8-2013

EFFECTS OF ONLINE REPUTATION MECHANISMS ON PERCEIVED CREDIBILITY AND HEALTH DECISION MAKING

Drew Link
Clemson University, dlink@clemson.edu

EFFECTS OF ONLINE REPUTATION MECHANISMS ON PERCEIVED CREDIBILITY AND HEALTH DECISION MAKING

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Applied Psychology

---

by
Drew M. Link
August 2013

Accepted by:
Dr. Leo Gugerty, Committee Chair
Dr. James McCubbin
Dr. Richard Pak

ABSTRACT

Reputation mechanisms and credibility are methods of adding additional information to forum posts, and are becoming more commonplace in online health forums. These systems provide users of forums additional information which can be used to evaluate the trustworthiness of the information being disseminated in community-run websites. The goal of the following studies is twofold. First, it is necessary to identify which elements of reputation systems and credibility participants use to make assessments of the trustworthiness, perceived credibility, and perceived accuracy of answers to health-related questions on a simulated web forum. Once the reputation mechanisms and credibility systems have been identified, the second study explored how high and low overall reputation affects decisions in a non-compensatory decision-making task. This study demonstrated a preference for a non-dominating alternative when it is associated with high reputation, and an overwhelming preference for a dominating alternative when associated with a high reputation. This study also showed that participants expressed higher levels of perceived credibility, trust, confidence, and accuracy of answers when making a decision based on reputation rather than utility.

TABLE OF CONTENTS

Page

Table of Contents (Continued)

Page

1. INTRODUCTION

Objective

Two studies were conducted to investigate how reputation systems and credibility ratings affect patient decision making on online health forums. The purpose of the first study was to gather user judgments on which factors of reputation and credibility of comments made by members of an online community-driven website seem likely to affect their decision making in a health context. This study used the factorial survey method to gather user judgments. The purpose of the second study was to incorporate the findings of the first and examine the effects of reputation and credibility ratings on actual user health-related decision making, using a multi-attribute decision task.

Health information searching

An increasing number of adults turn to the web to find information regarding their health (Pew Research Center, 2009). Their motivations for doing so include (to name a few) gathering general information regarding health conditions, information gathering for the purpose of self-educating, and participating in discussions related to various health-related topics. These individuals who use the internet as a tool for information gathering, known as "e-patients", tend to be healthy people simply looking for information for educational purposes, acutely ill patients investigating a new medical problem, or patients with chronic diseases utilizing online resources frequently (Ferguson, 2007). Pew Research Center's Internet & American Life Project and the California HealthCare Foundation estimates that 80% of adults consult the web for gathering information

regarding health information (Pew Research Center, 2010). Initially, e-patients tended to consult online resources maintained by doctors and research institutes for gathering information regarding health topics. WebMD, Mayo Clinic, and eMedicine Health include examples of sites that are maintained by professionals and experts in the field.

In an interesting trend in recent years, e-patients are consulting sites where other consumers are providing information regarding various health topics. Of the 80% of patients who consult the web, 59% engage in activities including reading other consumers' commentaries, consulting rankings of hospitals or healthcare providers, or browsing online forum discussions of health topics. Furthermore, 22% of e-patients actively contribute comments, reviews, and ratings to these sites. It is clear from this trend that patients are not only seeking medical advice from other sources beyond experts and medical professionals, but are contributing to the growing database of information online as well.

Community-driven websites

Many people search for health information on the web via querying a search engine for specific information and browsing through the results of that search to find an answer to the question. However, people often seek information, support, and advice using online forums and message boards to gather and disseminate information. An analysis of user behavior on online forums by Billman, Sciano, & Gugerty (2008) showed that the typical behaviors of users on community-driven sites included asking specific health-related questions, answering questions, and supporting other members

emotionally. Further, these sites are often used to provide users with advice and recommendations regarding medical professionals to seek treatment from (Coulson, Buchanan, & Aubeeluck, 2007). These findings suggest that the use of community-driven sites is increasing because these sites provide assistance beyond that which is given by informational websites, and that e-patients can ask questions on community sites that are more personalized and specific to their situation. However, the quality of information provided on these sites varies.

E-patients often seek out community-run forums when they or their family members are dealing with a chronic illness. While participation and reference of these forums may provide individuals with information, answers, and support, the fact that many of these forums are not moderated means that the information provided may be of dubious quality. Some studies have shown that using online forums may give patients the opportunity to explore gaps between evidence-based medicine and patients expectations, allowing e-patients to learn about the effectiveness of a treatment or intervention from reading personal accounts from others people with a similar condition (Coulson, Buchanan, & Aubeeluck, 2007). Largely, it is still inconclusive whether information provided on community-run forums has a positive or negative effect on patient behavior or decision-making. One reason is that it is difficult to discern when and if patients are utilizing these sites as sole resources for medical decision making and failing to complement their web research with professional consultation or other resources (Eysenbach, Powell, Englesakis, Rizo, & Stern, 2004). Eysenbach et al. also indicated that one of the reasons it was difficult to discern to what extent e-patients relied

on information from these sites was due to the lack of quantitative studies that evaluate the effects of the content of these sites on decision making and e-patient behavior.

Sillence, Briggs, Harris, & Fishwick (2007) conducted a longitudinal study that evaluated how patients undergoing hormone replacement therapy (HRT) were utilizing online resources. This comprehensive study showed that the likelihood online information of any kind is trusted by e-patients is affected by many subtle factors. Many of the participants in the study used heuristics to determine whether an informational site could be trusted, such as the design of the site, the use of professional language, and the inclusion of personalized content. The behavior of the participants in the Sillence et al., study would suggest that e-patients would make the same evaluations in order to gauge the trustworthiness of online forums as well. Eysenbach and Kohler (2002) observed and recorded participants as they searched the internet for health information, and then interviewed the participants about their search behavior. The participants in this study reported that website ownership and disclosure statements were important in their decision making process. In practice, participants behavior did not match their self-reports, since they did not read detailed "about us" portions of websites in order to determine the credibility of the information being presented.

In addition to the factors mentioned above, another method that online forums utilize to help inform users regarding the credibility of information is reputation mechanisms. These systems provide a means for users to assign credibility ratings to

authors and comments in an attempt to improve the quality of information disseminated in their community-run websites.

E-patient Trust in Online Information

Reputation mechanisms

Reputation mechanisms (or feedback systems) are elements of websites that allow users to rate, evaluate, comment on, or otherwise inform others about the accuracy of specific information (e.g., a website post), the pattern of information provided by a specific individual over time on a website, or a specific website. Originally developed for e-commerce websites, reputation mechanisms offered the means for users to provide input to inform other users which providers or products can be trusted and which cannot. Dellarocas (2003) states that reputation mechanisms are essential for ensuring good behavior of online sellers and service providers. Table 1 shows several examples of websites that utilize reputation mechanisms and the feedback mechanisms that are utilized to generate a reputation profile.

Table 1: Dellarocas' (2003) examples of online reputation mechanisms, adapted to include newer websites.

Web Site	Category	Summary of feedback mechanism	Type of feedback solicited	Type of reputation profiles
eBay	Online auction house	Buyers and sellers rate one another following transactions	Positive, negative or neutral rating plus comment	Sums of positive, negative and neutral ratings received during last 6 months
eLance	Professional services marketplace	Contractors rate their satisfaction with subcontractors	Numerical rating from 1-5 plus comment; subcontractor may post a response	Average of ratings received during last 6 months
Epinions	Online opinions forum	Users write reviews about products/services; other members rate the usefulness of reviews	Users rate multiple aspects of reviewed items from 1-5; readers rate reviews as "useful", "not useful", etc.	Averages of item ratings; % of readers who found a review "useful"
Reddit	Online discussion board	Postings are prioritized or filtered according to the ratings they receive from readers	Readers rate posted comments	No explicit reputation profiles are published
Yahoo Answers	Online Q&A	Users rate the member writing an answer and the answer provided. Users must provide a source for their answer.	Askers rate the profiles of people that answer and the content of an answer	Profiles have star ratings (1 – 5), answers are prioritized based on ratings

These reputation mechanisms allow users to give their own views about the quality of information on websites, and have expanded to include not only e-commerce sites such as eBay or Amazon, but are also websites that provide user-generated general information (e.g., Yahoo! Answers) and many health forums (e.g., healthboards.com, diabetesforums.com). The inclusion of reputation mechanisms on health-related websites is relatively new, and many sites may offer very few mechanisms and others may offer a variety of mechanisms. Still, these reputation mechanisms offer a way of allowing the user to rate information, decide what information or users are trustworthy, and a means of applying quick heuristics to filter information provided in these sites.

Credibility

Before discussing the importance of credibility on web forums, an important distinction must be made. Reputation mechanisms are community-controlled feedback tools that allow for several users of a website to rate the usefulness, trustworthiness, and reputation of an author or member, while credibility refers to aspects of the website or author that are not controlled by feedback mechanisms from the community, but are rather included in the content of the site. Several credibility factors can affect the degree to which a consumer trusts content on a website, be it a site providing general information or a community-run forum. Eysenbach & Kohler (2002) found a number of criteria that consumers noted were important for ascertaining the credibility of a website. After conducting an interview with consumers, holding focus groups, and conducting usability tests, Eysenbach and Kohler found several criteria that consumers identified as

affecting their credibility judgments, including: Authority of information sources, layout of the website, outbound links, site readability, pictures of the website owner, credentials and qualifications, and quality seals/third party endorsements. Some of these (e.g., source authority) apply to forums, as individuals who participate in online discussions are given the option of including information about themselves. Some forums simply give you the option to post your credentials and qualifications, while other forums actually require you to enter this type of information. Readability also applies to forum postings in the sense that the way answers are given can affect the perceived credibility of the answer and poster. During interviews, a common response was that websites should be readable and understandable to the layperson, while providing a sufficient amount of detail.

In addition to authority of the source and readability, other criteria mentioned in this study would apply to forums, including, photographs of people posting answers, and whether they provide outbound links. After conducting this study, Eysenbach & Thomson (2007) developed an internet fact-checking algorithm to assist consumers in finding credible information online. While this decision-making aid can be used for fact-checking forum responses, it is mainly used as a training tool and may or may not be used regularly by consumers when they are seeking information on a day-to-day basis.

As noted above, one method people use to perceive credibility in online content is the authority or expertise of the source. While Eysenbach and Kohler (2002) did find that consumers claim they rely on the credibility of the author of web information, studies

have also shown conflicting evidence on how source expertise or background knowledge of the author affects perceived credibility of information. Eysenbach and Kohler found that consumers did not check "About Us" sections of a website where references, citations, and details about authors is located, and in fact often did not recall specifically where they retrieved answers to questions on the web. Eastin (2001) found, however, that there was a relationship between source expertise as well as user domain knowledge on perceived credibility of information such that higher user knowledge of a health topic and higher expertise of author were correlated with higher perceived credibility of information. Similar results from other studies also suggest that people make their credibility judgments dependent upon the message contained and delivered within the site, rather than by the site's structure, when the expertise of the message communicator is perceived to be high (Flanagin & Metzger, 2007; Hong, 2005). Additionally, the implementation of education and training can improve the degree to which consumers evaluate the credibility of a website (Pirolli, Wollny, & Suh, 2009).

These studies have shown mixed results in determining to what degree people use credibility cues to determine if content of a website is trustworthy. However, with community-run forums, credibility is often conspicuous because it is built into the comments themselves. Consumers may be more likely to use credibility information in their decision making when the credibility information (e.g., source expertise and background knowledge) is a conspicuous part of forum comments than when the credibility info is less accessible, e.g., on a reference list on a separate page, in a separate "About Us" section, or in the fine print.

Overview of the Current Study

The goal of the present studies were to determine what effect (if any) reputation and credibility information in online forums have on participants' decision-making processes, subjective ratings of confidence, and perceived credibility ratings of the information sources relevant to a decision. In each study, reputation mechanisms and credibility were manipulated to reflect how they may exist in an online community.

The first study was conducted using the factorial survey method as a means of determining which aspects of online reputation and credibility mechanisms for user-generated content should be included in our second study. The first study examined two reputation mechanisms and one credibility factor: 1.) Overall rating of the comment author, represented using the common star-rating system that many forums and websites use to allow members to rate the general reputation of the authors of comments across all aspects of the forum (e.g., 1-star rating to 5-star ratings), 2.) comment rating, represented using the common "x number of people found this comment useful" method that allows members on the forum to make judgments about the usefulness or relevance of a particular comment made by a forum member (e.g., "7 people found this comment helpful"), and 3.) background knowledge of comment authors, represented by the author-controlled profile information that is often displayed on web-based forums (e.g., "Member Background: Expert in field"). This final variable is different from the first two variables in that it is not community controlled. The background knowledge is an aspect of forum profiles that allow the creators to enter their experience in a given field.

