2014 47th Hawaii International Conference on System Science

Why Aren't the Stars Aligned?
An Analysis of Online Review Content and Star Ratings

Susan M. Mudambi
Temple University
Susan.Mudambi@temple.edu

David Schuff
Temple University
David.Schuff@temple.edu

Zhewei Zhang
Temple University
zhang@temple.edu

Abstract

Consumer-generated product evaluations posted on online retailer or third party web-sites have been shown to increase buyer trust and aid consumer decision making. These online reviews typically have two components: star ratings and review text. This can communicate a complex, conflicting message to consumers, as the text of a review carries more nuance than what can be communicated through a simple numerical score. Misalignment between the star rating and the text can lead to increased consumer cognitive processing costs, suboptimal purchase decisions, and lower overall utility of the review site. This study seeks to understand where this misalignment is mostly likely to occur. We find that star rating/review text misalignment occurs more often for (1) experience goods, and (2) goods that receive high star ratings. Misalignment is especially pronounced for experience goods with high star ratings.

1. Introduction

As online shopping has emerged and expanded, so have consumer and academic interest in online product reviews and their impact on consumer behavior and sales. The ease of use of online review systems is a large part of the appeal, yet reviews routinely communicate a complex message, with conflicting signals found across and within reviews.

Online reviews typically have two components: star ratings and review text. The star ratings appear to be a straightforward, unambiguous way of communicating the consumer's overall assessment. However, research has established considerable heterogeneity in consumer interpretation and use of scales (e.g., [5]). One person's view of a "3" may be considerably different from another's. More generally, although the simple cue of a star rating can be easy to process, consumers can differ as to the quality of the cue and the usefulness of it for their own decision making.

The text comments accompanying the star ratings have the potential to alleviate star rating ambiguity by providing explanations and context for the rating. When the stars and text align, the reviews add depth and value to the consumer buying experience. Yet the opposite can happen, as text comments can increase ambiguity and uncertainty. When the stars are not aligned with the review text, this can be frustrating and annoying to consumers [3], and can reduce the value of online review systems, both to consumers and to firms. In this context, alignment is defined as *the equivalence between the star rating given by a reviewer and the perceived or predicted star rating of the review text content alone*. Why aren't the stars aligned? This research question drives this analysis of the drivers of misalignment between star ratings and online review content.

2. Literature and Theoretical Foundation

Consumer-generated product evaluations posted on company or third party web-sites have been shown to increase buyer trust [32], aid consumer decision making [34]), and increase product sales [6][8]. Numerous recent academic studies in information systems, computer science, and marketing have analyzed star ratings and review text (i.e., [9][10]). Research has utilized both star ratings and text characteristics to explain multiple aspects of online reviewing and consumer behavior.

Research on star ratings has acknowledged the role of attitude extremity. Chevalier and Mayzlin [4] found that a 1-star rating has a greater impact than a 5-star rating on Amazon book sales. Zhang, Dellarocas, and Awad [36] found that the average star rating predicts future movie revenues better than other measures. Mudambi and Schuff [19] examined the helpfulness of extreme ratings across different product types. Research on review text content has also generated insights into consumer attitudes and behavior. Textual comments regarding sellers on eBay can influence online reputation [23]. The rich information of text enables analysis of aspects such as emotionality, and positive or negative sentiment of the review. Multiple studies have found text

978-1-4799-2504-9/14 $31.00 © 2014 IEEE
DOI 10.1109/HICSS.2014.389

characteristics to be important factors in decisions (e.g., [7][10][18]).

Although past research has used star ratings and text characteristics to explain aspects of online reviewing and consumer behavior, the relationship between stars and text merits further analysis. Consumers have lamented star rating/review text misalignment, but past research has not fully examined the degree and drivers of misalignment.

