Age Differences in Credibility Judgment of Online Health Information

Q. Vera Liao
University of Illinois at UrbanaChampaign
201 N. Goodwin Ave
Urbana, IL 61801 USA
liao28@illinois.edu

Wai-Tat Fu
University of Illinois at UrbanaChampaign
201 N. Goodwin Ave
Urbana, IL 61801 USA
wfu@illinois.edu

ABSTRACT

To better support older adults' consumption of high quality health information on the Internet, it is important to understand how older adults make credibility judgments with online health information. For this purpose, we conducted two laboratory studies to explore how the credibility cues in message contents, website features, and user reviews could differentially impact younger and older adults' credibility judgments. Results from the first experiment showed that older adults, compared to younger ones, were less sensitive to the credibility cues in message contents, as well as those in the website features. Results from the second experiment showed that user reviews that were consistent with the credibility cues in message contents could reinforce older adults' credibility judgments. Older adults, compared to younger adults, seemed to be less swayed by user reviews that were inconsistent with the message contents. These results provided implications for designing health information websites that better support older adults' credibility judgments.

Categories and Subject Descriptors

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Experimentation, Human Factors

Keywords

Web credibility, online health information, aging, user review

1. INTRODUCTION

According to the Pew Internet Report [29], four out of five Internet users have searched the Web for healthcare information. Given older adults' higher needs for health information in support of their medical decisions, they are often found to be actively engaging in the consumption of online health information. Indeed, the Pew survey also showed that, for the older generation, visiting healthcare information websites was the second most popular one just behind using search engines.

A systematic review revealed that 70% of the studies that evaluated health information websites found that the quality of online health information to be unsatisfactory [7]. Given this situation, older adults' increasing exposure to the World Wide Web (WWW) may increase their vulnerability to low quality, even misleading information. Given that it is difficult to control what information will be available on the WWW, the ability to correctly judge the credibility of information is an important "self-defense" skill against being misled by unreliable information.

Credibility is particularly important for health information as misinformation for making medical decision can be hazardous, or even life threatening. This problem becomes more pressing considering studies have found that, unlike other routine online activities such as reading news or browsing entertainment sites, health information consumers often do not have a trusted brand name or Web portal in mind [7]. Instead, they often rely on search engines to reach for information concerning a particular health problem, and therefore can possibly bypass the source information on the front page, and encounter information in a random style.

On the other hand, a latest trend in quality control of online health related, especially medication information is the adoption of a user review system. Previous research has been quite optimistic about the social means to health information quality control via Web 2.0 technologies such as collaborative filtering and peer review[7][19][20]. However, they may add another layer of complexity to users' overall credibility assessment processes, since the largely anonymous and unfiltered user generated contents themselves demand credibility assessment, which may in turn interact with the original credibility assessment of the contents on the Web sites. This problem becomes even more pressing as merchants and even fraudsters start to realize the power of electronic Word of Mouth (eWOM), the manipulation of which could further obscure the boundaries between unbiased and commercial contents on the Internet.

In this paper we focus on the age differences in credibility judgment of online health information. Research has shown that, compared to younger adults, older adults tend to exhibit distinctive behavior and performance in terms of searching, evaluating, and comprehending information on the Internet [3][4][12]. The age differences, according to these studies, could be attributed to some unique characteristics of the older generation, such as their generally declined cognitive ability and inadequate experience with information technology. It implied that applications developed primarily for younger adults may be suboptimal, or even unusable for older adults. However, although important, there is a general lack of understanding on age differences in evaluating the quality of online health information. Therefore we conducted two systematic studies to explore age differences in credibility judgment of online health information. A better understanding of the age differences may help to provide design supports that enable older adults to benefit more from the freely available massive amount of health information on the Internet for better maintenance of health conditions.

This paper has two main goals. First, we were interested in investigating the differences in the credibility judgment processes of online health information by younger and older adults. To this end, we conducted a laboratory study to compare older and younger adults' credibility judgment of online health information,

and collected detailed verbal protocols during the process to reveal differences in the strategies that they used. Our second goal was to explore the effects of user reviews on credibility judgment, in order to understand whether these user reviews could induce further age differences in the credibility judging processes, and the extent to which they could help younger and older adults to make credibility judgment. Specifically, we were interested in whether older and younger adults would react differently to user reviews that are consistent or inconsistent with the credibility cues associated with the website contents.

