*Title Page (WITH AUTHOR DETAILS)

Negative Tweets and their Impact on Likelihood to Recommend

Jennifer B. Barhorst, College of Charleston

Alan Wilson, University of Strathclyde Business School

Joshua Brooks, Columbus State University

66 George Street

Charleston, SC 29424

Jennifer.barhorst@brandphd.com

Alan.wilson@strath.ac.uk

brooks_joshua1@columbusstate.edu

Declarations of interest: none

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Corresponding Author Information:

Jennifer B. Barhorst

Assistant Professor of Marketing

College of Charleston

66 George Street

Charleston, SC 29424

barhorstj@cofc.edu

+1 (864) 389-1280

*Manuscript (WITHOUT AUTHOR DETAILS)

Negative Tweets and their Impact on Likelihood to Recommend

1. Introduction

Customers' sharing of their negative experiences with brands within social media has received a great deal of media coverage in the last few years. One of the most memorable negative brand experiences in recent times was a video of passenger David Dao being forcibly removed from a United Airlines flight for refusing to give up his seat on an overbooked flight. The event, initially recorded and shared on Twitter, sparked global outrage with #UnitedAirlinesforcespassengeroffplane also becoming the top trending topic on Weibo, a Twitter alternative in China (Bowerman & Aulbach, 2017; Ohlheiser, 2017).

The shared experience of Dao with United Airlines not only put a spotlight on the increasing power that customers have through the use of microblogs such as Twitter and Weibo, but also raises questions about brand management in the digital domain. Specifically, does exposure to a negative brand experience on a microblog have any influence on receivers' likelihood to recommend the firm? Further, if a change in likelihood to recommend (LTR) does occur, what factors are particularly salient to a change in LTR? These are questions worth exploring as organizations now function in a world in which consumers increasingly share their everyday negative experiences about them just as they would tell a close friend or next-door neighbor in the past; only now, they tell the world about their experiences through the use of mobile technology (e.g., smart phones), the platforms that host social network sites (e.g., Twitter) and in a range of formats, including text, photographs, and videos.

To cope with the demands presented by this evolving digital landscape, firms such as American Airlines have dedicated social media teams working 24/7 to address complaints on microblogs (Comm, 2016), and use metrics such as LTR scores to consistently assess customers' overall perception of the brand (Matyszczyk, 2018). Further, a range of industries from airlines to financial services have invested in using LTR measures such as the Net Promoter Score to assess performance and perception of their brands (Bain & Company, 2019). Such investment gives one pause however as little is understood about the impact that negative brand experiences shared in microblogs have on the receivers of them. For example, research has increasingly expressed concerns about the risks posed to firms in the digital domain for the past several years (Berthon, Pitt, Plangger, & Shapiro, 2012; Cheng & Loi, 2014; De Maeyer, 2012; Grégoire, Salle, & Tripp, 2015; Rutter, Roper, & Lettice, 2016), yet research on the impact of negative valence microblog electronic word-of-mouth [MWOM] on an important actor in the digital domain—namely, receivers—is surprisingly lacking. Thus, this paper answers calls for academic research that is managerially relevant (Reibstein, Day, & Wind, 2009) and explores the potential risk that microblog users pose to firms and the factors that could influence a change in LTR upon exposure to negative MWOM. Doing so could help brand managers to better understand the potential risks to their brands in fast-moving environments such as Twitter, and any actions they can take to mitigate risks to their reputations.

This study therefore asks two primary research questions. First, does negative valence MWOM have an impact on receivers' likelihood to recommend the firm to others, and if so, to what extent? Second, which factors are particularly impactful to a change in LTR if changes occur? To address these questions, we operationalized quantitative research in the form of an online within-subject experiment in which 372 Twitter users took part. To determine whether a change in LTR occurred upon exposure to negative valence MWOM, we measured participants' perceptions of a range of U.S. airlines before and after exposure to negative tweets utilizing the practitioner based Net Promoter Score question. To confirm the factors that influence a change in

LTR on such exposure, we created three structural equation models (SEMs) using Smart-PLS to determine the comparative effects among three eWOM mediums: video, photo, and text.

2. Literature review and hypotheses

2.1. The Importance of the Likelihood to Recommend in the Digital Environment

As a result of today's digital environment, customer service operations have been transformed into 24/7 social listening operations (Comm, 2016) where brands continually monitor the risks and opportunities posed to their reputations as eWOM (electronic word of mouth) is shared about their brands. This environment has evolved as a result of the power that microblogs such as Twitter have brought to consumers and the subsequent transformation of the consumer-decision making process (Kumar, Choi, & Greene, 2017). Today, consumers increasingly rely on eWOM as a source of value to their decision-making process as reviews provided by consumers are seen to be more relevant than firm provided information (Allsop, Bassett, & Hoskins, 2007; Breazeale, 2009). From choosing an accommodation for their next holiday on TripAdvisor (Ganzaroli, De Noni, & van Baalen, 2017), to reading reviews on Twitter about new film releases (Hennig-Thurau, Wiertz, & Feldhaus, 2015), eWOM has widely been acknowledged as a powerful force. For example, one study determined that each ratings star added on a Yelp review of restaurants was equivalent to anywhere from 5 percent to 9 percent effect on revenues and that "Yelp is somewhat of a substitute for traditional forms of reputation" (Blanding, 2011, p.38). eWOM is thus a powerful force that not only helps to build brands through increased sales via MWOM (eWOM shared on microblogs) (Hennig-Thurau et al., 2015), it has also been widely acknowledged has having the potential to wreak havoc on brands' reputations (Berthon et al., 2012; Rutter et al., 2016). It is therefore no surprise that firms have developed robust systems to consistently monitor consumers' likelihood to recommend them at various stages of the consumption journey (Satmetrix, 2019) – including during check out at a register, via email, or pop-ups online. They often do so through the use of the Net Promoter Score question which asks a consumer *how likely are you to recommend Company X to a friend or colleague*? Use of the question has become so prominent amongst CEOs, it is said to have developed a cult-like following with NPS being cited more than 150 times in earnings conference calls by 50 S&P 500 companies in 2018 (Pacheco, 2019). Additionally it has evolved from a simple question to measure customer satisfaction, to an entire system used by companies to benchmark performance against other firms (Hyken, 2018; Marino-Nachison, 2018). Since its inception in 2003, it has not only been adopted by a range of firms, including Delta and Southwest Airlines (Bain & Company, 2019), but is also regularly used as an industry benchmark for brands' reputations. For example, firms with high NPS scores have been referenced as world-class companies (Hyken, 2018) and those which consumers consider the 'best of the best' (Griffith, 2018). Sametrix, in describing the use of the question, instructs practitioners to "use your NPS as the key measure of your customers' overall perception of your brand" by using a 0-10 scale to ask how likely is it that you recommend [brand] to a friend or colleague (Satmetrix, 2019)."

Although use of LTR metrics such as NPS have been widely adopted by firms, many questions remain. For example, it is not known whether a change in LTR occurs when receivers of negative valence MWOM are exposed in microblogs such a Twitter and we have little understanding of the magnitude of that change. Additionally, there is a dearth of research to understand whether there are any aspects of negative MWOM that may be more salient to a change in LTR if a change does occur. We therefore next explore microblogs as unique environments before moving to the literature in relation to the aspects of eWOM that make it most salient.

2.2. Microblogs as a unique environment

Microblogs are social media sites that enable users to send and read very short messages with a restriction on the number of characters that can be used in the message. Microblogging applications include Twitter, Weibo in China, and Me2day in South Korea. Many other social network sites have microblogging features called "status updates," including Facebook and LinkedIn (Stieglitz & Dang-Xuan, 2013). The most popular microblogging site in the United States is Twitter, which allows users to send tweets of up to 280 characters, also referred to as micro-sharing, micro-updating, "twittering," and "tweeting" (Hennig-Thurau et al., 2015; Jansen, Zhang, Sobel, & Chowdury, 2009). Consumers, journalists, celebrities, corporate executives, and political figures increasingly use Twitter to share stories, thoughts, and perceptions.

Twitter is a unique environment in which information is pushed to consumers, in contrast with other forms of eWOM such as review sites where it is pulled (Marchand, Hennig-Thurau, & Wiertz, 2017). Therefore, Twitter is inherently unique because of the limitation in the amount of information one can share and the potential to spread information quickly. Examples of the power of microblogging to spread information rapidly include events such as the Arab Spring and the 2012 U.S. presidential election (Hennig-Thurau et al., 2015). According to Rutter et al. (2016, p. 3097), microblogs such as Twitter represent "an honest and at times brutal feedback system, with offline [WOM] becoming online word of mouse, where brands engage with consumers and actively question, challenge and promote brands." Although such assertions are prominent in the literature, research on the impact of eWOM on recipients and their likelihood to recommend firms is scant. As such, the recent calls for more focus on microblogs in particular (Hennig-Thurau et al., 2015; Stieglitz & Dang-Xuan, 2013) are warranted.

