AIS Transactions on Human-Computer Interaction

Volume 9 | Issue 1 Article 1

3-30-2017

The Impact of Sentiment Analysis Output on Decision Outcomes: An Empirical Evaluation

Parisa Lak
Ryerson University, parisa.lak@ryerson.ca

Ozgur Turetken
Ryerson University, turetken@ryerson.ca

Recommended Citation

Lak, P., & Turetken, O. (2017). The Impact of Sentiment Analysis Output on Decision Outcomes: An Empirical Evaluation. *AIS Transactions on Human-Computer Interaction, 9*(1), 1-22. Retrieved from https://aisel.aisnet.org/thci/vol9/iss1/1
DOI:

Transactions on Human – Computer Interaction

Research Paper

The Impact of Sentiment Analysis Output on Decision Outcomes: An Empirical Evaluation

Parisa Lak
Ryerson University
parisa.lak@ryerson.ca

Ozgur Turetken
Ryerson University
turetken@ryerson.ca

Abstract:

User-generated online content serves as a source of product- and service-related information that reduces the uncertainty in consumer decision making, yet the abundance of such content makes it prohibitively costly to use all relevant information. Dealing with this (big data) problem requires a consumer to decide what subset of information to focus on. Peer-generated star ratings are excellent tools for one to decide what subset of information to focus on as they indicate a review's "tone". However, star ratings are not available for all user-generated content and not detailed enough in other cases. Sentiment analysis, a text-analytic technique that automatically detects the polarity of text, provides sentiment scores that are comparable to, and potentially more refined than, star ratings. Despite its popularity as an active topic in analytics research, sentiment analysis outcomes have not been evaluated through rigorous user studies. We fill that gap by investigating the impact of sentiment scores on purchase decisions through a controlled experiment using 100 participants. The results suggest that, consistent with the effort-accuracy trade off and effort-minimization concepts, sentiment scores on review documents improve the efficiency (speed) of purchase decisions without significantly affecting decision effectiveness (confidence).

Keywords: Sentiment Analysis, E-commerce, Customer Reviews.

The manuscript was received 09/06/2016 and was with the authors 3 months for 1 revision.

1 Introduction

Research has adequately established the essential role of information in decision quality (Keller & Staelin, 1987; Raghunathan, 1999; Stigler, 1961). Information is valuable but comes with the cost of the time and effort (and sometimes financial resources) spent to gather, analyze, and comprehend it. Further, after one reaches a critical "mass" of information, the value of additional information on decision quality diminishes. Purchase decisions are a typical area where these effects manifest themselves. Consumers are normally aware of the trade-offs between the costs and benefits of information (Stigler, 1961) and implicitly add the cost of information to the final value of the product or services that they purchase while making their decisions.

Word of mouth (WOM) represents one of the most important sources of information for consumers. The growing prominence of Web 2.0 technologies and the emergence of social media interactions have led to a form of WOM that is much more prevalent than classical WOM. This electronic form of WOM (eWOM) is generated continuously in the form of tweets, blog posts, news, reviews, and comments. Online reviews play a notably significant role in consumers' purchase decisions (Mudambi & Schuff, 2010) because of their accessibility and the variety of information that one can gather in near real-time fashion. Millions of people express their opinions about restaurants, hotels, products, and even their family physicians or university professors through online review websites such as Yelp (www.yelp.com), Tripadvisor (www.tripadvisor.com), Amazon (www.amazon.com), RateMds (www.ratemds.com), and Ratemyprofessors (www.ratemyprofessors.com). Individuals can use this user-generated content to make wiser decisions (Huang, Tu, Fu, & Amanzadeh, 2013). However, users often find the vastness of this content hard to digest, which gives rise to the challenge of information overload (Turetken & Sharda, 2005).

Star ratings are common decision aids that consumers use to address the overload problem in this domain. These ratings help decision makers to select what content to focus on by providing cues on the content's sentiment (polarity). These quick cues are often useful; however, as we detail later in this paper, they are limited in their availability and value as differing forms of user generated content such as tweets or posts on popular social media sites such as Facebook (www.facebook.com/) and LinkedIn (www.linkedin.com) becomes more common place. Meanwhile, online review content has high volume, high velocity, and high veracity; therefore, one can classify it as "big data", and it should benefit from analytics approaches as do other forms of big data. For this reason, we examine the potential of sentiment analysis, a text-analytics technique that generates scores, which can substitute for star ratings. Sentiment analysis uses various analytical techniques to determine whether a piece of text is positive, negative, or neutral (Liu, 2012). Sentiment analysis tools present the output either as a binary classification or on a continuous scale as a sentiment score. Some sentiment analysis tools can also express topic-specific polarities and a general polarity score (Liu, 2012).

Researchers have reported on various sentiment analysis applications in the literature (Bai, 2011; Balahur, Hermida, & Montoyo, 2011; Duric & Song, 2012; Go, Bhayani, & Huang, 2009; Huang et al., 2013; Jiang, Yu, Zhou, Liu, & Zhao, 2011; Kouloumpis, Wilson, & Moore, 2011; Pang, Lee, & Vaithyanathan, 2002; Reyes & Rosso, 2012), yet, to the best of our knowledge, they have not sufficiently tested whether those applications are successful in supporting decisions. In this study, we address the gap in the empirical evaluation of the usefulness of sentiment scores in assisting individuals in purchase decisions. More precisely, we investigate how sentiment scores impact individuals' purchase decisions when used with online reviews that are not supported by other decision aids such as star ratings. Specifically, we address the following research question:

> **RQ:** Does presenting sentiment scores with online user reviews improve decision outcomes compared to presenting the same reviews without decision aids?

To address this research question, we performed an empirical investigation through a controlled experiment with 100 subjects. We evaluated decision outcomes with both objective and subjective measures. The objective measure was the time that users spent to search, find, and analyze the information, while the subjective measure was users' evaluation of their level of confidence about their decision. The empirical findings provide evidence that sentiment analysis can help individuals to make more efficient decisions, which has implications for consumers and providers of product- or service-related information.

This paper proceeds as follows. In Section 2, we present the background for this study. In Section 3, we present our research model and hypotheses. In Section 4, we present our research methodology and, in Section 5, analyze the data that we extracted from the experiment. In Section 6, we discuss the findings and implications and directions for future work. Finally, in Section 7, we conclude the paper.

2 Background

In this section, we first briefly review the cost-benefit framework and information foraging theory, which, together, provide a theoretical grounding for our investigation into the impacts of sentiment scores on purchase decision outcomes. Then, we review two main areas related to this study (support of consumer purchase decisions and sentiment analysis technology) and review the empirical work in these areas.

2.1 Cost/Benefit Framework

Cost-benefit theory provides a conceptual foundation for studying human decision behavior. This framework asserts that individuals strive to choose a strategy to simultaneously acquire beneficial information and lower cognitive cost. For this, they weigh the benefits (i.e., the positive impact of using extra information) and costs of acquiring and processing that information for a decision making task (Christensen-Szalanski, 1980; Creyer, Bettman, & Payne, 1990; Payne, 1982). In cost/benefit literature, “cost” refers to individuals’ mental effort to acquire and compute information, while “benefit” refers to the positive impact that effort has on their decision outcomes.

The effort-accuracy framework of cognition (Payne, 1982), an extension to cost-benefit framework, proposes that decision maker primarily focus on maximizing accuracy (decision quality) and minimizing cognitive effort. Because these objectives often conflict, individuals make some trade-offs between the two. In a series of studies that investigated what strategies and choices decision makers who use decision support systems (DSS) made, Todd and Benbasat (1992, 1994, 1999, 2000) found that decision makers adapt their strategy selection in such a way as to maintain a low overall expenditure of effort. The results of studies in the domain of intelligent systems are consistent with this notion of “effort minimization” (Gregor & Benbasat, 1999).

While purchasing a product or service, consumers look for as much useful information as possible to make the best decision. This search for information follows the cost-benefit theory or effort-accuracy framework. One can consider the time that consumers spend to find and indulge information the cost and the satisfaction and confidence in the final decision the benefit of their decision. To lower the costs of searching for relevant information, individuals use certain information cues. Below, we review a theory that can be useful in framing the use of such cues.

