About It: Factors Influencing Word-of-Mouth Content

Chairs: Yu-Jen Chen, Lingnan University, Hong Kong, China
Rebecca Hamilton, University of Maryland, USA

Paper #1: Who’s Driving This Conversation? Systematic Biases in the Content of Online Consumer Discussions
Rebecca Hamilton, University of Maryland, USA
Ann E. Schlosser, University of Washington, USA
Yu-Jen Chen, Lingnan University, Hong Kong, China

Paper #2: When Do People Talk About and Why?
Evan Weingarten, University of Pennsylvania, USA
Jonah Berger, University of Pennsylvania, USA

Paper #3: The Content and Impact of Mobile Versus Desktop Reviews
Nicholas Lurie, University of Connecticut, USA
Sam Ransbotham, Boston College, USA
Hongju Liu, University of Connecticut, USA

Paper #4: Naive or Savvy: How Credible Are Online Reviews for Credence Services?
Shannon Lantz, University of Maryland, USA
Katherine Stewart, University of Maryland, USA
Rebecca Hamilton, University of Maryland, USA

SESSION OVERVIEW

Consumers often share information about their consumption experiences and discuss products and services with others online (e.g., by reviewing a service provider or participating in a discussion forum). Although previous research has provided insight about why consumers post content (e.g., Schlosser 2005) and how quantitative metrics such as review volume and ratings change over time (e.g., Moe and Schweidel 2012), we know much less about the factors that influence the qualitative content consumers share. An understanding of online content is critical because decades of research on social influence show that what consumers say influences others (e.g., strong vs. weak arguments: Petty and Cacioppo 1986; vividness effect: Herr, Kardes and Kim 1991).

The four papers in this session, all at an advanced stage, leverage both large datasets from controlled experiments to examine factors that influence word of mouth (WOM) content. Specifically, they examine whether what consumers say differs when they are:

- the first to post a response vs. the second or third?
- posting about the past or future instead of the present?
- posting a review on a mobile device vs. on their desktop computer?
- reviewing a provider whose services are harder vs. easier to evaluate?

Hamilton, Schlosser and Chen show that early responses to a post affect later responses to the post, suggesting that the first people to respond play a crucial role in driving the content of the discussion. Weingarten and Berger demonstrate that there is a strong present bias in what people talk about but that people tend to talk about the near future and distant past. This suggests that accessibility and emotional intensity affect when people talk about. The next two papers examine the effects of sharing devices and the type of experience. Lurie, Ransbotham and Liu find that restaurant reviews written on mobile devices are more affective and more negative than those written by the same reviewer on desktops. Notably, mobile reviews tend to be rated by others as less useful, perhaps because they are shorter and more emotional. Finally, Lantz, Stewart and Hamilton also examine reviews of service providers and find that negative reviews contain higher-quality arguments and more information than positive reviews for the same types of service providers. They show that reviews of credence service providers (e.g., doctors, auto mechanics) include more claims about easier-to-evaluate experience attributes than claims about harder-to-evaluate credence attributes.

Consumers rely on online content to make important decisions. We aim to “Make a Difference” by identifying systematic biases that affect the information being shared online. We expect this session to generate discussion about the emotional and affective nature of online content, the influence of previous posts (either by the same person or by others) on subsequent posts, and the various methods that can be used to study WOM. This session will appeal to researchers interested in the antecedents and consequences of WOM and social media marketing as well as those interested in social influence and communication.

Who’s Driving This Conversation? Systematic Biases in the Content of Online Consumer Discussions

EXTENDED ABSTRACT

Consumers rely on online word of mouth to make consumption decisions partly because they believe the information provided helps them make better-informed decisions. However, participants in online discussion forums may have multiple goals, only one of which is to provide others with accurate information. For example, participants may want to signal expertise (Schlosser 2005; Wojnicki and Godes 2012), convey uniqueness (Schlosser 2009) or build connections with others (e.g., Hennig-Thurau et. al. 2004; Schlosser 2009). Because online posting is context-dependent communication in which participants read previous posts within a thread before offering their own opinions (Moe and Schweidel 2012), the information shared in online forums will be influenced by previous posts. Therefore, it’s likely that early responses would drive the discussion towards or away from the information requested in the initial post.