2.3. Factors that affect eWOM effectiveness

Although recent studies continue to highlight the lack of research focused on MWOM (e.g., Hennig-Thurau et al., 2015; Marchand et al., 2017), researchers have provided a wealth of information and data on the factors that make eWOM effective.

2.3.1. Source style

Online reviews on social media can contain a combination of text and visual cues that exert an impact on users' online information acceptance (Teng, Khong, Goh, & Chong, 2014b). In relation to eWOM effectiveness, recent studies have demonstrated the impact of source style. For example, Lin, Lu, and Wu (2012) find that visual information such as photographs has a greater impact than text alone, and Wang, Cunningham, and Eastin (2015) show that benefit-centric reviews have a greater impact than attribute-centric reviews.

In the microblog Twitter, users can share their experiences with brands in multiple ways. Having increased the number of characters from 140 to 280 in November 2017, Twitter users can share their experiences of brands by simply tweeting a simple 280-character text message about their experiences or can choose to add a photograph or video to their tweet. With an aim of understanding the factors that influence a change in LTR from the receivers' perspective upon exposure to negative valence MWOM, this study includes the various types of tweets available. Formally stated:

H1. Negative MWOM that includes visual content will have a greater impact on likelihood to recommend than MWOM with text alone.

2.3.2. Information usefulness/involvement

The degree to which receivers are involved with a topic and whether they find the argument relevant has an impact on their opinions of the eWOM (Berger & Milkman, 2012; Cheng & Loi, 2014; Teng et al., 2014b). As Twitter users are exposed to a range of content in a fast-moving environment, the relevance of a tweet may influence whether a change in LTR occurs. Thus, we include a variable that represents these attributes and test the following hypothesis:

> **H2.** The extent to which receivers find negative MWOM relevant will negatively influence their likelihood of recommending the firm.

2.3.3. eWOM credibility

eWOM credibility, also referred to as message credibility (Cheung & Thadani, 2012) and argument quality/strength (Cheung, Luo, Sia, & Chen, 2009; Teng et al., 2014b), plays an important role in eWOM effectiveness. eWOM credibility captures how truthful or believable the recipient finds the eWOM to be (Cheung et al., 2009). Thus, if the receiver of MWOM finds the message believable and trustworthy, a change in LTR could occur depending on the argument being made. This is an especially salient assertion in an environment in which the proliferation of misinformation is making headlines. For example, M.I.T researchers Vosoughi, Roy, and Aral (2018) find that of 126,000 news items that were shared 4.5 million times by three million people, the news stories that were verified to be true rarely reached more than 1,000 Twitter users, while false news stories routinely reached more than 10,000 people. That study, and the ongoing discourse on misinformation, provides a sense of the challenges that receivers of tweets face when determining the credibility of a tweet; tweeted information can influence them, whether the information is true or not. Thus, we include a variable representing eWOM credibility in the model and hypothesize the following:

H3. The extent to which receivers find negative MWOM to be credible will negatively influence their likelihood to recommend the firm.

2.3.4. Emotions

The emotional aspect of eWOM is an area that has received scarce empirical research but is acknowledged to require further exploration (Berger & Milkman, 2012; Kim & Gupta, 2012; Standing, Holzweber, & Mattsson, 2016; Stieglitz & Dang-Xuan, 2013). Focusing on Twitter specifically, Stieglitz and Dang-Xuan (2013) analyze political data comprising 165,000 tweets and find that emotionally charged messages were retweeted more often and more quickly than neutral messages. In their case study of a Swedish firm and its communication campaigns, Standing et al. (2016) monitor comments after three campaigns on Kickstarter, Twitter, and Facebook and find that users expressing emotion were faster at spreading eWOM to their community.

Despite the limited research in the eWOM domain on emotions, the broader marketing literature provides some insight into the power of emotions to elicit actions. For example, advertisers have used the emotion *fear* to stimulate interest in products and services in advertising campaigns for life insurance, political parties, and public service announcements (LaTour & Zahra, 1988). In addition, the emotion *disgust* can enhance fear appeals and increase message persuasion (Morales, Wu, & Fitzsimons, 2012). The emotion *surprise* can be elicited through a divergence of perceptions and expectations, though a review of the marketing literature on surprise shows that research mainly focuses on the positive aspects of surprising consumers or the ability to attract them (e.g., Hutter & Hoffmann, 2011; Lindgreen & Vanhamme, 2003).

As Twitter users share their negative experiences with brands as they carry out their routine transactions, it is plausible that the emotions receivers feel about these shared brand experiences will play a role in changing LTR. We include variables representing these felt emotions and test the following hypothesis:

H4. The extent to which receivers experience emotions will influence a change in their likelihood to recommend the firm

2.3.5. Cognitive information processing

The way eWOM receivers process information can trigger attitude change (Cheung & Thadani, 2012; Teng et al., 2014b). The elaboration likelihood model of Petty and Cacioppo (1980) posits that the attitude formation process is dependent on the extent to which message receivers elaborate, or think about, issue-relevant arguments within a persuasive communication. When receivers of a persuasive communication have a relatively high motivation to process information, and that information is personally relevant (Petty & Cacioppo, 1986), they employ a central route with greater consideration of issue-relevant arguments. For example, research on blogs and cognitive information processing has demonstrated that blog members use the central route to persuasion when attending carefully to brand-related information on blogs by combining and integrating issue-relevant information into an overall evaluation reaction (Chu & Kamal, 2008). With regard to negative WOM specifically, its overall effect depends on the amount of issue-relevant information processing that it stimulates (Richins, 1984). In the microblog Twitter, users have a number of ways to consider information more closely when exposed to negative MWOM. For example, they can expand a tweet to read comments from others, or bookmark a tweet to return to it later. They can also retweet it to share with others who may then comment, thereby continuing to validate shared feelings. Conversely, the peripheral route to persuasion requires less thinking and consideration and includes cues such as emotions and source characteristics. In a fast-moving microblog environment such as Twitter, it is difficult to know how receivers of messages process MWOM. Therefore, we include both central route (giving an indication of a motivation to process more information) and peripheral route (giving an indication of less thinking and more heuristics involved with regard to information processing) variables in our model. In addition, we test the following hypotheses:

H5. The extent to which receivers find negative MWOM to be relevant will positively influence their likelihood of processing further issue-relevant information

H6. Upon exposure to negative MWOM, the likelihood of processing further issue-relevant information will negatively influence receivers' likelihood to recommend a firm

H7. Upon exposure to negative MWOM, variables representing both the central and peripheral routes to persuasion will influence receivers' likelihood to recommend a firm

2.3.6. Valence

Research indicates that the valence of WOM is an indicator of its effectiveness. For example, research demonstrates that both positive and negative WOM can influence consumers' decisions (Engel, Blackwell, & Kegerreis, 1969; Richins, 1983), where negative WOM is more helpful to receivers (Buttle, 1998). A stream of research also emphasizes the propensity of negative valence WOM and eWOM to more powerfully influence receivers than positive WOM (Buttle, 1998; Daugherty & Hoffman, 2014; Hennig-Thurau et al., 2015; Teng et al., 2014b). However, as this study is focusing on negative MWOM only, valence is not included in the model.

3. Conceptual model

With regard to MWOM and its potential influence on LTR, we created 11 different dependent variables for the model (see Table A.1 in the Appendix), using four categories identified in the literature: MWOM message relevance (H2, H5), MWOM message credibility

(H3), MWOM emotions (H4) and MWOM issue involvement (H6). For emotions, and consistent with pilot tests, we used a slightly modified version of Plutchik's (2001) eight basic emotions. These emotions included joy, sadness, anger, approval, disgust, fear, surprise, and not surprised. We examined the impact of source style (H1) by including three types of tweets available at the time of the study: a tweet with text only, a tweet with a photograph, and a tweet with a video. The conceptual model, depicted in Fig. 1, also provides an indication of the route to persuasion (H7), where "P" denotes the variables that are peripheral in nature (less information processing) and "C" denotes the variables that are central in nature (more information processing).

[INSERT FIGURE 1 HERE] with caption **Fig. 1. MWOM change in LTR conceptual model.**

4. Methodology

We undertook quantitative research in the form of an online experiment with a questionnaire through Qualtrics. Although recent studies have focused on examining MWOM and its effectiveness through aggregate data capture (Hennig-Thurau et al., 2015; Stieglitz & Dang-Xuan, 2013), research on receivers of MWOM and whether an individual instance of MWOM can influence their likelihood to recommend a firm is scarce. In addition, only limited research has examined receivers in particular and the factors related to MWOM that influence a change in LTR. As such, we operationalized a within-subject, pre-test/post-test experiment design because of the nature of the microblog environment and the way receivers are exposed to tweets about brands.

4.1. Experiment design

Research delineates the varying effects of environments (Marchand et al., 2017) on eWOM's effectiveness and the uniqueness of microblogs in the broader eWOM context (Hennig-Thurau et al., 2015). We thus designed the experiment not only to simulate what would happen in an actual Twitter environment but also to determine receivers' LTR both before and after exposure to negative valence MWOM. This form of experiment enabled us to glean useful insights into receivers of MWOM and MWOM's impact on LTR, which an aggregate data capture could not do.