2.1. The Effort/Accuracy Tradeoff

Most review sites include both star ratings and text content. The inclusion of both types of information reflects an acknowledgement of tradeoffs between effort and accuracy [12]. In the elaboration likelihood model (ELM) of information processing [26], consumers who are likely to cognitively process or mentally elaborate on information are said to take a central route, with a relatively systematic appraisal of information. This supports the provision of multiple and detailed text comments. For other consumers, mental elaboration and effort may be impractical or prohibitive. These visitors are more likely to use peripheral cues to make a decision. This supports the highlighting of star ratings, as they act as decision and comparison aids [30].

Most consumers who make the effort to locate and read online reviews are not simply aiming to minimize cognitive effort. Yet, site visitors can significantly differ in their motivation and ability to process review site content. Sometimes consumer behavior is guided by thoughtful information processing, but at other times decisions are immediate, and behavior is spontaneous [27]. ELM implies that retailers with a goal of producing enduring attitude and behavioral changes should actively encourage the central route to persuasion through extensive text content, while retailers with a goal of immediate but possibly short-lived attitude formation should emphasize the peripheral cues of star ratings [27]. Theory indicates a distinct role for each information type.

In practice, consumers see and process both stars and text. Consumers who are willing to put effort into the decision process act at least partially in expectation of a more accurate decision. However, processing both types of information can increase cognitive processing costs, especially when there is misalignment between stars and text comments. In this context, there is an unclear and unsatisfactory tradeoff between effort and accuracy.

2.2. Ambiguity

Star ratings are expected to be a simple, visual signal of perceived quality, and a clear peripheral cue. However, evidence suggests that star ratings contain ambiguity, defined as *information content open to multiple interpretations* [31]. Table 1 compares the rating systems of Amazon, CNET and Yelp. For example, 3 stars can indicate the product is "ok," "good," or "A-Ok." Guidelines for star ratings vary across review sites, as does the visibility of the explanations. Amazon's interpretation of the stars is not communicated on the review page, but can be located by a purposeful search. In contrast, on CNET and Yelp, moving the cursor over a star reveals the star guide. CNET also allows for .5 star ratings, thereby adding "abysmal," "poor," "ok," "very good," and "outstanding" to the options. These differences in the definitions indicate site interest in guiding consumer interpretations, while there is little evidence that providing the guidelines significantly affects how consumers use the star rating system. The existence of guidelines indicates an inherent ambiguity of star ratings.

Table 1. Comparison of star rating guidelines across sites

	Amazon	CNET	Yelp
1 star	I Hate It	Terrible	Eek! Methinks not.
2 star	I Don't Like It	Mediocre	Meh. I've experienced better
3 stars	It's Okay	Good	A-Ok.
4 stars	I Like It	Excellent	Yay! I'm a fan.
5 stars	I Love It	Spectacular	Woohoo! As good as it gets.

Text comments are by nature more ambiguous than numerical ratings. Yet the richer information in text content can provide helpful and nuanced explanations of the star ratings. The expectation is that the richer textual information aligns with the star cue. Analysis of the text alone should lead to the same numerical evaluation as analysis of the star rating. Review sites and customers generally expect reviews to be characterized by a close alignment of star ratings and text. Alignment is beneficial to the consumer, as it facilitates the cognitive processing of information and the decision process. Misalignment is detrimental to information processing and increases equivocality. Past research has linked ambiguity and risk (e.g., [13]). Since a key reason consumers turn to online reviews is to reduce purchase risk, ambiguity of reviews is undesirable. Review ambiguity and star/text misalignment can be expected to be generally undesirable to consumers. If common, this can lead to review or site dissatisfaction.

Given the potential importance of star/text misalignment, it is important to investigate how extensive is misalignment and what drives and explains its existence. The identification of clear and distinct 1, 2, 3, 4 and 5-star categorizations from text alone can indicate that review text and star ratings play the distinct roles implied by the central route and peripheral cues of ELM. In contrast, the presence of misalignment can indicate complementary or overlapping roles of the two information types.