2. RELATED WORK

Credibility has been examined as a multi-dimensional concept in the online environment. One major genre of the literature aims at uncovering the general factors of Web users' credibility assessment from an information processing perspective. Examples include Fogg's Prominence-Interpretation Theory [8] and Hilligoss's iterative process model [14]. Since persuasion is often considered a direct outcome of credibility assessment (as achieved by message acceptance), a large proportion of these studies were built on some forms of the dual processing model of persuasive communications. One widely cited model among these is the Elaboration Likelihood Model (ELM) [22]. The ELM explains attitudinal changes in individuals as they encounter two distinctive types of cues: central cues are related argument/message quality that requires systematic, deliberative processing. Persuasion through central cues is attained by users' careful consideration of the merits in the arguments in support of advocacy. Peripheral cues are related to surface features of the information, which can be processed in a heuristic way by relying on practical rules or experience (e.g. reputation of the source). With respect to Web credibility, central cues can be associated with the content of Web information, while peripheral cues may be associated with the contextual features of the websites (e.g., interface design, usability, source information, etc). In this paper, we will use content cues and contextual cues to refer to the two types of cues particularly for Websites.

Although there were only a limited number of studies that looked into how cues from these two different routes interact and impact users' credibility judgment [15], the interaction between content cues and contextual cues appears to be an interesting issue since more often than not, users may receive cues that contradict each other. This is especially important for online health information, as users tend to have high expectation for both information quality and source trustworthiness. In fact, one study showed that for healthcare websites, features of website credibility have only small, or at best moderate, correlation with the accuracy of information it provided [13]. This implies that apparently credible website may not necessarily provide reliable health related information.

Since the processing of central cues tends to be more cognitively demanding, studies proposing the stage model of credibility assessment have found that users tended to initiate credibility assessments by peripheral or heuristic processing [14][28]. However, there existed mixed results regarding the effects of specific peripheral cues on users' credibility assessment outcomes in previous studies. While most studies showed that notably better design appearances contributed to higher perceived credibility [1], users were often found to disregard certain specific website features such as third party endorsement, author information, advertisement, etc [15][26][30], when making credibility

judgment. These results implied that the application of dual processing model to Internet could be a complicated issue.

Although the impact of user reviews on credibility judgment is a relatively new topic in the area of health information, they have been extensively studied in the field of e-commerce. Previous studies on consumer trust indicated that social presence, the perceived social context, and the sense of the coexistent of other communicators, could increase consumers' trust in the e-Vendor [10]. Online user reviews could exert both informational and normative social influence upon individual's judgment in terms of product credibility and purchase intention [2]. When reading reviews, users may take the judgments of others as evidence and supplementary information for their decision making; on the other hand, they may also be susceptible to other social processes such as the norm of complying with other's opinion and conforming to "group opinion".

The primary focus of our study was to explore age differences in credibility judgment of online health information. While there were a number of studies looking into older adults' information search behavior with health related information, there were, however, few studies explored how older adults evaluate information credibility or quality on health information websites. Results from a survey-based study [4] indicated that older adults' trust with online health information was influenced by website identifiers (e.g. government agencies) and design features, which underscore the importance of contextual cues, i.e. website features that are not integral part of the message, to older adults' credibility judgments.

Considering that credibility judgment is often regarded as attitudinal change induced by content and contextual cues, previous studies regarding age difference in dual processing model may provide some insights into this particular issue [22]. Studies have provided robust evidence for age-related declines in the efficiency of systematic/ deliberative information processing of central cues, which could be associated with older adults' less efficient information processing, deficits in memory, decrease in the ability to inhibit false and irrelevant information, and less aware of factors that influence their judgments. However, there were less clear results regarding age difference in the processing of peripheral cues. By examining older consumers' attitude change within the Elaboration Likelihood Model framework, Yoon and her colleague's [32] found that the ELM effects, i.e. both central cue effects and peripheral effects on persuasion, were consistent with prior results obtained on younger adults.

Another stream of research relevant to our study concerning the effects of user review is age differences in sensitivity to persuasive influence on attitude change by external sources. Generally speaking, research on aging and attitude change show that resistance to external influence increases with age[16]. For example, studies that use the persuasion paradigm, in which detailed arguments are presented to convince a message recipient to adopt an advocated opinion, show that older adults were less deviated from their original attitude. The age difference is often attributed to skills developed by older adults to defend oneself against persuasion. In addition, it was shown that older adults have higher internal locus of control and tend to disregard external influence in the presence of internal reinforcement of their own opinion [16].