In the first study, participants read hypothetical forum comments that varied in terms of rating of the comment author, rating of the comment itself, and background of the comment author knowledge. For each post, participants rated their trust in the comment information and the comment author, the accuracy of the comment, and how likely they are to follow the advice in the comment.

The second study was an experiment conducted to evaluate how the reputation and credibility factors identified in the first experiment affected decision making. This study included a series of simulated forum pages. For each page, the participant engaged in a multi-attribute decision making task that had options varied between high or low levels of reputation and credibility for each given comment.

2. EXPERIMENT 1: ANALYSIS OF REPUTATION RATINGS VIA FACTORIAL SURVEY

The goal of Experiment 1 was to determine which methods of communicating reputation and credibility information in an online patient-centered health forum have the most effect on participants' perceptions of overall trust in the information they are given. The results from Experiment 1 determined which reputation mechanisms were used in the second study, and whether the credibility system was used. While prior studies indicate that design of a website and presentation of information have a large effect on whether users deem the information trustworthy (Eysenbach & Kohler, 2002; Sillence, Briggs, Harris, & Fishwick, 2007), these factors were not the focus of this study because they do not provide information about the reputation or credibility of individual forum authors. There is a lack of research investigating which methods of communicating reputation and credibility information affect consumers using web forums in the health domain.

Valence

Valence in framing and decision making research refers to the concept that an event or stimulus can elicit a positive or negative response from an individual. This term is often referred to in the study of emotion, as statements have been manipulated in studies to elicit a positive or negative response to a certain situation, in the study of cognition and emotion (e.g., Lerner & Dachner, 2000) or in performance appraisal research in organizational psychology (Crowe & Higgens, 1997). Positive valence can occur in the form of praise or approval of an individual or concept, while negative valence typically involves a stimulus that elicits an adverse response from an individual.

For instance, consumer reviews on e-commerce sites that employ the reputation mechanisms discussed previously often include some form of valence communication. These may be represented in some visual manner that indicates the "rating" of information provided on a website. Zou, Yu, and Hao (2011) conducted a study that evaluated the effect of valence on consumer decision making on e-commerce websites. The ratings that people gave on other people's comments had an effect on decision making, however it was moderated by the expertise of the consumer. Participants with a higher expertise in the relevant area were less influenced by valence than participants with less expertise.

In the current study, the use of reputation mechanisms in online web forums was an avenue for measuring the effects of valence on decision making regarding health information. Valence was manipulated via the inclusion of star ratings of authors and individual comments, each rating meant to represent the reputation mechanism allowing the evaluation by peers within the online forum. Valence was represented with a visual aid of a star-rating system, commonly incorporated into online forums. Positive valence was represented by a rating of four or five stars out of five and negative valence by a rating of one or two stars. The hypothesis is that, relative to negative valence, positive valence would be associated with more trust, perceptions of accuracy, and perceived believability in an answer, and less need to continue looking for information.

Amount of support

Not only can a positive or negative valence associated with a stimulus affect the decision making process of an individual, but the amount of support can have an effect as well. A study by Yaniv and Kleinberger (2002) showed that advice-taking behaviors can be influenced by the amount of support for a given opinion or suggestion. This study demonstrated that advice is likely to be taken when the source has a high reputation and large amount of support from others. Advice and judgments of these sources are weighted according to these factors, but may also be discounted by an individual based on their own personal experience on the matter. Often, this discounting occurs because of an individual's expertise on a given matter, however this is not always the case. Stanovich and West (1997) found that people who based decisions on a large amount of statistical support scored better on reasoning tasks than people who based their decisions on single anecdotes.

The research above suggests the following hypothesis. An author or comment rating with a high amount of support (i.e., high number of raters) will have more of an influence on decision making than ratings with less support. Amount of support should interact with valence in the following manner. When valence of an author or comment rating is high (positive), high support will lead to higher ratings of trust, believability, and accuracy and less need to continue looking compared to low support. In contrast, when valence of an author or comment rating is low (negative), high support will lead to lower ratings of trust, believability, and accuracy and more need to continue looking compared to low support. While I hypothesize that valence will show a main effect on perceptions

of trust in information presented on web forums, amount of support should not show a main effect because support works to magnify the positive or negative effects of valence. Therefore, the interaction above is hypothesized.

The hypothesis that credibility of the comment author and the author's reputation within the forums will have a strong positive relationship with perceived trustworthiness of information was tested in this study. The prediction that credibility information would be positively related to trust is based on the findings from Eysenbach and Kohler (2002) as well as Sillence, Briggs, Harris, & Fishwick (2007) suggesting that e-patients place value in credibility of online information in assessing the utility of websites. Our manipulation of the background knowledge of the post author attempts to replicate this effect. The justification for the strong positive correlations of comment ratings and comment author ratings is supported by research in e-commerce settings (Zou, Yu, & Hao, 2011), and more-so in the study by Yaniv and Kleinberger (2002) which, as mentioned previously, provided support for the idea that consultation of a high reputation source leads to more advice-taking behaviors from participants. While no research was found explicitly comparing comment rating with author ratings , Pavlou and Dimoka (2006) found that text comments with specific information led to higher perceived credibility and ratings of trust than simple rating systems. This finding supports the predicted effect of the comment rating, since ratings of a particular comment are relatively specific.

3. METHODS

Participants

Twenty adults (10 male and 10 female) participated in the first study. Participants were presented with all combinations of levels of the three reputation and credibility variables being manipulated. The scenarios used in this study involved questions posted to a web forum on the topic of pet health. Participants were therefore excluded if they had advanced knowledge of illnesses common in household pets. Participants were undergraduates recruited from Clemson University via the internal human participant pool.

Design

Experiment 1 was a factorial survey study. The design was a 4 (comment author rating: low with low number forum member ratings, low with high number of forum member ratings, high with low number forum member ratings, and with high number of forum member ratings) x 4 (comment rating: low with low number forum member ratings, low with high number of forum member ratings, high with low number forum member ratings, and with high number of forum member ratings) x 3 (background knowledge of comment author: low, moderate, and high) within-subjects design. As such, each of the 48 combinations of independent variables was given to the participants twice. The participants were given a total of 96 forum posts, and each post consisted of a question posed by a forum member asking a medical question regarding the care of a household pet. Each post included one comment from a participating member of the website. The content of the comment was a brief suggestion of a treatment or medication

to use. Additionally, each response contained information regarding valence and amount of support for the comment and author and author's background information in a information box to the left of the comment. See Figure 1 for an example of a post and comment.

After each presentation of a post and comment, the participant responded to four questions.

1. How trustworthy is the answer to the medical question asked? (Using a seven-point Likert scale ranging from 'very untrustworthy' to 'very trustworthy')
2. How likely would you be to continue looking for other potential answers after reading this answer? (Seven-point Likert scale ranging from 'very unlikely' to 'very likely')
3. How believable is the person who wrote this answer? (Seven-point Likert scale ranging from 'not a believable source' to 'a very believable source')
4. How accurate did you find this answer? (Seven-point Likert scale ranging from 'not at all accurate' to 'very accurate')

These four responses were dependent variables measured in the first study. The first, second, and third questions are designed to assess the participants' trust in the information provided in the sample forum post, and question four assesses the participants' perceived credibility of the member providing the answer. The Likert scale from 1 to 7 had a midpoint labeled as "Neutral" for each question.

The order in which participants were exposed to factors was counterbalanced. The three reputation and credibility factors (based on the 4x4x3 design) lead to 48 combinations of factors. Each of the 48 combinations was presented in each block of trials; and the blocks will be repeated twice using different questions (but the same format of response). The order in which each combination of factors presented was randomized within each block.

A power analysis was conducted using the software application G*power that reflected a conservative estimation used in multilevel models (MLM), using a conservative ANOVA for fixed effects. The sample size was calculated assuming a Cohen's d effect size of 0.25. Using the output from this power analysis, the number of participants deemed necessary was 20, including the 48 vignettes repeated twice.

Materials

Participants provided demographic information as per the form given in Appendix A. In addition to providing biographical information, participants answered questions regarding their level of education and experience with pets. High levels of knowledge with pet problems were evaluated further in an open-ended interview at the completion of the study, as participants may base their decisions of the quality of the information on aspects other than reputation mechanisms.

Scenarios

The experiment was created using Survey Monkey survey creation software. Traditionally, factorial survey studies involve the creation of a short scenario presented to participants. Across trials, different elements of the scenario are varied according to the factors being measured. A mock-up of a web forum will be created using Microsoft Powerpoint. It was used to simulate a post from the web, and as such participants were not able to interact with the forum posting other than reading the provided information. See Figure 1 for a sample scenario that was given.

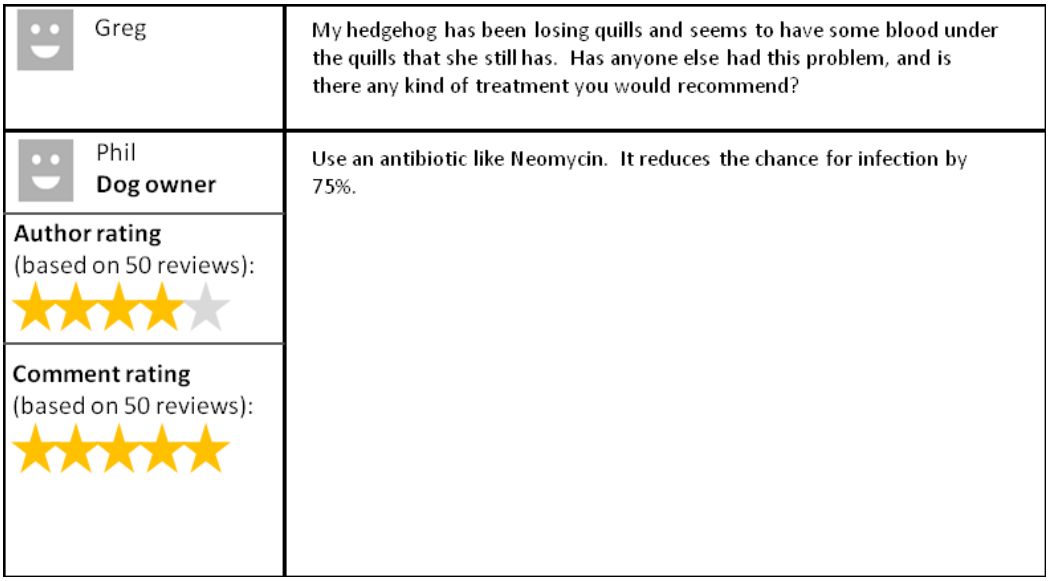


Figure 1: Sample scenario showing a member's question and a member's answer. Factors that would be manipulated include 1.) comment author's background knowledge, 2.) comment author's rating, and 3.) comment rating, displayed in that order under the member's username.

Each of the reputation and credibility factors were displayed on the left part of the answer space. Author and comment rating scores with a low valence were represented as either a one or two star rating, while high valence ratings were represented as either four or five star ratings. Amount of support was indicated as 'low' if the ratings were determined by 1 to 10 reviews, and indicated as 'high' if the ratings were determined by 41 to 50 reviews. Finally, the source background knowledge was displayed under the author's name and profile, either as being a veterinarian for a high rating, a specific-to-problem pet owner for a medium rating (e.g., hedgehog owner if post about hedgehogs), and an unrelated pet owner for a low rating (e.g., dog owner if post about hedgehogs).

The scenario presented an individual posting a question, with a response by a single forum member below the question. The content of the response consisted of one sentence indicating what the suggested course of action would be. This was to ensure that any aspects of wording and message structure as seen in the Sillence, Briggs, Harris and Fishwick (2007) or Hong (2005) studies will affect participants' assessments of credibility.

There were two scenarios written, one for each block of 48 trials formed by the 4x4x3 design.

Procedure

An experimental session lasted from 20 to 30 minutes. The participant first read an information letter regarding the study. Participants then read the initial instructions indicating that the study is meant to examine how people determine what information is

reliable when browsing the web. In the computerized portion of the study, participants listened to instructions on how to proceed with the study, and then were given a sample scenario. The instructions included an explanation of the various elements of the forum post (i.e., what a comment rating is, what a comment author rating is, and definitions of the elements of a forum-based website).

The next stage of the experimental session involved the participant reading a scenario presented. Upon reading the scenario, the participant answered four Likert-based questions that assess the participants trust in the information on the same page. All information was available to the participant on one page to allow for easy reference back to the scenario while answering the questions. After completing these questions, the participant clicked the next button and read the next scenario. Each participant completed 96 scenarios and were given a brief open-ended task-comprehension survey at the cessation of the experiment.

