Consumer Health Information on the Web: The Relationship of Visual Design and Perceptions of Credibility

David Robins
Kent State University, Information Architecture and Knowledge Management, Kent, OH.
E-mail: drobins@kent.edu

Jason Holmes
American Greetings Interactive, Cleveland, OH.

Mary Stansbury
University of Denver, Department of Library and Information Science, Denver, CO.

Consumer health information has proliferated on the Web. However, because virtually anyone can publish this type of information on the Web, consumers cannot always rely on traditional credibility cues such as reputation of a journal. Instead, they must rely on a variety of cues, including visual presentation, to determine the veracity of information. This study is an examination of the relationship of people's visual design preferences to judgments of credibility of information on consumer health information sites. Subjects were asked to rate their preferences for visual designs of 31 health information sites after a very brief viewing. The sites were then reordered and subjects rated them according to the extent to which they thought the information on the sites was credible. Visual design judgments bore a statistically significant similarity to credibility ratings. Sites with known brands were also highly rated for both credibility and visual design. Theoretical implications are discussed.

Introduction

Web sites that provide health-related information have proliferated since the World Wide Web's inception in the 1990s. Since virtually anyone can produce and publish a Web site, consumers of health-related information must be able to effectively judge the credibility of information found on the Web. Although the problem of credibility on the Web is not unique to consumer health information, health care is a need for most people at some point in their lives, and it is critical that information provided for consumers be accurate to avoid negative consequences of bad diagnoses or advice.

In October of 2006, a Pew Charitable Trust survey found that 8 of 10 respondents use the Web for health research (Fox, 2006). Of those, 48%–56% searched on behalf of another person. Fifty-eight percent of the respondents said that the information found on a Web site had an effect on a health care decision.

In this article, we lean heavily on two particular terms: visual design and credibility. Visual design is the rendering of information, in the case of this study, on a Web site. It includes structural features such as typography, images, color, and aesthetics. Visual design includes grid systems that divide pages into different sectors into which content is placed. Visual design is not so much about *what* is communicated but rather about *how* information is communicated (White, 2002; Arnheim, 1969). For example, the final visual design of a magazine includes the results of decisions about typography that determines fonts used in the publication's title, headings, and paragraphs. It includes decisions about how many columns to use in various situations throughout the publication (i.e., grid decisions). In short, there are countless decisions that ultimately determine how the content of a magazine will be presented. This presentation has a strong bearing on the communication of the content and the ideas contained in the magazine. Books, magazines, and newspapers all have a long history of publication in which their visual designs have settled around conventions of presentation. Books, for example, have certain structural elements that visual designs account for: Page grids must account for pagination elements, book covers must be designed to convey some aspect of the

book's content, and fonts must be employed that facilitate ease of reading.

Web sites, on the other hand, do not have a long history of the development of visual design conventions. For instance, one of the main design considerations that differentiate print and Web communication is the variable sizing of screens and browser windows on the user end. Because of this problem, Web site designers must consider solutions that either fix the width of content or allow content width to vary as users resize browser windows. However, just as print designers do, Web designers face many decisions on how to present content including the ones already mentioned. The end goal of all these efforts is to produce a visual design that effectively communicates the ideas expressed in content, and does so in a manner appropriate to the content.

Credibility is used in this study to define the extent to which information is perceived to be trustworthy or believable. Credibility is a complicated construct. It is generally assumed that credibility is a perceived quality rather than a property of an object (Petty & Cacioppo, 1981; Self, 1996). Credibility perceptions are the result of multiple dimensions of assessment, including the dimensions of trustworthiness and expertise perceived in information that is communicated (Stiff, 1994; Buller & Burgoon, 1996). For the purposes of this study, we define credibility as the perception of trustworthiness, believability or expertise perceived in the stimuli shown to subjects.

Why is visual design for Web sites important? Why is visual design important in the context of consumer health information Web sites? First, Web sites can be published by anyone without editorial or peer review controls. Therefore, site visitors must be prepared to sort through a number of cues that indicate the information's veracity. Visual design has been shown to have some influence on credibility judgments (Robins & Holmes, 2008; Fogg, 2003), and serves as a user's first impression of a Web site. It is possible that if a user's first impression is a positive one—that he simply likes the look of the site—that he may be more disposed to believing the information on the site. It is also possible that someone will visit a medical information site, be turned off by its appearance, immediately leave the site, and fail to retrieve potentially useful information.

Fogg et al. (2002) stated, "We found that when people assessed a real Web site's credibility they did not use rigorous criteria" (p. 6). Furthermore, they found what they described as a mismatch between what people say and what they actually do, with regard to making credibility judgments. That is, people may say they look at rigorous evaluation criteria such as authorship or sponsoring agencies, but they are relying quite heavily on the site's look and feel.

Because of this reported mismatch, this study continues prior research by this research team in the form of the second phase of exploratory investigation into the phenomenon of the relationship between visual design and credibility, and it is being carried out in the context of an important and heavily used domain: consumer health information. In this phase of the research, we isolate the initial impression of users form of a site's visual design and the credibility of the information on the site. Because a previous study (Robins & Holmes, 2008) found that credibility judgments were formed in less than three seconds, this phase of the study limits participants' exposure to the sites to that brief period of time.

Research Questions

This aim of this research is to investigate the relationship, if any, between visual design and credibility judgments on consumer health information Web sites. In addition, we wish to investigate the relationships that exist among brand (source authority or reputation), visual design, and credibility judgments of Web-based consumer health information. This line of research in an exploratory phase and will require us to use different methodological approaches over several iterations to begin to get a better picture of the intricate relationships we are investigating.

The research in this paper is guided by two research questions.

1. What is the relationship between the visual design of a consumer health information Web site and perceptions of the credibility of information found on it?
2. Is there a relationship between brand recognition, visual design preference and credibility judgments?

Based on these questions, we analyzed literature pertaining to work related to this area.

Related Literature

Credibility Theory and Framework

In an early landmark study on credibility, Hovland and Weiss (1951) exposed subjects to messages of the same topic delivered by sources of high and low credibility. Their subject pool included people of various opinions who listened to the same message delivered by both credible and non-credible sources. Predisposition to an opinion seemed to be the determining factor of agreement more so than the speaker's credibility. Their results indicated that after an initial exposure to messages from a credible source, those predisposed to that opinion felt the opinion to be "justified" over 70% of the time. Shown the identical message, subjects not predisposed to the opinion of a non-credible source, subjects disagreed with the message approximately 67% of the time. In other words, predisposition to an opinion seemed to be the determining factor of agreement (or "justification") more so than the speaker's credibility. Interestingly, these opinions changed over a 4-week period (i.e., the percentage who agreed with the high credibility source decreased and the percentage who agreed with the non-credible source increased). Whitehead (1968) further identified factors that constitute the credibility of a source: trustworthiness, professionalism, dynamism, and objectivity. Slightly different factors emerged from a study by Berlo, Lemert, and Mertz (1969), namely, safety, qualification and dynamism.

These early studies provide a basis for the current study by identifying aspects of source credibility, primarily human sources. The present study evaluates how people assess the credibility of Web sites, what Warnick (2004) describes as an "authorless" environment.

The notion of the Web as an authorless environment has implications for the present study as we are interested in people's perception of credibility as they have limited time to view a Web site. In previous studies by the present authors, Robins and Holmes (2008) found a possible "visceral" reaction to visual design that may influence credibility judgments. The present study is an attempt to further investigate that phenomenon and to see whether a longer viewing of a site might lead to a more "cognitive" interaction beyond the influence of visual design.

Metzger, Flanagan, Eyal, Lemus, and McCann (2003) acknowledge the added problems and dimensionality associated with evaluating Web-based information in their review of credibility theory and Web credibility. One of the problems addressed by Metzger et al. is that the Web blurs the distinction between source, message, and medium, which is based on communication between a person (source) who communicates a concrete idea (message) through a newspaper, television, or oration (medium). Web sites, on the other hand, do not always contain authorship attribution. Without authorship attribution (i.e., without a specified source), the reader is left to evaluate the message by way of cues provided by the medium itself (e.g. visual design or domain name).

Credibility on the Web

Fogg et al. (2003) looked for the cues people use to determine credibility on the Web in a study with some 2500 participants. Participants were asked to view Web sites, rate them according to their perceived credibility, and then answer questions about their ratings. Among the results pertinent to the present study, when asked what most influenced credibility ratings, people overwhelmingly made statements classed as "design/look." In fact, 46.1% of all participants indicated such followed by "information design" (28.5%), and then "information focus" at 25.1%. In fact, the results from this study provided motivation for the present line of research. It seems important to find out more about why (and even *if*) "design/look" is such a major criteria for judging credibility.

In describing Web credibility, Fogg (2003) frames Web credibility through four types of credibility: presumed credibility, surface credibility, earned credibility, and reputed credibility. Presumed credibility is a notion cognitively formed by a consumer based on cues such as domain identifiers (e.g., .org, .com). Surface credibility is also a cognitive or even affective view of a site's credibility based on cues such as visual design or the presence of advertising on the site. Surface credibility is a focus of the present study. Earned credibility is that which is proven to be useful and accurate over time through direct experience. The experience could be that of others who have compiled lists of credible sites for people to use. Reputed credibility may be referred to users by

trustworthy individuals or organizations. The present study accounts for reputed credibility by using a list of sites compiled by independent organizations (based on reputation) to compare against the results of our own examinations.