2.2 Information Foraging Theory

Information foraging theory explains how individuals adapt strategies and technologies for seeking, gathering, and consuming information to the flux of information in the environment. (Pirolli & Card, 1999). Pirolli and Card (1999) argue that information seeking in human mind is similar to food foraging behavior in animals and propose information foraging theory, which originated from optimal foraging theory or, more specifically, food foraging theory in anthropology (Winterhalder & Smith, 1992) and behavioral ecology (Stephens & Krebs, 1986).

Pirolli and Card (1999) note that: “The basic hypothesis of information foraging theory is that when feasible, a natural information system evolves toward stable states that optimize gains of valuable information per unit cost” (p. 643). The theory assumes that individuals, when possible, will modify their strategies of acquiring information or the structure of environment to maximize their rate of gaining valuable information. Optimal information foraging focuses on how people will best adapt themselves to their information environments and how information environments can best be shaped to serve their needs of getting the maximum amount of information with a limited amount of resource allocation (energy and time expenditure) (Pirolli, 2007).

Information foraging theory, as Pirolli and Card (1999) explain, attempts to specify the ways in which users search for information. According to Pirolli (1997), users are heavily influenced by the “information scent”. Pirolli (1997) indicates that cues in the immediate environment of information presentation will let out a “scent” about the nature of information. This scent will then direct the user to either choose and pursue that source of information or ignore it for another more promising information path.

Sundar, Knobloch-Westerwick, and Hastall (2007) investigated users' reliance on information cues to moderate information overload. They evaluated the impact of different cues on news websites and how they affected users' information foraging behavior. The found that individuals' main problem in information gathering and sense making was allocating attention (Pirolli & Card, 1999). Different types of cues have distinctive impact on users' attention. Sundar et al. (2007) compared the impact of different cues on users' news selection behavior and found evidence that different combinations of cues have different effects on users' information selection behavior.

Hyperlinked text on webpages is another example of information cues that can possess various levels of scent (a strong scent, weak scent, or no scent) based on the degree to which the hyper-linked words relate to the user's information goals (Pirolli, 1997). Pirolli, Pitkow, and Rao (1996) examined the impact of webpage clustering on information foraging and found support for the hypothesis that individuals who successfully use clustering may increase how effectively and efficiently they acquire information.

Further, as Khapre and Basha (2012) note, "Information cues, play a very important role in the process of directing the user to query information in the information foraging process" (p. 384). Individuals develop information feeding plans based on existing categories in their own minds (experience), their judgment of the available information (mental representation of information), and the specific tasks that they are trying to tackle. Sundar et al. (2007) also found that the variation in information foraging behavior depends on users' experience, information representation, and the information goal.

We believe that the quantitative data extracted from qualitative product/service reviews through sentiment analysis provides another cue that helps consumers navigate the information space and save them cognitive effort in choosing sources to read in a way to make the most informed decisions in limited time. The value that sentiment analysis scores provide in this context is similar to that of star ratings but has broader applicability since star ratings are not always available or sufficiently detailed to be effective. Also, as past research has suggested (Khapre & Basha, 2012; Sundar et al., 2007), we expect users' inherent characteristics to influence their information-foraging behavior with or without the presence of sentiment analysis scores.

2.3 Information Cues and Purchase Decisions

Researchers have traditionally thought that consumers search for new information mainly to reduce the uncertainty surrounding their decision (Cox, 1967; Hansen, 1972). They will search for information until they reduce their uncertainty to a tolerable level (Urbany, Dickson, & Wilkie, 1989). They can obtain such information from different sources. Economic and marketing studies have extensively shown that word of mouth (WOM) plays an important role in shaping consumer attitudes and behaviors (Buttle, 1998). More specifically, supplemental product information in the form of user and consumer feedback has increasingly begun to influence individuals' purchase decisions (Chevalier & Mayzlin, 2006; Dellarocas, Zhang, & Awad, 2007; Reinstein & Snyder, 2005).

One can consider reviews, either from a professional or a fellow consumer, as one of the best sources of product or service information (Dang, Zhang, & Chen, 2010). Because one can access online reviews regardless of time and distance, consumers often consider online reviews a better source of information than traditional paper-based reviews and, thus, prefer them (Mudambi & Schuff, 2010). Online reviews help consumers to decrease their decision time and effort, which contributes to a more satisfying purchase decision outcome (Schiffman & Kanuk, 2007). However, the rapid increase in the volume of Internet users and the growth of Web 2.0's (interactive Web) popularity among those users has given rise to massive collections of user-generated content (Turetken & Olfman, 2013). Hence, finding, gathering, comprehending, and using such information has become more challenging and time consuming.

Based on the theoretical background we discuss above, it is important to discern which reviews are the most useful in reducing consumers' purchase uncertainty. Star ratings that accompany reviews where consumers indicate their opinion about a product or service by writing comments along with a ranking (typically from 1-5) are popular among consumers. According to Chevalier and Mayzlin (2006), star ratings provide an excellent opportunity to measure the valence of comments without analyzing the comments themselves. Research has shown that consumers use decision and comparison aids (Todd & Benbasat, 1992) and numerical content ratings (such as star ratings) (Poston & Speier, 2005) to not only conserve cognitive resources and reduce energy expenditure to acquire information but also improve the purchase-decision process (Mudambi & Schuff, 2010).

Regardless of their popularity, star ratings can have limited utility in certain contexts. For example, some particularly long reviews have only an overall star rating assigned to the whole review, which is useful if the consumer decides to discard the review based on this overall rating. On the other hand, if a consumer finds the review to be potentially useful, the consumer then needs to decide which part of the overall review to read. A single star rating is not useful enough I such a case. This point is particularly relevant when comparing complex products and services with many features where it would be useful to have numeric scores for each specific feature separately.

Meanwhile, many other useful sources of reviews such as blog posts and social networking websites do not contain any numerical information that resembles star ratings. Therefore, the question arises as to which blog post or social networking website one should read given the limited (time) resources and the lack of additional cues such as star ratings on the products/services that these sources report on. Research from several disciplines (e.g., accounting, finance, consumer behavior) has found that, in the absence of adequate tools, such content would overload information users and lead to a decrease in the quality of the decisions they make (Abdel-Khalik, 1973; Chewning & Harrell, 1990; Shields, 1980; Snowball, 1980), an increase in the time required to make those decisions, and an increase in confusion regarding their decisions (Jacoby, Speller, & Berning, 1974; Jacoby, Speller, & Kohn, 1974; Malhotra, Jain, & Lagakos, 1982; Turetken & Sharda, 2004, 2005). This problem is even more relevant today when considering the big data paradigm, which not only points to the volume but also the dynamic and often inaccurate nature of available data. As we discuss in Section 2.4, researchers have advocated sentiment scores to improve information acquisition and to mitigate overload especially when other numerical ratings (e.g., star ratings) are not available.

2.4 Sentiment Analysis

Sentiment analysis (SA) is a data-summarization and opinion-mining technique that produces numeric sentiment scores in a spectrum from extremely negative to extremely positive. Many of the SA tools commonly used today rely on lexicons that store sentiment words and their polarity. Hatzivassiloglou and McKeown (1997) report that fully automated sentiment tools can identify sentiment words and their respective polarity in sentences with an accuracy level (compared to human experts) as high as 82 percent. Based on this promise, various researchers have focused on developing and improving sentiment dictionaries (Balahur et al., 2011; Maks & Vossen, 2012; Steinberger et al., 2012; Tufiş & Ştefănescu, 2012). This research has also strived to go beyond the typical sentiment-detection problems to emphasize multilingual sentiment analysis (Steinberger et al., 2012), actor subjectivity (Maks & Vossen, 2012), and irony detection (Reyes & Rosso, 2012). Recent research has also focused on going beyond supervised lexicon development to semi-automatic and automatic machine learning approaches (Bai, 2011; Duric & Song, 2012). This body of literature on sentiment analysis presents applications of the proposed approaches and their computational evaluations (Denis & Sagot, 2012; Go et al., 2009; Hajič, 2000; Huang et al., 2013; Jiang et al., 2011; Kouloumpis et al., 2011; Liu, Li, Zhou, & Xiong, 2011a, 2011b; Liu, Zhang, Wei, & Zhou, 2011; Pang et al., 2002); yet do not provide empirical support for the success of these applications on users in a typical decision making scenario such as purchasing a product.