4. RESULTS

The primary hypothesis for this study was that credibility (background information) of the author, ratings of comments, and ratings of comment authors will all have a strong positive relationship with measures of trust in information. Also, valence and trust were hypothesized to interact so that trust ratings would increase over the conditions of low valence and high support, low valence and low support, high valence and low support, and high valence and high support. The general construct of trust was measured by four questions where participants used seven-point scales to rate: trust in an answer, accuracy of an answer, believability of the comment author, and likelihood of looking for more information. The response labels for trust, accuracy and believability were scored so that higher numbers meant greater trust, accuracy or believability. The response labels for "likelihood of looking for more information" were scored so that lower numbers meant greater trust in information. Therefore, the prediction was that increases in the credibility, comment rating and author rating will be associated with increases in the dependent variables of trust, accuracy and believability, and decreases in likelihood of looking for more information. Effect sizes were calculated using Cohen's *f* (Cohen,1992). As such, the standard effect sizes of small, medium, and large are used as 0.1, 0.25, and 0.4 respectively.

For each of the four dependent variables, a separate multilevel model was run with each of the three independent variables, background knowledge, comment rating, and author rating as independent variables. This method included blocks 1 and 2 at first to ensure that there was no significant effect of block on the dependent measures. Since

no significant effect was found, this variable was removed from further analyses. For the comment and author ratings, we also ran multilevel models using valence and amount of support as independent variables.

Figure 2 shows the main effects of background knowledge, comment rating, and author rating on perceived trust of the information. The figure suggests that as background, comment rating, and author rating increase, trust increases. The MLM conducted with background, comment rating, and author rating as independent variables showed that background knowledge, F (2,1853) = 21.08, p=.000, f = 0.09, comment rating, F(3,1853) = 791.84, p=.000, f = 0.94, and author rating, F(3,1853) = 102.62, p=.000, f = 0.26, had significant effects on trust. This evidence supports our primary hypothesis. Additionally, comment rating showed a very large effect size, while author rating showed a medium effect size and background knowledge showed a small effect size.

Additional MLMs tested the relative contributions of valence and amount of support to the overall main effects of comment and author rating. For both of these ratings, I hypothesized a main effect of valence and an interaction of valence and amount of support such that high amount of support increases trust for high valence and decreases trust for low valence. For author rating and comment rating, Figure 2 suggests that some of these effects were present in the data. For comment ratings, there was a large main effect of valence, F(1, 937) = 440.12, p=.000, f = 0.91, on trust, and a significant interaction between valence and support, F(1, 937) = 29.12, p=.000, f = 0.22 , which

supported the hypotheses. The effect size for valence is much larger than the effect size for the amount of support interaction, suggesting that the valence of the comment contributed to greater differences in response to the question on trust.

For author ratings, there was a moderate main effect of valence, $F(1,937) = 74.13$, $p=.000$, $f = 0.25$ on trust, and no interaction between valence and support, $F(1,937) = 2.12$, $p=.146$, $f = 0.04$, which partially supported the hypothesis. There was only a detected effect size for valence, and there were no effects of amount of support or any evidence of an interaction. However, the hypothesis that valence would have more of an effect was supported.

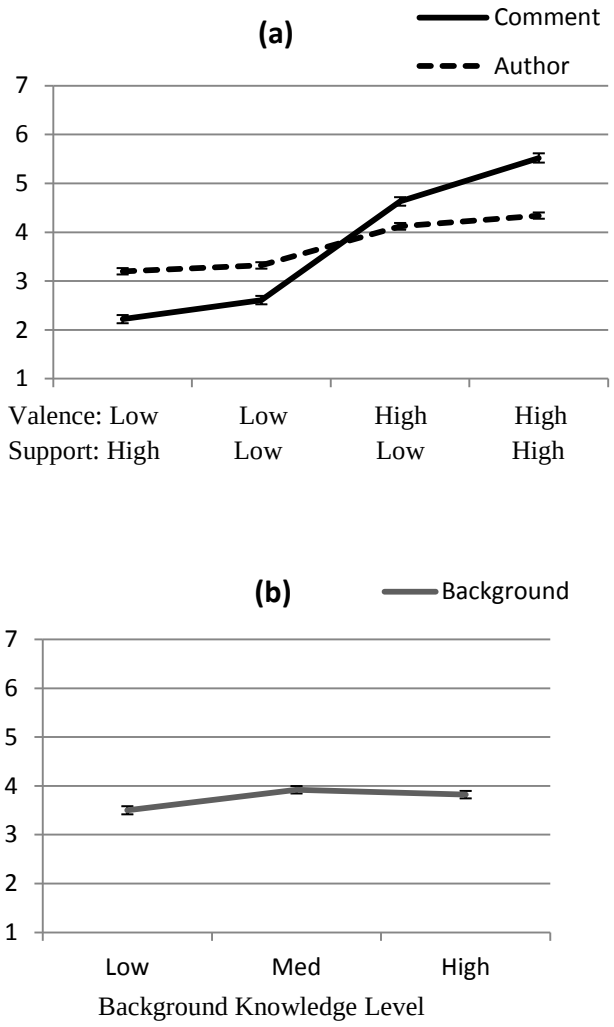


Figure 2: Scores on the question on trust plotted as a function of comment and author valence and amount of support (a) and level of background knowledge. Error bars represent standard error of estimate values.

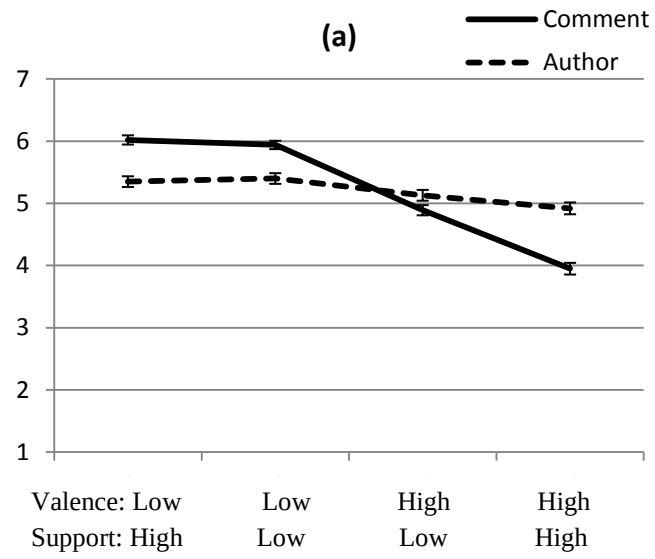
Figure 3 shows the main effects of background knowledge, comment rating, and author rating on likelihood of continuing to look for more information regarding the problem. The figure suggests that as background knowledge, comment rating, and author rating increase, the likelihood of continuing to look for more answer decreases. An MLM showed that background knowledge, $F(2,1853) = 4.28$, $p=.014$, $f = .05$, comment

rating, $F(3,1853) = 201.97$, $p=.000$, $f = .49$, and author rating, $F(3,1853) = 10.06$, $p=.000$, $f = .10$, had significant effects on the likelihood of continuing to look for more information. This evidence supports our primary hypothesis. Additionally, comment rating showed a large effect size, author rating showed a small effect size.

Additional MLMs tested the relative contributions of valence and amount of support to the overall main effects of comment and author rating. For both of these ratings, I hypothesized a main effect of valence and an interaction of valence and amount of support such that high amount of support decreases the likelihood of searching for more information for high valence and increases the likelihood for low valence. For author rating and comment rating, Figure 3 suggests that some of these effects were present in the data. For comment ratings, there was a large main effect of valence, $F(1,937) = 337.89$, $p=.000$, $f = 0.45$, on the dependent measure, and a significant interaction between valence and support, $F(1,937) = 36.45$, $p=.000$, $f = 0.15$, which supported the hypotheses. The effect size for valence is much larger than the effect size for the interaction of the amount of support, suggesting that the valence of the comment contributed to greater differences in response to the question on whether the participant would be likely to continue looking for more information.

For author ratings, there was a very small main effect of valence, $F(1,937) = 11.97$, $p=.001$, $f = 0.09$ on likelihood to continue looking for more information, and no interaction between valence and support, $F(1,937) = .538$, $p=.463$, $f = 0.02$, which partially supported the hypothesis. There was only a detected effect size for valence, and

there were no effects of amount of support or evidence of an interaction. However, the hypothesis that valence would have more of an effect on the DV than amount of support was supported.



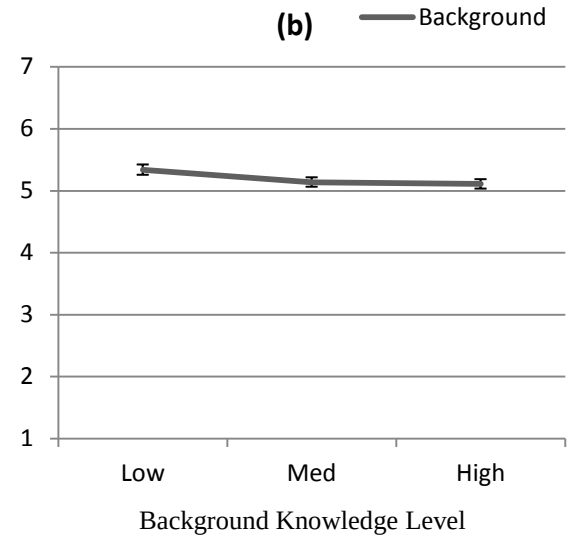


Figure 3: Scores on the question on the likelihood to continue looking for information plotted as a function of comment and author valence and amount of support (a) and level of background knowledge. Error bars represent standard error of estimate values.

Figure 4 shows the main effect of background knowledge, comment rating, and author rating on the perceived believability of the author. The figure suggests that as background knowledge, comment rating, and author rating increase, the perceived believability of the author increases. An MLM showed that background knowledge, $F(2,1853) = 61.36$, $p=.000$, $f = 0.16$, comment rating, $F(3,1853) = 112.45$, $p=.000$, $f = 0.27$, and author rating, $F(3,1853) = 670.20$, $p=.000$, $f = 0.79$, had significant effects on perceived believability. This evidence supports our primary hypothesis. Additionally, author rating showed a large effect size, while comment rating showed a medium effect size and background knowledge showed a small effect size.

Additional MLMs tested the relative contributions of valence and amount of support to the overall main effects of comment and author rating. For both of these ratings, I hypothesized a main effect of valence and an interaction of valence and amount of support such that high amount of support increases believability for high valence and decreases believability for low valence. For author rating and comment rating, Figure 4 suggests that some of these effects were present in the data. For comment ratings, there was a moderate main effect of valence, $F(1,937) = 78.05$, $p=.000$, $f = 0.26$, on believability, and a significant, but very small interaction between valence and support, $F(1, 937) = 10.24$, $p=.001$, $f = 0.09$, which supported the hypotheses. The effect size for valence is larger than the effect size for the amount of support interaction, suggesting that the valence of the comment contributed to greater differences in response to the question on believability.

For author ratings, there was a moderate main effect of valence, $F(1,937) = 966.99$, $p=.000$, $f = 0.78$ on believability, and a significant interaction between valence and support, $F(1,937) = 25.28$, $p=.000$, $f = 0.13$, which supported the hypothesis. There was a large effect size for valence, and there was a small effect size for the interaction. The hypothesis that valence would have more of an effect than amount of support was supported.