Lazar, Meiselwitz, and Feng (2007) reviewed literature that examined different dimensions of Web credibility in the context of two questions. First, "how do sites make themselves to appear credible?" and "how do users sort through and determine which sites are credible?" (p. 14). The line of research represented by the present study is concerned with both of questions. Lazar et al. note a crucial consideration when addressing the first question: that in which site developers make a site have the "design/look" of credibility, but the site itself is untrustworthy in one form or another. Because, according to Fogg et al (2003), "design/look" is the predominantly reported means of determining credibility, the second question could mean that people are vulnerable to unscrupulous hucksters.

Younger people, namely, students, are also consumers of Web information and, therefore, the question of how they assess credibility is relevant to this discussion. Mattus (2007) found that students judged credibility in ways that spanned Fogg et al.'s (2003) four types of credibility. Mattus, however, notes that students should receive extensive training to hone the ability to effectively weed out untrustworthy information.

Regardless of a person's age, credibility judgments are based on an array of criteria. It has been largely assumed in past studies on credibility that people base judgments primarily on cognitive authority such as that identified by Wilson (1983). For example, Rieh (2002) and Rieh and Belkin (2000) define cognitive authority as the extent to which users think or feel that they can trust the information with which they will interact. Rieh investigated cognitive authority in conjunction with information quality, which is defined as a user's perception of how truthful or trustworthy a given unit of information is. Rieh found that, far and away, information quality and cognitive authority were the primary means by which people judged credibility. In other words, credibility, according to Rieh, is established through trusted sources and high-quality information. This is in stark contrast to what was found by Fogg et al. (2003), who found that people relied on what might be considered less substantial criteria such as look and feel. In fact, Fogg and his associates at Stanford found that even when people have unlimited time to search for those cues, overwhelmingly, the reason for credibility judgments most expressed by the over 2500 participants in their study was a category of responses they defined as "Design look" (46.1%), and the next highest category of responses was "Information structure/design" (28.5%). Only 8.8% of the respondents stated that "Identity of site sponsor" was a factor in their credibility judgment. Among other categories identified by Fogg et al. were accuracy of the information (14.3%), and name recognition and reputation (14.1%).

Cognitive authority itself is not something that can be readily measured and objectively applied as a criterion for credibility judgments. Savolainen (2007) studied a group of people and how they assessed the credibility of information

using cognitive authority. He found that cognitive authority might vary quite a bit depending on one's perspective. For example, an environmental activist might rate the trustworthiness of newspapers to be low if he or she believes that corporately owned newspapers present a view divergent from that of environmentalists.

Reputation of a site is related to information quality and cognitive authority and was studied by Toms and Taves (2004). They define reputation as the "expectation of quality" (p. 292), and they found that people use a mixture of measures to determine reputation. The present study, however, represents an attempt to address their statement, "we do not know the extent to which the 'window dressing' contributed to the evaluation" (p. 314). They describe "window dressing" as visual design. It is the presumption in the present study that visual design is more than simply window dressing and, in fact, can form a strong initial impression for a user that can taint or enhance credibility judgments.

When visual design is added to the mix of variables used to evaluate the quality and credibility of Web-based information, the situation becomes even more complicated. Perhaps that is why this relationship between design and credibility has been only sparsely explored. For example, although Wathan and Burkell (2002) offer an array of factors that influence credibility judgments of Web sites, they do not mention visual design as one.

Lavie and Tractinsky (2004) found several types of aesthetic considerations to be important to visual designers for optimal human computer interaction, including what they term classical aesthetics, expressive aesthetics, usability, pleasure, and service quality. Classical aesthetics bring clarity, order, and orientation to a resource, whereas expressive aesthetics bring original and novel effects to design. Usability refers to the extent to which a resource is easy to use and navigate. Pleasurable interaction addresses the extent to which people experience joy and satisfaction, while using software and service quality refers to the confidence people have that the resource will provide them with trustworthy information and minimal trouble. This study represents a comprehensive view of the role of aesthetics in system design. The present study focuses mainly on what Lavie and Tractinsky refer to as classical and expressive aesthetics, but probably more to the former than the latter.

Robins and Holmes (2008) compared the same content from 20 Web sites, each with a "low aesthetic" design and a "high aesthetic" design and found that subjects rated the "high aesthetic" designs to be higher in credibility. Furthermore, credibility judgments in that study were made in 3.2 seconds unless three outliers were excluded, in which case judgments were made in 2.4 seconds. Although the time to credibility judgment here may seem short, Lindgaard, Fernandes, Dudek, and Brown (2006) found that visual design preferences were established in as little as 50 milliseconds. We found what was termed an "amelioration effect" of high aesthetic designs. That is, regardless of how people judged a Web site's content as to its credibility, a good design would ameliorate any negative effects of the site's perceived credibility.

In addition, because these judgments occurred so quickly, we posited this to be a visceral reaction (Norman, 2004) as opposed to a cognitive or more thought out reaction.

Credibility of Health Information on the Web

There is a growing body of literature related to credibility of consumer health information on the Web. The Pew Internet and American Life project has been conducting surveys over past few years. The last of these (Fox, 2006, 2008) indicate that people who access online consumer health information rarely check the source or date of the information they access. Pertinent to the present study, the Pew study found the most frequently searched health topic to be "specific disease or health problem" (p. 4). Because of this, part of our study's design involves a scenario in which people imagined themselves seeking information about diabetes for a friend.

Dutta-Bergman (2004) found that lack of completeness of consumer health information could lead consumers astray and negatively impact credibility judgments. He defines health information completeness as that which "not only presents the positive effects of a particular preventive or medical behavior, but also explains the process underlying the effects and expands upon the possible side effects" (p. 256). In other words, complete information would be best at helping a patient make a health-related decision. He investigated the interplay of various types of Web interaction (motivations) with information of varying completeness and found that the completeness was a much stronger factor in credibility judgments than Web motivation. It would be of interest to see if completeness or lack thereof was coupled with an equally incomplete, or poor, visual design that might also have negatively impacted credibility judgments.

Noting that health professionals do not author much health information on the Web, Eastin (2001) found little discrimination on credibility among subjects when he varied source expertise. Other criteria, however were found to have an affect on credibility judgments of health information, namely, street addresses and links to external sources (Freeman & Spyridakis, 2003).

Eysenbach (2008) criticized much of the literature touting the extent to which the Internet is being used to search for health information, claiming that people may report using the Internet for such purposes in a survey because it is socially acceptable. He claims that "apomediaries," or sites that guide people to high quality health information sites, have replaced intermediaries.

What do these apomediaries say about health information sites? In the present study, we sought to find a list of high-quality sites for use in the data analysis. We used three such sources that provided lists of quality consumer health information: Consumer and Patient Health Information Section (CAPHIS), The Medical Library Association (MLA), and *Consumer Reports*. Although CAPHIS is a section of MLA, they produced two different, but similar lists of recommended consumer health information sites. Consumer Reports provided a third listing of recommended sites. We reviewed the

TABLE 1. Consolidation of occurrences of top-rated sites included in this study.

	Name	Domain	Occurrences	Site Traffic Rank
1	WebMD	.com	1[c]	806
2	MedicineNet	.com	1[b]	1,785
3	MayoClinic.com	.com	3[abc]	1,914
4	Med Help	.org	1[b]	3,695
5	Hardin M.D.	.edu	1[b]	5,917
6	FamilyDoctor	.org	2[ab]	9,382
7	Cleveland Clinic	.com	1[b]	22,054
8	NetWellness	.org	1[b]	44,383
9	IntelliHealth (Aetna)	.com	2[bc]	65,558
10	MedlinePlus	.gov	2[ab]	75,635
11	Healthfinder	.gov	2[ab]	137,492
12	MD Choice	.com	1[c]	397,220

[a]Medical Library Association, [b]CAPHIS, [c]Consumer Reports (2002).

list of recommended sites on each of these resources and developed a list of 12 sites that occurred on at least one of the resources (12 was the maximum number of recommended sites and was found on CAPHIS). Table 1 shows the results of this review. Site traffic ranks for each site is included to show whether the site's popularity on the Web is in sync with the recommendations of CAPHIS, MLA, and *Consumer Reports*.

Research Design

This study is designed to address the two research questions stated earlier in this report. Specifically, we asked:

- What is the relationship between the visual design of a consumer health information Web site and perceptions of the credibility of information found on it?
- Is there a relationship between brand recognition, visual design preference, and credibility judgments?

In the following sections, we provide descriptions of our participants, the stimuli used and the two phases of the research design of the project.

Participants

For this project, we sought participants over the age of 35 to observe a population that is more likely to seek consumer health information (an assumption on our part made in the absence of supporting data). We used 34 participants to achieve statistical power of >0.80.

Stimuli

Thirty-one consumer health information Web sites were chosen for this study from the results of a Google search on the terms "consumer health information." The sites were captured as screen shots, saved in JPEG format, and arranged as a slide show to be shown to participants. Table 2 provides an overview of the stimuli used for this study. Of the sites shown in Table 2, it should be noted that 71% were sponsored by United States-based organizations, 16% were Australian, and the remaining were from New Zealand, Switzerland, or

were of unknown origin. Thirty-five percent were ".com," 26% ".gov," 26% ".org," 6% ".edu," 3% ".net," and 3% ".ch" (ch is the DNS code for Switzerland).