Our studies were based on previous studies regarding Web credibility and studies on age differences in using Internet. In particular, the Elaboration Likelihood Model indicates that people's credibility judgments with online information are influenced by two types of cues: cues in the message contents and cues in the contextual features of the Website. Considering older adults' generally declined cognitive abilities and inexperience with Web features, there may be age differences in processing credibility cues from both of the two sources. In addition, the influence of user reviews on older and younger adults' credibility judgments is a particularly interesting question because these days it is inevitable for health information seekers to encounter user review system when looking for information about medication or other health related products. While previous studies suggested that user reviews would exert influence on people's credibility judgment about the original contents, our focus was to explore age differences in such influence. For these purposes, we conducted two experiments. In the first experiment, we manipulated the strength of credibility cues in message contents and website features for a number of Web pages, and asked a group of older adults and a group of younger adults to perform the same credibility judgment task. By comparing their credibility ratings as well as their credibility judging process using verbal protocols, we investigated age differences in making correct judgments with cues from the two types of sources (i.e., content cues and contextual cues). In the second experiment, we added user reviews to the original Web pages used in the first experiment, and manipulated their consistency to the credibility cues in the message contents. Then we randomly assigned younger and older participants to tasks with and without user reviews. By comparing the credibility ratings of the four groups, we explored the age difference in subjectivity to the influence of user reviews on credibility judgments.

3. EXPERIMENT 1

In this study, we conducted a laboratory experiment to explore age differences in making credibility judgments with medication information found on the Internet. Specifically, we were interested in how content argument strength (content cues) and presence or absence of credibility-related Web page features (contextual cues) would differentially influence younger and older adults' credibility judgment.

3.1 Participants

Sixteen older adults (between 62 and 80 years old, Mean=69.38, SD=5.81; 66.7% were female) and sixteen younger adults (between 19 and 26 years old, Mean=21.56, SD=2.10, 50% were female) participated in our study. All participants were recruited from the Champagin-Urbana area in Illinois, USA. There was no significant difference between the younger group and older group in education levels, self-reported experiences in seeking health-related information online, and health related domain knowledge.

3.2 Experimental Design and Materials

We used a 2×2×2 mixed factorial design to study older and younger adults' credibility judgment of health information Web pages. There were two within subject variables: content cue strength (strong/weak) and contextual cue strength (strong/weak), and one between subject variable: age (young/old). All participants performed eight credibility judgment tasks, with each task composed of four Web pages that corresponded to the four possible combinations of strong/weak contextual and content cues.

Content Cue Manipulation

In our experiment, articles of alternative medicines were presented on Web pages we created. For content cue manipulation, we adopted the empirical method used by Petty and Cacioppo [22] to verify the credibility of contents shown on the Web pages. First, we selected materials from a popular healthcare website (www.revolutionhealth.com). It has articles of alternative medicine for different diseases, with ratings and reviews provided by users and professionals. Based on the review ratings we selected articles to be materials with "strong" and "weak" content cues. We further modified their use of evidence, argument rigor, information quality and bias, all of which were identified to be information credibility indicators [9][11][23]. We also balanced their length and amount of information to create a set of articles. The medicine names were modified such that they could not be recognized. To further verify our manipulation, we asked a group of 7 pilot participants naïve to the experiment to rate the credibility by reading the article, which means only processing the content cues. We filtered out articles that had the lowest consistencies among the pilot participants and ended up with 8 sets of documents.

Contextual Cue Manipulation

To manipulate the contextual cue, i.e. website features, we focused on features from two categories: design look and source features. Fogg et al.[9] identified that design look, including layout, typography, white space, images, etc, to have the largest impact on web credibility evaluation. Source features are features that indicate the source authority and expertise, which include references, contact information, privacy policy statement, thirdparty endorsements, site ownership, displayed awards, commercial features, etc [15]. First we selected web page templates from highly recognized and professionally designed healthcare websites (e.g., medlineplus, livestrong, mayo clinic), based on their public reputation, website traffic, and endorsement by authoritative third party such as the Health on the Net Network (HON). We adopted the design look and source features of these Web pages to act as "strong" contextual cues, and randomly chose to deliberately remove 3-5 features from these two categories to create Web pages that had "weak" contextual cues. This approach allowed us to avoid focusing on specific Website features that repeatedly happen to every task we used.