Regarding the population of interest, the Pew Research Center (2018) indicates that the highest percentages of Twitter users (40%) are between the ages of 18 and 29 years, followed by 30–49 years (27%). In addition, the United States is one of the most prominent locations of Twitter users, with 57% having some college education. Thus, we identified college-educated Twitter users between the ages of 18 and 49 years, who resided in the United States, as the population of interest. We chose an access panel as the sampling frame. Use of the panel facilitated flexible pre-screening and enabled us to operationalize a study comprised of 372 male and female college-educated Twitter users between the ages of 18-24 years (62%), 25-34 years (31%), 35-44 years (6%) and 45-54 years (1%). Participants who took part in the experiment had a maximum of 25 minutes to complete the survey and were paid the equivalent of $9 an hour.

As the overarching objective of this study was to explore the microblog environment and the impact of negative valence MWOM on receivers' LTR, we determined that focusing on the U.S. airline industry in the experiment portion of the study would be pertinent. As a standard industry practice, the airline industry engages with customers via social media and specifically through Twitter in the United States (Sachs, 2014). For example, with approximately 2,500–2,600 Twitter mentions every day, JetBlue, a U.S.-based airline, employs a Twitter social media team of customer service, corporate communications, and marketing representatives to engage with customers and to manage the volume of tweets received each day (Kolowich, 2014).

4.2. Operationalization of the experiment

4.2.1. Step 1: Initial evaluation of the airlines

We took three steps to meet the objectives of the experiment and to simulate the Twitter environment. First, we asked participants to provide an evaluation of eight U.S.-based airlines by means of a Net Promoter Score (NPS) question. Participants rated how likely they would be to recommend these airlines to a friend or colleague. We used the ratings firm Skytrax to determine which airlines to include in the experiment. As we wanted to obtain a composite representation of study participants' LTR included in steps 2 and 3, we decided that airlines rated as 3 stars, or *average*, would be optimal to curtail the variable effects of prior knowledge about and familiarity and involvement with the airlines. Skytrax describes 3-star airlines as "airlines delivering a fair quality performance equating to an industry 'average' of acceptable product and service standards[1]." The rationale for this step was not only to capture participants' evaluations of airlines rated as average, but also to minimize the risk of recalling their answers later after exposure to the MWOM and being asked the NPS question again. The airlines in the initial question included American Airlines, Delta Airlines, United Airlines, US Airways, Hawaiian Airlines, Virgin America, JetBlue Airways, and Southwest Airlines.

4.2.2. Step 2: Exposure to negative MWOM

Second, all participants were exposed to three actual negative valence tweets (see Appendix for examples) for three airlines included in step 1 (American Airlines, US Airways and Delta), inclusive of the different types of tweets (i.e., tweet with text for American Airlines, a tweet with photograph for Delta Airlines, and tweet with video for US Airways). Inspired by Daugherty and Hoffman's (2014) use of Pinterest content in their study on consumer attention in

[1] https://skytraxratings.com/about-airline-rating

social media, we selected three actual tweets from Twitter to use in the experiment. We used Twitter's Advanced Search feature to find tweets that were about airlines, negative in valence, and in the United States. To confirm whether the tweets were considered negative in valence, we searched for tweets that had smiling or frowning face emoticons and emotional words that could be considered negative in valence (Kim & Gupta, 2012). Two pilot studies with 20 participants enabled us to further confirm the valence of the tweets, to make any necessary changes to the experimental design, and determine the average amount of time for participants to take part in the experiment. We reduced carry-over effects, or order effects, by randomizing the order of presentation of the types of tweets (Field & Hole, 2002).

4.2.3. Step 3: Questionnaire

Third, after exposure to each tweet, the participants answered a series of questions about the tweets, including a follow-up NPS question. We employed a range of scales to assess the factors that influenced a change in LTR.

4.3. Measures

We measured participants' perceptions of the airlines at two points: before exposure to the MWOM and after through the use of the NPS question. Although some researchers question the validity of the NPS to measure company profitability and brand loyalty (Grisaffe, 2007; Keiningham, Cooil, Andreassen, & Aksoy, 2007), we chose the NPS to represent LTR for two reasons. First, due its extensive use in the practitioner domain, and the fast-moving nature of environments such as Twitter where users are exposed to information at an unprecedented speed (Hennig-Thurau et al., 2015), we utilized the NPS question to represent LTR to obtain a snapshot inference of LTR both before and after participants were exposed to the tweets. As noted previously, use of the NPS system has been widely adopted by firms to continually evaluate their performance and brand perception. They often do so through the use of the Net Promoter Score question which asks a consumer *how likely are you to recommend Company X to a friend or colleague*? Individual customers give an answer between 0 (least likely to recommend) to 10 (most likely to recommend). These particular answers are then separated into three groups: 0-6 are termed detractors, 7-8 are denoted as passives, and 9-10 are called promoters. The overall score is calculated by taking the percentage of promoters minus the percentage of detractors. As our study focuses on individual experiences moving along this scale, we use the NPS scale to represent a particular individual's response to the NPS question and their likelihood to recommend the firm. Hence, we examined MWOM's impact on receivers' LTR by assessing whether there was any movement in the scale before and after exposure to the tweets utilized in this study. We did so by utilizing Paired-Samples T-Tests. After calculating whether there was any movement of the scale upon exposure to negative MWOM, our dependent variable in this study is the individual change in contribution to NPS score – i.e. we calculated NPS change by subtracting the NPS score provided after exposure to the MWOM from the NPS provided before exposure for each participant in our study. Secondly, other disciplines recognize the utility of one question in terms of brevity and simplicity and of making fewer demands on respondents; it is also reliable and valid (Bowling, 2005; Brinberg, 1995). In addition to participants' LTR, we measured other constructs that influence eWOM effectiveness. To keep the experiment as brief as possible, we employed a combination of single-item and modified multi-item scales to explore the factors that influenced any change in LTR. Table A.1 in the Appendix provides the constructs measured and scale and non-scale items.

4.4 Data analysis

To determine whether there was any impact on receivers' evaluation of LTR after exposure to the tweets, we ran a paired-samples t-test using SPSS. We selected this t-test as the measurement instrument because the same participants were asked the same NPS question on two occasions, before and after exposure to the negative tweets. Employing a paired-samples t-test enabled us not only to determine whether there was an actual change in LTR, but also to understand the extent of the change. It was also an applicable method to use because the data was normally distributed, independent, and interval based (Tabachnick & Fidell, 2013). For robustness, we also included the Wilcoxon signed-rank test. To provide further context to our initial analyses, we conducted a One-Way ANOVA to assess the impact of source style (tweet with photo, tweet with video, or tweet with text only) on NPS scores.

To determine the factors that influenced a change in LTR after exposure to the tweets, we employed partial least squares structural equation modeling (PLS-SEM). According to Hoyle (1995, p. 1), SEM "is a comprehensive statistical approach to testing hypotheses about relations among observed and latent variables." This method has also become "quasi-standard in marketing and management research when it comes to analyzing the cause-effect relations between latent constructs" (Hair, Ringle, & Sarstedt, 2011, p. 139). PLS-SEM was suitable for our study because the theoretical model includes a mix of reflective and formative indicators (Lowry & Gaskin, 2014) and the models being tested are exploratory in nature (Hair, Hult, Ringle, Sarstedt, & Thiele, 2017; Lowry & Gaskin, 2014; Sarstedt, Hair, Ringle, Thiele, & Gudergan, 2016).

5. Results

5.1 Paired-samples t-test and ANOVA results

To understand which variables influence a change in LTR upon exposure to negative valence MWOM, it was first necessary to understand whether a change in the perception of the airline had actually occurred and, if so, to what extent. As Table 1 shows, the significance value of the t-tests was .000 ($p < .01$), indicating a significant difference between the NPS score given by survey participants both before and after exposure to the negative valence MWOM.