2.3. Product Type

Product attributes are relevant to misalignment. Products exist along a continuum of search goods to experience goods, a distinction established by Nelson [20][21] to reflect differences in the ability to obtain diagnostic information on product quality before trial or purchase. Search goods have more attributes that can be objectively measured and easily compared before trial, while experience goods have more attributes that are subjective, difficult to compare, and require consumer use of their senses for the evaluation of quality [19]. Online consumer behavior has been shown to differ across search and experience goods [33].

Since search goods have more quantifiable attributes, it follows that an overall evaluation of a search good is relatively easy to summarize with a numeric rating, or peripheral cue. For example, a laser printer can be evaluated in terms of throughput speed, jam rate, and print quality, and this facilitates an overall quantified assessment. In contrast, since experience goods such as music CDs are dominated by attributes that are difficult to objectively measure, this creates ambiguity in how to quantify the overall evaluation. Reviewers provide clarifying information through the central cue of the review text, but this requires more mental elaboration by consumers. Given that consumers find the differences between a 4-star CD and a 5-star CD to be fuzzy, it can be expected that, across reviews, the language used to evaluate product attributes and experiences fit less neatly with the star ratings. This leads us to our first proposition:

P1. Product type affects the degree of overall star/text misalignment. Reviews of experience goods have greater star/text misalignment than reviews of search goods.

2.4. Valence

In addition to product attributes, review attributes are also expected to play a role in misalignment. In particular, review valence, the degree of positive or negative sentiment, is relevant. Reviews posted on sites such as Amazon.com are predominantly positive. This is not surprising, since consumers who purchase a product were predisposed to like it. In addition, consumers who purchase a product and then bother to post online content on the product indicate a relatively high level of product involvement.

A 5-star rating clearly signals positive affect about the product. The text content of 5-star reviews should be overwhelmingly positive, just as text content of 1-star reviews should be overwhelmingly negative. However, consumers have multiple incentives to post negative comments, even alongside a 5-star product evaluation. They often have altruism, reciprocity, and community-building motivations [28], with a general intention of being helpful to other consumers, or helpful to the company's new product development efforts. In addition, positive evaluations are seen as less diagnostic and credible than negative evaluations. Research has shown that people who make negative comments are seen as more intelligent than those who make only positive comments [1]. This implies that reviews with strongly positive peripheral cues (star ratings) require more elaboration. Reviewers, out of a desire to provide useful information, introduce more negative comments into positive reviews than they introduce positive comments into negative reviews. This complicates the interpretation of highly positive reviews. Therefore, we propose:

P2. Review valence affects the degree of star/text misalignment. Highly positive reviews, as indicated by star rating, are more likely to have negative textual content than highly negative reviews are to have positive textual content.

Purchase-specific involvement ranges across product categories [16], with experience goods such as computer games typically characterized as higher involvement than search goods (such as toasters), although they are in a similar price range. Therefore, a key difference between search and experience goods is the connection consumers have with experience goods. Yet, some consumers become so deeply connected with music CDs and other experience goods that they can be seen as star-struck "groupies" and therefore not objective. Because it is easier to dismiss comments from fans, a strongly positive review of an experience good can strain credibility more than a strongly positive review of a search good. We would expect that the association described in P2 (greater misalignment in highly positive reviews) would be even greater for experience goods than search goods. Therefore, we propose:

P3. Highly positive reviews, as indicated by star rating, are more likely to have negative textual content when the products are experience goods.

To investigate the extent and drivers of star/text misalignment, we apply a text mining algorithm to determine if the text of the review is aligned with the reviewer-designated numerical star rating.

3. Methodology

3.1. Text Analysis and Classification

Conventional text analysis examines basic characteristics such as length (word count). Recent studies utilize computational tools such as sentiment analysis to a gain better understanding of the key aspects of a body of text. One research stream on sentiment analysis relies on traditional natural language processing techniques, which normally use pre-defined linguistic-based models to determine the sentiment or emotion intended by the text. This approach has been used to analyze sentiment on Twitter and other social platforms (e.g., [2][22]). Another research stream benefits from the recent adoption of machine learning-based approaches. This approach does not rely on a pre-defined model to determine the sentiment of text. Instead, the software generates a predictive model of the text's sentiment through learning from a corpus of training data and then uses that model to determine the sentiment of an additional set of test data.