We closely examine the role of early responses on later responses in online discussion threads in which the consumer seeking information explicitly tells other participants which decision criteria are important. For example, a consumer posting on TripAdvisor.com might describe an upcoming trip and ask for hotel recommendations, noting her preferences for location, price level and other services such as parking. Despite the objective value of the information that has been requested, we predict that the discussion content will show systematic biases. For example, we expect already mentioned attributes to be more rather than less likely to be mentioned in subsequent responses. Research on social influence has shown that group members tend to repeat shared information (Larson, Foster-Fishman and Keys 1994; Mojzisch et al. 2010; Stewart and Stasser 1995). Thus, we expect information provided in online discussion forums to be driven by early responses to the query: instead of equally reflecting attributes mentioned in the query, we expect the content to focus more on attributes already mentioned. We suggested that this is because consumers exhibit conformity to express affiliation toward others.

We test our predictions by analyzing public data from online discussion forums and conducting controlled experiments. Our analysis of discussion threads from three online discussion forums (on...
Tripadvisor.com and DISboards.com) provides empirical evidence that early responses to a query drive the subsequent discussion. We examined two different consumption categories (hotels and restaurants) in three cities (Seattle, Washington DC, and Orlando). Our sample included 324 posts from 85 discussion threads with at least three participants and the goal of seeking advice about a specific decision (e.g., Should I stay at hotel A or B?) between January 2011 and January 2013. After identifying the attributes most frequently mentioned for each product category (e.g., location, price), two raters coded each response to indicate whether each attribute was mentioned. Because each response could mention multiple attributes, we used a Generalized Estimating Equations approach (GEE; Liang and Zeger 1986) to analyze the data. Controlling for the importance of the product attributes, we find that whether previous responses mention an attribute is more predictive of whether subsequent responses mention the attribute than whether the initial post mentions the attribute. This finding suggests that previous responses to a post play a critical role in influencing the attributes discussed in subsequent responses.

In Study 2, participants were given information about the attributes of two restaurants. Using this information, participants were asked to respond to a posted question asking about the appropriateness of the two restaurants for an occasion involving a large group of friends. The number of previous responses to the question was varied so that participants were either providing the first, second or third response. The content of the previous responses was also varied so that previous posts either mentioned the critical attribute (the suitability of the restaurant for groups) or did not (and mentioned the atmosphere of the restaurant instead). Consistent with our manipulation of the importance of the critical attribute, when there were no previous responses, more participants mentioned the restaurants’ suitability for groups (M = 70%) and fewer mentioned atmosphere (M = 43%). Importantly, more participants mentioned the restaurants’ suitability when all other previous posts mentioned suitability for groups (M = 71%) than when no previous posts mentioned suitability for groups (M = 48%). Notably, when some responses mentioned suitability for groups and others did not, participants were also less likely to mention the important attribute (M = 52%). As in the real discussion threads we analyzed, participants in this study seem to be more strongly influenced by the initial responses to the post than to the query.

Study 3 allowed participants to communicate their own knowledge in their response to a post about a real event in which they were familiar. All participants read an inquiry post and a first response about the event. We manipulated the content of the first response in terms of the attribute being mentioned. The results were consistent with Study 2 such that participants were more likely to be influenced by the first response on attribute that is critical (M_{mention} = 50% vs. M_{not mention} = 13%). However, when the attribute is not critical, whether the first response mentioned did not change participants’ response (M_{mention} = 45% vs. M_{not mention} = 49%)

Study 4 examined whether the perceived expertise of the first responder moderates the effect. The procedure was identical to study 2 except that we manipulated expertise by describing the first responder as a frequent or first-time traveler to the Boston area. Although we replicated the key finding of our first lab study, showing that whether the first response mentions the key attribute has a significant effect on whether the participant mentions it (M = 79% vs. 46%), we did not observe an attenuation of the effect when the first responder was perceived to be less expert. This suggests that the influence of the first responder derives not from his or her expertise but from social incentives to focus on shared rather than unshared information.

In our last study, we manipulate the participant’s goals to provide accurate information in response to the query or form relationships with group members. The procedure was similar to study 2 except that participants were encouraged to adopt either a group goal or an accuracy goal. As predicted, we replicated the early response effect only in the group goal condition, suggesting that participants’ goals moderate the effect, suggesting that participants are responding to normative influence within the group setting.