[INSERT TABLE 1 HERE]

Panel A of Table 1 shows the mean NPS values for all three negative valence tweets decreased after exposure to the MWOM:

- **Tweet with video:** decrease from an NPS1 mean of 4.81 to an NPS2 of 3.39
- **Tweet with photo:** decrease from an NPS1 mean of 5.30 to an NPS2 mean of 3.73
- **Tweet with text only:** decrease from an NPS1 mean of 5.17 to an NPS2 of 3.69

The results in Panel B of Table 1 show the extent of change for the NPS scores. Correspondingly, the Wilcoxon test shows that the changes are highly statistically significant. Based on Cohen's criteria (1988), all effects were large with the following results. For the tweet with a photo, the mean difference between the scores was 1.57, with a 95% confidence interval from a lower bound of 1.35 to an upper bound of 1.79. The effect size was $0.36 – 14.29^2/[14.29^2 + (372 – 1)]$. For the tweet with a video, the mean difference between the scores was 1.42, with a 95% confidence interval from a lower bound of 1.18 to an upper bound of 1.66. The effect size was $0.27 – 11.719^2 /[11.719^2 + (372 – 1)]$. For the tweet with text only, the mean difference between the scores was 1.48, with a 95% confidence interval from a lower bound of 1.25 to an upper bound of 1.71. The effect size was $0.30 – (12.62^2 /[12.62^2 + (372 – 1)]$. We then conducted a One-Way ANOVA to understand whether the type of tweet was significant. Panel C of Table 1 demonstrates that when ignoring other factors present in our model, the changes in NPS across different types of negative tweets are similar in magnitude. We therefore fail to reject the null hypothesis of equal means - indicating that without controlling for additional variables, the medium of tweets does not seem to have an impact on their ability to affect NPS. Hence, these results indicate that H1 is not supported. Instead of collapsing our data into a single model, we maintain separate models for each tweet medium in anticipation that the key drivers of these changes may also vary by type of tweet.

5.2. PLS-SEM results

We created three PLS-SEM models to determine the "change in LTR" predictors for the negative valence tweets. In each of these models, unless otherwise noted, the dependent variable was the change in NPS scale. We tested 11 different independent variables (see Table A.1 in the Appendix) in the model. Summary statistics for each model's variables are shown in Tables A.2 A.3 and A.4 of the Appendix; correlations are available in the supplementary materials for this article.

5.2.1. Validation of the measurement models

To assess the validity of the measurement models, we used the methods Wong (2013) and Hair, Hult, Ringle, and Sarstedt (2016) describe. Discriminant validity was established when the factor loading coefficients for the items that constituted each latent variable were greater than their cross-loadings on alternative latent variables. We assessed the cross-loadings for the models, and all three models fit the criteria.

Convergent validity was established when the average variance extracted (AVE) by the multiple indicators of each latent variable was greater than 0.5. Internal consistency and reliability were established when the composite reliability coefficient was greater than 0.6. Table A.5 in the Appendix illustrates that convergent validity was established as the AVE by the multiple indicators of each latent variable was greater than 0.5. Table A.5 also shows that internal consistency and reliability were established, as all the composite reliability coefficients for the latent variables were greater than 0.6.

5.2.2. Evaluation of the structural model

We estimated the statistical significance of each path coefficient (β) through bootstrapping. We randomly sampled the raw data 5,000 times and computed the mean of each β coefficient. To avoid overfitting our model, we use the Akaike information criteria (AIC) and the Bayesian information criteria (BIC). These values are calculated using finite mixture segmentation (using 2 segments, 5,000 iterations, and 10 repetitions). We use a general-to-specific methodology to arrive at a more parsimonious model. We estimated each model using the bootstrapping procedure and then eliminated the least significant variable. We compared each model's AIC and BIC until a minimum was reached. The results obtained using AIC and BIC to remove extraneous variables closely modeled those indicated by Cronbach's alphas and the composite reliability scores (those not shown here are available on request).

5.2.3. MWOM SEM model results

Table 2 displays the significant predictors and hypotheses results of a change in LTR of the three PLS-SEM structural models based on the data for the three negative valence tweets. As the table shows, *MWOM issue involvement (–), fear (–),* and *surprise (–)* were predictors in the model with the photo-only tweet, thus H4 and H6 are partially supported. Similarly, *MWOM issue involvement (–) and the emotion disgust (–)* were predictors in the tweet with video model, again partially supporting H4 and H6. *MWOM message relevance (–),* and the emotions and *joy*

(+) and *surprise (–)* were predictors in the tweet with text-only model, thus H2 and H4 are partially supported. *MWOM message relevance (-)* was a predictor of MWOM issue involvement in all three models, thus H5 is supported. Additionally, we note that the relative size of the coefficient on *MWOM message relevance* was highest for text compared to the other mediums. All three models had a combination of central and peripheral variables as predictors, thus H7 is supported. Finally, the effect size (R^2) indicated the proportion of the variance explained in the dependent variable by the predictor variables with the photo at 0.150, the video at 0.240 and the text only at 0.246.

[INSERT TABLE 2 HERE]

In regressions (not shown), we test a priori whether demographics have any impact on NPS before, NPS after, and the change in NPS. According to the p-values, participants' age, education level, and sex are statistically meaningless. Daily use is statistically significant, but only in the initial NPS. To allow for the possibility that these demographic factors are statistically relevant, we add them to our SEM-PLS model to further evaluate the robustness of the findings.

As demonstrated in Table 3, the demographics were found to have little value with a couple of exceptions. Education seems to significantly reduce the effects of negative videos. Daily use increases the impact of negative photographic and text information. Correspondingly, these demographic coefficients are found to have a similar order of magnitude to our previous variables of interest. In order to check for overfitting, we apply both the Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC) to our model. We iteratively fit our model and remove the least significant variable at each step to arrive at the best AIC and BIC models. These additional criteria also serve to reinforce the robustness of our original findings.

[INSERT TABLE 3 HERE]

6. Discussion

In this study, we explored the impact that negative microblog word of mouth (MWOM) could have on receivers' likelihood to recommend (LTR) firms by using an industry-standard metric, the NPS question to do so. The results suggest that negative valence tweets have an impact on LTR, and they do so with a statistically large effect based on Cohen's (1988) criteria. The results for microblogging, as indicated in this study, are aligned with previous assertions and studies on the impact of negative eWOM on brands. The study also sought to establish which variables predict a change in LTR after receivers are exposed to negative valence tweets. The findings suggest that the relevance of a negative valence tweet, the emotions receivers feel after exposure to the tweet, and issue involvement all play a key role in whether a change in LTR occurs after exposure. Additionally, our study suggests that the relevance of a tweet is a strong predictor of issue involvement across all types of tweets, signifying its importance when Twitter users are exposed to negative valence MWOM. These findings support extant literature and add to the burgeoning research on microblogs specifically.

The findings also underscore the impact of source style and the routes to persuasion that receivers employ. In the negative valence models with complex visual cues (e.g., a photo, a video), significant predictors included the peripheral route to persuasion variables (*fear*, *disgust*, and *surprise*) and a central route to persuasion variable (*MWOM issue involvement*). This finding suggests that tweets with complex visual cues motivate receivers to process issue-relevant information about a negative brand experience portrayed in a tweet through the elicitation of high arousal emotions. Both *fear* and *disgust* elicit avoidance tendencies (Morales et al., 2012), while surprise is evoked from a divergence of perceptions and expectations (Hutter & Hoffmann, 2011). In the negative valence model with text only, and thus an absence of complex visual cues, emphasis was placed on how receivers felt after exposure to the tweet (*joy* and *surprise*) and the relevance of the tweet (*MWOM message relevance*). This finding again highlights the key role of source style and suggests that the absence of complex visual cues does not motivate receivers to become involved in the message by processing more information about the tweet. Rather, how a tweet makes receivers feel, and how relevant it is to them personally, influence a change in LTR. This finding is consistent with prior research in the marketing discipline and adds additional context to the microblog environment. For example, emotions generate various levels of psychological arousal or activation. Anger and anxiety are associated with states of heightened arousal or activation, whilst contentment and sadness are associated with low arousal, or deactivation. Further, low arousal is characterized by relaxation and high arousal with activity (Berger & Milkman, 2012). In the two negative valence models with complex visual cues (i.e. the video and photo) the high arousal emotions of fear (-), surprise (-) and disgust (-) were significant. A high arousal and a low arousal emotion were significant predictors in the text only model however – surprise (-) and joy (+) respectively. It can therefore be suggested that the significance of only high arousal emotions generated issue involvement and action to process more information in the tweets with complex visual cues (the photo and video), whereas the addition of a low arousal emotion, joy (+), evoked by the text-only tweet garnered a feeling of personal relevance rather than action. The type of media associated with the tweet however, or source style, does not affect receivers' LTR. Whether exposed to a tweet with a video, or tweet with text only, all tweets had a large effect on a change in LTR, and we found a marginal difference among all three types of tweets.

With regard to emotions, there were a couple of unexpected findings. *Joy* as a *positive* predictor of LTR and *surprise* as a negative predictor in the text-only model was unanticipated.

This outcome could potentially be explained by research that suggests that consumers experience *schadenfreude*, or pleasure at others' misfortunes (Heider, 1958), when they feel their choice of brand and subsequent misfortune are deserved (Yucel-Aybat & Kramer, 2017). The text-only tweet may have induced schadenfreude as the MWOM shared in the tweet concerns the sender thinking that the airline had made some changes (see Figure A.1 in the Appendix) and therefore a sense of *joy* positively influenced a change in LTR. Further, and as noted previously, research suggests that *surprise* is a result of a divergence of expectations (Hutter & Hoffmann, 2011). The lack of change by the airline may have been expected and thus *surprise* negatively influenced a change in LTR.