For example, Huang et al. (2013) propose that one can use sentiment analysis results to make wiser and faster decisions based on the idea that the extensive amount of product or service information extracted from user generated content (mainly retrieved from Web 2.0) is not easy to digest otherwise. (Huang et al. (2013) explain that “feature-sentiment information can help users digest user-generated reviews more efficiently”. Their work relies on the results from a previous study by Yatani, Novati, Trusty, and Truong (2011) on “feature sentiments” and their impact on decisions. Neither of these studies evaluates the impact of “sentiment scores” or “polarity detections” on decision making.

In similar work, Cambria, Song, Wang, and Howard (2014) suggest that one can use sentiment analysis tools to extract useful information from unstructured data. Their system uses a blend of common and common sense knowledge to build a comprehensive resource in an attempt to emulate how tacit and explicit knowledge is organized in human mind and use this resource to design an opinion mining and sentiment analysis system. However, they fails to provide any evidence to support the claimed usefulness of the system they introduce.

Liu (2012) lists a set of application domains for sentiment analysis ranging from evaluating consumer products, services, healthcare, and financial services to analyzing social events and political elections. He particularly believes that one can use the information derived from sentiment analysis to make predictions.

He argues such analyses can predict sales performance, volume of comments in political blogs or box-office success of movies as well as characterizing social relations. However, he does not study these claims further and does not investigate the actual impact of the polarity of reviews on these human metrics. An understanding of the impact of review polarity on each individual is the first step in predicting aggregate outcomes such as sales performance that Liu (2012) discusses.

Schumaker, Zhang, Huang, and Chen (2012) developed a two-stage approach to detecting sentiment by first analyzing the subjectivity of financial news articles. They used these subjectivity scores and the sentiment scores of the subjective articles to predict stock prices and found that subjective news did not improve predictive ability.

From surveying the sentiment analysis literature, Rosas, Mihalcea, and Morency (2013) list some applications for sentiment analysis technology: 1) branding and product analysis (Hu & Liu, 2004), 2) analysis of political debates (Carvalho, Sarmento, Teixeira, & Silva, 2011), 3) question answering (Yu & Hatzivassiloglou, 2003), and 4) tracking sentiment timelines in online forums and news (Lloyd, Kechagias, & Skiena, 2005). Among these applications, scholars have empirically evaluated only the sentiment analysis of political debates (e.g., Tumasjan, Sprenger, Sandner, & Welpe, 2010; Thomas, Pang, & Lee, 2006). Likewise, Liebmann, Hagenau, and Neumann (2012) found evidence that using sentiment analysis to organize unstructured qualitative e-commerce data can help financial analysts and investors make better decisions in allocating resources in e-commerce. These studies stand out in their empirical support for the usefulness of sentiment analysis, but their focus is different from the more common consumer product decisions that we investigate in this paper.

The review of the empirical literature suggests that sentiment analysis is a promising technique to generate scores that can substitute explicit cues as star ratings in the absence of those ratings or when star ratings cannot sufficiently help one navigate a collection of reviews. However, to date, research in sentiment analysis has emphasized developing tools and increasing the accuracy of automatically detected sentiment scores but paid little attention to collecting empirical evidence on the usefulness of sentiment scores. We believe that the current maturity level of sentiment analysis more than warrants rigorous empirical work on its success in supporting decision makers. What we report in the rest of this paper contributes to filling that research gap.

3 Research Model and Hypotheses

Given the background we present in Section 2, we propose:

> **H1:** Sentiment scores that accompany user reviews improve consumers' purchase decision outcomes compared to online reviews without decision aids.

As with past research (e.g. Tan, Tan, & Teo, 2012), we evaluated purchase decision outcomes through efficiency and effectiveness with objective and subjective measures, respectively. We investigated efficiency via the time that consumers spend to acquire information and make a decision (surrogate for effort) and effectiveness via individuals' confidence in their decision.

As the extant literature that Tan et al. (2012) summarizes indicates, when a decision aid supports users, they benefit from lower search effort (Häubl & Murray, 2006; Häubl & Trifts, 2000; Swaminathan, 2003) and lower search time (Pedersen, 2000; Vijayasarathy & Jones, 2001). According to Tan et al. (2012), they benefit in this way partly because the decision aid provides users with more confidence. On the other hand, reflecting on our discussion on cost benefit trade-offs and the notion of effort minimization (Todd & Benbasat, 1992, 1994, 1999, 2000), we suggest that such aid will benefit users by reducing the effort they need to make (and, hence, by increasing the speed with which they will make decisions) regardless of whether they report higher confidence levels or not. As Tan et al. (2012) argue, consumers take time to not only inform themselves on the features of the decision choices they are faced with but also self-justify their final decisions. The presence of decision aids will shorten the time for both activities. Therefore, we hypothesize:

> **H1:** Product reviews with sentiment scores as a decision aid help individuals make their purchase decisions faster than decisions made with reviews without decision aids.

Confidence is an appropriate surrogate for effectiveness for this study because the decisions we examined presented no "dominating" alternatives to users; therefore, unlike what Tan et al. (2012) presented to their users in their experiment, we presented no right or wrong decisions. Therefore, one cannot define a strict accuracy measure; rather, what matters is for the users to make their decisions based on sufficient information. Users' relative confidence in their decisions indicates their subjective assessment of the "well informedness" of their decisions (Sniezek, 1992). In fact, as Russo, Schoemaker, and Russo (1989), providing appropriate information (i.e., information supported by decision aids) may make decision makers more confident even if that information is not directly useful for their decisions. Therefore, we hypothesize:

H2: Product reviews with sentiment scores as a decision aid help individuals make their decisions with a higher level of confidence than decisions made with the help of product reviews without decision aids.

As we discuss above, users' characteristics influence their information-foraging behavior (Khapre & Basha, 2012; Sundar et al., 2007). Other research on human behavior (e.g., Bucklin & Gupta, 1992; D'Souza, Taghian, & Khosla, 2007; Slama & Taschian, 1985) has also established that "user characteristics" play a significant role in decision outcomes. For example, decision makers' experience about a product may affect both their time to make a decision (Newman & Staelin, 1971) and their level of confidence in that decision (Washburn & Plank, 2002). Consumers who have previously consumed or used a product will make purchase decisions (in most cases) of the same type of product faster and are typically more satisfied with their choice. Therefore, we measure and control for participants' familiarity (familiarity with subject) with the decision context.

Likewise, demographical characteristics might affect decision outcomes both directly and through their effect on information-foraging behavior (Murphy & Olaru, 2009). Many researchers in the marketing and consumer behavior fields have extensively investigated gender differences in online purchase decisions (e.g., Berni, 2001; Brody, 1984; Chiger, 2001; Gutteling & Wiegman, 1993; Peter, Olson, & Grunert, 1999). Some earlier studies have focused on differences in perception of risk in performing online shopping, while more recent research has focused on consumer behavior or attitude in terms of information processing in specific situations. For example, Brody (1984), Gutteling and Wiegman (1993), Stern, Dietz, and Kalof (1993) found that women perceive greater risks in a wide variety of domains including financial, medical, and environmental. In more recent work, Homburg and Giering (2001) indicate that gender strongly influences the sales process and purchase intention satisfaction. Further, Konrad, Ritchie, Lieb, and Corrigall's (2000) results indicate that females, when involved in shopping decisions, pay more attention to personal interaction and interpersonal relationships than men. Peter et al. (1999) found that men and women process information differently. Chiger (2001) and Berni (2001) confirm as much in showing that men approach shopping tasks differently from an information processing perspective and are more independent and confident than women when it comes to purchase decisions (Darley & Smith, 1995). Given these findings, we included gender as another control variable in our model.