In order to test our propositions, we used a machine learning approach to categorize the review text as a 1, 2, 3, 4, or 5-star review. We adopted this approach because this provides an objective way to classify the text by looking at its syntactic content. In this study we are not seeking a subjective assessment of the review content's sentiment; instead, we want to determine that the text of the review is most typical of reviews of "n" stars. Further, previous research confirms that better classification performance can be achieved through a machine learning based approach [24][25]. The analysis identifies the valence of the text and identifies similarities across documents (in this case, reviews). The result is that the classification algorithm can identify a "typical" 5-star review, 3-star review, and so on.

Past research has analyzed the text of online reviews. For example, Titov and McDonald [29] and Yu, Zha, Wang, and Chua [35] identified important product characteristics and developed an aspect-based ranking system. Lim, Nguyen, Jindal, Liu, and Lauw [17] used machine learning to classify reviews as spam. Kim, Pantel, Chklovski, and Pennacchiotti [14] and Ghose and Ipeirotis [9] used this technique to predict the helpfulness of online reviews.

It is important to note that our goal is not to predict the star rating of a particular review. Instead, we aim to use the ability of the software to correctly classify the reviews as a measure of the alignment of the review text with its designated star rating. We make the assumption that, in general, reviews with positive sentiment in the review text have higher star ratings than those with negative sentiment in the text. It follows that the more aligned the sentiment (i.e., valence) of the text of the reviews is with the ratings, the more accurate the categorization will be. If there is a great degree of ambiguity in the text of a particular review, then the algorithm will be less reliable in categorizing the set of reviews "cleanly" into a star rating.

3.2. The Data Set

We collected our data from the U.S. Amazon.com site using a custom software agent. For each product, we collected each review's text, star rating, and product type. The data set contains 1,734 randomly selected reviews from a set of 23 products across eight product categories. The product set is representative of both product types (search and experience goods). Search goods comprised 11 of the products and included cameras, coffee makers, grills, and toasters. Experience goods comprised 12 of the products, and included books, music CDs, MP3 players, and diapers. We deliberately selected a broad, representative set of products within each category with appeal to diverse consumer demographics. The list of products, and the number of reviews selected, is provided in Table 2.

Table 2. List of products used in this study

Product Name	Product Category	# of Reviews
Born to Run	Book	53
The Girl with the Dragon Tattoo	Book	197
Sarah's Key	Book	80
Outliers: A Story of Success	Book	110
Metallica	CD	112
Back in Black	CD	64
Demon Days	CD	30
Recovery	CD	32
Pampers Baby Dry	Diapers	86
Luvs Premium	Diapers	54
Pampers Cruisers	Diapers	48
Apple iPod Classic 160GB	MP3 Player	30
Zune HD 32GB	MP3 Player	68
Sanyo Grill	Grill	74
George Foreman Grill	Grill	82
Cuisinart DCC-2000	Coffee Maker	96
BUNN Velocity Brew	Coffee Maker	79
Cuisinart TOB-195	Toaster	56
Breville BOV800XL	Toaster	84
Nikon D90	Camera	63
Canon SD780IS	Camera	129
Canon EOS T1i	Camera	55
Panasonic Lumix	Camera	52

4. Analysis and Results

To perform the analysis we used TagHelper, an add-on to the Weka platform [11]. Weka is a commonly used machine learning software suite. TagHelper transforms the text into vectors that describe each document. It uses collections of unigrams (single words), bigrams (two word phrases), part-of-speech bigrams (linguistic categories of words), punctuation, and line length as input for the algorithm. A "training set" of reviews and their star ratings were input into TagHelper to build the classification model based on the text characteristics of those reviews. The resulting model was then used to classify an additional "test set" of reviews to see how well the model performed at predicting their star rating.