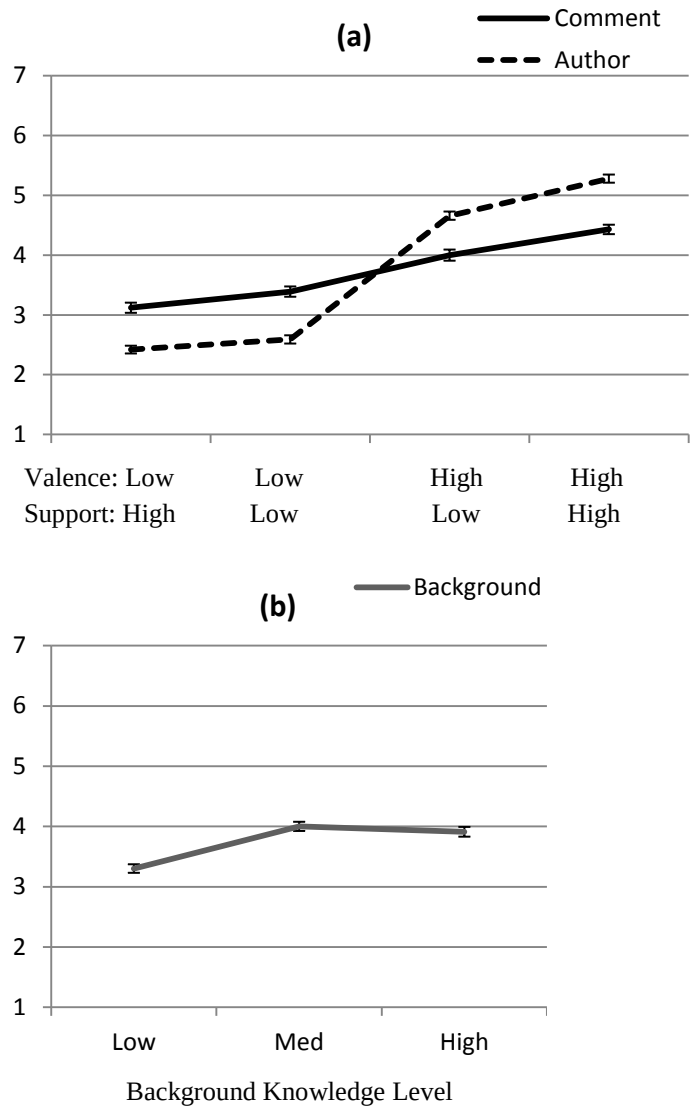


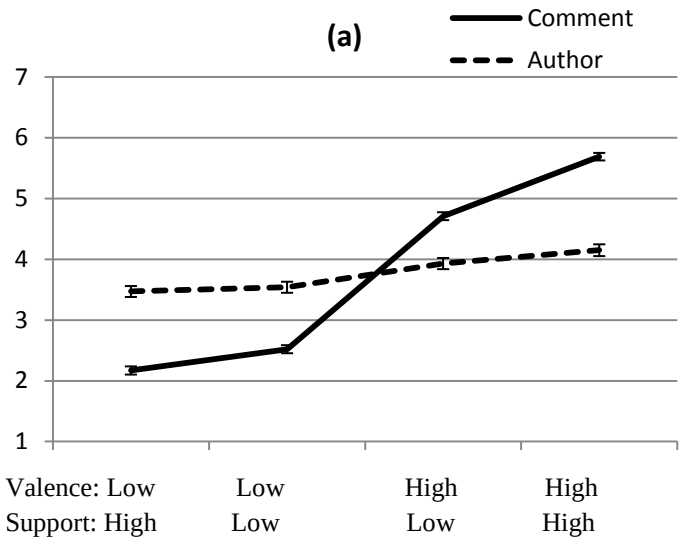
Figure 4: Scores on the question on perceived believability of the author plotted as a function of comment and author valence and amount of support (a) and level of background knowledge. Error bars represent standard error of estimate values.

Figure 5 shows the main effect of background knowledge, comment rating, and author rating on the perceived accuracy of the comment. The figure suggests that as background knowledge, comment rating, and author rating increase, the perceived accuracy of the comment increases. An MLM showed that background knowledge, $F(2,1853) = 10.29$, p=.000, f = .06, comment rating, $F(3,1853) = 910.56$, p=.000, f = 1.04,

and author rating, $F(3,1853) = 33.25$, $p=.000$, $f = 0.14$, had significant effects on perceived accuracy. This evidence supports our primary hypothesis. Additionally, comment rating showed a large effect size, while comment rating showed a small effect size.

Additional MLMs tested the relative contributions of valence and amount of support to the overall main effects of comment and author rating. For both of these ratings, I hypothesized a main effect of valence and an interaction of valence and amount of support such that high amount of support increases perceived accuracy of the answer for high valence and decreases perceived accuracy for low valence. For author rating and comment rating, Figure 4 suggests that some of these effects were present in the data. For comment ratings, there was a large main effect of valence, $F(1,937) = 1567.99$, $p=.000$, $f = 1.01$, on accuracy, and a significant, but moderate interaction between valence and support, $F(1, 937) = 85.63$, $p=.001$, $f = 0.25$, which supported the hypotheses. The effect size for valence is larger than the effect size for the amount of support interaction, suggesting that the valence of the comment contributed to greater differences in response to the question on accuracy.

For author ratings, there was a small main effect of valence, $F(1,937) = 20.48$, $p=.000$, $f = 0.13$ on accuracy, but not a significant interaction between valence and support, $F(1,937) = 1.52$, $p=.217$, $f = 0.03$, which supported the hypothesis. There was a small effect size for valence, and there was no effect size for the interaction.



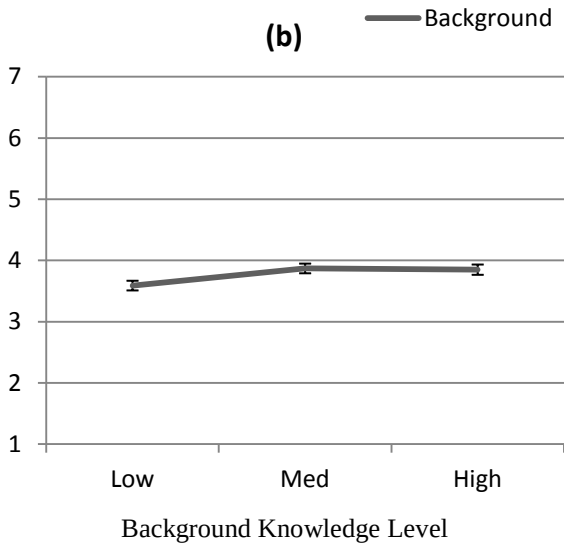


Figure 5: Scores on the question on perceived accuracy of the comment plotted as a function of comment and author valence and amount of support (a) and level of background knowledge. Error bars represent standard error of estimate values.

5. DISCUSSION

The purpose of the first experiment was to determine which methods of communicating reputation and credibility information in an online patient-centered health forum have the most effect on participants' perceptions of overall trust in the information they are given, and potentially how that might guide advice-taking behaviors in the early stages of the decision-making process. This was accomplished by employing a factorial survey approach, which helps determine which factors may be worth inclusion for a follow-up study, and to determine if a reasonable argument can be made for potential interactions between variables. Initial analyses focused on main effects of credibility (background information) and reputation (community ratings of the author and comments). Additional analyses investigated whether the amount of support can change how a valence rating is used as a decision making aid.

For the primary hypothesis, significant effects were found in that background knowledge, credibility, and amount of support all significantly contributed to perception of trust in the author, the information, and in the likelihood of engaging in advice-taking behaviors. These data are compelling in many ways. First, they suggest that each of these aspects of credibility and reputation are significantly related to perception of the information in such a way that a strong justification could be made to incorporate all of the independent variables into the second study. Second, this study suggests that, on average, people are sensitive to the cue of amount of support and use it in appropriate ways, such that high support increases overall trust when paired with high valence and decreases trust when paired with low valence. This relationship is also apparent from the

task comprehension survey responses, showing that participants were aware of this relationship. Finally, an argument could be made that the information displays for credibility and reputation were salient enough to participants; thus these displays can inform the design of the mock forum posts in the second study. This presents the possibility of keeping the design the same or similar from the first study to the second study, allowing for more control over our manipulations for the second study. While participants were instructed to attend to certain parts of the forum post, nevertheless an argument could be made for the salience of the information in that people responded similarly throughout the entire 96 trials, and background information, author and comment ratings, valence and amount of support all had significant effects on participants ratings that were in accord with the hypotheses.

Another interesting result found in the first study is that the content of the questions about perceived accuracy, perceived trust, and likelihood to continue searching for information all produced larger effect sizes with for comment ratings than author ratings. Conversely, a larger effect size was found for the author rating on the question of perceived believability. This was likely due to the phrasing of each of these questions. For the former, perceived accuracy and trust prompted the participants to think of the accuracy and trust of the content of the forum post. This may have resulted in a tendency to be more affected by the actual response rather than the author of that response. For the latter, the question about perceived believability was worded to ask for the believability of the author of a given comment. Therefore, it is likely that participants were more affected by the author valence and amount of support than the comment valence and

amount of support. This provides further evidence that people are taking both comment and author ratings into consideration when making an assessment of the overall trust in the information.

For the secondary hypothesis, significant effects were found more frequently in valence than with amount of support and the interaction of amount of support and valence. These data suggest that, as hypothesized, valence was taken into consideration more in determining a participant's overall trust in information than the amount of support. In every context, as valence changes from high to low for both comment and author, the change in the amount of trust decreases more than the overall effect of the amount of support. This difference of effect size is reduced from comment ratings to author ratings in the same dependent measures as above: Perceived trust, perceived accuracy, and likelihood to continue looking for more information. Again, this provides support for the idea that the wording of the question reduced the effect of author ratings for these measures. And for the dependent measure of perceived believability of the author, the main effect of valence is smaller for the comment rating than the author rating.

The primary justification for conducting the first study was to inform the design of the second study. It was necessary to evaluate which aspects of credibility and reputation impact responses on questions involving overall trust of online information in order to aid in the design of the independent variables of the second study. However, while the first study measures people's perceptions of information on a simulated web

forum, it does not include a true behavioral measure. Therefore, it is necessary to conduct a study that requires a behavioral response in addition to an assessment of the perceived trust in the information and authors of information online.

6. EXPERIMENT 2: ANALYSIS OF REPUTATION RATINGS VIA A NON-COMPENSATORY DECISION MAKING TASK

The goal of Experiment 2 was to assess how reputation and credibility information in a web forum affects decision making. The findings of Experiment 1 were used to choose the indicators that were rated as providing the strongest reputation or credibility information. Each of the manipulations made in Experiment 1 showed significant main effects in overall trust, with background ratings showing weaker effects overall compared to reputation and credibility. However, background information was used in Experiment 2 regardless of this finding. Eysenbach and Kohler (2002) found that participants frequently used authority of the source as a method of assessing a website's credentials. Similarly, Yaniv and Kleinberger (2002) found that advice-taking occurred more frequently when the information agreed with the participant's own knowledge and the level of perceived expertise of the advisor was higher. These two studies provide a theoretical justification for including information regarding the background knowledge of the authors. However, pilot testing revealed that participants were unsure of the term "veterinarian" as a description of background knowledge. This term may have been too broad compared to a specific and relevant pet owner, therefore the background knowledge level of "veterinarian" was not included in Experiment 2.

Participants used web forum information to make a multi-attribute non-compensatory decision in a healthcare context. This kind of decision is representative of everyday healthcare decision making. When patients are seeking health information online in order to help them make a decision often there is not a clear cut, singular

decision alternative. In these more difficult compensatory decisions, the patient is presented with multiple alternatives and therefore must weigh the costs and benefits of each alternative across the various attributes used to evaluate the alternatives. In other cases, one alternative may be a better option to take, because it is evaluated highly on all the attributes. In this case, this dominating option is chosen without need for controlled, deliberate decision-making techniques. This easier decision, which is called a non-compensatory decision, was used in Experiment 2.

The justification for using non-compensatory decisions in this study is that if one decision alternative is dominating in all attributes, but is associated with low reputation and credibility, it could affect how people make their decisions. If people choose not to opt for the dominating alternative when that alternative is associated with low reputation and credibility, then an effect of reputation and credibility can be inferred. A key question addressed in this study is: When consumers use a health forum to make a non-compensatory decision, will they use or ignore reputation information and perceived credibility in making their decision?

Experiment 2 included a non-compensatory decision making task where three decision alternatives were presented, with each alternative evaluated on two attributes. The attribute outcomes for the two alternatives will be specified so that, without reputation and credibility information, it will be simple to determine which alternative is the best choice, that is, the dominating alternative will be clear. For example, Figure 6 shows a decision where, before reputation or credibility information is considered,

Gentamycin is the best alternative in terms of infection rate and side effects, so it is the dominating alternative.

However, if participants take into account the reputation and credibility information available regarding each outcome, they may find the decision on whether to trust the dominating alternative more difficult. For example, in Figure 7 the dominating alternative, Gentamycin, has the most positive outcomes regarding chance of infection rate and side effects but both of these attributes are coupled with low credibility and reputation ratings. Therefore if participants choose Cefradoxil, which is the alternative with the second highest expected utility but is the only alternative associated with high credibility and reputation, then their decision may have been influenced by credibility and reputation information.

Question: My hedgehog has been losing quills and seems to have some blood under the quills that she still has. Has anyone else had this problem, and is there any kind of treatment you would recommend?			
Attribute: Chance of infection (higher is better)	Use an antibiotic like **Gentamycin**. It **reduces chance for infection by 90%**	Use an antibiotic like **Cefradroxil**. **It reduces chance for infection by 70%**	Use an antibiotic like **Kanamycin**. **It reduces chance for infection by 50%**
Attribute: Level of side effects (lower is better)	Use an antibiotic like **Gentamycin**. There are **virtually no** side effects associated with it.	Use an antibiotic like **Cefradroxil**. There are **very few** side effects associated with it.	Use an antibiotic like **Kanamycin**. There are **moderate** side effects associated with it.

Figure 6: Non-compensatory decision-making where the dominating alternative has high utility based on two attributes.

Most of the decisions participants completed in Experiment 2 will be similar to the decision in Figure 7 in that the dominating alternative has low credibility and reputation. In these circumstances, either the second highest utility option or the third highest utility option will be paired with high credibility. For a few of the decisions participants completed, the dominating alternative will have high credibility and reputation. Comparing participants' decisions on these two types of problems will allow us to see how strongly credibility and reputation influence their decisions.