In this study, we gave participants only reviews and product specifications (both written documents) for them to make decisions about what to buy. Therefore, we used individuals' reading speed as a determinant of how effectively and efficiently they processed information for decision making purposes. If faster readers chose to consume comparable amounts of information as slower readers regardless of their efficiency advantage, then one should see effects in the speed with which they complete the decision task. If, on the other hand, they used this advantage to consume more of the available information, then they would be likely to feel more informed about the decision task and more confident of the end results. Although one cannot predict a priori how faster readers would choose to use this advantage in completing the decision task, one should nevertheless include reading speed as a control variable in the model.

Education is another factor that might affect information seeking behavior and consequently both decision making speed and confidence. Murphy and Olaru (2009) found significant differences between foraging styles in different education levels. Users with higher levels of education voraciously seek a range of information sources, while users with a lower education level wait for a few convenient sources to come their way (Murphy & Olaru, 2009). Our research model (see Figure 1) shows these demographical characteristics and their effect on decision making outcomes.

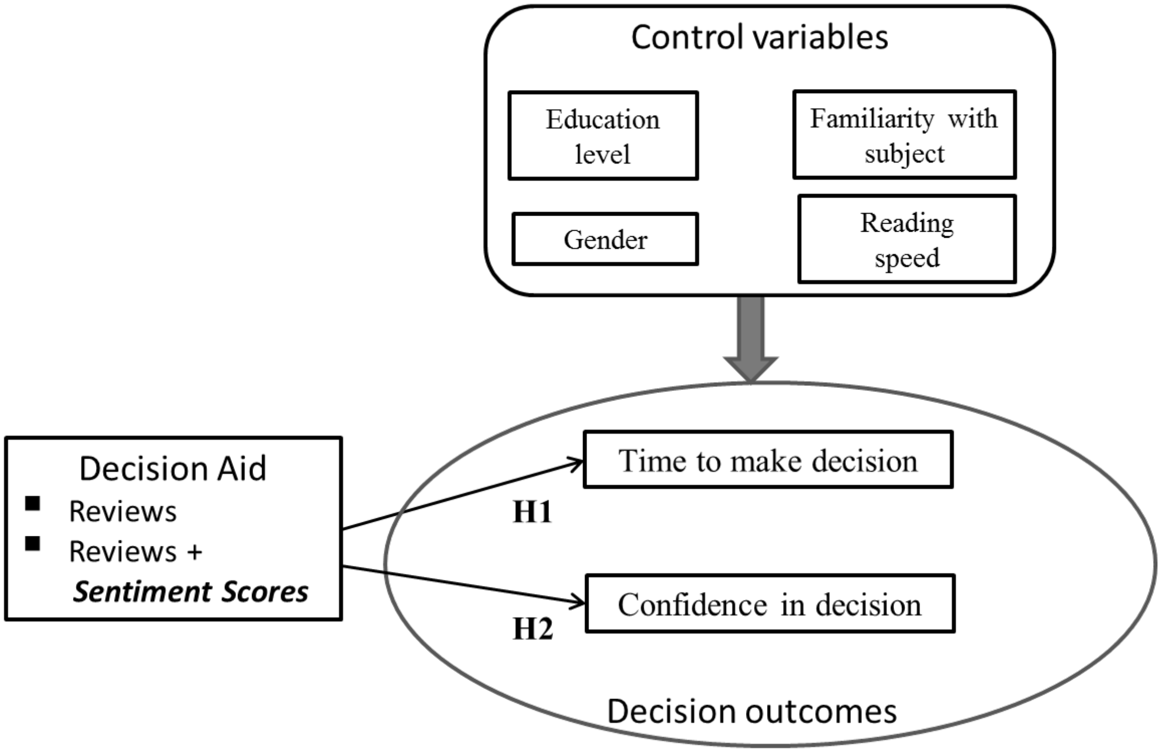

Figure 1. Research Model

4 Methodology

4.1 Sentiment Analysis Tool

In this study, we do design or develop a new sentiment analysis tool but rather assess the effectiveness of a tool that we consider to represent the available state-of-the-art sentiment analysis technology. This approach is consistent with many studies in literature (e.g., Schumaker et al., 2012) that have used an off-the-shelf commercial tool to represent a given technology in general. There are various open source and commercial text analytics tools that can perform sentiment analysis. The most commonly used one in scholarly papers is OpinionFinder (mpqa.cs.pitt.edu/opinionfinder/opinionfinder_1). Other common off-the-shelf sentiment analysis systems include SentiStrength (sentistrength.wlv.ac.uk), and Sentiment140 (www.sentiment140.com). These popular tools focus primarily on tweets and cannot analyze long documents. For our purposes, we chose Lexalytics (https://www.lexalytics.com/), which is a publicly available tool that can handle longer texts and delivers a sentiment score in three decimal places in the continuum between -1 and +1.

Like many other SA algorithms, the Lexalytics algorithm is lexicon based and depends on the sentiment scores of sentiment-bearing parts of the text. Lexalytics identifies these sentiment-bearing words and phrases through a well-known natural language processing technology called parts of speech (POS) tagging and combines their scores in a particular piece of text through a technique called “lexical chaining”.

We do not review Lexalytic’s algorithm here because it falls outside the study’s scope; however, Lexalytic (n.d.) provides one. To verify its performance, we compared the results from Lexalytics to those from another popular system, Lymbix (Gînscă et al., 2011), that analyzes and yields analytically equivalent results (i.e., the outcome for both systems is a specific polarity score rather than a binary classification). To conduct this comparison, we randomly selected 90 paragraphs from a collection of product reviews and applied sentiment analysis to this sample using each tool. Correlation analysis results show (Table 1) that the quantitative values that the two systems provided were significantly correlated ($p < 0.01$) and that the distributions of the scores from each tool were not significantly different ($p = 0.297$) from each other, which confirm that Lexalytics well represents common sentiment analysis tools that deliver a sentiment score (rather than a binary classification) on variable length text. As such, we found it appropriate for the study.

Table 1. Relationship between Alternative Sentiment Analysis Tools (Lexalytics and Lymbix)

Correlations		Lymbics	Lexalytics
Lexalytics	Pearson correlation	1	.328**
	Sig. (two-tailed)		.002
	N	119	85
Lymbix	Pearson correlation	.328**	1
	Sig. (two-tailed)	.002	
	N	85	85
Chi-Square Tests			
	Value	**Df**	**Asymptotic significance (2-sided)**
Pearson chi-square	4562.311	4560	.488
Likelihood ratio	616.180	4560	1.000
Linear-by-linear association	9.018	1	.003
Number of valid cases	85		

4697 cells (100.0%) had an expected count less than 5. The minimum expected count was 0.01.
** Correlation was significant at the 0.01 level (two-tailed).

4.2 Experimental Design and Subjects

We conducted the experiments following a two-group between-subjects design: we provided the control group with only online reviews and the treatment group with online reviews and sentiment scores that we obtained in advance using Lexalytics. We recruited experimental subjects via announcements made through posters in a large urban university campus and invitations sent through Eventbrite website (www.eventbrite.com). Even though recruiting subjects in this manner made the sessions open to public, students were the main participants in the experiment. The sample comprised 117 subjects, most of whom were students with different levels of education. Further, 61 percent of the participants were male and 39 percent were female; 30 percent of the participants were high school graduates and 26 percent had a postgraduate degree. Therefore, we are confident that our sample was diverse from an educational point of view. Given the task domain as we discuss in Section 4.3, one can also reasonably assume that this sample represents the general population of potential consumers for the products we presented in the experiments.