The lack of significance of *MWOM message credibility* is counter to the findings in extant literature. Although it might be plausible that the credibility of a tweet would influence a change in LTR, as the literature stresses the importance of credibility with regard to eWOM effectiveness (Cheung & Thadani, 2012; Cheung et al., 2009; Lin et al., 2012; Teng, Khong, & Goh, 2014a; Teng et al., 2014b), our findings suggest that the credibility of the message, or how believable and trustworthy the recipient found the message to be, was not an influencing factor in a change in LTR. Considering that Twitter users have a propensity to share false information more than true information (Vosoughi et al., 2018), it is clear that there are challenges in the Twitter environment related to message credibility. This finding therefore further adds to the growing body of literature on eWOM in microblogs and the lack of significance of MWOM credibility.

6.1. Theoretical contributions

A key contribution of this study to the literature is the provision of empirical data that confirm the change in LTR when receivers are exposed to negative valence MWOM. Although research highlights the risk to firms as a result of eWOM (Berthon et al., 2012; Cheng & Loi, 2014; De Maeyer, 2012; Grégoire et al., 2015; Rutter et al., 2016), empirical research on MWOM from the receiver's perspective and their impact on LTR is scarce. Furthermore, this study contributes by providing initial exploratory evidence of the factors that influence a change in LTR upon exposure to negative valence MWOM. Finally, this study shows that source style, or the type of tweet that receivers are exposed to, plays a key role in determining which factors influence a change in LTR.

6.2. Managerial implications

From a managerial perspective, the results of this study have several useful implications. First, managers must be *diligent*, as a change in LTR does indeed occur when receivers are exposed to negative valence MWOM. As such, managers should try to ensure that the customer experience meets expectations and work to obtain feedback during the service encounter to reduce the likelihood of customers issuing negative tweets. Second, managers must be *vigilant*, as we show that complex visual cues (photographs and videos) bring about action-eliciting emotions and the motivation to process more issue-relevant information about the tweets. These findings emphasize the importance of managers using social listening platforms to track sentiment and engagement with tweets. Doing so could signal risks to their reputations and provide the opportunity to respond to negative valence tweets in order to provide information to receivers as they search for more details. Finally, it is important for managers to understand that the credibility of MWOM does not play a key role in receivers' LTR. Again, this finding underscores the importance of using a brand's voice after negative brand experiences have been shared, as the credibility of a tweet (whether receivers believe it or not) does not play a key role in predicting a change in LTR.

6.3. Limitations

Limitations associated with this study may pave the way for future research. Three limitations to our research concern the research methods utilized. First, although we examined a range of airlines rated as 3-star, or average, by the firm Skytrax at the time of the study, we did not include all airlines rated as 3-stars in the study. Future studies could conduct research that includes all airlines within a particular rating to conduct further comparative analyses. Secondly, due to the nature of microblogs and the literature's reference to the potential damage to brands that negative mentions can have, we examined negative valence MWOM in our study. Further studies could include positive MWOM in order to discern the influence this type of MWOM could have on receivers' LTR. Thirdly, due to the exploratory nature of our study and our emphasis on examining the impact of various types of MWOM on receivers in a simulated Twitter environment, we did not include a control group in our experiment.

Another limitation is our focus on the airline industry. We chose this industry because it was at the epicenter of microblog members' tweets and also offered a wealth of data. Although the use of this industry was a practical one for a sound execution of the study, research could undertake a similar study with another industry or even a range of industries to determine whether similar outcomes would occur. A further limitation is the location of the experiment (the United States). Given that social network site users can share their experiences of brands with anyone around the world, research should undertake a similar analysis with receivers in other countries. Finally, the study focused on one social media platform, Twitter. It would be worthwhile for research to undertake a similar study using other social media platforms to determine whether our results hold.

7. References

Allsop, D. T., Bassett, B. R., & Hoskins, J. A. (2007). Word-of-mouth research: Principles and applications. *Journal of Advertising Research, 47*(4), 398-411.

Bain & Company. (2019). How and why did you develop the net promoter score? Retrieved from http://www.netpromotersystem.com/about/why-net-promoter.aspx

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research, 49*(2), 192-205.

Berthon, P. R., Pitt, L. F., Plangger, K., & Shapiro, D. (2012). Marketing meets web 2.0, social media, and creative consumers: Implications for international marketing strategy. *Business Horizons, 55*(3), 261-271.

Blanding, M. (2011). The yelp factor: Are consumer reviews good for business? Retrieved from http://hbswk.hbs.edu/item/the-yelp-factor-are-consumer-reviews-good-for-business

Bowerman, M., & Aulbach, L. (2017). *United airlines under fire after man is dragged off overbooked flight.* Retrieved from https://www.usatoday.com/story/travel/nation-now/2017/04/10/united-under-fire-after-man-dragged-off-overbooked-flight/100287740/.

Bowling, A. (2005). Just one question: If one question works, why ask several? *Journal of Epidemiology and Community Health, 59*(5), 342-345.

Breazeale, M. (2009). Word of mouse an assessment of electronic word-of-mouth research. *International Journal of Market Research, 51*(3), 297-318.

Brinberg, D. (1995). The multiples of science. Presidential address, Society for Consumer Psychology Winter Conference.

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

Cheng, V. T., & Loi, M. K. (2014). Handling negative online customer reviews: The effects of elaboration likelihood model and distributive justice. *Journal of Travel & Tourism Marketing, 31*(1), 1-15.

Cheung, C. M., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems, 54*(1), 461-470.

Cheung, M. Y., Luo, C., Sia, C. L., & Chen, H. (2009). Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. *International Journal of Electronic Commerce, 13*(4), 9-38.

Chu, S., & Kamal, S. (2008). The effect of perceived blogger credibility and argument quality on message elaboration and brand attitudes: An exploratory study. *Journal of Interactive Advertising, 8*(2), 26-37.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2d ed.). Hillsdale, NJ: Erlbaum.

Comm, J. (2016). *How American Airlines won twitter.* Retrieved from https://www.inc.com/joel-comm/how-american-airlines-won-twitter.html.

Daugherty, T., & Hoffman, E. (2014). eWOM and the importance of capturing consumer attention within social media. *Journal of Marketing Communications, 20*(1/2), 82-102.

De Maeyer, P. (2012). Impact of online consumer reviews on sales and price strategies: A review and directions for future research. *Journal of Product & Brand Management, 21*(2), 132-139.

Engel, J. F., Blackwell, R. D., & Kegerreis, R. J. (1969). How information is used to adopt an innovation. *Journal of Advertising Research, 9*(4), 3-8.

Field, A., & Hole, G. (2002). *How to design and report experiments*. London: Sage.

Ganzaroli, A., De Noni, I., & van Baalen, P. (2017). Vicious advice: Analyzing the impact of TripAdvisor on the quality of restaurants as part of the cultural heritage of Venice. Tourism Management, 61(C), 501–510.

Gottschalk, S. A., & Mafael, A. (2017). Cutting through the online review Jungle—Investigating selective eWOM processing. *Journal of Interactive Marketing, 37*, 89-104.

Grégoire, Y., Salle, A., & Tripp, T. M. (2015). Managing social media crises with your customers: The good, the bad, and the ugly. *Business Horizons, 58*(2), 173-182.

Griffith, E. (2018). Consumer recommended 2018: The tech brands you love most. Retrieved from https://www.pcmag.com/news/365392/consumer-recommended-2018-the-tech-brands-you-love-most

Grisaffe, D. B. (2007). Questions about the ultimate question: Conceptual considerations in evaluating Reichheld's net promoter score (NPS). *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior, 20*, 36-53.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., & Thiele, K. O. (2017). Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods. *Journal of the Academy of Marketing Science, 45*(5), 616-632.

Hair, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). A primer on partial least squares structural equation modeling (PLS-SEM) Sage Publications.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice, 19*(2), 139-152.

Heider, F. (1958). *The psychology of interpersonal relations*. NewYork:Wiley.

Hennig-Thurau, T., Wiertz, C., & Feldhaus, F. (2015). Does twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies. *Journal of the Academy of Marketing Science, 43*(3), 375-394.

Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage.

Hutter, K., & Hoffmann, S. (2011). Guerrilla marketing: The nature of the concept and propositions for further research. *Asian Journal of Marketing, 5*(2), 39-54.

Hyken, S. (2018). The best and worst companies to do business with. Retrieved from https://www.forbes.com/sites/shephyken/2018/07/12/the-best-and-worst-companies-to-do-business-with/#5cef04d03a48

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Micro-blogging as online word of mouth branding. In *Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 3859-3864). New York: ACM.

Keiningham, T. L., Cooil, B., Andreassen, T. W., & Aksoy, L. (2007). A longitudinal examination of net promoter and firm revenue growth. *Journal of Marketing, 71*(3), 39-51.

Kim, J., & Gupta, P. (2012). Emotional expressions in online user reviews: How they influence consumers' product evaluations. *Journal of Business Research, 65*(7), 985-992.

Kolowich, L. (2014). *Delighting people in 140 characters: An inside look at jet blue's customer service success.* Retrieved from http://blog.hubspot.com/marketing/jetblue-customer-service-twitter#sm.0001cwqfdky9kf05115v60e3zsu9l.