Overview of the Research Design

In a previous study, we found that study participants made credibility judgments in 2.4–3.2 seconds (depending on whether outliers were included in the results; Robins & Holmes, 2008). So, in this study, we limited the viewing time of our stimuli (consumer health information Web pages) to 2.8 seconds before soliciting ratings. By limiting the viewing time, we are interested in isolating and observing the brief period in which people begin to form credibility judgments about a Web site. We are currently conducting a study of the same stimuli in which people will be able to spend as much time as they like looking at them. Those results, in the interest of time and space limitations, will be reported in subsequent papers.

Ratings were done on two criteria: (a) visual design preferences and (b) perception of credibility. This study reports comparisons of the ratings of visual design preference with the ratings or rankings for credibility perception done after very brief viewing. A subsequent study will report rankings done with unlimited viewing time.

Participants were shown each stimulus in a random order for 2.8 seconds. After each image was shown, a blank screen appeared allowing the participant to make a judgment (on visual design preference in part 1, and on perception of credibility of information in part 2). The same participants made both visual design preference judgments and credibility judgments on the same set of stimuli (presented in different order for visual design and credibility judgments). Stimuli order was reversed for every other participant to mitigate ordering effects.

The rating scale ranged from -1 to -4 (-4 being the worst) for negative judgments of visual design and credibility, and $+1$ to $+4$ ($+4$ being the best) for positive judgments of visual design and credibility. Participants entered their rankings on a keypad specifically designed for experimental situations.

Each participant, after signing an informed consent waiver, was shown and read a prepared script that explained they were to view a set of images/slides of Web pages (screen shots of real Web pages). After each image, they were to rate the image "on whether [the participant] liked the visual design, or aesthetics, of the page." Participants were given a practice session of four slides of non-medical sites to get a feel for the procedures of the study. The order of event for each slide was to (a) show a white screen with a crosshair in the middle to focus their attention to the middle of the screen, (b) show a stimulus image for 2.8 seconds, and (c) show a blank, black screen during which the participant makes their rating. This procedure was repeated for all 31 slides in the stimulus set.

After all 31 slides were judged, participants were asked what, if anything, they remembered about pages that caused them to make positive visual design ratings and similarly,

TABLE 2. Listing and description of stimuli included in this study.

Sponsoring Organization (Source/Authority)	URL (at the time of the study, 4/2008)	Domain	Stimulus Abbreviation	Nationality	Site Traffic Ranking
Aetna	http://www.intelihealth.com/	com	Aetna InteliHealth	US	65,558
Australian Prescription Products Guide	http://www.appgonline.com.au/	com	APP Guide	Australia	569,202
Medical Library Association (Consumer and Patient Health Information Section)	http://caphis.mlanet.org/	org	CAPHIS	US	424,737
Cleveland Clinic	http://my.clevelandclinic.org/health/	org	Cleveland Clinic	US	22,054
American Academy of Family Physicians (AAFP)	http://familydoctor.org/online/famdocen/	org	FamilyDoctor	US	9,382
Columbia University, New York, NY	http://www.goaskalice.columbia.edu/search.html	edu	Go Ask Alice	US	2,396*
Hardin Library for the Health Sciences, University of Iowa	http://www.lib.uiowa.edu/hardin/md/	edu	Hardin MD	US	5,917*
National Health Information Center	http://www.healthfinder.gov/	gov	Health Finder	US	137,492
Public Library of Charlotte & Mecklenburg County	http://www.healthlinkplus.org/	org	Health Link Plus	US	5,039,001
Health On the Net Foundation	http://www.hon.ch/MedHunt/	ch	Health On The Net	Switzerland	81,140
Medi-Span, through HealthTouch online	http://www.healthtouch.com/level1/search.htm	com	Health Touch	US	1,489,174
Mayo Clinic	http://www.mayoclinic.com/	com	Mayo Clinic	US	1,914
The HealthCentral Network, Inc.	http://www.mdchoice.com/pt/	com	MD Choice	US	379,220
Unclear	http://www.medguide.net/	net	MedGuide	Unclear	no rank
MedHelp, through partnerships with medical institutions	http://www.medhelp.org/	org	MedHelp	US	3,695
WebMD	http://www.medicinenet.com/	com	Medicine Net	US	1,785
INFOtrieve	http://www4.infotrieve.com/newmedline/search.asp	com	Medline Infotrieve	US	367,404
U.S. National Library of Medicine, National Institutes of Health, Department of Health & Human Services	http://www.nlm.nih.gov/medlineplus/	gov	Medline Plus	US	75,635
New Zealand Ministry of Health	http://www.medsafe.govt.nz/Consumers/CMI/CMIForm.asp	govt.nz	MedSafe 2	New Zealand	176,194
New Zealand Ministry of Health	http://www.medsafe.govt.nz/Consumers/cons.asp/	govt.nz	MedSafe	New Zealand	176,194
MIMS Consumer Health Group, a division of the global healthcare publishing company, CMP Medica	http://www.mydr.com.au/drugs/drugs.asp	com.au	MyDR	Australia	106,283
Case Western Reserve University, The Ohio State University, University of Cincinnati	http://www.netwellness.org/	org	NetWellness	US	44,383
National Health Information Center, U.S. Department of Health and Human Services	http://www.health.gov/nhic/	gov	NHIC	US	174,143
National Institutes of Health, U.S. Department of Health and Human Services	http://health.nih.gov/	gov	NIH	US	360
Novartis	http://www.novartis.com.au/consumer.html	com.au	Novartis	Australia	2,301,270
National Prescribing Service Limited	http://www.nps.org.au/search_by_medicine_name	org.au	NPS ltd	Australia	798,835
U.S. National Library of Medicine and the National Institutes of Health	http://www.ncbi.nlm.nih.gov/PubMed/	gov	PubMed	US	360
The Royal Australian College of General Practitioners	http://www.racgp.org.au/medicineinformation	org.au	RACGP	Australia	317,186
U.S. Department of Health and Human Services Office of the Surgeon General	http://www.surgeongeneral.gov/	gov	USHHS	US	294,339
University of Iowa Hospitals and Clinics	http://www.uihealthcare.com/vh/	com	Virtual Hospital	US	93,127
WebMD	http://www.webmd.com/	com	WebMD	US	806

what, if anything, they remembered that caused them to make negative ratings. This debriefing was intended to stimulate conversation to elicit statements by participants that would reveal the reasons why they preferred some designs and not others.

After this debriefing, we elicited ratings on credibility judgments. We gave participants a scenario to keep in mind while making credibility judgments. The script is meant to keep consistency of criteria for credibility judgments among the participants in the study. The scenario was as follows:

> Your friend has just been diagnosed with diabetes. You've just had a long talk with your friend and you are concerned and want to know more about the disease. You've done a search on Google and the Web pages you are about to see are the results of that search. In this next, and last, part of the study, after viewing each page, we'd like you to rate the page on whether you think the information on the page is trustworthy, credible, or believable.

After the scenario was read to participants, we used the same procedure using crosshair slides, stimuli, blank slides, and −4 to +4 rating scales to gather credibility judgments. Stimuli were presented in different order than they were for the visual design judgments. Immediately after the completion of all 31 credibility judgments, participants were debriefed in a similar manner as they were following visual design judgments. They were asked to recall, if possible, what caused them to rate pages high in credibility or low in credibility. After the debriefing, participants were given a gift certificate to a department store. The time for each participant to complete the study ranged from 20 to 30 minutes.

After all data were collected, we converted all ratings to positive numbers. That is, the rating scale presented to participants was −1 to −4 for negative ratings and 1 to 4 for positive ratings. For more simplicity of analysis and presentation, we converted those numbers to a scale of 1 to 8. All presentation of our results was accomplished on this positive scale.

Because of the exploratory nature of this project, a variety of analyses were performed on the data we collected:

- Calculate correlations between visual design ratings and credibility ratings for each stimulus
- Calculate means of ratings for:
  - Visual design preference
  - Credibility of information
- Compare means of visual design preferences and credibility of information in the following ways:
  - Rank all stimuli on visual design preference rating, and them compare means (t test) of credibility ratings for top half and bottom half of stimuli
  - Rank all stimuli on credibility ratings and then compare means (t test) of visual design ratings for the top half and bottom have of stimuli
  - Rank all stimuli on site traffic rank and then compare means (t test) for both visual design preference and credibility rankings on the top and bottom half of stimuli
- Examine the cooccurrence of stimuli in top and bottom half ratings when ranked by visual design, credibility, and site traffic

TABLE 3. Spearman Rho correlations between visual design and credibility ratings ($N = 34$).

Visual design/ credibility	Correlation coefficient	Significance (2 Tail)	$P<$
RACGP	0.660	0.000	0.01
HealthTouch	0.575	0.000	0.01
WebMD[a]	0.484	0.004	0.01
APPGuide	0.452	0.007	0.01
Medline Infotrieve	0.409	0.016	0.05
HealthOnTheNet	0.373	0.030	0.05
CAPHIS	0.371	0.031	0.05
NetWellness[a]	0.350	0.042	0.05
HardinMD[a]	0.316	0.069	
ClevelandClinic[a]	0.313	0.072	
MyDR	0.307	0.077	
USHHS	0.306	0.079	
GoAskAlice	0.301	0.083	
NPS ltd	0.292	0.094	
HealthFinder[a]	0.252	0.151	
MedicineNet[a]	0.251	0.152	
MedGuide	0.239	0.173	
MedSafe 2	0.235	0.181	
MDchoice[a]	0.213	0.226	
VirtualHospital	0.203	0.251	
MedlinePlus[a]	0.188	0.288	
MedSafe	0.177	0.316	
FamilyDoctor[a]	0.165	0.352	
NIH	0.155	0.382	
AetnaInteliHealth[a]	0.148	0.404	
Mayo[a]	0.144	0.417	
MedHelp[a]	0.137	0.438	
PubMed	0.122	0.491	
Novartis	0.104	0.557	
HealthLinkPlus	0.067	0.706	
NHIC	0.063	0.722	

[a]Sites that were rated as top health information sites by one or each of MLA, CAPHIS, or Consumer Reports.