4.3 Material Used in the Experiment

As Wang and Benbasat (2007, 2008) suggest, we used digital cameras as the focus of the decision task. We believe digital cameras are a reasonable choice because they are valuable enough for consumers to spend a nontrivial amount of effort researching but too important for them to seek additional offline advice as would likely be the case if they sought to purchase a house or a car. We selected three choices of competing digital single lens reflect (DSLR) cameras from three different brands as the target of the experimental purchase decision. These cameras were potential substitutes for each other not due to their cost but to their general capability according to both the BestBuy and Amazon websites. The final document presented to the experimental subjects in the control group was approximately 80 pages of reviews from camera professionals and from (potential) buyers with a hyperlinked directory that provided the readers with direct access to any part of the document. The treatment group received the same material supplemented by sentiment scores. The decision task was choosing a camera among the given alternatives.

4.4 Experimental Procedure

The experimental sessions took place in a large urban North American university. We provided a US$10 incentives for completing the experiment to increase the probability of receiving more accurate input and realistic feedback. We organized multiple time slots for flexibility. In a typical session, we gave participants information on the basics of the research such as the study procedure and the content and limitations of the information we provided. After we collected participants' signed consent forms, they started their session by completing the questions regarding their background and demographics information (see Table 2). We then provided the participants with a brief tutorial, which described the task and the relevant information given to them; we provided participants in the treatment group with addition information on what the sentiment scores accompanying the reviews mean. Subsequently, we directed all participants to the camera review information (with or without sentiment scores). Once they were convinced that they had sufficient information to make their decision, they made their simulated purchase and completed the session by answering questions regarding their confidence in their purchase decision and their thoughts about the information provided to them.

4.5 Measurement Scales

The online survey tool (Qualtrics) automatically captured "time to make decision" from the second subjects started reading the reviews to when they made their decision and got back to the questions. We also used this time-measurement tool to measure and monitor participants' natural reading speed while they were completing the background and demographic questions. We extracted decision makers' confidence questions from O'Connor's study on "validation of a decisional conflict scale" (O'Connor, 1995). We measured user characteristics by direct questions on age, gender, and education level. We used questions on prior experience with the product (considered for purchase in the study) and owning or intention to purchase it in a one-year timeframe to evaluate the participant's familiarity with the subject. Table 2 summarizes the pre- and post-test questionnaire.

Table 2. The Questionnaire Items

1. Background
What is your gender? • Male • Female
What is the highest level of education you have completed? • Some high school • High school graduate • Some college • Trade/Technical/Vocational training • College graduate • Some post graduate work • Postgraduate degree
Please specify the extent you agree with the below statement (1 = strongly disagree; 7 = strongly agree) • I am confident that my choice of camera is the best of the three • I have a strong feeling that the camera I chose is the best • I am confident about my choice of camera

5 Data Analysis and Results

5.1 Preliminary Analysis

After we removed outliers (belonging to users who spent five minutes or less reviewing the file before making their decisions) from the dataset, 100 data points remained, which were distributed equally between the treatment and control groups. We used this reduced sample in subsequent analyses. We used IBM's statistical analysis tool SPSS (ver. 21) to calculate a single confidence variable obtained from confirmatory factor analysis conducted using principal component analysis (MacCallum & Browne, 1993) (see Table 3). Table 4 provides descriptive statistics for our variables. Table 5 shows the distribution of the data for each variable. Most of our participants claimed they had moderate knowledge about the subject (i.e., DSLR cameras), while 17 percent reported absolutely no familiarity, and 6 percent reported strong familiarity. In our analysis, we considered a familiarity level lower than 2 (out of 7) as unfamiliar and a familiarity level higher than 5 (out of 7) as familiar.

As Table 6 shows, we found no significant correlation between our two dependent variables. Therefore, we used two separate analysis of covariance (ANCOVA) models (one for each dependent variable)

(Keppel, 1991) to test the difference between the two groups supported by different levels of decision aids while controlling for the other user characteristics (familiarity, education, reading speed, and gender) that may have had an effect on the dependent variables of time to make decision and confidence in decision.

We first assessed the assumptions of ANCOVA on our dataset. We investigated the normality assumption through Q-Q plots. Both dependent variables followed a normal distribution with a slight deviation for lower values of time (see Figure 2), which might increase the type 1 error (which would bias the analysis results, making the significance estimates more conservative). We checked the independence of the covariates and treatment effects assumption via a t-test of the covariates (control variables).

We tested the homogeneity of regression slope and the regression relationship between the dependent variable and concomitant variables through scatter plots and investigated the independence of error terms through an independent sample t-test on all control variables. The results provide support that this technique was appropriate to test our hypotheses.

Table 3. Factor Analysis on Confidence Items

<table>
<tr><th colspan="7">Communalities</th></tr>
<tr><th>Item</th><th colspan="3">Initial</th><th colspan="3">Extraction</th></tr>
<tr><td>Confidence 1</td><td colspan="3">1.000</td><td colspan="3">.706</td></tr>
<tr><td>Confidence 2</td><td colspan="3">1.000</td><td colspan="3">.815</td></tr>
<tr><td>Confidence 3</td><td colspan="3">1.000</td><td colspan="3">.603</td></tr>
<tr><td colspan="7">Extraction method: principal component analysis</td></tr>
<tr><th colspan="7">Total Variance Explained</th></tr>
<tr><td>Component</td><td colspan="3">Initial eigenvalues</td><td colspan="3">Extraction sums of squared loadings</td></tr>
<tr><td></td><td>Total</td><td>% of variance</td><td>Cumulative %</td><td>Total</td><td>% of Variance</td><td>Cumulative %</td></tr>
<tr><td>1</td><td>2.124</td><td>70.802</td><td>70.802</td><td rowspan="3">2.124</td><td rowspan="3">70.802</td><td rowspan="3">70.802</td></tr>
<tr><td>2</td><td>.586</td><td>19.524</td><td>90.325</td></tr>
<tr><td>3</td><td>.290</td><td>9.675</td><td>100.000</td></tr>
<tr><td colspan="7">Extraction method: principal component analysis</td></tr>
<tr><th colspan="7">Component matrix</th></tr>
<tr><th rowspan="2" colspan="3">Item</th><th colspan="4">Component</th></tr>
<tr><td colspan="4">1</td></tr>
<tr><td colspan="3">Confidence 1</td><td colspan="4">.840</td></tr>
<tr><td colspan="3">Confidence 2</td><td colspan="4">.903</td></tr>
<tr><td colspan="3">Confidence 3</td><td colspan="4">-.777</td></tr>
<tr><td colspan="7">Extraction method: principal component analysis</td></tr>
</table>

Table 4. Descriptive Statistics

	N	Minimum	Maximum	Mean	Standard dev.
Decision aid	100	0	1	.50	.503
Gender	100	1	2	1.39	.490
Education level	100	2	7	4.49	2.028
Familiarity with subject	100	1.000	3.000	2.15	0.880
Reading speed	100	1.451	219.829	51.018	36.888
Time to make decision	100	63.200	2,945.131	1,097.952	785.486
Confidence in decision	100	-2.532	1.393	0.000	1.000

Table 5. Frequencies

Decision aid	
Review + Sentiment score	50
Review	50
Gender	
Male	61
Female	39
Familiarity with subject	
Low	32
Medium	21
High	47
Education level	
High school graduate	30
Some college	12
College graduate	21
Some post graduate work	11
Post graduate degree	26

Table 6. Dependent Variable Correlations

	Time to make decision	Confidence in decision
Time to make decision	1.000	0.012 (0.906)
Confidence in decision	0.012 (0.906)	1.000