When Do People Talk About and Why?

EXTENDED ABSTRACT

People talk about things that occur over various time horizons into the past, present, and future. They talk about the television show they are currently watching, the groceries they bought yesterday, and the restaurant they are going to tomorrow.

But, when are people more likely to talk about, and why? That is, how does the likelihood people talk about something vary based on its temporal position from the present, and how does this pattern inform us about the drivers of interpersonal communication?

We suggest two key factors shape when people talk about. The first is accessibility. People are more likely to talk about things that are accessible (Berger and Schwartz 2011). Temporally distant events are less accessible than temporally close events (D’Argembeau and Van der Linden 2004). In the context of when people talk about, that suggests that (1) individuals should share more about the present than the past and future, (2) that temporally close events should be shared more frequently than temporally distant events, (3) that people should talk more about the future later in the week when they get closer to the weekend, and (4) that people should talk more about the past 24 hours later in a given day. Second, arousal impacts sharing (Berger 2011; Berger and Milkman 2012) and emotional arousal should also shape when people talk about. Past research has found the future to be more emotionally intense than the past (D’Argembeau and Van der Linden 2004; Van Boven and Ashworth 2007), more positive (Newby-Clark and Ross 2003), and less constrained (Van Boven, Kane, and McGraw 2008). Consequently, while people may talk about the near past than near future (because it is more accessible), they may talk about the distant future rather than distant past (because it is more emotionally intense).

To provide a naturalistic examination of when people talk about, we collected a large sample of everyday conversations. We randomly selected about 3,000 tweets posted in the public twitter stream, 3,000 Facebook posts taken from an online database of Facebook data (Kosinski and Stillwell 2011), and hundreds of pages of transcribed in-person conversations from a written corpus. Next, two research assistants removed spam posts from the Twitter and Facebook data set. Then, three research assistants individually coded each message in Excel on a continuous scale in terms of fraction of a day from the present (-7 = a week ago, -1/24 = an hour ago, 0 = now, 1 = a day from now, 365 = a year from now), and resolved disputes together. The intraclass correlation coefficient for the datasets was high (>.95).

Four key results emerged. First, as predicted, interpersonal communication is very focused on the present. Approximately 43% of all rated posts were about exactly now, 29% were about the past, and 28% were about the future (χ² (2, N = 3812) = 165.64, p < .001). The past and future did not differ significantly overall in their frequencies (χ² (1, N = 2169) = 2.324, p = .13). If the range that counted as “present” was expanded (from 0) to include posts within a half
hour into the past and future, 60% of posts were about the present. This number grows to 64% if broadened to an hour into the past or future. This result shows that many online social media posts, if not focused on the very instant of posting, are focused heavily on the present, which is the most accessible.

Second, as predicted, time of day affects when people talk about. As the day progresses, individuals talk more about what happened within the previous twenty-four hours (excluding the past hour). This result holds in terms of hours (B = .015, Wald = 4.032, \( p = .045 \)) and minutes (B = .000217, Wald = 3.171, \( p = .075 \)) since midnight. Overall, this result is consistent with the accessibility explanation: individuals have more accessible information from earlier in the day as it progresses.

Third, as predicted, day of week also affects when people talk about. As people move further into the week (and away from the weekend), people talk more about the upcoming twenty-four hours. This progression holds for days since Saturday (B = .054, Wald = 7.131, \( p = .008 \)) and trends in the right direction for hours since Saturday (B = .00134, Wald = 2.476, \( p = .12 \)), and holds when reviewing days (B = .055, Wald = 7.113, \( p = .008 \)) and hours (B = .00195, Wald = 4.854, \( p = .028 \)) since Sunday. This result is also consistent with the accessibility mechanism if the future becomes more accessible as Friday gets closer.

Fourth, people talk more about the near past and distant future. In other words, the relative frequency of past and future fluctuates based on temporal distance (\( \chi^2 (1, N = 2169) = 63.041, p < .001 \)).