Results

These procedures were preformed to address the research questions asked earlier in the article. The results that follow report the analyses done within the context of the research questions.

RQ1: What is the relationship between the visual design of a consumer health information Web site and perceptions of the credibility of information found on it?

The relationship between visual design, a viewer's perception of visual design, and any perception of credibility of information presented through a visual design is complicated. At this stage of our understanding, we present only the results of ratings and comments provided by participants. These ratings indicate, as in our prior study (Robins & Holmes, 2008), that visual design ratings and credibility ratings bear some similarity.

In a first look at this relationship, correlations between visual design and credibility ratings for each stimulus were calculated. Table 3 shows the results of a Spearman Rho analysis of such correlation. Eight of the 31 (26%) stimuli demonstrated a positive and statistically significant correlation
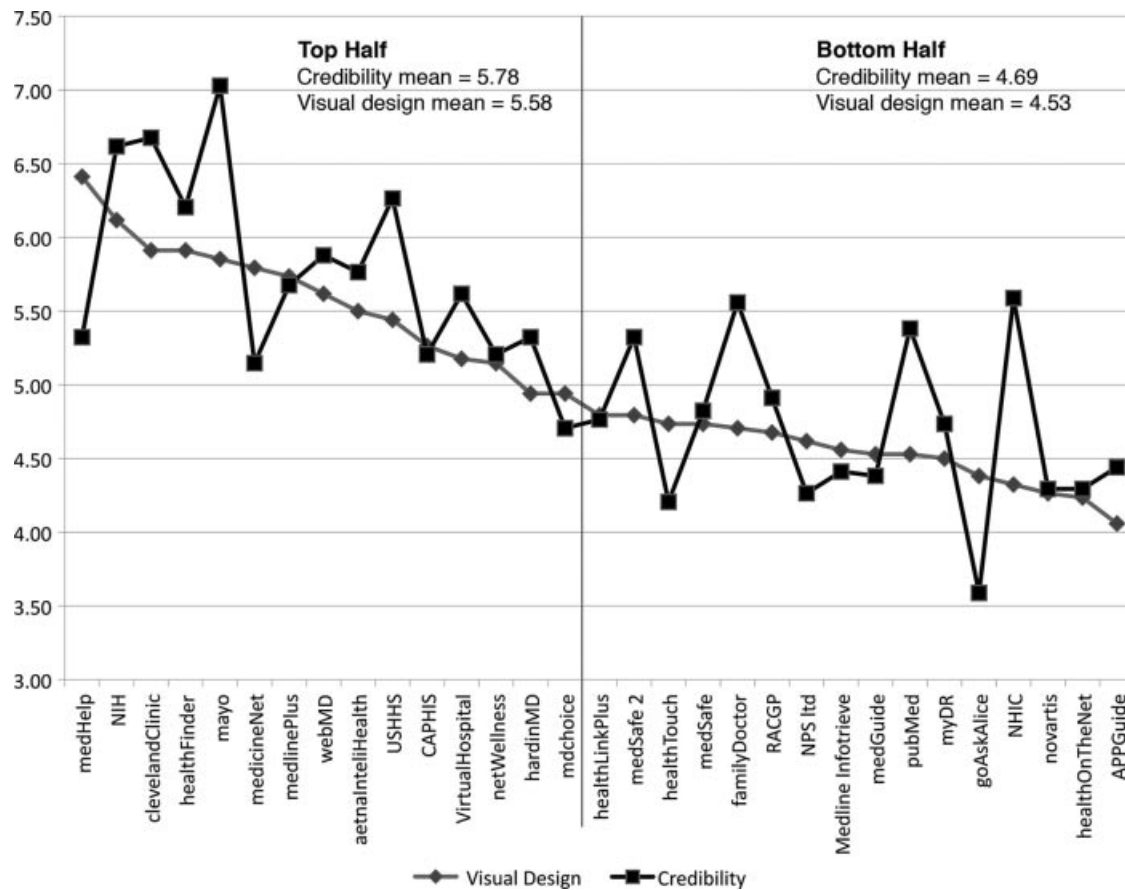
FIG. 1. Stimuli sorted by descending visual design ratings shown with corresponding credibility ratings.

between visual design and credibility ratings among our participants. A statistically significant correlation indicates that ratings among the 34 participants were similar for both visual design preference and perception of credibility. For stimuli without statistically significant correlations, ratings were less similar or even random in their manifestations.

Of the eight stimuli found to have statistically significant correlations between visual design and credibility ratings, only two were among the 12 top-rated sites identified earlier. From that, it might be possible to say that without brand recognition, visual design plays a more important role in how credibility is assessed. In other words, when visual design is rated high, so will credibility. This is not necessarily so in the case of sites with good reputation but perhaps with lower-rated visual designs. More will be said of this in a discussion of the second research question.

In this section, visual design preferences, credibility judgments, and site traffic are observed in turn. In three separate observations, each is ranked in turn and the corresponding data are observed. For example, if all 31 stimuli are sorted from highest to lowest rated visual design, it is possible to observe whether similar ratings on credibility and site traffic follow the visual design ratings. Figure 1 shows how credibility ratings follow visual design ratings when visual design ratings are ranked from highest rated to lowest rated.

If this configuration of the data is divided into top and bottom halves (that is, after ranking by visual design ratings, there is an arbitrary division between the top 15 and bottom 16 stimuli), there is a statistically significant difference between the corresponding credibility ratings in the top and bottom halves. This difference between top and bottom half credibility ratings when data are ranked by visual design preference shows an association between sites highly rated for visual design and sites highly rated for credibility. Although these data do not show why this relationship exists, they show a significant relationship, nonetheless. Table 4, item 1, shows the results of the t tests used to analyze various rankings of these relationships.

In a similar manner to the analysis just reported, the data are ranked from highest to lowest rated credibility, and then observed the corresponding visual design ratings for each stimulus. Figure 2 shows the results of this ranking and how visual design ratings follow credibility ratings. When the corresponding visual design ratings were grouped into top and bottom halves, there was also a statistically significant difference between the means in the top and bottom half (see Table 4, item 2).

Figure 2 and the resulting t test demonstrate another type of relationship between visual design and credibility perception. In this case, when all stimuli are ranked by credibility

TABLE 4.   Results of t-tests between top and bottom half means when stimuli are ranked by visual design, credibility, or site traffic.

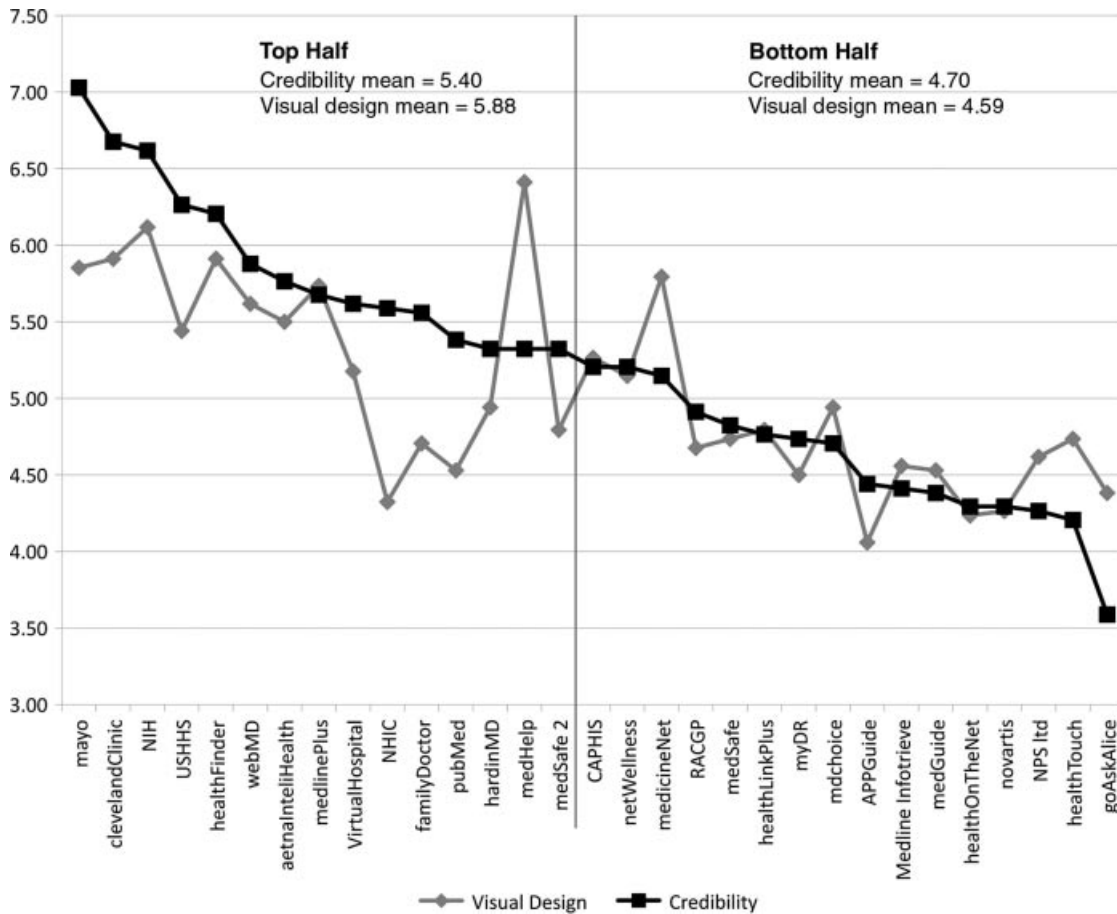| Item	Stimuli ranked by	Comparison of top/bottom half by	Top half average	Bottom half average	t-test statistic	Sig.
1	Visual design	Credibility	5.78	4.69	4.96	$p < 0.01$
2	Credibility	Visual design	5.40	4.70	3.63	$p < 0.01$
3a	Site traffic	Credibility	5.54	4.91	2.29	$p < 0.05$
3b		Visual design	5.34	4.76	2.82	$p < 0.01$