Question: My hedgehog has been losing quills and seems to have some blood under the quills that she still has. Has anyone else had this problem, and is there any kind of treatment you would recommend?			
Attribute: Chance of infection (higher is better)	Use an antibiotic like **Gentamycin**. It **reduces chance for infection by 90%** Background: Dog owner Reputation: Low	Use an antibiotic like **Cefradroxil**. **It reduces chance for infection by 70%** Background: Hedgehog owner Reputation: High	Use an antibiotic like **Kanamycin**. **It reduces chance for infection by 50%** Background: Cat owner Reputation: Low
Attribute: Level of side effects (lower is better)	Use an antibiotic like Gentamycin. There are **virtually no** side effects associated with it. Background: Dog owner Reputation: Low	Use an antibiotic like Cefradroxil. There are **very few** side effects associated with it. Background: Hedgehog owner Reputation: High	Use an antibiotic like Kanamycin. There are **moderate** side effects associated with it. Background: Dog owner Reputation: Low

Figure 7: Non-compensatory decision-making task of experiment 2 presented in a matrix format where the alternative with the second highest overall utility has high credibility and reputation.

In Experiment 2, each factor of reputation mechanism and credibility were not varied separately. Instead, both attributes for a decision alternative (chance of infection and side effects) always had either high credibility and reputation or low credibility and reputation. Also, in each problem, there was always one alternative with high credibility and reputation and two with low credibility and reputation, as in Figure 7.

Pilot studies showed that participants almost always chose the decision alternative with high credibility and reputation, even when this alternative had the lowest utility. To avoid a ceiling effect when using credibility and reputation, another manipulation was included in the design of Experiment 2. For the majority of the problems, participants were asked to make a decision when the difference between reputation was large (e.g., 5 stars vs. 2 stars). However, some problems were added where the reputation difference was smaller (e.g., 5 stars vs. 3 stars). This difference was balanced so that the large and small reputation difference occurred when compared with each decision alternative (i.e., the highest, middle, and lowest utility decision alternative). This manipulation of large and small reputation differences was not implemented for credibility, because the levels of expertise did not have enough precision to allow large and small differences between credibility levels.

Thus, Experiment 2 used a factorial design with two independent variables: whether the decision alternative with high credibility and reputation was associated with the highest, middle or lowest utility alternative; and whether the difference between high and low reputation alternatives was large or small. After making a choice of which decision option they would prefer, participants answered four questions assessing the perceived credibility and trustworthiness of the comments and comment authors for decision alternative. In addition, participants answered a fifth question to assess their level of confidence in the decision they made regarding the selection of treatment.

It is my hypothesis that participants would choose the treatment alternative that is associated with high credibility and reputation even when this alternative has lower expected utility from other alternatives. When the high credibility and reputation option also had the highest utility, I expected participants to always choose this option; since credibility, reputation and utility all favor this option. However, for problems like Figure 7 where credibility and reputation conflict with utility, participants were expected to frequently choose the alternative with the second highest expected utility (i.e., Cefradroxil) if they are considering credibility and reputation as well as utility. However, I also predict that the frequency with which participants choose the alternative with high credibility and reputation will decrease as it is paired with decreasing levels of utility. In other words, credibility should be used less often as the conflict between credibility and utility is greater.

Regarding the independent variable of reputation difference, I hypothesize that reputation would influence choice more when the reputation difference is high than when it is low. Thus, when high credibility and reputation was paired with the second- or third-best utility, a high reputation difference should lead to more frequent choice of these options than a low reputation difference. When high credibility and reputation matches maximum utility, reputation difference should not matter since people are expected to always choose the high credibility option. Therefore, I predict an interaction of the independent variables of reputation difference and the utility of the alternative with high credibility and reputation, such that as the utility decreases, the frequency of selecting the

high credibility and reputation alternative decreases more strongly with a small reputation difference than a large one.

Regarding the confidence and trust ratings, it is my prediction that participants will feel more confident in their decisions and they will rate higher levels of perceived credibility and trustworthiness when they do not choose the alternative with dominating utility and instead make a choice based on the overall high reputation and credibility of the comment author.

7. METHODS

Participants

Fifty-five adults (27 male and 28 female) participated in this study. The scenarios used in this study involved questions posted to a web forum on the topic of pet health, using the same wording as the first study. Participants were therefore excluded if they had advanced knowledge of illnesses common in household pets. Participants were undergraduates recruited from Clemson University via the internal human participant pool.

Design

In Experiment 2, participants completed fourteen decisions. One independent variable is the association of credibility and reputation with the decision options: high reputation associated with the dominating (highest utility) alternative vs. the second best vs. the third best alternative). For three decisions, the higher credibility and reputation scores were associated with the dominating alternative. For the other eleven decisions, the higher credibility and reputation scores were associated with the other two alternatives that have a lower overall expected utility (see Table 2 for full design). Figure 7 showed an example of a forum post in which the option with the second highest overall expected utility was associated with high credibility and reputation scores. The second independent variable is the difference in credibility (large vs. small). For the credibility ratings, either the difference will be large (e.g., 5 star rating vs. 2 star rating) or small (e.g., 5 star rating vs. 3 star rating). This independent variable will be manipulated across five of the fourteen questions, balanced across each level of the first independent

variable. See Table 2 for reference regarding the location of the levels of this independent variable. Background knowledge was not manipulated in this manner, as there were only two levels of background knowledge used in the second study (relevant pet owner and unrelated pet owner).

The fourteen questions of each type (i.e., 3 when high reputation is with the dominating alternative, and 11 when high reputation is with the second or third best decision alternative) were shown in random order. Additionally, each question was designed such that order in which the utility levels of each option, the levels of credibility difference, and the location of the decision option the highest credibility was associated with, were mixed. This was done to ensure that any preference for decisions based on the order in which they were presented was equal across all combinations of variables. Table 2 shows the specifics of how pairing will occur for high credibility and reputation with the different levels of utility, and occurred as follows: High credibility was paired with the dominating alternative three times, high credibility was paired with the second highest utility option six times, and high credibility was paired with the third best utility option five times. Additionally, there were four problems that had small differences from high to low credibility, and ten problems that had large differences. The four small differences in credibility were divided as follows: One problem was included in the dominating decision alternative, two problems were included in the second best utility option, and one problem was included in the third best utility option. The ten large differences in credibility were divided as follows: Two problems were included in the dominating decision alternative, four problems were included in the second best utility

option, and three problems were included in the third best utility option. Participants were presented with all combinations of levels of pairing high credibility with the maximum, middle, and lowest utility option and each level of the amount of credibility difference.

Table 2: Design for experiment 2 decision-making problems. Each column represents the level of expected utility, each cell indicates the location of high credibility and reputation and which utility it is associated with, and each row is the question number. * Questions that have a small difference between high and low credibility.

	Overall expected utility of decision alternative		
	Best utility	Second best utility	Third best utility
Question 1	**High Credibility**	Low Credibility	Low Credibility
Question 2	Low Credibility	Low Credibility	**High Credibility**
Question 3*	Low Credibility	Low Credibility	**High Credibility**
Question 4	Low Credibility	**High Credibility**	Low Credibility
Question 5	**High Credibility**	Low Credibility	Low Credibility
Question 6	Low Credibility	**High Credibility**	Low Credibility
Question 7	Low Credibility	**High Credibility**	Low Credibility
Question 8*	Low Credibility	**High Credibility**	Low Credibility
Question 9	Low Credibility	**High Credibility**	Low Credibility
Question 10	Low Credibility	Low Credibility	**High Credibility**
Question 11*	**High Credibility**	Low Credibility	Low Credibility
Question 12	Low Credibility	Low Credibility	**High Credibility**
Question 13*	Low Credibility	**High Credibility**	Low Credibility
Question 14*	Low Credibility	Low Credibility	**High Credibility**

After being presented with the decision question and the six comments from the web in the matrix format of Figures 6 and 7, the participants chose which of the three decision alternatives is the best. After selecting the alternative, the participant answered these five questions regarding the trust in the information:

1. Regarding the treatment you selected, how trustworthy is the answer to the medical question asked? (Using a seven-point Likert scale ranging from 'very untrustworthy' to 'very trustworthy')
2. Regarding the treatment you selected, how likely would you be to continue looking for other potential answers after reading your selected answer? (Seven-point Likert scale ranging from 'very unlikely' to 'very likely')
3. How believable are the people who wrote the answers you chose as the best answer? (Seven-point Likert scale ranging from 'not a believable source' to 'a very believable source')
4. How accurate did you find the information from the treatment you chose? (Seven-point Likert scale ranging from 'not at all accurate' to 'very accurate')
5. How confident are you in your choice? (Seven-point Likert scale ranging from 'not at all confident' to 'very confident')

Materials

Equipment

A mock-up of a web forum was created using Microsoft Power Point. The information was arranged similar to Figures 6 and 7, with a question requiring a decision

at the top and six responses arranged in a matrix format. Participants were not able to interact with the forum posting in the form of clicking on links or manipulating scores. The experiment was designed using Survey Monkey web survey creation software.

Questionnaires

Prior to beginning the experiment, participants completed the same demographic information from Experiment 1 (Appendix A). In addition to providing biographical information, participants answered questions regarding their level of education and experience with pets. Again, scores of high levels of knowledge with pet problems were used as an exclusion criterion, as participants may have based their decisions of the quality of the information on aspects other than reputation mechanisms.

A 20-item inventory measuring tolerance for ambiguous information and domain-specific risk-taking was given at the conclusion of the study. This inventory was adapted from the Domain-Specific Risk Taking (DOSPERT) Inventory (Weber, Blais, & Betz, 2002) and the Need for Closure Scale-Short Version (Roets & Van Hiel, 2011). The purpose of this inventory was to determine if some participants are more willing to gamble on medicines or treatments, or if the participants were likely to choose options and respond with a neutral tendency due to being uncomfortable working with incomplete or lacking information. The inventory is included in Appendix D.

Procedure

An experimental session lasted roughly 30 minutes. The participant first read and signed consent forms. The initial instructions informed the participant that the study is meant to examine how people make decisions regarding medical information online. The experimenter then provided the participant with a brief training on the structure and description of the mock forum posts in the experiment. This included discussion of the difference between the study's matrix format presentation of forum responses from actual web forum presentation of information, an overview of what type of information is available in a response (i.e., information regarding the author, comment, and an explanation of the comment showing different attributes of a given option), and a sample problem using a scenario similar to an e-commerce forum posting. The training did not direct participants' attention to specific elements of credibility or reputation mechanisms explicitly, as in the first experiment. The focus of the second experiment was to determine if participants were aware of this information and whether they use it to influence their decisions. Therefore, only a general explanation was given about the structure of a forum post for the second study. This portion of training used visual aids presented gradually as to show each aspect of a forum post incrementally. Then the participant completed the demographics questionnaire.

On each problem during the experimental session, participants read the information in the six forum comments and select one of the three treatments. The treatment options, as well as the five questions regarding the choice, were presented on the same screen so that participants could refer to the problem easily without having to

scroll or navigate between more than one window. After finishing all the decision problems, each participant was given the post-test inventory and were debriefed on the study.

8. RESULTS

Frequency of Choosing Decision Alternatives

The main manipulation in this study was that the decision alternative with high credibility and reputation was paired with either the maximum, middle or lowest utility alternative. The other two alternatives always had lower credibility and reputation. Since credibility and reputation were always either both high or both low, in presenting the results we will use the term credibility to stand for credibility and reputation. The main hypothesis of this study was that participants would choose the high credibility alternative more often when this option had higher utility and less often when it had lower utility. We expected 100% choice of the high credibility alternative when it matched maximum utility, but as the high credibility option was associated with lower utility (second best, then third best), we expected the frequency of choosing the high credibility alternative to decrease.

The other manipulation was that the credibility difference between the single high credibility option and the two low credibility options in a decision was sometimes large (e.g., 5 vs. 1 stars) and sometimes small (e.g., 4 vs. 2 stars). I expected that credibility would influence choice more when this credibility difference is high than when it was low Thus, when high credibility was paired with the second- or third-best utility, a high credibility difference should lead to more frequent choice of these options than a low credibility difference. Finally, I hypothesized an interaction such that as the high utility option is paired with maximum, second best, and then third best utility, the frequency of

choosing the high credibility option should decrease less sharply with a high credibility difference than with a low credibility difference.

These hypotheses both suggest that participants will use credibility as a decision-making aid, especially when high credibility is paired with higher utility or the difference between high and low credibility options is large. All effect sizes were calculated using Cohen's *f* (Cohen, 1992).

Further, I hypothesized that as high credibility is paired with lower utility, the level of confidence the participant has in that decision will decrease. In other words, as credibility and utility become more discrepant, confidence will decrease. Also, I hypothesize that participants will feel more confident in their decision when the difference between the high and low credibility ratings are large compared to when the difference between credibility ratings is small. The construct of confidence in a decision was measured using responses to a question about the participant's perceived confidence in their selection of a treatment. The responses were measured on a seven-point scale with high scores indicating more confidence and lower scores indicating less confidence.