3.3 Task and stimuli

Participants started the task by reading the instruction and the scenario description of each task. Participants were instructed to imagine that they were helping a friend to evaluate some alternative medicines randomly collected from the Internet. The concern for potentially ineffective or even fake medicine was mentioned to implicitly induce their motivation for judging the credibility of each page. Participants were then presented with the task interface, which presented the 4 web pages under each of the 8 diseases on a regular Web browser. Participants could then click one of the disease names and browse any of the four web pages, each of which described an alternative medicine. Each of the articles had four parts: introduction, side effects, interaction, and dosage information of the medicine. These parts were typical medicine information on most medication websites. Participants could then click on "Rate" button on the interface and submit their ratings of the medicine based on a scale from 1 to 7 (1 =not recommend; 7=highly recommend).

To understand the evaluation process, verbal protocols were collected from participants by asking them to "think aloud" during two of the eight tasks[6]. All the protocols were recorded as digital files by the computer and later transcribed and analyzed.

3.4 Results

Credibility Rating Analysis

To understand the effects of age difference on credibility judgment of online health information, we performed a three-way ANOVA on their credibility ratings, with content cue strength and contextual cue strength as within-subject variables, and age as between- subject variable. The results showed that the main effects of content cue $(F(1,30)=22.04,\ p<0.01)$ and contextual cue $(F(1,30)=41.81,\ p<0.01)$ were significant. Also the interaction between content cue and age $(F(1,30)=4.18,\ p=0.05)$, and interaction between contextual cue and age $(F(1,30)=5.60,\ p=0.03)$ were significant. No other effect was significant.

The main effects of content cue and contextual cue validated our manipulation of content and contextual cues, and higher credibility ratings were given to Web pages with strong content cue and strong contextual cue rather than those with weak content cue and weak contextual cue. Interestingly, the two-way interaction between central cue and age, and that between peripheral cue and age were caused by the fact that older adults gave closer credibility ratings between strong and weak content, and between strong and weak contextual cues (see Figure 1). It suggested that older adults were less sensitive to credibility cues in both Web page contents and contextual features. Planned comparisons in each age group showed that while younger adults could successfully differentiate between strong and weak contextual cues $(\dot{F}(1,15)=19.10, p < 0.01)$, old adults were not able to do so (F(1,15)=3.90, p=0.07). The result indicated that older adults had difficulties in differentiating between strong and weak contextual cues.

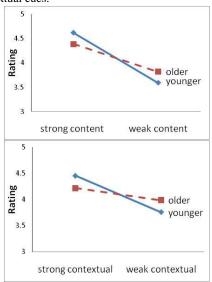


Figure 1. Credibility ratings for pages with strong or weak content/contextual cues

Protocol Analysis

To investigate the causes of older adult' lower sensitivity to credibility cues, we collected verbal protocols from each participant for two of the eight tasks (altogether 8 pages). We classified the transcribed protocols based on whether they were

about the content or contextual cues. For content cue processing, we further differentiated protocols showing that participants were passively reading facts or statements on a page from those showing that they were actively deliberating on information they read. This differentiation has important practical implications because more passive reading than active deliberation of information could imply that participants were less sensitive to the information quality or logical deficiencies of the page contents, and could potentially be misled or misinformed. We further broke down the three major categories of protocols (facts reading, deliberation, and contextual cues) into subcategories (Table 1).

We investigated which major categories of the protocols showed significant age differences. We calculated the percentages of cues mentioned in each category for each participant, to understand whether there were age differences in the relative weightings of different types of cue to make credibility judgment. First we conducted a two-way ANOVA by using age and type of cues as independent variables. It showed that the main effect of type of cues was significant (F(1,90)=46.33, p<0.01), and the interaction between age and type of cues was significant (F(2,90)=14.55, p <0.01). As Figure 2 showed, older and younger adults processed the types of cues differently, which led to the significant interaction. Post-hoc analysis with Bonferroni correction was performed to test the age difference in each type of cue. Results showed that younger adults had higher percentages under the category and contextual cue processing (p=0.016), while older adults had a higher percentage in the category of facts reading (p=0.001). The results suggested that contextual cue processing contributed more to younger adults' final ratings than that of older adults, while older adults' tended to rely more on accepting the facts they read to make their final ratings.