Looking at nearby time periods, people share more about the previous twenty-four hours than the next twenty-four hours (\( \chi^2 (1, N = 1823) = 20.858, p < .001 \)), but for distant time periods (i.e., greater than a day away), people talk more about the future than the past (\( \chi^2 (1, N = 346) = 44.439, p < .001 \)). Whereas people talk more about the recent past than near future, they talk more about the distant future than the distant past. This latter finding is in line with the emotional intensity explanation inasmuch as the more evocative, distant future is shared more than the distant past.

Overall, these results are consistent with our predictions about the impact of accessibility and emotional intensity on the sharing from temporally close and distant time frames. People share more nearby events, and more distant future than distant past events. We are currently conducting various lab studies to further explore and provide evidence for our proposed mechanisms as drivers of these effects.

The Content and Impact of Mobile Versus Desktop Reviews

EXTENDED ABSTRACT

Although there is a growing body of research on the content and impact of word-of-mouth on consumer behavior, little is known about differences between mobile and traditional desktop reviews.

Recent empirical research shows that desktop reviews are generally positive (Chevalier and Mayzlin 2006), that product ratings in online reviews increase over time but decline as the number of reviews increases (Godes and Silva 2012), and that later reviewers are influenced by earlier ones (Moe and Trusov 2011). Other research shows that the valence, dispersion, and volume of consumer reviews predict sales in traditional (i.e., desktop) online environments (Chevalier and Mayzlin 2006; Duan, Gu, and Whinston 2008; Godes and Mayzlin 2004). Other research suggests that the perceived value of consumer-created content depends on characteristics such as contribution length, contributors’ prior posting behavior, contributor expertise, and the perceived similarity of the creators to the readers of this content (Forman, Ghose, and Wiesenfeld 2008; Weiss, Lurie, and MacMillan 2008). However little is known about differences in the characteristics and relative influence of consumer content created in mobile versus traditional settings. In addition, there is substantial debate over the pros and cons of allowing consumers to write real-time reviews with some fearing that mobile reviews will not benefit from reflection and that consumers will affectively (and negatively) respond to their current experiences. This has led some review sites (e.g., Yelp) to allow consumers to start reviews on their mobile devices but require them to use a desktop computer to finish them.

This research addresses these issues by 1) identifying potential differences in content created on mobile versus non-mobile devices and 2) assessing how these factors affect the relative influence of mobile versus desktop reviews.

Given that mobile reviews are more likely to be written during or immediately after service experiences, they should be more accurate, and less optimistically biased, than retrospective (i.e., desktop) evaluations (e.g., the rosy view; Mitchell et al. 1997; Novemsky and Ratner 2003; Wirtz et al. 2003). This means that mobile reviews will be more negative than desktop reviews. However, because they are likely to be written in real-time, and therefore reflect “hot” reasoning (Ariely and Loewenstein 2006), mobile reviews may be more emotional and heuristic decreasing their perceived usefulness relative to desktop reviews. Further, time pressure and tradeoffs between physical and cognitive effort on mobile devices (Lurie, Song, and Narasimhan 2009) means that mobile reviews should be shorter; potentially also reducing their perceived usefulness (Weiss et al. 2008).

To gain initial insights into these issues, we collected 299,806 reviews from 125,147 users on 144,231 restaurants on the website Urban Spoon. Of these, 136,306 reviews (45%) were written on mobile devices and 163,500 (55%) were written on non-mobile devices. To control for user characteristics that might drive differences between mobile and non-mobile reviews, we focused on 4,449 users who wrote both mobile and non-mobile reviews providing a sample of 48,610 (20,616 mobile and 27,994 non-mobile) reviews. Review text was analyzed using the LIWC program (Pennebaker, Booth, and Francis 2007). To measure influence we examined differences in the number of people that “liked” a review. Results show that mobile reviews are shorter (as expected), more negative, involve a greater percentage of words expressing positive emotion, are less cognitive, and more affective. Mobile reviews are also less influential (i.e., fewer readers “like” these reviews) even after controlling for differences between mobile and desktop reviews.

Naive or Savvy: How Credible Are Online Reviews for Credence Services?