FIG. 2.   Stimuli sorted by descending credibility ratings shown with corresponding visual design ratings.

ratings, there was a direct relationship in the way our participants rated visual design. Specifically, sites perceived to have high credibility also were perceived to have better visual designs. This can be seen in the graph itself and is reinforced by the t-test.

Another way to analyze the relationship between visual design and credibility is to compare the site traffic ranking with corresponding visual design and credibility ratings. In a fashion similar to the two previous analyses, stimuli were ranked by their site traffic rank (as measured by Alexa, http://www.alexa.com/, an online service that keeps track of site traffic), and then examined the corresponding ratings for both visual design and credibility. Figure 3 shows that although there is large variability in the visual design and credibility ratings as they follow site

traffic rank, there is a trend for them both to decline as sites decline in traffic rank. A t-test confirms this relationship and the results are shown in Table 4, items 3a and 3b.

The relationship between the visual design of information on a Web site and the perceived credibility of that information is complicated. It has been shown here that health information sites that are rated higher for visual design tend also to be rated higher for perceived credibility. It should be noted that the relationship is not exact. For example, when visual design ratings are ranked high to low and credibility ratings are observed (as in Figure 1), the credibility ratings do not exactly match the line formed by visual design ratings: there is variability around the visual design line. Nevertheless, the tendency is there.
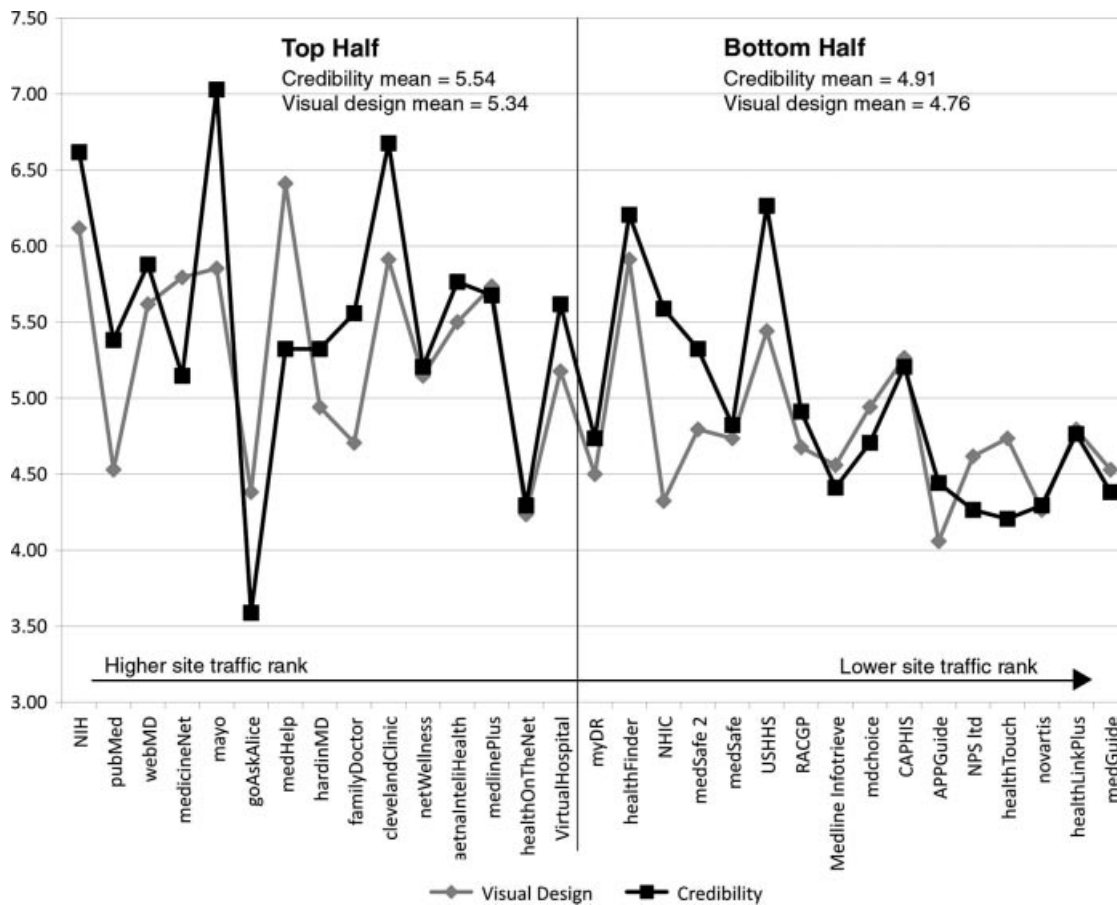
FIG. 3.   Stimuli sorted by descending site traffic rank shown with corresponding visual design and credibility ratings.

So far, this analysis has not shown why such a relationship might exist. In the following section, one possible influence on both visual design preference and perception of credibility is discussed: brand recognition.

RQ2: Is there a relationship between brand recognition, visual design preference, and credibility judgments?

Earlier, sites were identified that are independently recognized for their authority. Here, we look at how the name recognition of these sites might influence judgments about the relative merits of a site's visual design and credibility. In addition, site traffic rankings are compared with visual design and credibility ratings overall and among the independently top-rated sites. Table 5 provides a summary of this analysis.

Table 5 shows a tabular configuration of the data shown in Figure 3, but in addition, it highlights the 12 sites identified in Table 1 as rated high in authority. Ten of those 12 sites are in the top half of the stimuli rankings for site traffic (i.e., 10 of the top 15 stimuli ranked by site traffic are also rated high in authority by at least one independent rater). This finding is not particularly startling on its own, but it is interesting to note that the mean ratings for the "higher authority" sites are higher than the overall means for both visual design and credibility. The same is true for the top and bottom half of the table. The differences between the overall ratings and the ratings for the higher authority sites are not statistically

significant, but the numbers suggest a possible influence of the brand name associated with the site.

What is not clear from this analysis is whether brand name influenced visual design judgments. Visual design preference judgments were always conducted prior to credibility ratings and before any explanation that credibility was to be rated. Therefore, it is reasonable to question whether visual design ratings were positively influenced by a recognizable brand. Because it was the intent of this study to examine whether visual design influenced credibility and not the other way around, this may be a confounding factor in the study. We did their best to control for this possible problem by explaining in no uncertain terms that in the first round of judgments, participants were to rate strictly on their preference for the visual design or aesthetics of the site. In future studies, we plan to disguise the brand but keep the design in order to reduce the influence of brand.

One particular stimulus reflects problems that can arise when observing the effect of brand, and it should be noted here. The site called "Go Ask Alice" elicited feelings of mistrust among participants in this study. Several participants used the same question to express their mistrust of the site: "Who is Alice?" The name of the site, it is assumed, refers to lyrics from a Jefferson Airplane song written by Grace Slick (1967) called "White Rabbit." It may also be based on

TABLE 5. Stimuli ranked by site traffic with corresponding means for visual design preference ratings and credibility perception ratings.

	Mean visual design rating for each stimulus	Mean for top & bottom 15 in visual design	Mean credibility rating	Mean for top & bottom 15 in credibility for each stimulus	Traffic ranking
NIH	6.12	Overall = **5.34**	6.62	Overall = **5.54**	360
Pubmed	4.53		5.38		360
Webmd[a]	5.62		5.88		806
Medicinenet[a]	5.79	Top-rated sites = **5.56**	5.15	Top-rated sites = **5.76**	1,785
Mayo[a]	5.85		7.03		1,914
Goaskalice	4.38		3.59		2,396
Medhelp[a]	6.41		5.32		3,695
Hardinmd[a]	4.94		5.32		5,917
Familydoctor[a]	4.71		5.56		9,382
Clevelandclinic[a]	5.91		6.68		22,054
Netwellness[a]	5.15		5.21		44,383
Aetnaintelihealth[a]	5.50		5.76		65,558
Medlineplus[a]	5.74		5.68		75,635
Healthonthenet	4.24		4.29		81,140
VirtualHospital	5.18		5.62		93,127
MyDR	4.50	Overall = **4.76**	4.74	Overall = **4.91**	106,283
HealthFinder[a]	5.91		6.21		137,492
NHIC	4.32		5.59		174,143
MedSafe 2	4.79	Top-rated sites = **5.43**	5.32	Top-rated sites = **5.46**	176,194
MedSafe	4.74		4.82		176,194
USHHS	5.44		6.26		294,339
RACGP	4.68		4.91		317,186
Medline Infotrieve	4.56		4.41		367,404
Mdchoice[a]	4.94		4.71		379,220
CAPHIS	5.26		5.21		424,737
APPGuide	4.06		4.44		569,202
NPS ltd	4.62		4.26		798,835
HealthTouch	4.74		4.21		1,489,174
Novartis	4.26		4.29		2,301,270
HealthLinkPlus	4.79		4.76		5,039,001
MedGuide	4.53		4.38		No rank

[a]Sites that were rated as top health information sites by one or each of MLA, CAPHIS, or Consumer Reports.

a controversial, anonymously written book about a girl with drug problems. Go Ask Alice is, in fact, a medical information site directed at a younger audience on issues such as sex, drug use, acne, or any other medical-related issue to younger people.