Kumar, V., Choi, J. B., & Greene, M. (2017). Synergistic effects of social media and traditional marketing on brand sales: Capturing the time-varying effects. Journal of the Academy of Marketing Science, 45(2), 268–288. https://doi.org/10.1007/s11747-016-0484-7

LaTour, M. S., & Zahra, S. A. (1988). Fear appeals as advertising strategy: Should they be used? *Journal of Services Marketing*, *2*(4), 5-14.

Laczniak, R. N., & Muehling, D. D. (1993). The relationship between experimental manipulations and tests of theory in an advertising message involvement context. Journal of

Advertising, 22(3), 59-74.Lin, T. M., Lu, K., & Wu, J. (2012). The effects of visual information in eWOM communication. *Journal of Research in Interactive Marketing, 6*(1), 7-26.

Lin, Tom M., Kuan-Yi Lu, and Jia-Jhou Wu, (2012), "The Effects of Visual Information in eWOM Communication," Journal of Research in Interactive Marketing, 6(1), 7-26.

Lindgreen, A., & Vanhamme, J. (2003). To surprise or not to surprise your customers: The use of surprise as a marketing tool. *Journal of Customer Behaviour, 2*(2), 219-242.

Lowry, P. B., & Gaskin, J. (2014). Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: When to choose it and how to use it. *IEEE Transactions on Professional Communication, 57*(2), 123-146.

Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing, 34*(2), 336-354.

Marino-Nachison, D. (2018). Nike: Have Investors Left Their 'Skepticism' Behind? - Barron's. Retrieved from: https://www.barrons.com/articles/nike-have-investors-left-their-skepticism-behind-1529590462

Matyszczyk, C. (2018). American airlines admits it has no brand purpose. Retrieved from https://www.inc.com/chris-matyszczyk/american-airlines-desperately-wants-passengers-to-like-it-more-why-should-they.html

Morales, A. C., Wu, E. C., & Fitzsimons, G. J. (2012). How disgust enhances the effectiveness of fear appeals. *Journal of Marketing Research, 49*(3), 383-393.

Ohlheiser, A. (2017). *The full timeline of how social media turned united into the biggest story in the country.* Retrieved from https://www.washingtonpost.com/news/the-intersect/wp/2017/04/11/the-full-timeline-of-how-social-media-turned-united-into-the-biggest-story-in-the-country/?noredirect=on&utm_term=.399e042536d2.

Pacheco, K. S. and I. (2019). The Dubious Management Fad Sweeping Corporate America. Wall Street Journal. Retrieved from https://www.wsj.com/articles/the-dubious-management-fad-sweeping-corporate-america-11557932084

Park, C., Wang, Y., Yao, Y., & Kang, Y. R. (2011). Factors influencing eWOM effects: Using experience, credibility, and susceptibility. *International Journal of Social Science and Humanity, 1*(1), 74-79.

Petty, R. E., & Cacioppo, J. T. (1980). Effects of issue involvement on attitudes in an advertising context. *Proceedings of the Division, 23*, 75-79.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. Communication and persuasion (pp. 1-24) Springer.

Pew Research Center. (2018). Social media fact sheet. Retrieved from http://www.pewinternet.org/fact-sheet/social-media/

Plutchik, R. (2001). The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist, 89*(4), 344-350.

Reibstein, D. J., Day, G., & Wind, J. (2009). Guest Editorial: Is Marketing Academia Losing Its Way? Journal of Marketing, 73(4), 1–3.

Richins, M. L. (1983). Negative word-of-mouth by dissatisfied consumers: A pilot study. *The Journal of Marketing, 47*(1), 68-78.

Richins, M. L. (1984). Word of mouth communication as negative information. *ACR North American Advances, 11, 697-702.*

Rutter, R., Roper, S., & Lettice, F. (2016). Social media interaction, the university brand and recruitment performance. *Journal of Business Research, 69*(8), 3096-3104.

Sachs, A. (2014). *Now boarding passengers on air social media.* Retrieved May 05, 2015, from https://www.washingtonpost.com/lifestyle/travel/now-boarding-passengers-on-air-social-media/2014/04/03/a755440e-b390-11e3-b899-20667de76985_story.html.

Satmetrix. (2019). What is net promoter? Retrieved from https://www.netpromoter.com/know/

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research, 69*(10), 3998-4010.

Skytrax (2018). Retrieved from http://www.airlinequality.com/

Standing, C., Holzweber, M., & Mattsson, J. (2016). Exploring emotional expressions in e-word-of-mouth from online communities. *Information Processing & Management, 52*(5), 721-732.

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems, 29*(4), 217-248.

Tabachnick, B., & Fidell, L. (2013). Using multivariate statistics (6th international ed.). London: Pearson.

Teng, S., Khong, K. W., & Goh, W. W. (2014a). Conceptualizing persuasive messages using ELM in social media. *Journal of Internet Commerce, 13*(1), 65-87.

Teng, S., Khong, K. W., Goh, W. W., & Chong, A. Y. L. (2014b). Examining the antecedents of persuasive eWOM messages in social media. *Online Information Review, 38*(6), 746-768.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146-1151.

Wang, S., Cunningham, N. R., & Eastin, M. S. (2015). The impact of eWOM message characteristics on the perceived effectiveness of online consumer reviews. *Journal of Interactive Advertising, 15*(2), 151-159.

Wong, K. K. (2013). Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS. *Marketing Bulletin, 24*(1), 1-32.

Yucel-Aybat, O., & Kramer, T. (2017). Comparative advertisements and schadenfreude: when and why others' unfortunate choices make us happy. Marketing Letters, 28(4), 579-589.

8. Appendix

[INSERT TABLEs A.1-A.5 HERE]

[INSERT FIGURE A.1 HERE] with caption **Fig. A.1. Negative MWOM with Text Only**

[INSERT FIGURE A.2 HERE] with caption **Fig. A.2. Negative MWOM with Photograph**

[INSERT VIDEO A.1 HERE] with caption **Vid. A.1. Negative MWOM with Video**

Figure

Figure A.1 – Negative MWOM with Text Only

@AmericanAir poor customer service exper., unfortunately. Agents weren't apologetic at all. :(I thought the airline had made some changes.

Figure A.2 – Negative MWOM with Photograph

Waiting over an hour and a half for my friend's luggage on @**Delta** for a flight delayed over 5 hours >:(@DeltaAssist

Table

Table 1. Change in NPS. Panel A shows summary statistics of our key variables. Correlations are the correlation between the initial and final NPS. The significance of the correlations are also shown. Panel B presents paired sample t-tests of the NPS along with the Wilcoxon test. Panel C demonstrates the one-way ANOVA[1] results for the individual changes in NPS between each medium.

A. Summary						
	Video		**Photo**		**Text**	
	Initial NPS	Final NPS	Initial NPS	Final NPS	Initial NPS	Final NPS
Mean	4.810	3.390	5.300	3.730	5.170	3.690
SD	2.088	2.138	2.424	2.182	2.262	1.896
Correlation	0.581		0.389		0.418	
Significance	0.000		0.000		0.000	
B. Before and After Differences						
	Video		**Photo**		**Text**	
Pairs T-test	1.658		1.786		1.712	
p-value	0.000		0.000		0.000	
Wilcoxon	5202		11464.5		8255.5	
p-value	0.000		0.000		0.000	
C. ANOVA on Change in NPS for Tweet Type						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-Value*	*F Crit*
Between Groups	4.260	2	2.130	0.424	0.655	3.004
Within Groups	5590.632	1113	5.023			
Total	5594.892	1115				

[1] The ANOVA, or analysis of variance, tests for whether the variation of observations is driven by the differences between groups or the variation within particular groups. This test enables us to see if the type of tweet has any explanatory power in the size of the observed differences in individual NPS.

Table 2. Negative PLS-SEM Model Results. MWOM is delineated by type of tweet. Coefficients and t-statistics are found by bootstrapping 5,000 times.

		Photo			Video			Text-only		
		Coeff.	**P-Val.**	**Hypothesis**	**Coeff.**	**P-Val.**	**Hypothesis**	**Coeff.**	**P-Val.**	**Hypothesis**
	MWOM Constructs									
H2	Message relevance	-0.015	0.701	Not supported	-0.083	0.221	Not supported	-0.217	0.008	Supported
H3	Message credibility	-0.086	0.095	Not supported	-0.013	0.760	Not supported	0.036	0.432	Not supported
H6	Issue involvement	-0.133	0.020	Supported	-0.162	0.011	Supported	-0.051	0.263	Not supported
H5	MI → II	0.691	0.000	Supported	0.561	0.000	Supported	0.663	0.000	Supported
	Emotions									
H4	Anger	-0.049	0.392	Not supported	-0.099	0.130	Not supported	-0.085	0.212	Not supported
H4	Approval	0.096	0.098	Not supported	0.095	0.081	Not supported	0.089	0.154	Not supported
H4	Disgust	0.005	0.905	Not supported	-0.152	0.038	Supported	-0.101	0.141	Not supported
H4	Fear	-0.166	0.030	Supported	0.034	0.402	Not supported	-0.058	0.401	Not supported
H4	Joy	0.032	0.443	Not supported	0.035	0.464	Not supported	0.182	0.018	Supported
H4	Not surprised	0.090	0.117	Not supported	0.054	0.267	Not supported	-0.022	0.558	Not supported
H4	Sadness	-0.058	0.238	Not supported	-0.094	0.124	Not supported	-0.117	0.062	Not supported
H4	Surprise	-0.132	0.040	Supported	-0.074	0.169	Not supported	-0.188	0.002	Supported
	R^2	0.150			0.240			0.246		

Table 3. SEM-PLS model results with demographics. MWOM is delineated by type of tweet. Coefficients and t-statistics are found by bootstrapping 5,000 times.