Multilevel models were run using generalized estimated equations (GEE) for dependent measures of frequency of choosing a decision option and using the SPSS mixed model procedure for dependent measures of Likert-scale responses on the question of confidence in a decision. These models used both of the independent variables (the association of high credibility with the level of utility in an answer and the credibility difference between responses) as predictors.

Figure 8 shows the frequency of choosing the option associated with the highest credibility ratings as a function of the association of high credibility with differing levels of utility. The figure suggests that when the high credibility option was associated with the maximum utility, participants always chose the high credibility option, as expected. Then, as high credibility was associated with lower utility options, the frequency of choosing those options decreases. Also, participants seemed to use credibility and reputation more when the difference between high and low reputation alternatives was small than when it was large. Finally, the decrease in using credibility and reputation as high credibility is paired with lower utility seems larger with a low credibility difference than a high one, as expected. A 2 by 3 multilevel model would not run using the two cells of this design where high credibility matched maximum utility because these cells had 0 variability. Thus, a 2 by 2 multilevel model was run using only the two levels where high credibility matched medium or lowest utility. A GEE showed that the pairing of high credibility with decreasing levels of utility had a significant main effect on frequency of choosing the option with high credibility χ^2 (1) = 51.88, p = .000. Additionally, the GEE showed that the degree of difference between high and low credibility decision options had a significant main effect on the frequency of choosing the option with high credibility χ^2 (1) = 42.73, p = .000. Contrary to the proposed hypothesis, there was not a significant interaction of credibility difference and association with levels of utility of responses, χ^2 (1) = 1.95, p = .162. These findings support the primary hypothesis that high credibility will influence people to choose options with suboptimal utility, but that this tendency will decrease as the utility paired with high

credibility becomes lower. They also show that people used credibility to choose an option with suboptimal utility more frequently when the difference between high and low credibility in a decision was high than when it was low. The evidence does not support the hypothesis that there is an interaction present in the data.

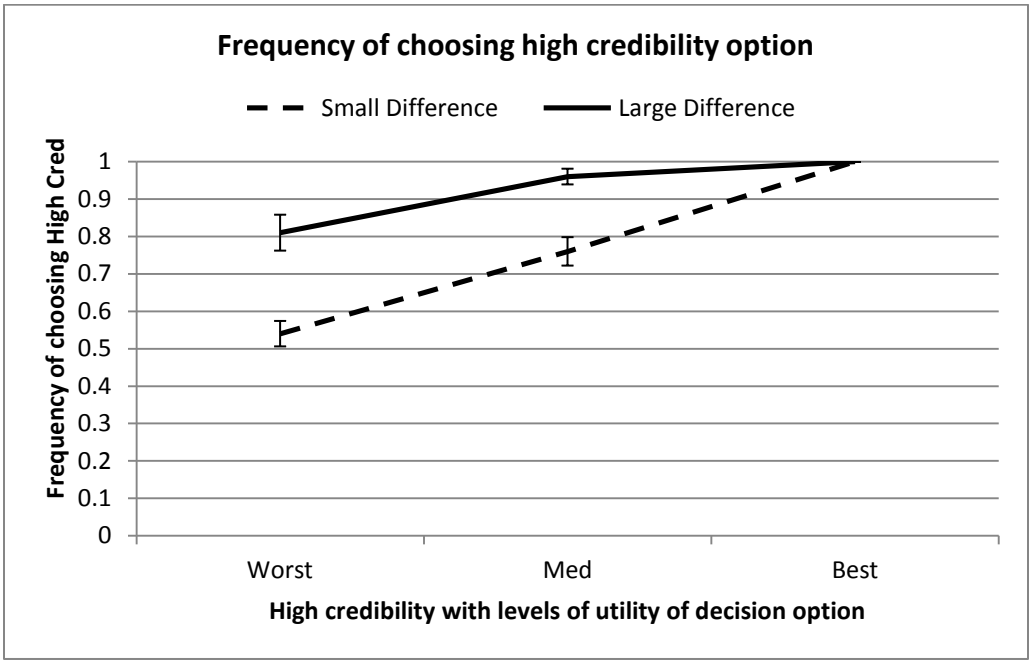


Figure 8: Frequency of selecting the option associated with the highest credibiltiy plotted as a function of the pairing of high credibiltiy with the worst, middle, and best utility decision options. The different lines represent instances where the difference between high and low credibiltiy were either large or small. Error bars represent standard error of estimate values.

Confidence and Related Variables

I hypothesized that participants would feel more confident in their decisions and they will have more trust in authors and comments when they do not choose the alternative with dominating utility and instead make a choice based on the overall high reputation and credibility of the comment author. To test this hypothesis, a main effect of

reputation score difference and whether participants chose based on high credibility or not was tested for confidence and trust ratings. Figure 10 shows the effect of reputation score difference and selection of high credibility options on ratings of confidence in participant's selected treatment. In order to calculate this main effect, it was necessary to exclude cases where participant either chose the option where high credibility was matched with the highest utility option or cases where participants chose an option that was neither the highest credibility option nor the highest utility option (i.e., choosing an option that is less desirable on all aspects). The former cases were excluded because participants could choose between credibility and utility when one decision alternative has the highest credibility and utility. The latter cases were excluded because participants seemed to ignore both credibility and utility. These cases did not happen frequently; of the 770 scenarios across 55 participants, only 9 indicated a situation where participants chose in this manner. These cases were excluded for each of the analyses of confidence and trust ratings.

Figure 9 suggests that confidence is higher when participants choose based on credibility (i.e., choose high credibility option) than when participants choose based on utility (i.e., choose low credibility option). This effect was evident both when the difference between low and high credibility was large and small. An MLM showed that selecting an option based on credibility instead of utility had a significant effect on confidence when the credibility difference was high, $F(1, 376) = 49.27$, $p = .000$, $f = 0.55$, and when it was low, $F(1, 216) = 18.28$, $p = .000$, $f = 0.29$. The choice of the high credibility option when the credibility difference was high yielded a large effect size,

compared to the moderate effect size seen when the credibility difference was smaller. The hypothesis that basing decisions on credibility instead of utility would lead to increased confidence was supported.

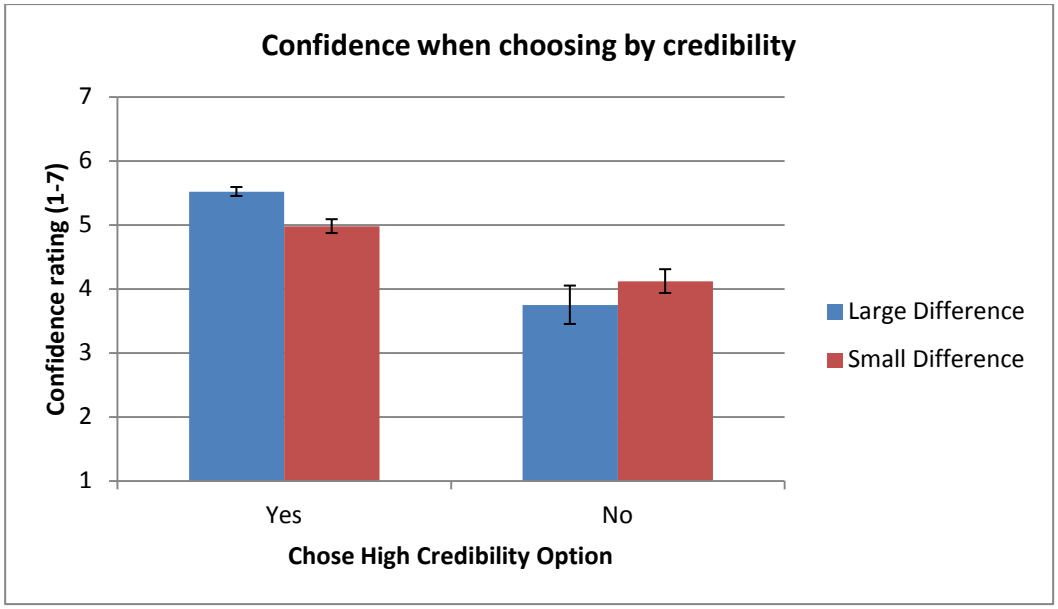


Figure 9: Confidence ratings plotted as a function of whether the participant selected the option associated with highest credibility (Yes) vs. highest utility (No). The different bars represent instances where the difference between high and low credibility scores were either large or small. Error bars represent standard error of estimate values.

Figure 10 suggests that trust is higher when participants choose based on credibility (i.e., choose high credibility option) than when participants choose based on utility (i.e., choose low credibility option). This effect was evident both when the difference between high and low credibility was large and small. An MLM showed that selecting an option based on credibility instead of utility had a significant effect on trust when the credibility difference was high, $F(1, 376) = 64.78$, $p = .000$, $f = 0.66$, and when the credibility difference was low, $F(1, 216) = 41.12$, $p = .000$, $f = 0.44$. Similar to the

findings regarding confidence, the choice of high credibility option when the credibility difference was high yielded a large effect size, compared to the moderate effect size seen when the credibility difference was smaller. The hypothesis that basing decisions on credibility instead of utility would lead to increased trust was supported.

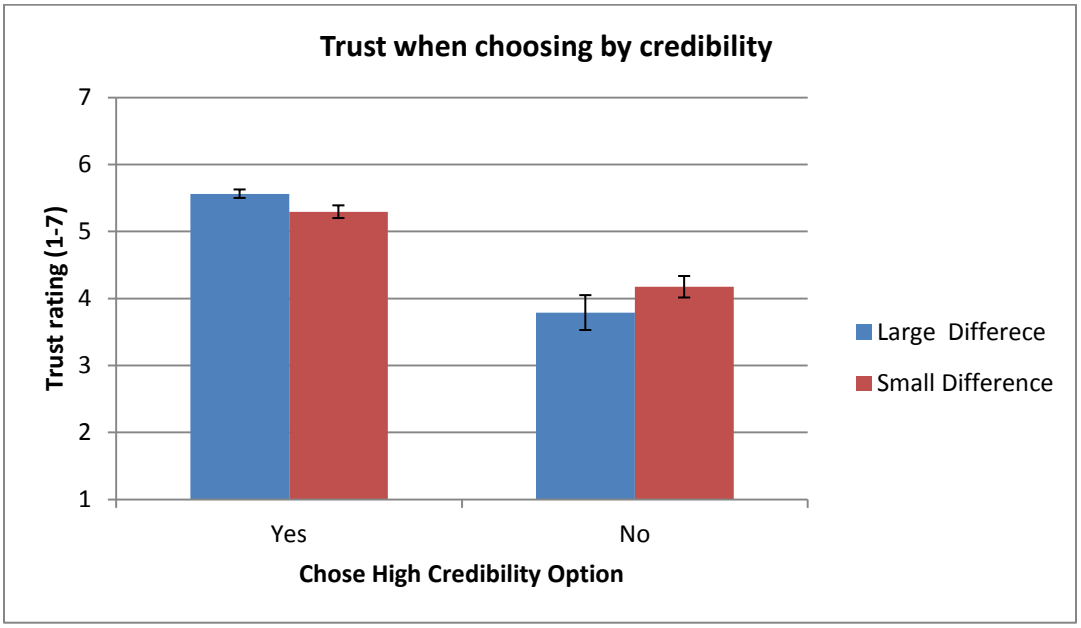


Figure 10: Trust ratings plotted as a function of whether the participant selected the option associated with highest credibility (Yes) vs. highest utility (No). The different bars represent instances where the difference between high and low credibility scores were either large or small. Error bars represent standard error of estimate values.

Figure 11 suggests that likelihood of searching for more information is lower when participants choose based on credibility (i.e., choose high credibility option) than when participants choose based on utility (i.e., choose low credibility option). This effect was evident both when the difference between low and high credibility was large and small. An MLM showed that selecting an option based on credibility instead of utility

had a significant effect on likelihood of continuing to search for more information when the credibility difference was high, $F(1, 376) = 12.39$, p = .000, $f = 0.34$, and when the credibility difference was low, $F(1, 216) = 4.70$, p = .031, $f = 0.15$. This effect was the weakest effect size found for the dependent measures of overall trust in the information. A choice of high credibility option when the credibility difference was large yielded a moderate effect size, compared to the small effect size seen when the credibility difference was small. Nevertheless, the hypothesis that basing decisions on credibility instead of utility would lead increased trust expressed as the likelihood of continuing to look for more information was supported.