This stimulus presents some interesting issues for the study of branding and credibility. First, the reference to the song lyrics that constitute the name of the site are quite dated and there is some question whether young people will recognize it. Although our participants were over the age of 35, few of them seemed to know the reference. Second, if the reference to the lyrics was recognized, the lyrics do have a connotation drug abuse, and that seems an odd choice of branding association for a Web site focused on providing health information to teens and young adults.

Third, although "Go Ask Alice" rated 31st of 31 stimuli in credibility perception ratings in our study, their sponsor is Columbia University Health Services (CUHS). The banner for Go Ask Alice includes the a Go Ask Alice logo on the left and a textual CUHS logo on the right connected by a band of color (see Figure 4). None of our participants mentioned seeing that the site was created by CUHS, probably because of visual design issues such as left hand placement

and the more noticeable color and typeface of the Go Ask Alice logo. A more noticeable CUHS presence may have mitigated participants' negative reaction to the site. The URL for the site was present in the stimulus (the stimuli were all screen shots of pages as they are displayed in a browser including the browser toolbars). The URL for Go Ask Alice is http://www.goaskalice.columbia.edu/, another indication of the sponsoring agency, but that was not noticed either.

The rather high site traffic ranking by Alexa (2,396 of millions of Web sites) is somewhat misleading and represents an inherent problem of site traffic ranks. These rankings are done at the domain level. In the case of Go Ask Alice, the domain is "columbia.edu" and all Web sites with "columbia.edu" at the end of their URL have a site traffic ranking of 2,396. Therefore, only by examining Web analytics at the server level, which are not publically available, would it be possible to know the exact visitation rate at the Go Ask Alice site.

Ad-Hoc Analysis

After reviewing the results of this study, we wanted to account for the comments made by participants when they were asked why they rated credibility judgments as they did.
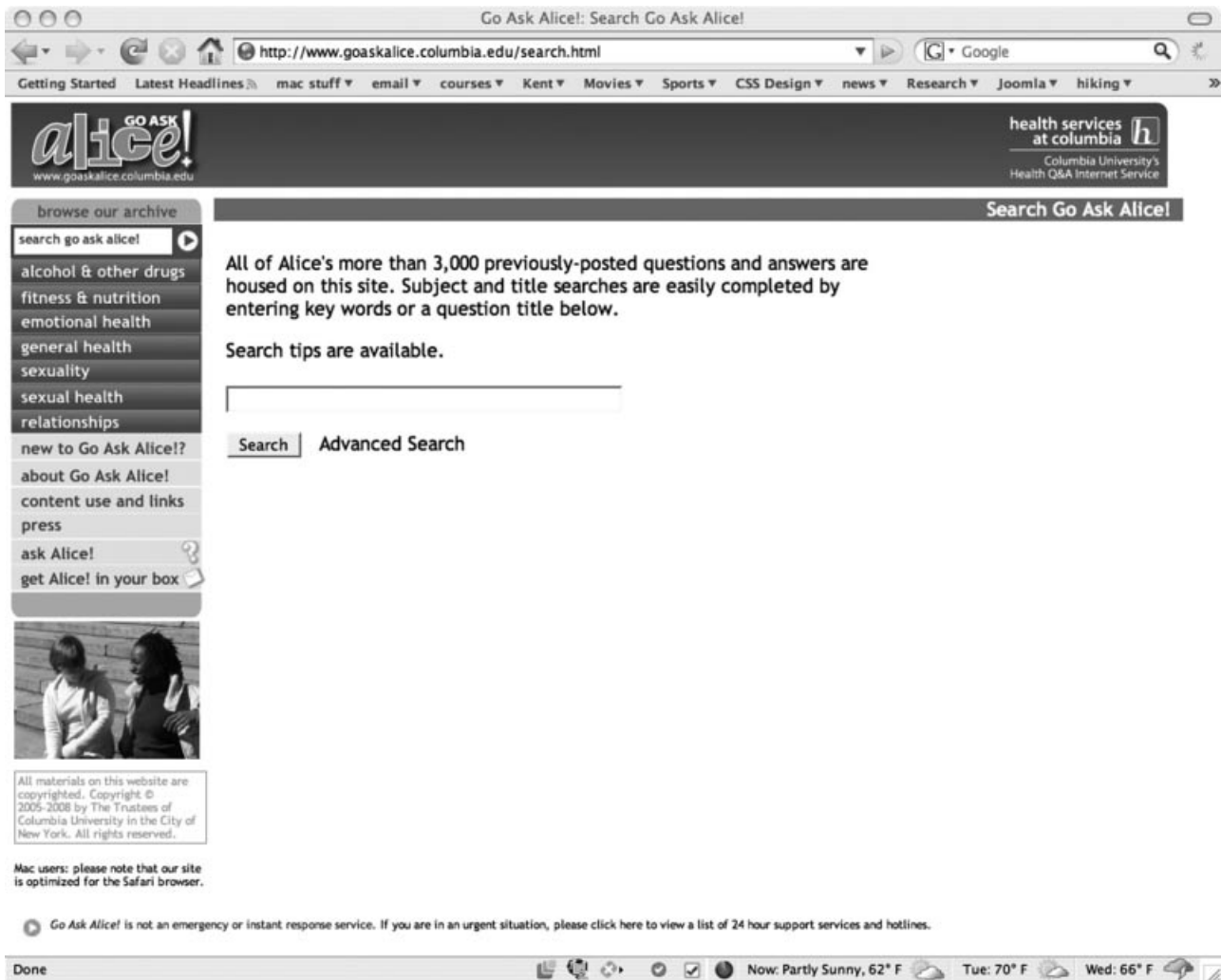
FIG. 4.   Search page for "Go Ask Alice" Web site.

TABLE 6.   Types of credibility judgments made by participants ($N = 34$).

	Coder 1 observations	Coder 2 observations	% Agreement	N Agreements	N Disagreements
Presumed	24	24	82.4%	28	6
Surface	21	27	82.4%	28	6
Reputed	15	20	61.8%	21	13
Earned	4	6	88.2%	30	4
Total	64	77			

After each participant had completed all 31 ratings for credibility, the researcher asked them: "Thinking back over the process of rating these sites, what caused you to: (a) rate a site's credibility high and (b) rate a site's credibility low. All comments were noted by the researcher on a form and kept in a file cabinet filed by the participant's number. Then, the comments were entered into a spreadsheet and coded by two coders. One coder was on the research team and the other was a graduate student not affiliated with the project. The coding scheme used was Fogg's (2003) four types of credibility (surface, earned, presumed, and reputed) as discussed in the literature review of this article. The purpose of this analysis was to try to understand more completely why participants rated these sites for credibility as they did, and whether these comments shed any light on whether there was anything to be learned about the implications of these ratings for the

credibility of information on consumer health information Web sites.

We wrote each string of text during data collection as a separate line in the spreadsheet. Coders were instructed to assign an "S" if the line referred to surface credibility, an "E" if it referred to earned credibility, a "P" if it referred to presumed credibility, and an "R" if it referred to reputed credibility. Coders were instructed what each of these types of credibility meant. If a coder could not determine what the line referred to, then they were instructed to leave it blank. Table 6 shows the results of this coding, as performed by the two coders. Note that a coder could apply any or all of the four types of credibility to the responses given by an individual subject. That is, it is possible that the participant could, in the course of their response, might talk about any or all of the four types of credibility. It is also possible that several lines written by a researcher might have been for surface credibility, for example. In that case, only one "S" was counted. This was because it was not appropriate to record added weight to comments that might have been recorded in an idiosyncratic way by the researcher. This method of coding also accounts for the fact that each coder ended up with a different total of codes (64 and 77). In addition, each coder did not understand all comments, and so that was another reason for the different totals.

Table 6 shows the number of codes applied by each coder to the comments made by subject, the percentage of agreement in the application of codes, the number of agreements and disagreements for each code, and the total number of subjects. The agreement is particularly high (i.e., above 80%) in all categories except "reputed," (61.8%). The low agreement in the reputed category is not surprising because it may have been difficult for coders to distinguish "reputed" from "presumed." However, the important thing about this analysis is that whether a coder designated a category as "reputed" or "presumed," they are responding to a subject's statement that is not about visual design, but rather about some aspect of the site related to its presumed or reputed authority.

What Table 6 also shows is that within the brief time of exposure to the stimuli, participants performed credibility judgments on a variety of criteria. Visual design (surface), considerations of the source of the site (presumed), a site's reputation (reputed), and sites with credibility won through prior use (earned) all figured into credibility judgments performed in 2.8 seconds.