		Video				Photo		Text	
		AIC-Sel.		BIC-Sel.		AIC/BIC-Sel.		AIC/BIC-Sel.	
		Coeff.	P-Values	Coeff.	P-Values	Coeff.	P-Values	Coeff.	P-Values
MWOM Constructs									
Issue involvement	H6	-0.157	0.019	-0.157	0.017	0.041	0.358	-0.042	0.321
Message credibility	H3	0.01	0.792	0.010	0.801	-0.087	0.088	0.043	0.37
Message relevance	H2	-0.108	0.148	-0.108	0.155	-0.002	0.971	-0.216	0.01
MI -> II	H5	0.691	0.000	0.691	0.000	0.561	0.000	0.663	0.000
Emotions									
Anger	H4	-0.108	0.115	-0.107	0.115	-0.044	0.446	-0.085	0.215
Approval	H4	0.095	0.084	0.095	0.085	0.114	0.067	0.089	0.158
Disgust	H4	-0.15	0.044	-0.151	0.039	0.006	0.899	-0.094	0.157
Fear	H4	0.036	0.371	0.035	0.380	-0.177	0.021	-0.066	0.338
Joy	H4	0.049	0.325	0.049	0.328	-0.111	0.07	0.185	0.015
Not surprised	H4	0.053	0.268	0.053	0.270	0.083	0.142	-0.033	0.407
Sadness	H4	-0.104	0.111	-0.104	0.101	-0.062	0.218	-0.124	0.06
Surprise	H4	-0.076	0.167	-0.075	0.162	-0.133	0.036	-0.185	0.004
Demographics									
Age		-0.004	0.887			0.061	0.183	0.029	0.441
Daily use		-0.036	0.283	-0.036	0.289	-0.079	0.079	-0.083	0.056
Education		0.121	0.020	0.119	0.012	-0.031	0.440	0.034	0.418
Sex		-0.044	0.217	-0.044	0.214	0.005	0.872	0.037	0.299
R^2 for ΔNPS		0.266		0.257		0.163		0.267	
R^2 for II		0.676		0.478		0.528		0.624	

Table A.1 – Constructs and scale items. Table A.1 demonstrates the constructs, routes to persuasion, single and multi-item scale items, and the supporting references.

Construct Measured	**Route to Persuasion**	**Scale Items**	**Major References**
MWOM Message Credibility	Peripheral (Petty & Cacioppo, 1980)	Thinking about this Tweet, to what extent do you agree or disagree with the following statements? • The tweet is believable. • The tweet is informative. • The tweet is trustworthy. • The tweet is reliable.	Park, Wang, Yao, & Kang, 2011; Cheung et al., 2009
MWOM Message Relevance	Central (Richens, 1984; Petty & Cacioppo, 1980)	Thinking about this tweet, to what extent do you agree or disagree with the following statements? • The tweet is useful. • The tweet is interesting. • The tweet is worth remembering. • I liked the tweet.	Laczniak & Muehling, 1993
MWOM Issue Involvement	Central (Richens, 1984; Petty & Cacioppo, 1980)	After seeing this tweet, how likely are you to do the following? • Retweet it • Expand to read (airline)	Petty, & Cacioppo, 1980

		response • Expand to read comments of others • Favorite it	
MWOM Emotions	Peripheral (Petty & Cacioppo, 1980)	Reflecting on this tweet, to what extent do you feel the following emotions? • Joy • Sadness • Surprise • Not surprised • Anger • Fear • Disgust • Approval	Plutchik, 2001

Table A.2 – Summary statistics by model. The table below shows summary statistics for each variable used to build our constructs separated by medium. NPS ranges from 0 (least likely) to 10 (most likely to recommend). The other Likert variables ranged from 0 to 5. Scale items and values for age include: 18-24 years (2), 25-34 (3), 35-44 years (4) and 45-54 years (5). Scale items and values for Sex include: female (1) and male (2). Education is represented by the following: High School/GED (2), some college (3), 2-year college degree (4), 4-year college degree (5), master's degree (6), doctoral degree (7), and (8) professional degree (JD, MD). Daily use is divided into the following categories: (2) for 30 minutes or less, (3) for 30-60 minutes, (4) for 60-90 minutes, and (5) for more than 90 minutes per day.

A. Video model summary statistics.								
		Mean	*SD*	*Min*	*1Q*	*Median*	*3Q*	*Max*
NPS	Final NPS	3.387	2.138	0	2	4	5	10
	Initial NPS	4.806	2.088	0	5	5	5	10
Issue Involvement	Retweet	2.753	1.357	1	2	3	4	5
	Favorite	2.747	1.346	1	2	3	4	5
	Read Response	4.212	1.046	1	4	4	5	5
	Comments	3.952	1.197	1	4	4	5	5
Message Credibility	Believable	4.089	0.818	1	4	4	5	5
	Informative	3.860	0.947	1	3	4	4	5
	Trustworthy	3.766	0.915	1	3	4	4	5
	Reliable	3.731	0.930	1	3	4	4	5
Message Relevance	Useful	3.664	0.997	1	3	4	4	5
	Interesting	4.089	0.837	1	4	4	5	5
	Remember	3.852	1.005	1	3	4	5	5
	Like	3.559	1.051	1	3	4	4	5
Emotions	Joy	1.430	1.281	0	0	1	2	5
	Sadness	2.094	1.499	0	1	2	3	5
	Surprise	2.987	1.530	0	2	3	4	5
	Anger	2.191	1.561	0	1	2	3	5
	Fear	0.892	0.982	0	0	1	1	5
	Disgust	2.097	1.660	0	1	2	3	5
	Not Surprised	1.559	1.414	0	0	1	2	5
	Approve	1.180	1.153	0	0	1	2	5
Demographics	Age	2.468	0.678	2	2	2	3	5
	Sex	1.570	0.496	1	1	2	2	2
	Education	4.032	1.333	2	3	4	5	8
	Daily Use	2.511	0.796	2	2	2	3	5

Table A.3 – Summary statistics by model (continued). The table below shows summary statistics for each variable used to build our constructs separated by medium. NPS ranges from 0 (least likely) to 10 (most likely to recommend). The other Likert variables ranged from 0 to 5. Scale items and values for age include: 18-24 years (2), 25-34 (3), 35-44 years (4) and 45-54 years (5). Scale items and values for Sex include: female (1) and male (2). Education is represented by the following: High School/GED (2), some college (3), 2-year college degree (4), 4-year college degree (5), master's degree (6), doctoral degree (7), and (8) professional degree (JD, MD). Daily use is divided into the following categories: (2) for 30 minutes or less, (3) for 30-60 minutes, (4) for 60-90 minutes, and (5) for more than 90 minutes per day.

B. Photo model summary statistics.								
		Mean	SD	Min	1Q	Median	3Q	Max
NPS	Final NPS	3.731	2.182	0	2	4	5	10
	Initial NPS	5.301	2.424	0	5	5	7	10
Issue Involvement	Retweet	2.008	1.032	1	1	2	2	5
	Favorite	2.110	1.121	1	1	2	3	5
	Read Response	4.054	1.090	1	4	4	5	5
	Comments	3.761	1.188	1	3	4	5	5
Message Credibility	Believable	4.089	0.801	1	4	4	5	5
	Informative	3.788	0.838	1	3	4	4	5
	Trustworthy	3.409	0.824	1	3	3	4	5
	Reliable	3.363	0.804	1	3	3	4	5
Message Relevance	Useful	3.535	0.930	1	3	4	4	5
	Interesting	3.524	0.981	1	3	4	4	5
	Remember	3.376	1.139	1	2.75	4	4	5
	Like	3.003	0.870	1	2	3	4	5
Emotions	Joy	1.255	0.981	0	1	1	2	5
	Sadness	1.946	1.359	0	1	2	3	5
	Surprise	2.175	1.309	0	1	2	3	5
	Anger	2.005	1.433	0	1	2	3	5
	Fear	1.177	1.026	0	0	1	2	5
	Disgust	1.995	1.418	0	1	2	3	5
	Not Surprised	2.599	1.572	0	1	3	4	5
	Approve	1.277	1.164	0	1	1	2	5
Demographics	Age	2.468	0.678	2	2	2	3	5
	Sex	1.570	0.496	1	1	2	2	2
	Education	4.032	1.333	2	3	4	5	8
	Daily Use	2.511	0.796	2	2	2	3	5

Table A.4 – Summary statistics by model (continued). The table below shows summary statistics for each variable used to build our constructs separated by medium. NPS ranges from 0 (least likely) to 10 (most likely to recommend). The other Likert variables ranged from 0 to 5. Scale items and values for age include: 18-24 years (2), 25-34 (3), 35-44 years (4) and 45-54 years (5). Scale items and values for Sex include: female (1) and male (2). Education is represented by the following: High School/GED (2), some college (3), 2-year college degree (4), 4-year college degree (5), master's degree (6), doctoral degree (7), and (8) professional degree (JD, MD). Daily use is divided into the following categories: (2) for 30 minutes or less, (3) for 30-60 minutes, (4) for 60-90 minutes, and (5) for more than 90 minutes per day.