We tried several alternative classification algorithms including Naïve Bayes and Support Vector Machine. Of those, the Naïve Bayes classifier had the best rate of prediction with our dataset as it better handles highly skewed data. Most online reviews tend to be positive; the average star rating for our randomly selected set was 4.25 out of 5, and this is consistent with the findings in other studies (for example, [4][19]. The Weka software was trained to classify a review based on the characteristics of each review "document," based on the reviewer's original star rating. The output from this process is called a confusion matrix, a table that cross-tabulates the actual categorizations (the star rating the reviewer gave the product) with the predicted categorizations (the star rating the software assigned to the product based on the text of the review). This provides a snapshot of how well the valence of the review sentiment matches with the reviewer's star rating.

We used Cohen's kappa, a common test for inter-rater agreement, to assess the alignment between text sentiment and the reviewer's star rating. The reviewer's star rating is treated as the first "rater" and the algorithm's classification is treated as the second "rater." The higher the level of agreement, the more accurately the algorithm was able to guess the star rating the reviewer gave the product based solely on the content of the review text.

To test our first proposition (product type affects the degree of overall star/text misalignment), we compared the level of agreement between the star rating and text of search goods to experience goods. We proposed that search goods would have more alignment between the reviewer-designated star rating and the valence of the text. The cross tabulations for search and experience goods are in Tables 3 and 4.

Both kappa statistics were significant at the 0.01 level, indicating association. The analysis reveals that, as proposed, the algorithm's ability to correctly predict the valence of the review through the review text was much higher for search goods (kappa = 0.728) than experience goods (kappa = 0.328). According to Landis and Koch's [15, p. 165] benchmark scale for kappa, 0.728 is a "substantial" association while 0.328 is considered a "fair" association. This implies that the sentiment of the review text is more aligned with the star rating for search goods than experience goods. If this was not the case, then the algorithm would have an equally difficult time differentiating between the valences of the review text for both types of goods

To test our second proposition (review valence affects the degree of star/text misalignment) and third proposition (highly positive reviews are more likely to have negative textual content when the products are experience goods), we compared the distribution of the predicted valence for positive and negative reviews. To more clearly see the differences, we calculated the percentage of reviews classified as each star rating level for search and experience goods. This is displayed in Tables 5 and 6.

In our second proposition we argued that review valence affects the degree of star/text misalignment. We expected greater alignment overall for negative reviews than positive reviews when comparing the reviewer-designated star rating to the assessed valence of the review text by the algorithm. We find this to be the case. For reviews where the reviewer gave the product one star, the text was aligned with that rating 94% of the time for search goods and 79% for experience goods. However, for reviews where the reviewer gave the product five stars, there was alignment only 85% of the time for search goods (about 9.5% lower than search good one star reviews) and 64% for experience goods (about 20% lower than experience good one star reviews). This suggests that alignment is lower when the valence of the reviewer-assigned product star rating is high. This misalignment means that enough negative information is included in the review to influence the algorithm's assessment of that review's valence.

Table 3. Analysis of star/text alignment for search goods

	Predicted (Based on text)					
Actual (Star)	1	2	3	4	5	Total
1	**50**	1	1	1	0	53
2	1	**28**	0	2	2	33
3	0	0	**45**	3	4	52
4	1	1	2	**120**	33	157
5	0	0	8	62	**405**	475
Total	52	30	56	188	444	770
Measure of agreement (Kappa) = 0.728 Approximate significance = 0.000						

Table 4. Analysis of star/text alignment for experience goods

	Predicted (Based on text)					
Actual (Star)	1	2	3	4	5	Total
1	**65**	10	2	1	4	82
2	6	**28**	1	3	6	44
3	6	11	**38**	1	15	71
4	25	38	3	**82**	56	204
5	43	95	4	62	**359**	563
Total	145	182	48	149	440	964
Measure of agreement (Kappa) = 0.398 Approximate significance = 0.000						

Table 5. Analysis of star/text alignment for search goods (as %)