In addition to what Table 6 shows, 10 of the 34 participants (29%) indicated specifically that if a site contained advertisements, any indication of drug company or insurance company sponsorship, or if the site appeared to be a ".com" site, the participant reacted negatively to the site. Similarly, 4 of the 34 participants expressed suspicion of consumer health information on sites with non-US sponsors. That is, they clearly indicated that they would not consider health information on sites outside the United States to be as trustworthy as those from the United States. Implications of this analysis will be discussed in the following section.

Discussion

The first aim of this article was to investigate the relationship between people's preferences for visual design and their perception of credibility of the information on consumer health information Web sites. The second aim of the article was to investigate whether factors such as brand (i.e., reputation) had any influence over credibility perceptions in a brief viewing of a Web site. As for the second aim, because site identities (i.e., the names of the sites, or brands) were not disguised in any way, we investigated the relationship between brand, visual design preference, and credibility judgments. Furthermore, this research is situated within earlier work on source credibility in which messages are generated from a source through a channel or medium for consumption. Both of these aims were pursued in the context of consumer health information Web sites.

Because consumer health information is the object of searches for approximately 113 million adults in the United States alone (or 80% of all internet users in 2006; Fox, 2006), it seems important to understand better some of the behaviors exhibited by people seeking this sort of information. The high volume of searching on this topic, in addition to the importance of health information for individual consumers, was one of the primary motivations of this topic. Another motivation was the results of Fogg's (2003) study in which there was overwhelming evidence that visual design and "look and feel" accounted significantly for perceptions of credibility. Finally, we wanted to further investigate findings by the research team in prior studies that initial credibility judgments were made at a visceral level in less than 3 seconds, and that those judgments were significantly linked to changes in visual design.

The results of this study, seen in light of previous studies in this area, seem to indicate that visual design preferences play a significant role in peoples' perception of credibility. The fact that this study showed sites with preferred visual designs tended to be sites perceived to be more credible (and also more popular in terms of site traffic) could be coincidence but that is unlikely. Over the course of this study, 2,108 ratings were made (i.e., 34 participants judged 31 Web sites twice—once for visual design and again for credibility—hence, $34 \times 62 = 2,108$), which reduces the chance of random results.

However, it would be inaccurate to say that visual design alone influenced credibility judgments. In 2.8 seconds, participants were able to identify brand and even site sponsors, nationalities, and advertisements, all of which were reported to influence credibility perceptions. Therefore, in a very short period of time, multiple assessments were occurring.

The implications for consumer health information Web sites are many. As mentioned earlier, people rely heavily on Web-based health information (Fox, 2006). Web sites, however, can be created and published by virtually anyone without any presumed peer review, expert review or editorial process of any kind. Consumers of Web-based information, therefore, must rely on cues to a site's credibility based on

each consumer's experience. If people are making decisions about whether to trust consumer health information on which they are basing actions concerning their health, and those trust decisions are based on surface credibility (i.e., look and feel), then this is a phenomenon worthy of investigation. At the very least, health care providers should be concerned that health information being used by their patients should be accurate and that people should know how to judge accuracy of information on criteria more substantial than visual design.

This study did show that credibility judgments are made on the basis of more substantial criteria, such as reputation, or the lack of for-profit site sponsors, such as drug companies. Even so, our results show that although our participants reported presumed credibility (based on the source of the site), surface credibility (look and feel), and to a slightly lesser extent reputed credibility (based on reputation of the source), only 10 of 34 identified specific potential credibility problems, such as drug company or insurance advertisements. The research team is currently conducting a study in which people are allowed as long as they wish to view these same stimuli. In that phase of the study, we will be able to compare results in which participants' viewing time is less than 3 seconds with a situation in which the viewing time is unlimited. A comparison of how people judge credibility when they have more time to consider the full array of cues present on a Web page versus a 2.8 second exposure seems a fair and useful one. What this study does not show is how initial impressions formed in the first 3 seconds of viewing might change as people continue to interact with a site (assuming they do). Future studies should also be longitudinal to the extent that they observe the entire search and use process so that it will be possible to see whether initial judgments about a resource correlate to whether the resource was ultimately used.

In any case, participants in this study demonstrated the use of multiple criteria for judging the credibility of information on health information Web sites. Judging the credibility of information of health information is particularly thorny because of the complexity of the information and because of the potential consequences of misinformation. The difficulty of providing accurate and understandable health information in a medium that allows anyone to publish would suggest a need for consumer training on how to evaluate consumer health information on the Web. Perhaps vetting of existing health information Web sites by a federal agency might be one way of providing people with a list of credible sites on the subject.

That being said, participants in this study did rate highly sites such as Mayo Clinic, WebMD, Aetna IntelliHealth, and VirtualHospital were all highly rated (1, 6, 7, and 9, respectively) and are ".com" sites. Although 10 of the participants expressed suspicion about medical information being skewed by the profit motive, participants overall rated four ".com" sites in the top 10 for credibility. In addition, 6 of the 12 sites listed in Table 1 (sites highly reputed by independent organizations) are ".com" sites. This anomaly may represent a difference of opinion among participants. It could also mean that the visual design of these sites suggest different things to different people. For example, a visual design of one site might place advertisements for drug companies in such a way that some people notice them, and to others, it simply looks like more information about a health condition. An interesting area of research would be to look more at the impact of advertisements on consumer health information sites and their impact on credibility judgments. Among other things, it would be informative to observe how people who express suspicion about the presence of advertisements perceive the credibility of sites with advertising on them.

Among the many cues that people may use to determine a health information Web site's credibility, this study investigated visual design as a cue. Robins and Holmes (2008) demonstrated visual design to be a significant factor in a consumer's perception of credibility in non-health related information. Similarly, this study has shown that the presentation of a message has a significant relationship to its perceived credibility. The present study differs from the 2008 study in that there was no manipulation of the visual designs in the present study's stimuli. The intent was to measure immediate, visceral reactions to visual designs before subjects were instructed to consider credibility. Then, when they were later instructed to consider a similarly visceral reaction to the messages' credibility, we looked for relationships. In the previous study, the original visual designs of the stimuli were muted, forming pairs of visual designs: original and muted. These pairs were randomly ordered and shown to participants who rated them as quickly as they perceived any notion of credibility. In both the 2008 study and the present study, a statistically significant relationship existed between visual design and credibility.

In the former study, subjects rated muted visual designs lower in credibility than they did the original visual designs. In the present study, when stimuli were ranked by subjects' visual design ratings, subsequent credibility ratings were statistically similar. When responses were ranked by credibility ratings, visual design rankings were also similar in a statistically significant manner. Similar results were found when stimuli were ranked by site traffic.

The findings of this study, however, do not suggest a clear distinction between a visceral reaction to stimuli and a cognitive interaction with the content viewed. Subjects in this study were able to recognize brands and mention some of them by name after viewing all 31 stimuli. This clearly represents processing beyond a visceral reaction to the stimuli. However, it could be that phenomena, such as presumed and earned credibility (Fogg, 2003) are, at least in part, visceral reactions. On the other hand, visual design preferences may be accomplished on a more visceral level. Our participants had little time to perform any sort of formal analysis of the visual designs they saw. They were able to get only a general impression on a gestalt level, perhaps. Few expressed any criticisms of specific formal qualities of visual design (except for mentions of one site that presented images of people with skin disease) such as typography or layout. Ratings seemed to be made on a "gut level." Now that the link between visual design preferences, credibility,

and a site's popularity (as measured by site traffic) has been established, future studies should focus more on the amount of time people spend in visceral reactions to site presentation and how much effort is necessary to perceive cues to better determine credibility. This knowledge will help visual designers design sites that more effectively present credibility cues.

Future studies might focus on this integration of brand into visual design through the effective use of cues that may reach users at a visceral level. Some of the designs in this study's stimulus set were presumably able to imprint their brand on participant's memory long enough for them to repeat it in debriefings. For example, Mayo clinic and Go Ask Alice were two of the most repeated brands. The former was consistently referred to as presumed to be trustworthy while the latter as not.

In addition, because the interplay of visual design and brand is probably not as linear as the processes in our study (i.e., people might perceive visual design and brand in tandem or in once very quickly after another), we are planning a future study in which existing designs will be manipulated to hide the name of the sponsor. For example, instead of Mayo Clinic, we will write "Health Information," but maintain the visual design of the page as much as possible. That way, brand bias can be eliminated, but it will be possible to keep the varieties of visual design from the current study. This will allow us to measure the effect of brand when compared with how people rate visual design and credibility without brand. We will use the same stimuli with the brand eliminated in that future study.

Knowledge of how to influence people to recognize brand as trustworthy may also make less credible information appear to be more credible, perhaps in an effort to sell advertising. Similar to the way "knockoff" brands tend to water down the prestige and market value of known brands, trusted health information brands' reputations might be undermined by sites with visual designs that mimic them. Subsequent studies might look more closely at sites that practice deception in such ways.

These findings also suggest a broader range of possible theoretical and practical implications. First, with respect to source credibility theory, it is often difficult to determine authors of Web site content, and participants in this study had little to go on but visual design and brand to judge credibility. Visual design preferences were in fact similar to those of credibility judgments on Web sites overall. This suggests that in the absence of more traditional methods of determining credibility, such as a trusted channel (e.g., a known medical journal) and a trusted source (e.g., a trusted doctor as an author), a user is left to look for cues such as a professional look and feel (Fogg et al., 2002) or a known brand such as the Mayo Clinic.