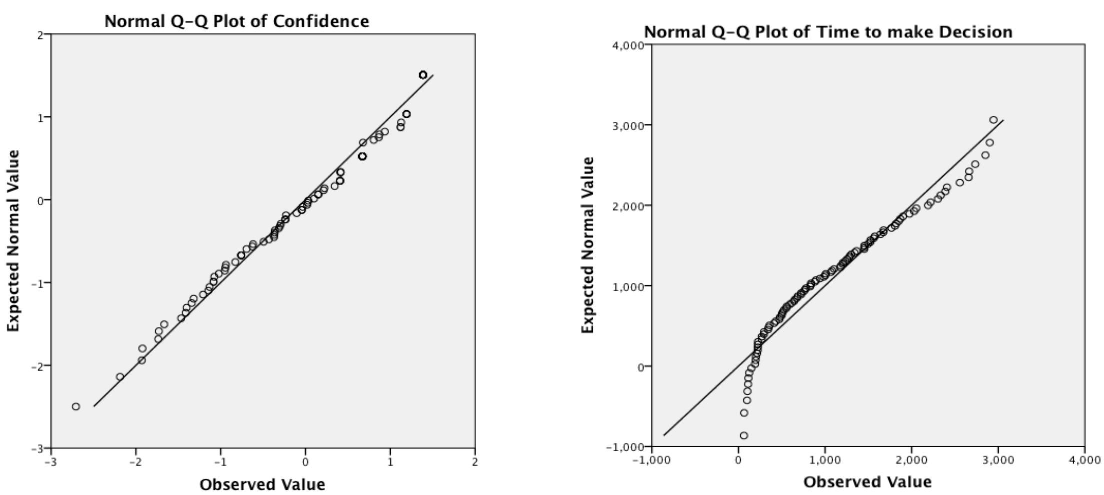

Figure 2. Test for Normality Using Q-Q Plots

5.2 Test of the Hypotheses

We tested the hypotheses one at a time. Table 7 provides the results from the ANCOVA on time to make decision. We observed a significant difference between the two groups (p = 0.004, one-tailed). The lower panel of Table 6 shows that the difference was in the hypothesized direction (time to make decision (reviews only) > time to make decision (reviews + sentiment scores)), which supports H1.

Table 7. ANCOVA Test for H_1

Source		Type III sum of squares	df	F	Significance
Decision aid	Hypothesis	5482844.487	1	10.028	.004
	Error	5799729.993	10.067		
	Error	12457926.856	23.606		
Decision aid * education	Hypothesis	4396629.267	4	2.410	.059
	Error	27825135.921	61		
Decision aid * familiarity with subject	Hypothesis	1335493.920	2	1.464	.239
	Error	27825135.921	61		
Decision aid * gender	Hypothesis	30741.111	1	.081	.777
	Error	27825135.921	61		
Dependent variable: time to make decision					
Decision aid	Mean	Std. error	95% confidence interval		
			Lower bound	Upper bound	
Reviews + sentiment score	778.690	175.917	426.923	1130.458	
Reviews only	1313.444	153.391	1006.721	1620.167	
We evaluated covariates appearing in the model at the following values: reading speed = 51.01799.					

Table 8 provides the analysis on confidence. As the table shows, we found no main effect for decision aid (p = 0.367, one-tailed), but the significant interaction term (decision aid x gender) (p = 0.019, two-tailed) suggests that the results partially support H2 (for one of the genders). For that, we tested H2 separately for each gender and found no significant effect for decision aid for male (p = 0.235, one-tailed) or female (p = 0.099, one-tailed) subjects (Tables 9 and 10). However, although not significant, the effect of decision aid on confidence was in opposite directions for male (-) and female subjects (+).

Table 8. ANCOVA Test for H_2

Source		Type III sum of squares	df	F	Significance
Decision aid	Hypothesis	0.264	1	0.144	.367
	Error	4.438	2.414		
Decision aid * education	Hypothesis	4.802	4	1.398	.245
	Error	52.389	61		
Decision aid * familiarity with subject	Hypothesis	1.238	2	0.721	.490
	Error	52.389	6		
Decision aid * gender	Hypothesis	4.966	1	5.782	.019
	Error	52.389	61		

Table 9. ANCOVA Test for H_2 for Male Subjects Only

Source		Type III sum of squares	df	F	Significance
Decision aid	Hypothesis	0.547	1	0.541	.235
	Error	19.034	2.414		
Decision aid * education	Hypothesis	5.138	4	1.477	.232
	Error	27.832	32		
Decision aid * familiarity with subject	Hypothesis	2.344	2	1.347	.274
	Error	27.832	32		

Table 10. ANCOVA Test for H_2 for Female Subjects Only

Source		Type III sum of squares	df	F	Significance
Decision aid	Hypothesis	0.777	1	1.832	.099
	Error	5.858	13.819		
Decision aid * education	Hypothesis	2.887	4	0.923	.488
	Error	7.819	10		
Decision aid * familiarity with subject	Hypothesis	0.111	1	.141	.715
	Error	7.819	10		

6 Discussion of the Results and Future Directions

Our results indicate that sentiment scores improve the efficiency of individuals' purchase decisions but do not affect their effectiveness. This finding is important because it provides empirical evidence on the usefulness of sentiment analysis technology especially in the absence of other common cues such as star ratings. We can explain the fact that we found significant speed improvement without a significant change in effectiveness by the "effort minimization" concept as discussed by Todd and Benbasat (1992), who argue that, when decision makers have decision aids that expand their information-processing capabilities, they use them not necessarily to analyze problems in more depth to make better decisions but to reduce their effort to achieve a similar level of outcome quality. By showing the applicability of the "effort minimization" concept in the use of text analytics tools, this study makes a contribution to research in the area of DSS impact by suggesting that the concept of DSS should be broadly defined to involve contemporary (big) data analytics tools.

Meanwhile, our results show a significant effect of the interaction between decision aid and gender on decision confidence, which may be how the impact of sentiment analysis technology differs from that of other decision support tools that previous research has explored where neither the theory nor empirical findings point to an interaction between the form of decision support and gender. However, we did not focus on this interaction in this study; as such, we did not include this interaction term in our research model a priori and did not focus on gender balance in the study sample while recruiting our participants. Therefore, our sample does not provide sufficient power to thoroughly test this interaction. Future research that evaluates the impact of sentiment analysis should focus on gender as a moderator.

We did not include explicit user ratings in the review material we used in our experiments; therefore, our results may generalize to reviews of various kinds including those done in blogs and other forms of social media posts. We suggest that businesses can benefit from integrating sentiment analysis tools along with reviews (with or without explicit (star) ratings) on their e-commerce websites. If consumers choose to use these aids, the expected increase in their decision speed will likely increase their satisfaction with the overall experience and improve the popularity of the products and brands offered through the website in question.

Sentiment scores as used in our experiments would also have practical managerial value especially if one could complement them with further text analytics techniques such as clustering. Product and brand managers can use clustering to identify emerging themes in positive or negative reviews. Managers could then direct their efforts to strengthen the themes (e.g. easy-to-use products) that evoke positive sentiments while modifying the attributes of their offerings (e.g., slow customer service) that evoke negative sentiments. There is potential for significant future research in this area of managerial decision support.

Meanwhile, theories of task technology fit (e.g., Goodhue & Thompson, 1995) argue that the success of technology such as sentiment analysis tools depends on the tasks that they are expected to support. We expect that services and experience goods may be a better fit for leveraging sentiment analysis technology. Therefore, we plan to further investigate the success of sentiment analysis in supporting decision tasks such as selecting doctors, planning travel, and selecting schools.

Lastly, the relatively low familiarity of sentiment scores may have limited their impact. Virtually every consumer who has shopped (or even only looked at product reviews) online is familiar with the presentation of user ratings in the form of stars or a score out of 5. In the experiments, the presentation of sentiment scores deviated from this standard where we provided sentiment scores as a non-integer number with 3 decimal points. Although this presentation has higher precision (a continuous versus a discrete score) and, hence, conveys more information, the presentation of scores in the form of star ratings, which are familiar to users, might have been easier to comprehend. In future work, we will test different alternative visualizations of sentiment scores to determine their most effective presentation.

7 Conclusion

The big data phenomenon is pervasive. User-generated content in social media conforms to the commonly articulated characteristics of big data and, as with other forms of big data, should be processed with data analytics techniques to leverage its true potential. In this study, we used sentiment analysis, which automatically detects the tone of text. To our knowledge, our paper represents the first attempt to evaluate the effectiveness of sentiment analysis technology via a rigorous user study. Our results show promise in sentiment scores' usefulness as a decision aid, which leads to practical implications and future research opportunities. We believe this paper provides a good framework for how research can further evaluate sentiment analysis technology by identifying the technological and other contextual factors that enable one to realize its true value.