	Predicted (Based on text)					
Act. (Star)	1	2	3	4	5	Total
1	94%	2%	2%	2%	0%	100%
2	3%	85%	0%	6%	6%	100%
3	0%	0%	87%	6%	8%	100%
4	1%	1%	1%	76%	21%	100%
5	0%	0%	2%	13%	85%	100%
Total	7%	4%	7%	24%	32%	100%

Table 6. Analysis of star/text alignment for experience goods (as %)

	Predicted (Based on text)					
Act. (Star)	1	2	3	4	5	Total
1	79%	12%	2%	1%	5%	100%
2	14%	64%	2%	7%	14%	100%
3	8%	15%	54%	1%	21%	100%
4	12%	19%	1%	40%	27%	100%
5	8%	17%	1%	11%	64%	100%
Total	15%	19%	5%	15%	46%	100%

In line with the third proposition, we see a greater tendency to include negative information in positive search good reviews than in positive experience good reviews. The analysis for proposition 2 above indicates that although there is less alignment for positive reviews than negative reviews, for experience goods this difference is especially pronounced. In fact, the classification algorithm never assessed the textual content of a five star search good review as negative (i.e., the algorithm never gave a predicted value for those reviews as 1 or 2 stars). Even for four star reviews, only 2% of those reviews were classified as 1 or 2 star reviews based on their text. However, for five star experience good reviews, the text analysis classified about 25% of those reviews as negative. For four star experience good reviews, 31% were classified as negative. This indicates that there is far more ambiguity in the text of high star rating experience good reviews than high star rating search good reviews. In other words, good reviews that carry a star rating indicating a strongly positive overall assessment are more likely to include negative information when the product is an experience good.

5. Conclusions

This theory-grounded analysis has generated three significant findings. First, overall star rating/review text misalignment occurs about two-and-a-half times more often for experience goods (40.7%) than for search goods (15.9%). This implies there is less ambiguity in the reviews for search goods, potentially making reviews for those products more useful objective evaluations of quality. Second, misalignment appears to occur more often in reviews where the reviewer-designated star rating is high. Both search and experience good reviews had higher misalignment for 4 and 5-star reviews (15.4% for search, 21.2% for experience) than 1 and 2-star reviews (2.5% for search, 14.7% for experience). This implies that there is greater ambiguity in positive product assessments than in negative product assessments. Third, we found evidence that misalignment is especially pronounced for goods with both of those characteristics. Experience goods with high star ratings have a misalignment rate of 21.2%, compared to only 2.5% for search goods with low star ratings). This leads to a clear and practical suggestion: for highly rated experience goods, consumers should pay close attention to the text of the review to get a truly accurate view of product quality.

5.1 Limitations

First, even though our study does not focus on how to improve the text classification algorithm, a closer examination of the Naïve Bayes and SVM classification algorithms can only increase the confidence in our findings. Through further tuning and the use of a larger data set, future studies can verify that the classification is reliable and the algorithm is performing optimally.

Second, more study is needed to increase the generalizability of our findings. Our set of 23 products, while representative of products sold in high volume on Amazon.com, obviously does not cover all product categories (i.e., food, video games). Also, our results for Amazon reviews raise the question as to whether the relationships will also be found on other retail sites and on non-retailing sites. Sites such as Yelp or CNET are especially important to consider, as they are not online retailers, but are third-party sites that provide evaluations of products and services. They also use different rating categorizations (refer to Table 1). Our study could be replicated with reviews from other sites with different business models and different star rating schemes.

5.2 Future Research

Extensions of this work can explore these and other relationships in more depth. First, the study can be replicated using a larger data set or a set of reviews with a different product mix. This would enable an analysis to determine if our results still hold and to test possible theories that may explain the misalignment.

Second, future research could increase the robustness of the analysis by conducting a manual classification of reviews done not by machine learning, but by "real people." This manual classification can be compared to the automated classification to see if the results are consistent. The drawback of human classification is that it is difficult to scale to a large data set. However, human classification is a potential validity check on the algorithm-based classification.