C. Text model summary statics.

		Mean	*SD*	*Min*	*1Q*	*Median*	*3Q*	*Max*
NPS	Final NPS	3.688	1.896	0	3	4	5	9
	Initial NPS	5.169	2.262	0	5	5	6.25	10
Issue Involvement	Retweet	1.863	0.969	1	1	2	2	5
	Favorite	1.917	1.002	1	1	2	2	5
	Response	3.718	1.298	1	3	4	5	5
	Comments	3.368	1.328	1	2	4	4	5
Message Credibility	Believable	3.758	0.788	1	3	4	4	5
	Informative	3.355	1.042	1	3	4	4	5
	Trustworthy	3.315	0.869	1	3	3	4	5
	Reliable	3.274	0.878	1	3	3	4	5
Message Relevance	Useful	3.102	1.054	1	2	3	4	5
	Interesting	2.917	1.052	1	2	3	4	5
	Remember	2.734	1.100	1	2	3	4	5
	Like	2.535	0.947	1	2	3	3	5
Emotions	Joy	0.720	0.809	0	0	1	1	4
	Sadness	1.583	1.430	0	0	1	3	5
	Surprise	1.392	1.339	0	0	1	2	5
	Anger	1.360	1.340	0	0	1	2	5
	Fear	0.758	0.955	0	0	1	1	5
	Disgust	1.446	1.360	0	0	1	2	5
	Not Surprised	1.938	1.638	0	0	2	3	5
	Approve	0.809	0.945	0	0	1	1	5
Demographics	Age	2.468	0.678	2	2	2	3	5
	Sex	1.570	0.496	1	1	2	2	2
	Education	4.032	1.333	2	3	4	5	8
	Daily Use	2.511	0.796	2	2	2	3	5

Table A.5 - Composite reliability and AVE. Convergent validity was established as the average variance explained (AVE) by the multiple indicators of each latent variable was > than 0.5. Internal consistency reliability was established, as all of the composite reliability coefficients for the latent variables were > 0.6 (single-item constructs measure at 1). Please see Table A.1 for an overview of how these constructs were measured.

Variable	Composite Reliability			AVE		
	Video	Photo	Text	Video	Photo	Text
Anger	1	1	1	1	1	1
Approval	1	1	1	1	1	1
Disgust	1	1	1	1	1	1
Fear	1	1	1	1	1	1
Joy	1	1	1	1	1	1
Not surprised	1	1	1	1	1	1
Sadness	1	1	1	1	1	1
Surprise	1	1	1	1	1	1
MWOM message relevance	0.912	0.86	0.918	0.722	0.607	0.736
MWOM message credibility	0.927	0.867	0.919	0.762	0.623	0.74
MWOM issue involvement	0.856	0.8	0.866	0.601	0.507	0.618
NPS change	1	1	1	1	1	1

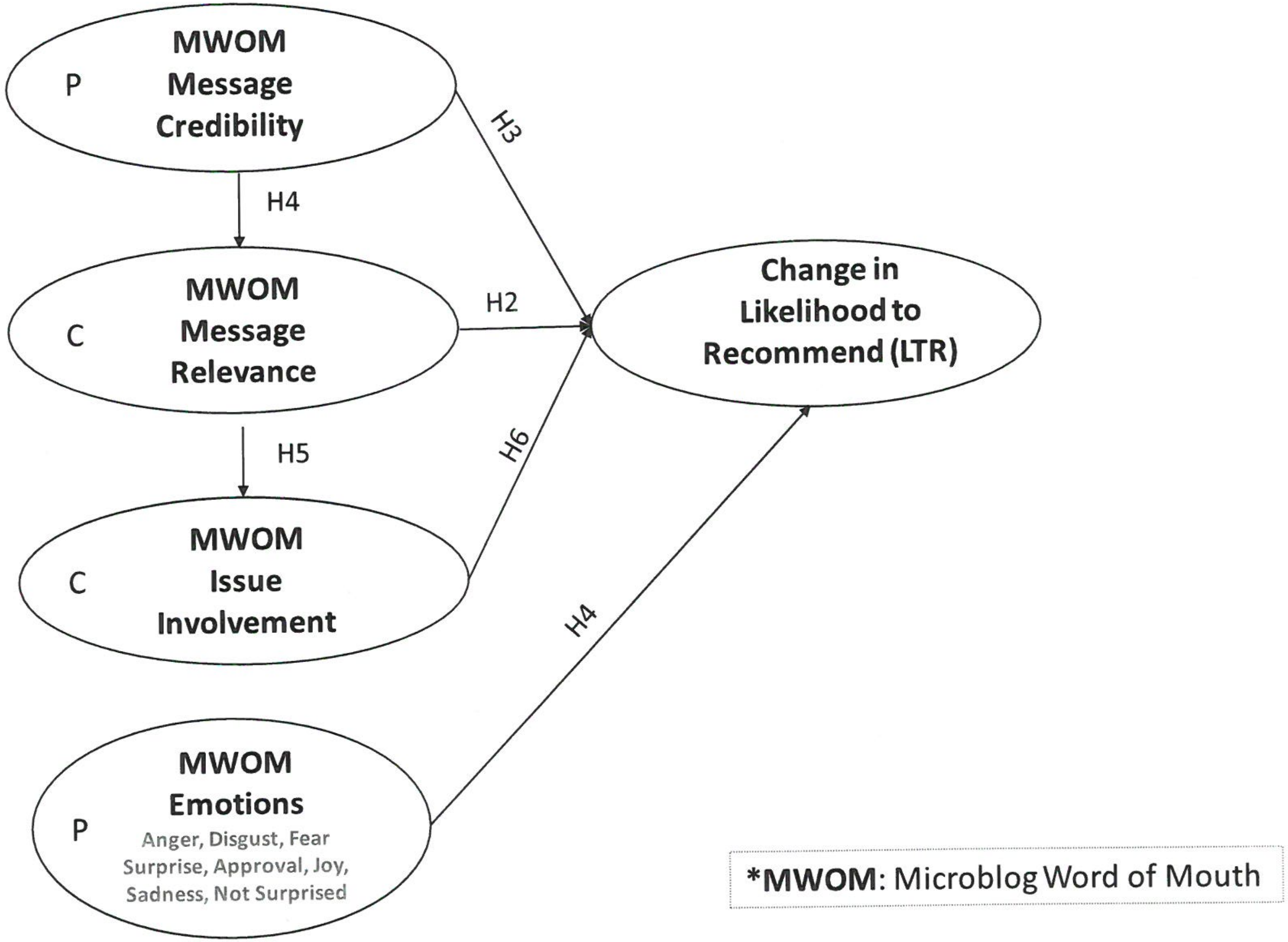
Video

Click here to download Video: Video Negative MWOM.mp4

e-component

Click here to download e-component: Correlations by Type.xlsx

Graphical Abstract



Author Biography

Jennifer B. Barhorst is an Assistant Professor of Marketing at the College of Charleston. Prior to completing her Ph.D. in Marketing at the University of Strathclyde, she owned a brand management consulting firm and spent several years as a brand management and technology consultant. Her research interests comprise emerging technologies and brand management.

Alan Wilson is a Professor of Marketing at the University of Strathclyde Business School, Glasgow, UK. He has published widely in the areas of Branding, Service Delivery and Corporate Reputation. His research and teaching interests are in services and hospitality marketing, corporate branding and reputation, digital marketing, customer experience management and marketing research. Prof Wilson is a Visiting Professor of Services Marketing at Ecole Hoteliere Lausanne, Switzerland and was a former Head of the Department of Marketing, and Vice Dean at Strathclyde Business School.

Joshua Brooks is an Assistant Professor of Finance at Columbus State University, Columbus, GA. He completed his Ph.D. in Finance at The University of Alabama and published in derivatives valuation. In addition to his teaching and research responsibilities, he has consulted in swap valuation, options strategies, performance attribution, and enterprise risk management. His research and teaching interests are investments, risk management, and applied empirical methods.