Acknowledgments

This work was partially supported by a Natural Sciences and Engineering Research Council-NSERC (of Canada) Discovery Grant.

References

Abdel-Khalik, A. R. (1973). The effect of aggregating accounting reports on the quality of the lending decision: An empirical investigation. *Journal of Accounting Research*, 104-138.

Bai, X. (2011). Predicting consumer sentiments from online text. *Decision Support Systems, 50*(4), 732-742.

Balahur, A., Hermida, J. M., & Montoyo, A. (2011). *Detecting implicit expressions of sentiment in text based on commonsense knowledge.* Paper presented at the Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis.

Berni, S. (2001). He shops, she shops. *Progressive Grocer, 80*(3), 51-53.

Brody, C. J. (1984). Differences by sex in support for nuclear power. *Social Forces, 63*(1), 209-228.

Bucklin, R. E., & Gupta, S. (1992). Brand choice, purchase incidence, and segmentation: An integrated approach. *Journal of Marketing Research, 29*(2), 201.

Buttle, F. A. (1998). Word of mouth: understanding and managing referral marketing. *Journal of Strategic Marketing, 6*(3), 241-254.

Cambria, E., Song, Y., Wang, H., & Howard, N. (2014). Semantic multidimensional scaling for open-domain sentiment analysis. *IEEE Intelligent Systems, 29*(2), 44-51.

Carvalho, P., Sarmento, L., Teixeira, J., & Silva, M. J. (2011). *Liars and saviors in a sentiment annotated corpus of comments to political debates.* Paper presented at the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*(3), 345-354.

Chewning, E. G., & Harrell, A. M. (1990). The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task. *Accounting, Organizations and Society, 15*(6), 527-542.

Chiger, S. (2001). Consumer shopping survey. *Catalog Age, 18*(9), 57-60.

Christensen-Szalanski, J. J. (1980). A further examination of the selection of problem-solving strategies: The effects of deadlines and analytic aptitudes. *Organizational Behavior and Human Performance, 25*(1), 107-122.

Cox, D. F. (1967). Risk handling in consumer behavior: An intensive study of two cases. In D. F. Cox (Ed.), *Risk taking and information handling in consumer behavior* (pp. 34-81). Boston: Harvard University Press.

Creyer, E. H., Bettman, J. R., & Payne, J. W. (1990). The impact of accuracy and effort feedback and goals on adaptive decision behavior. *Journal of Behavioral Decision Making, 3*(1), 1-16.

D'Souza, C., Taghian, M., & Khosla, R. (2007). Examination of environmental beliefs and its impact on the influence of price, quality and demographic characteristics with respect to green purchase intention. *Journal of Targeting, Measurement and Analysis for Marketing, 15*(2), 69-78.

Dang, Y., Zhang, Y., & Chen, H. (2010). A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems, 25*(4), 46-53.

Darley, W. K., & Smith, R. E. (1995). Gender differences in information processing strategies: An empirical test of the selectivity model in advertising response. *Journal of Advertising, 24*(1), 41-56.

Dellarocas, C., Zhang, X. M., & 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.

Denis, P., & Sagot, B. (2012). Coupling an annotated corpus and a lexicon for state-of-the-art POS tagging. *Language Resources and Evaluation, 46*(4), 721-736.

Duric, A., & Song, F. (2012). Feature selection for sentiment analysis based on content and syntax models. *Decision Support Systems, 53*(4), 704-711.

Gînscă, A.-L., Boroş, E., Iftene, A., TrandabĂţ, D., Toader, M., Corîci, M., Perez, C. A., & Cristea, D. (2011). *Sentimatrix: Multilingual sentiment analysis service.* Paper presented at the Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis.

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford, 1*(12).

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, *19*(2), 213-236.

Gregor, S., & Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Quarterly*, *23*(4), 497-530.

Gutteling, J. M., & Wiegman, O. (1993). Gender-specific reactions to environmental hazards in the Netherlands. *Sex Roles, 28*(7-8), 433-447.

Hajič, J. (2000). Morphological tagging: Data vs. dictionaries. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference.*

Hansen, F. (1972). Consumer choice behavior: A cognitive theory. New York: The Free Press.

Hatzivassiloglou, V., & McKeown, K. R. (1997). *Predicting the semantic orientation of adjectives.* In *Proceedings of the 8th Conference on European Chapter of the Association for Computational Linguistics.*

Häubl, G., & Murray, K. B. (2006). Double agents: Assessing the role of electronic product recommendation systems. *Sloan Management Review, 47*(3), 8-12.

Häubl, G., & Trifts, V. (2000). Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing Science, 19*(1), 4-21.

Homburg, C., & Giering, A. (2001). Personal characteristics as moderators of the relationship between customer satisfaction and loyalty—an empirical analysis. *Psychology & Marketing, 18*(1), 43-66.

Hu, M., & Liu, B. (2004). *Mining and summarizing customer reviews.* In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Huang, S.-W., Tu, P.-F., Fu, W.-T., & Amanzadeh, M. (2013). *Leveraging the crowd to improve feature-sentiment analysis of user reviews.* In *Proceedings of the 2013 International Conference on Intelligent User Interfaces.*

Jacoby, J., Speller, D. E., & Berning, C. K. (1974). Brand choice behavior as a function of information load: Replication and extension. *Journal of Consumer Research, 1*(1), 33-42.

Jacoby, J., Speller, D. E., & Kohn, C. A. (1974). Brand choice behavior as a function of information load. *Journal of Marketing Research*, *11*(1), 63-69.

Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011). *Target-dependent twitter sentiment classification.* In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.

Keller, K. L., & Staelin, R. (1987). Effects of quality and quantity of information on decision effectiveness. *Journal of Consumer Research, 14*(2), 200-213.

Keppel, G. (1991). *Design and analysis: A researcher's handbook*. New York: Prentice-Hall.

Khapre, S., & Basha, M. S. (2012). Advancement in information foraging theory. *Intelligent Information Management, 4*(6), 383.

Konrad, A. M., Ritchie, J. E., Jr., Lieb, P., & Corrigall, E. (2000). Sex differences and similarities in job attribute preferences: A meta-analysis. *Psychological Bulletin, 126*(4), 593-641.

Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg! In *Proceedings of the 5th International Conference on Weblogs and Social Media* (pp. 538-541).

Lexalytic. (n.d.). *Sentiment analysis: What is sentiment analysis?* Retrieved from https://www.lexalytics.com/technology/sentiment

Liebmann, M., Hagenau, M., & Neumann, D. (2012). Information processing in electronic markets: Measuring subjective interpretation using sentiment analysis. In *Proceedings of the International Conference on Information Systems.*

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*(1), 1-167.

Liu, X., Li, K., Zhou, M., & Xiong, Z. (2011a). Collective semantic role labeling for tweets with clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence.*

Liu, X., Li, K., Zhou, M., & Xiong, Z. (2011b). Enhancing semantic role labeling for tweets using self-training. In *Proceedings of the AAAI Conference on Artificial Intelligence.*

Liu, X., Zhang, S., Wei, F., & Zhou, M. (2011). *Recognizing named entities in tweets.* In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies.*

Lloyd, L., Kechagias, D., & Skiena, S. (2005). Lydia: A system for large-scale news analysis. In *Proceedings of the International Symposium on String Processing and Information Retrieval*.

MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin, 114*(3), 533-541.

Maks, I., & Vossen, P. (2012). A lexicon model for deep sentiment analysis and opinion mining applications. *Decision Support Systems, 53*(4), 680-688.

Malhotra, N. K., Jain, A. K., & Lagakos, S. W. (1982). The information overload controversy: An alternative viewpoint. *The Journal of Marketing*, 46, 27-37.

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

Murphy, J., & Olaru, D. (2009). How information foraging styles relate to tourism demographics and behaviours. *Journal of Vacation Marketing, 15*(4), 299-309.

Newman, J. W., & Staelin, R. (1971). Multivariate analysis of differences in buyer decision time. *Journal of Marketing Research*, *8*(2), 192-198.