Third, future research can investigate other possibilities that may affect the misalignment between star rating and text. The misalignment is determined by comparing the star rating, which is based on reviewer's overall evaluation of the product, and perceived rating based on review content, which is consumer's assessment of the meaning of the review text. Consumer perception is likely influenced by many factors. One important factor is prior experience, which can be captured several ways. When looking at the human-coded assessments, one can look for differences in the perceived star rating due to each coder's self-assessed prior experience. Other possible factors include product popularity, brand reputation, and price of the products in the sample.

Lastly, although we have clearly established the theoretical and practical reasons for why star/text misalignment is undesirable for consumers and firms, additional research is needed on the consequences of misalignment for consumers and firms. For example, to what degree are "misaligned" reviews less helpful to the decision-making process? Are those reviews less engaging? Future research can identify different forms of misalignment, and the resulting consequences on consumer behavior and on firm effectiveness at encouraging site engagement and site-based sales.

5.3. Contributions

For academics, our study draws on the elaboration-likelihood model, and literature in consumer behavior and information processing to provide a theory-based explanation for the causes that might lead to misalignment between the peripheral cue of the star rating and the central cue of the review's text. Additionally, we applied a text-mining tool to measure alignment between human and computer-based assessments in a novel way.

For practice, this study sheds light on a potentially significant problem for online review sites and for consumers. Misalignment between the star rating and the review text could lead to a systemic misrepresentation of product quality to consumers. Closely reading all the reviews is not a viable alternative, as many products have thousands of reviews. Online review sites, including online retailers, should look closely at the potential inconsistencies in information on their sites and consider remedies, such as displaying an "adjusted star rating" for each review that reflects the sentiment in the text. This computed value could be averaged across all reviews and presented as a complement to information already provided to consumers. When the stars are aligned, online review sites and consumers benefit.

6. References

[1] Amabile, T., "Brilliant But Cruel…Perceptions of Negative Evaluators", *Journal of Experimental Social Psychology* (19), 1983, pp. 146-156.

[2] Bollen, J., H. Mao, and A. Pepe, "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena", in *Proceedings of the International AAAI Conference on Weblogs and Social Media,* Washington, DC, 2011.

[3] Celis, D. "Why I Hate Five-Star Ratings", 2012. Retrieved from http://davidcel.is/blog/2012/02/01/why-i-hate-five-star-ratings/

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

[5] Churchill, G.A. Jr., "A Paradigm for Developing Better Measures of Marketing Constructs", *Journal of Marketing Research*, 16(1), 1971, pp. 64-73.

[6] Cui, G., H. Lui, and X. Guo, "The Effect of Online Consumer Reviews on New Product Sales", *International Journal of Electronic Commerce*, 17(1), 2012, pp. 39-58.

[7] Feldman, G., "Techniques and Applications for Sentiment Analysis," *Communications of the ACM*, 56(4), 2013, pp. 82-89.

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

[9] Ghose, A. and P.G. Ipeirotis, "Designing Novel Review Ranking Systems: Predicting the Usefulness and Impact of Reviews", in *Proceedings of the Ninth International Conference on Electronic Commerce,* New York, NY, 2007, pp. 303–310.

[10] Ghose, A. and P.G. Ipeirotis, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics", *IEEE Transactions on Knowledge & Data Engineering*, 23(10), 2011, pp. 1498-1512.

[11] Hall, M., E, Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The WEKA Data Mining Software: An Update", *SIGKDD Explorations*, 11(1), 2009, pp. 10-18.

[12] Johnson, E. and J. Payne, "Effort and accuracy in choice", *Management Science*, 31(4), 1985, pp. 395-415.

[13] Kahn, B.E. and R.K. Sarin, "Modeling Ambiguity in Decisions Under Uncertainty", *Journal of Consumer Research*, 15(2), 1988, pp. 265-272.

[14] Kim, S.M., P. Pantel, T. Chklovski, and M. Pennacchiotti, "Automatically Assessing Review Helpfulness", In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing,* Stroudsburg, PA, USA, 2006, pp. 423–430.