EXTENDED ABSTRACT

Online reviews allow consumers to share information about their service experiences with other consumers. Theoretically, sharing this information should reduce market information asymmetries between potential consumers and service providers because many attributes of a service experience cannot be evaluated prior to consumption (i.e., they are credence or experience attributes rather than search attributes; Huang, Lurie, and Mitra 2009). For example, although a consumer might be able to evaluate the prices of entrees (a search attribute) prior to eating at a restaurant, it is very difficult to evaluate their tastiness (an experience attribute) without having consumed a meal, and even after eating them, a consumer cannot verify the claim that they are made with organic ingredients (a credence attribute). Because consumers cannot assess the quality of credence attributes, consumer reviews of services that are dominated by cre-
dence attributes (e.g., doctors, auto mechanics) are of dubious credibility. Indeed, many doctors have strongly resisted the legitimacy of consumer reviews (Andrews 2008; Jain 2010), in some cases even requiring patients to sign documents promising never to review their doctors (ElBoghdady 2012).

Although consumers have numerous forums to review credence services (e.g., RateMDs.com, Angie’s List, Yelp.com), extant research has focused only on reviews of search and experience goods (Huang et al. 2009; Mudambi and Schuff 2010; Park, Kim, and Han 2008). We address this gap by examining the claims typically included in reviews of credence service providers and whether consumers rationally discount claims about credence attributes within reviews. We pose three interrelated research questions: 1) Do the claims made in reviews of credence services on forums such as Angie’s List and Yelp.com substantively differ from the claims made in reviews of experience services, and if so, how? 2) Which attributes do consumers believe are most important for evaluating credence service providers? and 3) When evaluating credence service providers, do consumers weigh credence claims more heavily than experience claims because credence attributes are more important for these services, or do they weigh experience claims more heavily because these claims are more credible?

We address these questions by conducting a content analysis of real online reviews and a series of lab experiments. First, we design and test a rigorous classification method for identifying credence and experience services based on their core and peripheral attributes. The economics of information literature categorizes goods and services into search, experience and credence categories, but does not discuss the possibility that credence services may have search, experience and credence attributes (e.g., a doctor’s office location is a search attribute, bedside manner is an experience attribute, and accurate diagnosis is a credence attribute). Our categorization method fills a gap in the literature by explicitly linking a product’s relevant search, experience, and credence attributes (Ford, Smith, and Swasy 1988) to the categorization of the product as a search, experience, or credence good (Dulleck and Kerschbamer 2006). Previous work has simply assumed the classification of goods as search, experience, or credence a priori, which has led to conflicting classifications. For example, in Nelson’s original work (1970) and in subsequent papers a camera was classified as a search good, but other authors have argued that cameras should be labeled experience goods (Huang et al. 2009). We propose a replicable and rigorous method for classifying services based on their core and peripheral attributes.

Next, we content analyze online reviews of service providers to compare their content and structure across services that are dominated by credence vs. experience attributes. Using Toulmin’s (1958) theoretical framework for evaluating arguments and claims, we compare the structure of claims made in reviews of experience vs. credence service providers. We find that negative reviews contain higher-quality arguments as well as more information than positive reviews for the same types of service providers. This finding provides an alternative explanation for the greater impact of negative reviews in online word of mouth that has been demonstrated in earlier work (Cao, Duan, and Gan 2011; Laczniak 2001). This pattern of results also may explain why review length seems to mediate the effect of review valence on helpfulness ratings (Wu, van der Heijden, and Kort�atis 2011).

Further, we code claims about specific service provider attributes that are mentioned in reviews. Our results indicate that reviews of credence services include a mix of claims about both credence and experience attributes, but claims about experience attributes appear more frequently than claims about credence attributes. Thus, we find that although reviews of credence services contain claims about experience attributes (e.g., a doctor’s bedside manner), they also contain claims about credence attributes that consumers cannot accurately evaluate (e.g., whether the doctor’s diagnosis was accurate).

Finally, we are conducting a series of experimental studies to explore consumers’ perceptions of reviews, manipulating factors such as the type of service provider (experience or credence), the type of attribute in the review (experience or credence), and the valence of the review (positive or negative). We expect to find that although consumers reading reviews may acknowledge the low credibility of credence claims when they are explicitly asked to evaluate the claims, claims about credence attributes significantly influence their evaluations of credence service providers.

REFERENCES
Advances in Consumer Research (Volume 41) / 119


Kosinski, Michal and David Stillwell (2011) “my Personality Research Wiki,” myPersonality Project.