O'Connor, A. M. (1995). Validation of a decisional conflict scale. *Medical Decision Making, 15*(1), 25-30.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up? Sentiment classification using machine learning techniques.* In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*.

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin, 92*(2), 382-401.

Pedersen, P. E. (2000). Behavioral effects of using software agents for product and merchant brokering: an experimental study of consumer decision-making. *International Journal of Electronic Commerce, 5*(1), 125-141.

Peter, J. P., Olson, J. C., & Grunert, K. G. (1999). *Consumer behavior and marketing strategy*. London: McGraw-Hill.

Pirolli, P. (1997). Computational models of information scent-following in a very large browsable text collection. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems.*

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*. Oxford: Oxford University Press.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review, 106*(4), 643-675.

Pirolli, P., Pitkow, J., & Rao, R. (1996). *Silk from a sow's ear: Extracting usable structures from the web.* In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems.*

Poston, R. S., & Speier, C. (2005). Effective use of knowledge management systems: A process model of content ratings and credibility indicators. *MIS Quarterly*, *29*(2), 221-244.

Raghunathan, S. (1999). Impact of information quality and decision-maker quality on decision quality: A theoretical model and simulation analysis. *Decision Support Systems, 26*(4), 275-286.

Reinstein, D. A., & Snyder, C. M. (2005). The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *The Journal of Industrial Economics, 53*(1), 27-51.

Reyes, A., & Rosso, P. (2012). Making objective decisions from subjective data: Detecting irony in customer reviews. *Decision Support Systems, 53*(4), 754-760.

Rosas, V., Mihalcea, R., & Morency, L.-P. (2013). Multimodal sentiment analysis of Spanish online videos. *IEEE Intelligent Systems, 28*(3), 38-45.

Russo, J. E., Schoemaker, P. J., & Russo, E. J. (1989). *Decision traps: Ten barriers to brilliant decision-making and how to overcome them*. New York: Doubleday.

Schiffman, L. G., & Kanuk, L. L. (2007). *Consumer behavior.* Upper Saddle River, NJ: Pearson.

Schumaker, R. P., Zhang, Y., Huang, C.-N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems, 53*(3), 458-464.

Shields, M. D. (1980). Some effects on information load on search patterns used to analyze performance reports. *Accounting, Organizations and Society, 5*(4), 429-442.

Slama, M. E., & Tashchian, A. (1985). Selected socioeconomic and demographic characteristics associated with purchasing involvement. *The Journal of Marketing*, *49*(1), 72-82.

Sniezek, J. A. (1992). Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behavior and Human Decision Processes, 52*(1), 124-155.

Snowball, D. (1980). Some effects of accounting expertise and information load: An empirical study. *Accounting, Organizations and Society, 5*(3), 323-338.

Steinberger, J., Ebrahim, M., Ehrmann, M., Hurriyetoglu, A., Kabadjov, M., Lenkova, P., Steinberger, R., Tanve, H., Vazquez, S., & Zavarella, V. (2012). Creating sentiment dictionaries via triangulation. *Decision support systems, 53*(4), 689-694.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton, NJ: Princeton University Press.

Stern, P. C., Dietz, T., & Kalof, L. (1993). Value orientations, gender, and environmental concern. *Environment and Behavior, 25*(5), 322-348.

Stigler, G. J. (1961). The economics of information. *The Journal of Political Economy*, *69*(3), 213-225.

Sundar, S. S., Knobloch-Westerwick, S., & Hastall, M. R. (2007). News cues: Information scent and cognitive heuristics. *Journal of the American Society for Information Science and Technology, 58*(3), 366-378.

Swaminathan, V. (2003). The impact of recommendation agents on consumer evaluation and choice: The moderating role of category risk, product complexity, and consumer knowledge. *Journal of Consumer Psychology, 13*(1), 93-101.

Tan, W.-K., Tan, C.-H., & Teo, H.-H. (2012). Consumer-based decision aid that explains which to buy: Decision confirmation or overconfidence bias? *Decision Support Systems, 53*(1), 127-141.

Thomas, M., Pang, B., & Lee, L. (2006). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing.*

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

Todd, P., & Benbasat, I. (1994). The influence of decision aids on choice strategies: An experimental analysis of the role of cognitive effort. *Organizational Behavior and Human Decision Processes, 60*(1), 36-74.

Todd, P., & Benbasat, I. (1999). Evaluating the impact of DSS, cognitive effort, and incentives on strategy selection. *Information Systems Research, 10*(4), 356-374.

Todd, P., & Benbasat, I. (2000). Inducing compensatory information processing through decision aids that facilitate effort reduction: An experimental assessment. *Journal of Behavioral Decision Making, 13*(1), 91-106.

Tufiş, D., & Ştefănescu, D. (2012). Experiments with a differential semantics annotation for WordNet 3.0. *Decision Support Systems, 53*(4), 695-703.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In *Proceedings of the AAAI Conference on Artificial Intelligence (pp.* 178-185).

Turetken, O., & Olfman, L. (2013). Introduction to the special issue on human-computer interaction in the Web 2.0 era. *AIS Transactions on Human-Computer Interaction, 5*(1), 1-5.

Turetken, O., & Sharda, R. (2004). Development of a fisheye-based information search processing aid (FISPA) for managing information overload in the web environment. *Decision Support Systems, 37*(3), 415-434.

Turetken, O., & Sharda, R. (2005). Clustering-based visual interfaces for presentation of web search results: An empirical investigation. *Information Systems Frontiers, 7*(3), 273-297.

Urbany, J. E., Dickson, P. R., & Wilkie, W. L. (1989). Buyer uncertainty and information search. *Journal of Consumer Research, 16*(2), 208-215.

Vijayasarathy, L. R., & Jones, J. M. (2001). Do Internet shopping aids make a difference? An empirical investigation. *Electronic Markets, 11*(1), 75-83.

Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems, 23*(4), 217-246.

Wang, W., & Benbasat, I. (2008). Attributions of trust in decision support technologies: A study of recommendation agents for e-commerce. *Journal of Management Information Systems, 24*(4), 249-273.

Washburn, J. H., & Plank, R. E. (2002). Measuring brand equity: An evaluation of a consumer-based brand equity scale. *Journal of Marketing Theory and Practice, 10*(1), 46-62.

Winterhalder, B., & Smith, E. A. (1992). Evolutionary ecology and the social sciences. *Evolutionary Ecology and Human Behavior*, 3-23.

Yatani, K., Novati, M., Trusty, A., & Truong, K. N. (2011). *Review spotlight: A user interface for summarizing user-generated reviews using adjective-noun word pairs.* In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

Yu, H., & Hatzivassiloglou, V. (2003). *Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences.* In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*.

About the Authors

Parisa Lak is a PhD candidate at Data Science Laboratory, Department of Mechanical and Industrial Engineering at Ryerson University, Toronto, Canada. Parisa is also IBM's research partner involved in IBM Watson Analytics projects. Her research focus is on topics related to machine learning, data mining and information retrieval, and more specifically on recommender systems. She received her MBA from Sharif University of Technology and MMSc in Management of Innovation and Technology from Ted Rogers School of Management at Ryerson University. She is also a member of IEEE and ACM Special Interest Group of Information Retrieval.

Ozgur Turetken is a professor at and the director of the Ted Rogers School of Information Technology Management at Ryerson University. His research and teaching interests are in applied analytics and its impacts on individual information use and decision making. His prior works have been published in many peer reviewed international journals and conferences such as *ACM Database, Association for Information Systems Transactions on Human Computer Interaction, Communications of the ACM, Decision Support Systems, Information & Management, IEEE Transactions on Professional Communication,* and *International Conference on Information Systems.* Ozgur was the chair of the Association for Information Systems Special Interest Group on Decision Support and Analytics, and is on the editorial board of *Association for Information Systems Transactions on Human Computer Interaction*. He received a PhD in management science and information systems from Oklahoma State University. He also holds BSc and MBA degrees from Middle East Technical University in Ankara, Turkey.