[15] Landis, J.R. and G.G. Koch, "The measurement of observer agreement for categorical data", *Biometrics*, 33(1), 1977, pp. 159–174.

[16] Laurent, G. and J. Kapferer, "Measuring Consumer Involvement Profiles", *Journal of Marketing Research* (22)1, 1985, pp. 41-53.

[17] Lim, E.P., V.A. Nguyen, N. Jindal, B. Liu, and H.W. Lauw, "Detecting Product Review Spammers Using Rating Behaviors", In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management,* New York, NY, USA, 2010, pp. 939–948.

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

[19] Mudambi, S. and D. Schuff, "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com", *MIS Quarterly*, 34(1), 2010, pp. 185-200.

[20] Nelson, P., "Information and Consumer Behavior", *Journal of Political Economy* (78)20, 1970, pp. 311-329.

[21] Nelson, P., "Advertising as Information," *Journal of Political Economy* (81)4, 1974, pp. 729-754.

[22] O'Connor, B., R. Balasubramanyan, B. Routledge, and N. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series", In *Proceedings of the International AAAI Conference on Weblogs and Social Media,* Washington, DC, 2010.

[23] Pavlou, P. and A. Dimoka, "The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation," *Information Systems Research*, 17, 2006, pp. 392-416.

[24] Pang, B. and Lee, L. "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Stroudsburg, PA, 2004. Article 271.

[25] Pang, B., Lee., L., and Vaithyanathan, S. "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, 10, 2002, pp. 79-96.

[26] Petty, R.E., J.T. Cacioppo, and D. Schumann, "Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement", *Journal of Consumer Research,* 10(2), 1983, pp. 135-146.

[27] Petty, R.E., B. Pabro, and J.R. Priester, "Mass Media Attitude Change: Implications of the Elaboration Likelihood Model of Persuasion," In J. Bryant and M.B. Oliver (eds.), *Media Effects: Advances in Theory and Research,* Routledge, 2009, pp. 129-164.

[28] Schau, H., A. Muñiz, and E. Arnould, "How Brand Community Practices Create Value", *Journal of Marketing*, 73(5), 2009, pp. 30-51.

[29] Titov, I., and R. Mcdonald, 2008. "A Joint Model of Text and Aspect Ratings for Sentiment Summarization", In *Proceedings of Annual Meeting of the Association for Computational Linguistics: Human Language Technologies,* 2008, pp. 308–316.

[30] Todd, P., and I. Benbasat, "The use of information in decision making: An experimental investigation of the impact of computer-based decision aids", *MIS Quarterly*, 16(3), 1992, pp. 373-393.

[31] Trevino, L. K., Daft, R. L., and Lengel, R. H. "Understanding media choices: A symbolic interactionist perspective." In J. Fulk & C. W. Steinfield (Eds.), *Organizations and Communication Technology*, Newbury park, CA: Sage Publications, 1990, pp. 71-94.

[32] Utz, S., P. Kerkhof, and J. van den Bos, "Consumers Rule: How Consumer Reviews Influence Perceived Trustworthiness of Online Stores," *Electronic Commerce Research & Applications*, 11(1), 2012, pp. 49-58.

[33] Weathers, D., S. Sharma, and S.L. Wood, "Effects of Online Communication Practices on Consumer Perceptions of Performance Uncertainty for Search and Experience Goods", *Journal of Retailing,* 83(4), 2007, pp. 393-401.

[34] Xinxin, L, and L. Hitt, "Self-Selection and Information Role of Online Product Reviews", *Information Systems Research*, 19(4), 2008, pp. 456-474.

[35] Yu, J., Z.J. Zha, M. Wang, and T.S. Chua, "Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews", In *Proceedings of the Annual Meeting of the Association of Computational Linguistics,* 2011, pp. 1496-1505.

[36] Zhang, M., C. Dellarocas, and N. Awad, "The Impact of Online Movie Reviews on Box Office Performance", In *Workshop on Information Systems and Economics (WISE)*, College Park, MD, 2004.