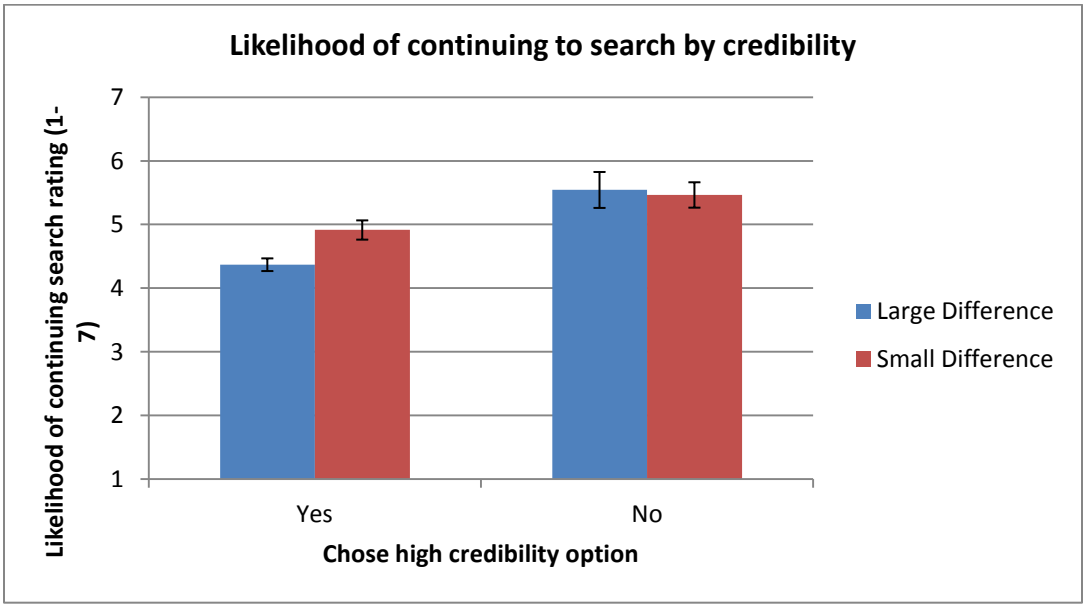


Figure 11: Likelihood of continuing to search for more answer ratings plotted as a function of whether the participant selected the option associated with highest credibility (Yes) vs. highest utility (No). The different bars represent instances where the difference between high and low credibility scores were either large or small. Error bars represent standard error of estimate values.

Figure 12 suggests that believability is higher when participants choose based on credibility (i.e., choose high credibility option) than when participants choose based on utility (i.e., choose low credibility option). This effect was evident both when the difference between high and low credibility was large and small. An MLM showed that selecting an option based on credibility instead of utility had a significant effect on believability when the credibility difference is high, $F(1, 376) = 159.84$, $p = .000$, $f = 1.13$, and when the credibiltiy difference was low, $F(1, 216) = 68.37$, $p = .000$, $f = 0.58$. This effect was the strongest effect size found for the dependent measures of overall trust in the information. The choice of high credibility option when the credibility difference was large and small yielded a large effect size. The hypothesis that basing decisions based on credibility instead of utility would lead to increased believability was supported.

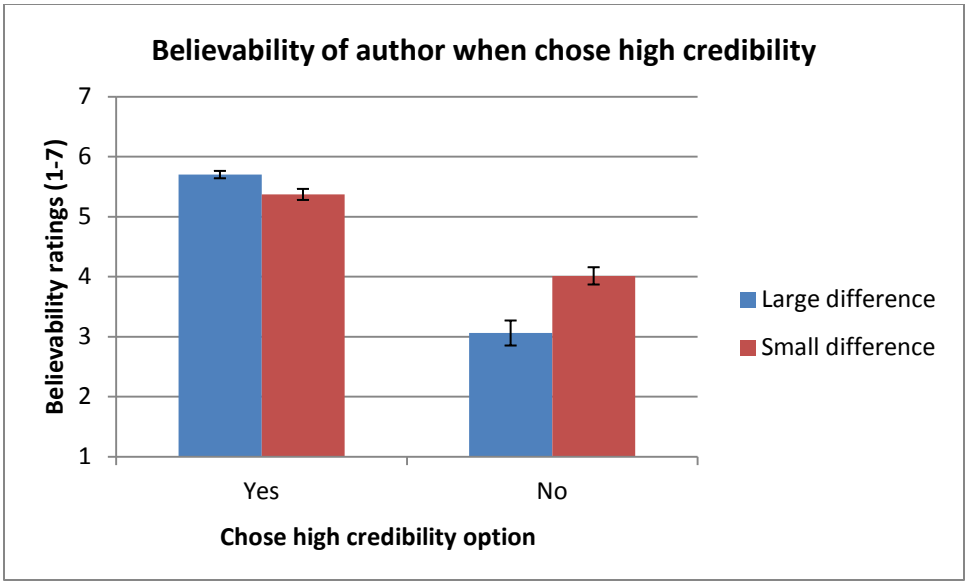


Figure 12: Believability of author ratings plotted as a function of whether the participant selected the option associated with highest credibility (Yes) vs. highest utility (No). The different bars represent instances where the difference between high and low credibility were either large or small. Error bars represent standard error of estimate values.

Figure 13 suggests that perceived accuracy is higher when participants choose based on credibility (i.e., choose high credibility option) than when participants choose based on utility (i.e., choose low credibility option). This effect was evident both when the difference between low and high credibility was large and small. An MLM showed that selecting an option based on credibility had a significant effect on perceived accuracy in the option selected when the credibility difference was high, $F(1, 376) = 86.67$, $p = .000$, $f = 0.72$, and when the credibility difference was low, $F(1, 216) = 43.56$, $p = .000$, $f = 0.43$. The choice of the high credibility option when the credibility difference was large and small yielded a large effect size. The hypothesis that basing

decisions on credibility instead of utility would lead to increased perceived accuracy was supported.

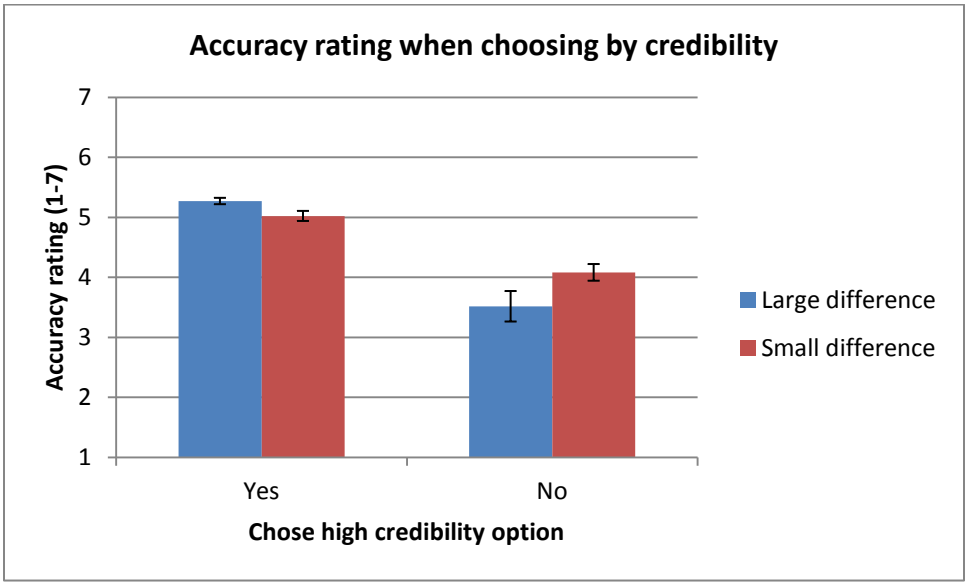


Figure 13: Perceived accuracy of comments plotted as a function of whether the participant selected the option associated with highest credibility (Yes) vs. highest utility (No). The different bars represent instances where the difference between high and low credibiltiy were either large or small. Error bars represent standard error of estimate values.

An unhypothesized result is shown in Figure 14, which shows the effects of pairing high credibility with high vs. low utility and the degree of difference between high and low credibility on the rating of confidence the participant has in his or her choice. The figure suggests that as high credibility is associated with lower utility, the confidence in the participants' choice decreases both when the difference between low and high credibility is large and small. An MLM showed that pairing of high credibility with different levels of utility had a significant effect on the confidence in the option selected, $F(2, 710) = 111.073$, $p = .000$, $f = 0.46$. Additionally, the MLM showed that

whether the difference between high and low credibility was large and small had a significant effect on the confidence in the option selected, $F(1, 710) = 15.79$, $p = .000$, $f = 0.15$. The MLM also showed a significant interaction between high credibility associated with levels of utility and the credibility difference, $F(2, 710) = 3.80$, $p = .023$. The high credibility associated with different levels of utility had a large effect size, compared to the small effect size seen for high and low credibility differences. The hypothesis regarding the main effect of both high credibility association with overall expected utility, and the degree of credibility difference were supported. However, the interaction was not predicted.

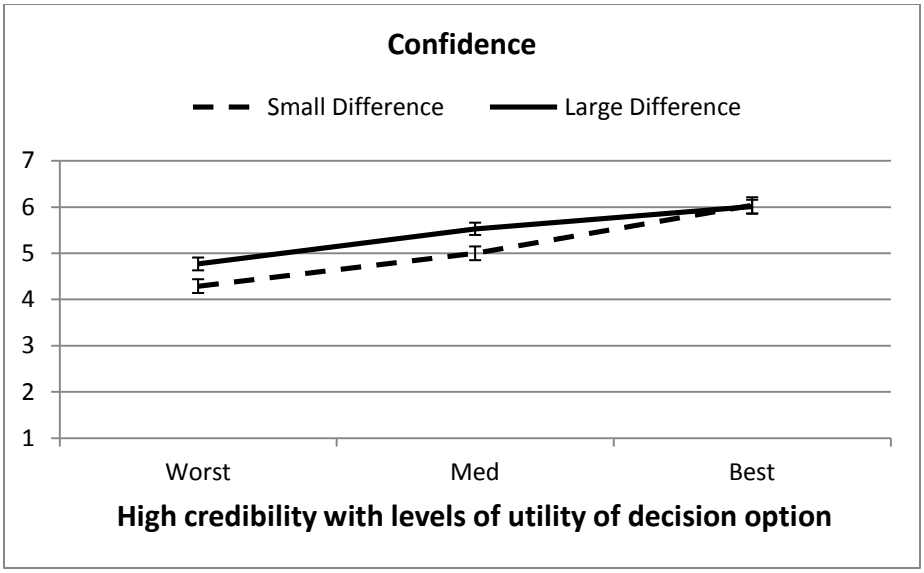


Figure 14: Ratings of confidence in option selected plotted as a function of the pairing of high credibiltiy with the worst, middle, and best utility of the decision option. The different lines represent instances where the difference between high and low credibiltiy were either large or small. Error bars represent standard error of estimate values.

9. DISCUSSION

The purpose of the second experiment was to build on the first experiment to see if behaviors of participants reflect their self-report information regarding their perception of the overall trustworthiness of credibility and reputation of comments and authors on a web forum. The incorporation of a non-compensatory decision-making task allowed for the collection of behavioral data regarding decisions on what particular treatment would be preferred and which authors and comments participants take under advisement when making their decision. This method was important in creating a simulation of a real-world situation in which different treatments are recommended by different sources, and also allowed for the inclusion of credibility and reputation manipulations to act as implicit weights for each available decision alternative. Initial analyses focused on the main effect of high credibility and reputation paired with different levels of overall expected utility of alternatives, and indicated that a significant effect of this pairing on the frequency of choosing the alternative associated with high credibility and reputation. Additionally, analyses indicated that when the difference between reputation scores was large, participants much more frequently chose the alternative with the highest credibility and reputation support than when the difference between scores was small.

For the primary hypothesis, significant effects were found in that high credibility and reputation associated with the different levels of overall expected utility significantly increased the frequency of selecting an option based on reputation information. In other words, participants were much more likely to choose an alternative that was associated with high reputation if that alternative was also higher in terms of overall expected utility.

Also related to the primary hypothesis, participants were much more likely to make a decision that agreed with high reputation and credibility when the high reputation scores were much larger than the low reputation scores. When the reputation scores were closer together, participants much less frequently chose the options where high credibility was associated with the second or third best decision alternative. These findings primarily suggest that participants are not only aware of reputation and credibility information on web forums, but are also affected by this information such that they are more likely to take the advice of a comment or comment author if they are reputable within the online community.

For the secondary hypothesis, significant main effects were found supporting the prediction that when participants choose an answer with high credibility and reputation scores, they are more likely to feel confident in their answer. Furthermore, when the high reputation score answer is associated with a larger difference from the low reputation score answer, participants are more likely to feel confidence in their answer. This suggests a main effect of both pairing high reputation and credibility with higher levels of expected utility and of the manipulation of the degree of difference in reputation scores. An unexpected interaction was observed, the nature of which seems to indicate that the increase in confidence is stronger when the difference between reputation ratings is large compared to when the difference is small. This is similar to the trend observed in the primary hypothesis, although no significant interaction was detected in this manner for the primary hypothesis.