This study also addresses Fogg's (2003) framework of four types of credibility. This study focused mainly on the relationship of surface credibility cues to presumed credibility (brand), earned credibility (participants recognized the site as one they had used before), and reputed credibility (those sites endorsed as credible by independent and knowledgeable sources). Surface credibility turns out to be important because credibility judgments and visual design preferences happen in parallel. That is, visual design preference was shown as a cooccurring factor to credibility judgments as evidence to confirm Fogg et al.'s finding that his survey respondents used visual design to determine credibility. Our subjects, in debriefings after making visual design judgments and after making credibility judgments, did in fact mention if they recognized sites in the stimulus set and rated them higher in credibility if they had used them before (i.e., earned credibility). They also indicated if they thought a site might be credible because of its name (e.g., Mayo Clinic), although they had not used it before (presumed credibility).

Relating this study to source credibility theory, it is possible to draw parallels between dynamism and professionalism as properties of credibility judgments (Whitehead, 1968; Berlo et al.,1969) and visual design. That is, visual design sets a mood or tone to some degree of positive to negative to which people react and carry forward with them as they interact with a Web site. These reactions might be to the lack of a professional tone or the high degree of organization in which the information is presented. The present study demonstrated that a site that elicits positive preferences for visual designs are likely to elicit similarly positive credibility judgments. Although one cannot draw a direct parallel between the dynamism of public speaker and a Web site's visual design, it is possible to posit a relationship between a visual design that viscerally elicits a positive feeling about a site through its aesthetics and structure and the visceral connection one feels with a dynamic speaker.

The notion of authority or trustworthiness as properties of credibility judgments (Rieh, 2002; Whitehead, 1968; Hovland & Weiss, 1951) was a definite factor in credibility judgments in this study. Even with only 2.8 seconds of viewing time, people were not only able to discern well-known sites such as Mayo Clinic, but they were also able to determine subtle distinctions such as whether a site was based in the United States (which several participants indicated to be a factor in their judgments).

Given the number of subtleties necessary to make credibility judgments, the distinction among different properties in the source credibility theoretical framework becomes blurred (Metzger et al., 2003). For example, because it is not always clear who authors are on these Web sites, consumers rely on other cues for credibility, such as organizational identity and country in which the site is based. Therefore, the message and source are not always distinguishable, and the medium itself provides cues in the form of visual design and logos to influence people's feelings about the credibility of content. Specifically, if one considers an article on a consumer health information Web site as a message carrying health-related information of varying accuracy, then its source is either an individual or group author (source) named or unnamed. The medium is a Web site and it has properties such as visual design, loading speed, and screen resolution and window size.

The Web as a medium through which consumer health information is used presents a unique set of issues for

consumers, designers, and other stakeholders. This study has established a relationship among visual design, brand, and credibility judgment. Future studies should focus on what visual design cues are most important when it comes to influencing credibility perception. This will help designers to use the medium in such a way to avoid designs that can harm the perception of credibility of an otherwise credible site.

Conclusion

This study has shown a relationship in the judgments made about visual designs of health information Web sites and judgments concerning the credibility of the information found on those sites. It has also shown that trustworthiness and authority as demonstrated by a recognized and trusted brand were rated higher (although not with statistical significance) not only in credibility but also in visual design. This study was limited by the fact that not all properties that influence credibility could be included or varied in the study. For example, it would be interesting to eliminate the brand of each stimulus but retain the visual design. This would remove the impact of known brands and would provide a more direct way to observe the relationship between visual design and credibility. It might also be of use to design a study in which two large subject groups of the same demographic make independent judgments of visual design and credibility in order to eliminate any possibility of bias from judgments made for both.

Nevertheless, this study has provided a start to this line of investigation. Among future studies either underway or planned are ones that investigate changes, if any, in credibility judgments as people have more time to interact with a site, and studies in which we will mask the brands and, therefore, isolate visual design as a cue for credibility. In addition, the impact of advertising and the perception of objectivity would make a useful study. It would be of particular interest to know how people interact with health information over time in not simply information seeking but also information use. Finally, we would like to know more about the visual language of credibility, if it is possible to elucidate such a thing. It would be valuable for visual designers to know more specifically how to employ tools that promote credibility in Web sites whose health information is worthy of such a presentation.

Acknowledgments

References

Arnheim, R. (1969). Visual thinking. Berkeley: University of California Press.

Berlo, D.K., Lemert, J.B., & Mertz, R.J. (1969). Dimensions for evaluating the acceptability of message sources. The Public Opinion Quarterly, 33(4), 563–576.

Buller, D.B., & Burgoon, J.K. (1996). Interpersonal deception theory. Communication Theory, 6(3), 203–242.

Dutta-Bergman, M. (2004). The impact of completeness and Web use motivation on the credibility of e-Health information. Journal of Communication, 54, 253–269.

Eastin, M. (2001). Credibility assessments of online health information: The effects of source expertise and knowledge of content. Journal of Computer-Mediated Communication, 6(4). Retrieved April 6, 2009, from http://jcmc.indiana.edu/vol6/issue4/eastin.html

Eysenbach, G. (2008). Credibility of health information and digital media: New perspectives and implications for youth. In M.J. Metzger & A.J. Flanagin (Eds.), Digital media, youth and credibility (pp. 123–154). Cambridge, MA: MIT Press.

Fogg, B.J. (2003). Persuasive technology: Using computers to change what we think and do. Amsterdam: Morgan Kaufmann Publishers.

Fogg, B.J., Soohoo, C., Danielson, D.R., Marable, L., Stanford, J., & Tauber, E.R. (2003). How do users evaluate the credibility of Web sites? A study with over 2,500 participants. Proceedings of the 2003 Conference on Designing for User Experiences (pp. 1–15). New York: ACM Press.

Fox, S. (2006). Online health search 2006 (Pew Internet & American Life Project, October 29, 2006). Retrieved April 6, 2009, from http://www.pewinternet.org/Reports/2006/Online-Health-Search-2006.aspx

Fox, S. (2008). The engaged e-patient population. (Pew Internet & American Life Project, August 26, 2008). Retrieved April 6, 2009, from http://www.pewinternet.org/Reports/2008/The-Engaged-Epatient-Population.aspx

Freeman, K.S., & Spyridakis, J.H. (2003). An examination of factors that affect the credibility of online health information. Technical Communication, 51(2), 239–263.

Hovland, C.I., & Weiss, W. (1951). The influence of credibility on communication effectiveness. The Public Opinion Quarterly, 15(4), 635–650.

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of Web sites. International Journal of Human-Computer Studies, 60(3), 269–298.

Lazar, J., Meiselwitz, G., & Feng, J. (2007). Understanding Web credibility: A review of the research literature. Foundations and Trends in Human-Computer Interaction, 1(2), 139–202.

Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention Web designers: You have 50 milliseconds to make a good first impression! Behaviour & Information Technology, 25(2) 115–126.

Mattus, M. (2007). Finding credible information: A challenge to students writing academic essays. Human IT, 9(2), 1–28.

Metzger, M.J., Flanagin, A.J., Eyal, K., Lemus, D.R., & McCann, R. (2003). Credibility in the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. In P. Kalbfleisch (Ed.), Communication Yearbook 27 (pp. 293–335). Mahwah, NJ: Lawrence Erlbaum.

Norman, D.A. (2004). Emotional design: Why we love (or hate) everyday things. New York: Basic Books.

Petty, R., & Cacioppo, J. (1981). Attitudes and persuasion: Classic and contemporary approaches. Dubuque, IA: Brown.

Rieh, S.Y. (2002). Judgment of information quality and cognitive authority in the Web. Journal of the American Society for Information Science and Technology, 53(2), 145–161.

Rieh, S.Y., & Belkin, N.J. (2000). Interaction on the Web: Scholars' judgment of information quality and cognitive authority. Proceedings of the Annual Meeting of the American Society for Information Science, 37, 25–38.

Robins, D., & Holmes, J. (2008). Aesthetics and credibility in Web site design. Information Processing and Management, 44(1), 386–399.

Savolainen, R. (2007). Media credibility and cognitive authority. The case of seeking orienting information. Information Research, 12(3). Retrieved March 14, 2009, from online: http://informationr.net/ir/12-3/paper319.html

Self, C.S. (1996). Credibility. In M. Salwen & D. Stacks (Eds.), An integrated approach to communication theory and research (pp. 435–456). Mahwah, NJ: Erlbaum.

Slick, G. (1967). White Rabbit [Recorded by Jefferson Airplane]. On Surrealistic Pillow [LP]. New York: RCA Victor.

Stiff, J. (1994). Persuasive communication. New York: Guilford.

Toms, E., & Taves, A. (2004). Measuring user perceptions of Web site reputation. Information Processing and Management, 40(2), 291–317.

Warnick, B. (2004). Online ethos: Source credibility in an "authorless" environment. American Behavioral Scientist, 48(2), 256–265.

White, A.W. (2002). The elements of graphic design: Space, unity, page architecture and type. New York: Allworth Press.

Whitehead, J.L. (1968). Factors of source credibility. Quarterly Journal of Speech, 54, 59–63.

Wilson, P. (1983). Second-hand knowledge: An inquiry into cognitive authority. Santa Barbara, CA: Greenwood Publishing Group.