10. GENERAL DISCUSSION

The experiments conducted in this study were designed to assess whether inclusion of reputation and credibility scores in an online health forum would be perceived and subsequently used in a decision-making task. The first study focused on determining which common aspects of forum credibility and reputation mechanisms affected overall trust in information, and which aspects didn't. The first study also applied concepts of decision-making theory to the way people interpret reputation of a source using the concepts of valence and overall amount of support. It was evident upon analysis of the first study that each of the reputation mechanisms strongly affected the degree to which participants trust information and authors of that information, with more of an effect observed regarding online reputation mechanisms compared to self-reported background knowledge or credentials. As shown in previous studies, participants and patients are likely to report that a general information websites are more trustworthy when the content of the site is maintained by a professional source and that there is evidence of quality control via information updated by experts in the field (Eysenbach & Kohler, 2002; Pirolli, Wollny, & Suh, 2009; Sillence, Briggs, Harris, & Fishwick, 2007). Even if people respond that this degree of credibility and reputation information is important, they still may not actually investigate a website or source's credentials when gathering information. The first study was important in identifying which aspects of credibility and reputation influence trust, but was not sufficient for determining if participants would actually attend to and use this information in decision making.

The second study addressed issues addressed by researchers such as Eysenbach and Kohler (2002) by providing a decision-making task in which participants may or may not use reputation and credibility information in determining which decision alternative is the most preferred. If participants state that reputation and credibility information is important, but do not use it, we should have observed behaviors that would indicate that participants most frequently chose alternatives with the highest expected utility regardless of credibility and reputation information. Instead, we observed that high reputation and credibility greatly influenced participants' decisions regarding which treatment they would accept as being their preferred choice. These findings are compelling, and suggest that people are not only aware of this information, but use it as a tool for making decisions regarding health-related information.

It is possible that participants only used credibility and reputation information because it was readily available and highly salient to the participant. In the Eysenbach and Kohler (2002) mentioned previously, participants used actual websites to gather information regarding health topics. On these websites, any credibility or expertise information was not immediately available to the participants. Rather, this information had to be specifically searched for on different pages of a site. This salience of the credibility and reputation information may have increased the likelihood of participants using this information for decision making, and if the credibility and reputation information was separated from the comments themselves this effect may have been less pronounced. However, in this study the design of the forum posting was meant to represent the way in which this information is often presented on community-run forums.

Reputation within a community is managed by forum users; therefore this information is often readily available and viewable for each comment and author. Future research may benefit from analyzing the effect of reputation and credibility of sources in a non-compensatory decision-making task when this information is not immediately available or noticeable to a participant.

Demand characteristics could have played a factor into some of the decision-making strategies of participants as well. Although efforts were made to make the presentation of the manipulation discreet by varying amount of support and valence within the high and low levels of the manipulation, participants may have still detected that this study was examining high credibility and reputation and its effects on decision-making. If participants perceived a strong demand to primarily use credibility and reputation in making their decisions, it would be expected that the responses based on credibility would be inflated such that participants would choose the option with high credibility much more frequently than choosing based solely on utility of the alternatives. However, as Figure 8 shows, the decrease in selecting based on high credibility when the high credibility is associated with progressively lower utility would not be as pronounced if this demand characteristic were truly dominating participants' decision making strategy. Furthermore, it would be expected that the difference between large and small reputation scores would not be significant either. Both of these effects were detected, therefore the argument for strong effects of demand characteristics is more difficult to make. Also, Figure 14 shows that as the high credibility and reputation option is associated with progressively lower utility, participants become less confident in their

choices. This suggests that participants are using not just credibility and reputation but also utility in making their decisions.

According to responses from participants in the study conducted by Sillence, Briggs, Harris, and Fishwick (2007), people often rely on the content of information regarding health topics rather than the credibility or expertise of the source. In the first and second experiment, the content of each message was the same in terms of grammatical structure from comment to comment. However, Sillence et al. found that this had an impact on people's perception of the credibility of the information. If the websites used more professional language (communicated clearly), participants were more likely to trust that information. Further research could also be done to evaluate the content of an answer in addition to the credibility of the author and the reputation scores received from the online community.

Finally, responses from pilot subjects at the beginning of the second experiment suggested that a primary strategy for choosing an answer regarding a health topic online was to find information from one website and compare that information to other websites, particularly if the first source was a community-run forum. This commonly cited strategy would explain the consistently high responses on the question regarding the likelihood of participants to continue looking for information regarding the answer on other websites, and may have affected participants' judgments of the accuracy of the information (e.g., if they can't compare the information to other sources, how can they be sure if the information is accurate?). John Flavell (1979) discussed the concept of double-checking

information as an important metacognitive strategy for monitoring the accumulation of information from a source or sources. While the second experimental allowed for participants to compare information across two different comment authors regarding a single decision alternative, they were not able to check the information with other sources on the web. Despite the fact that the study had this limitation, the process of searching for information regarding a specific health topic must have an initial starting point. The second experiment makes no claim that a participant would make their final decision based on one source of information, rather this study provides evidence suggesting that the decision for the best possible treatment may be an important initial step in gathering information regarding a particular question.

REFERENCES

Billman, D., Schiano, D., & Gugerty L. (2008). Patient problem-solving on the web: How do patients use web forums to cope with chronic disease? September 2008. Medicine 2.0 Toronto, Canada

Coulson, N. S., Buchanan, H., Aubeeluck A. (2007) Social support in cyberspace: A content analysis of communication within a Huntington's disease online support group. *Patient Education & Counseling,* 68(2), 173-178

Crowe, E., Higgens, E. T. (1997) Regulatory focus and strategic inclinations: Promotion and prevention in decision making. *Organizational Behavior and Human DecisionProcesses,* 69(2), 117-132

Dellarocas, C. (2003) The digitization of word-of-mouth: Promise and challenges of online reputation systems. *Management Science,* 1407-1424.

Eastin, M. S. (2001) Credibility assessments of online health information: The effects of source expertise and knowledge of content. *Journal of Computer Mediated Communication,* 6(4), Retrieved Feb. 29 2012 from: http://jcmc.indiana.edu/vol6/issue4/eastin.html

Eysenbach, G., Kohler C. (2002) How do consumers search for an appraise health related information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews, *BMJ,* 324, 573-577.

Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., Stern, A. (2004) Health related virtual communities and electronic support groups: Systematic review of the effects of online peer to peer interactions. *British Medical Journal,* 328, 1-6.

Eysenbach, G., Thomson, M. (2007) The FA4CT Algorithm: A new model and tool for consumers to assess and filter health information on the internet. *MedInfo 2007,* 129, 142-146

Ferguson, T. (2007) E-patients: How they can help us heal health care. Retrieved Aug. 15, 2010 from http://e-patients.net/e-Patients_White_Paper.pdf

Flavell, J.H. (1979) Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist,* 34(10), 906-911

Flanagin, A. J., Metzger, M. J. (2007) The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media Society,* 9(2), 319-342.

Hong, T. (2005) The influence of structural and message features on web site credibility. *Journal of the American Society for Information Science and Technology,* 57(1), 114-127.

Hunink, M., Glasziou, P., Siegel, J., et al. (2001) *Decision making in health and medicine: Integrating evidence and values.* Cambridge: Cambridge University Press.

Lerner, J., Dachner, K. (2000) Beyond Valence: Toward a Model of Emotion-Specific Influences on Judgment and Choice, *Cognition and Emotion,* 14(4), 473-493

Pavlou, P. A., Dimoka, A. (2006) The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research,* 17(4), 392-414.

Pirolli, P., Wollny, E., Suh, B. (2009) So you know you're getting the best possible information: A tool that increases Wikipedia credibility. *Proceedings of the 27th international conference on human factors in computing systems.* (pp. 246-247). Boston, MA: CHI 2009.

Pew Research Center (2009) 61% of American adults look online for health information. Retrieved Dec. 12, 2011 from http://www.pewinternet.org/Press-Releases/2009The-Social-Life-of-Health-Information.aspx

Pew Research Center (2010) What internet users do online. Retrieved from http://www.pewinternet.org/Static-Pages/Trend-Data/Online-Activities-Total.aspx]

Roets, A., Van Hiel, A. (2011) Item selection and validation of a brief, 15-item version of the need for closure scale. *Personality and Individual Differences, 50,* 90-94.

Sillence, E., Briggs, P., Harris, P. R., Fishwick, L. (2007) How do patients evaluate and make use of online health information? *Social Science & Medicine*, 64, 1853-1862.

Weber, E.U., Blais, A.R., Betz, E. (2002) A Domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15,* 263-290.

Yaniv, I., Kleinberger, E. (2000) Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes, 83(2),* 260-281.

Zou, P., Yu, B., Hao, Y. (2011) Does the valence of online consumer reviews matter for consumer decision making? The moderating role of consumer expertise. *Journal of computers*, 6(3) 484-489

APPENDICES

Appendix A
Demographics:

1.) Male _____ Female _____

2.) Age _____

3.) Pets owned: _____

a. List types of pets owned _____

4.) How informed are you on the topic of pet health? (circle one)

1_____2_____3_____4_____5_____6_____7
Uninformed **Expert on topic**

Appendix B
Overall Trust Questionnaire for Experiment 1

1. How trustworthy is the answer to the medical question asked?

   1_____2_____3_____4_____5_____6_____7
   Very Untrustworthy **Neutral** **Very Trustworthy**

2. How likely would you be to continue looking for other potential answers after reading this answer?

   1_____2_____3_____4_____5_____6_____7
   Very Unlikely **Neutral** **Very Likely**

3. How believable is the commenter?

   1_____2_____3_____4_____5_____6_____7
   Not a believable source **Neutral** **Very believable source**

4. How accurate did you find the information on the page?

   1_____2_____3_____4_____5_____6_____7
   Not at all accurate **Neutral** **Very accurate**

Appendix C
Sample Scenario for Factorial Study

 Greg	My hedgehog has been losing quills and seems to have some blood under the quills that she still has. Has anyone else had this problem, and is there any kind of treatment you would recommend?
 Phil Hedgehog owner for five years User rating 5 people rated this comment as correct	Use an antibiotic like Neomycin.

Appendix D
Overall Trust Questionnaire for Experiment 2

1. How confident are you in the decision to choose the option that you chose?

   1_____2_____3_____4_____5_____6_____7
   Not at all confident Neutral Very Confident

2. Regarding the treatment you selected, how trustworthy is the answer to the medical question asked?

   1_____2_____3_____4_____5_____6_____7
   Very Untrustworthy Neutral Very Trustworthy

3. Regarding the treatment you selected, how likely would you be to continue looking for other potential answers after reading this answer?

   1_____2_____3_____4_____5_____6_____7
   Very Unlikely Neutral Very Likely

4. How believable are the people who wrote the answers you chose as the best answer?

   1_____2_____3_____4_____5_____6_____7
   Not a believable source Neutral Very believable source

5. How accurate did you find the information regarding the treatment you chose?

1_____2_____3_____4_____5_____6_____7

Not at all accurate **Neutral** **Very accurate**

Appendix E

Post-Test Questionnaire

ID: _____

For each of the following statements, please indicate the **degree** in which you agree with the statement. Provide a rating from 1 to 6, using the following scale:

1	2	3	4	5	6
Strongly Disagree					Strongly Agree

1.) _____ I'd rather know bad news than stay in a state of uncertainty.

2.) _____ I would rather make a decision quickly than sleep over it.

3.) _____ I feel uncomfortable when someone's meaning or intention is unclear to me.

4.) _____ It's annoying to listen to someone who cannot seem to make up his or her mind.

5.) _____ I don't like situations that are uncertain.

6.) _____ When I have made a decision, I feel relieved.

7.) _____ I dislike it when a person's statement could mean many different things.

8.) _____ When I am confronted with a problem, I'm dying to reach a solution very quickly.

9.) _____ I like to know what people are thinking all the time.

10.)_____ In most social conflicts, I can easily see which side is right and which is wrong.

11.)_____ When I am confused about an important issue, I feel very upset.

12.)_____ I feel uncomfortable when I don't understand the reason why an event occurred in my life.

13.)_____ I almost always feel hurried to reach a decision, even when there is no reason to do so.

14.)_____ Even if I get a lot of time to make a decision, I still feel compelled to decide quickly.

15.)_____ I would quickly become impatient and irritated if I would not find a solution to a problem immediately.

For each of the following statements, please indicate the **likelihood** of engaging in each activity. Provide a rating from 1 **to** 5, using the following scale:

1	2	3	4	5
Extremely Unlikely		Not Sure		Extremely Likely

1.) _____ Eating 'expired' food products that still 'look okay'.

2.) _____ Ignoring some persistent physical pain by not going to the doctor.

3.) _____ Taking a medical drug that has a high likelihood of negative side effects.

4.) _____ Never using sunscreen when you sunbathe.

5.) _____ Eating high cholesterol foods.