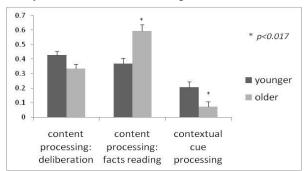


Figure 2. Percentage of cues in each category mentioned by each participant

We were also interested in what cues older and younger adults processed first when they evaluated a Web page. A two-way ANOVA (with type of cues and age as independent variables) on the proportions of the type of cue in the first cue mentioned by each participant was conducted. It showed there were significant maineffect of type of cue (F(1,90)= 10.52, p<0.01) and significant interaction effect between age and type of cues (F(2,90)= 11.99, p<0.01). The results indicated that older and younger adults processed different types of cues when they first evaluated a Web page. Post-hoc analysis with Bonferroni correction was performed to compare the results of younger and older adults for each type of cue. As Figure 3 showed, younger adults had a higher tendency to process contextual cues than older adults(p=0.010), while older adults had a higher tendency to read facts of the medicine than younger adults(p=0.001).

Categories	Subcategories	Criteria	Examples
	1.1Checking evidence	Checking studies, data, etc	"There is high quality scientific evidence", "The research looks only preliminary"
	1.2Evaluating information quality	Commenting on the completeness, accuracy, writing tone, or bias of information,etc	"There is way too much information devoted to healthy lifestyle, not the medicine itself",
Delibera- tion	1.3 Reasoning	Commenting on logical problems, contradictory facts, unclear explanations, etc; doubting claims, motives, etc;	" It said few adverse effects in the first part, but listed numerous ones in side effects part"
	1.4Relating to personal experience	Talking about personal experience and preference	"I took similar fiber product before and it helps" "I would not recommend OTC products"
	1.5Comparson	Comparing with other medicines read	"It works the same way with last one"
	2.1Introduction	Reading claims in the introduction part, including	"It lowers cholesterol",
	(effectiveness)	treaging efficacy, ease of use, history,background, etc	"It has been used in Asia for 1000 years"
Facts	2.2 Side effects	Reading claims in the side effects part	"Side effects included dizziness"
reading	2.3 Interactions	Reading claims in the interactions part	"Caution advised in people who take drugs lowering
			blood pressure"
	2.4 Dosage	Reading claims in the dosage part	"The dosage is 25 mg"
	3.1Design feeling	Aesthetical quality, layout, color, structure,etc	"layout is pretty simple, easy to read"
	3.2Reference features	Reference literature, resource links, suggestions for	"It lists some decent referecne",
Contextual		relevant information, etc	"Have links to read abstract of research"
processing	3.3Website source	Features indicating website reliability, e.g. sponsor	"The website is a non-profit organization",
	features	information, contact, endorsement	"American Heart Association recommended"
	3.4Commertial features	Advertising, promotion, donate button, etc	"The site is covered by advertisements"

Table 1. Coding scheme for protocol analysis

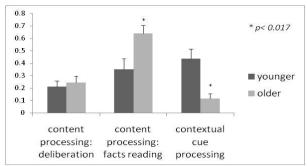


Figure 3. Percentage of different category mentioned for the first cue

According to the stage model of Web credibility assessment, users tended to first engage in preliminary assessment with the site by processing contextual cues before performing a more systematical, in-depth evaluation of the information on that site (Sillence, 2006; Sillence, 2007; Hilligoss & Rieh, 2008). However, it is worth noting that, inconsistent with this model, we found that older adults seemed to have a higher tendency to start by reading facts than deliberation (p<0.01) and processing contextual (p<0.01). This may imply that older adults tended to deviate from "common" behavioral patterns as observed in "regular" Internet users (who were likely younger adults in the studies that validated the stage model).

To summarize, from experiment 1 we found that older adults were in general less able to make correct credibility of online health information. Specifically, we found that they were less sensitive to the credibility cues in message contents as well as the website surface features, both of which are considered primary sources of cues for credibility judgment. By performing verbal protocol analysis to understand participants' credibility judging process, we found that these age differences could be explained by the differences in the strategies that older and younger adults adopted when processing online health information: 1) For content cue processing, older adults tended to rely on the facts they read in the

message contents, but less frequently engaged in deliberating on the information quality; 2) For contextual cue processing, older adults would less likely pay attention to features or attributes of the website during credibility judgment. Moreover, while younger adults tended to start with processing contextual cues on the website, older adults would more likely start by directly reading the text on the Web page. The set of results seem to confirm the intuition that older adults' were less adapted to the Web environment than younger adults, as they appeared to be browsing Web pages as if they were processing traditional forms of text such as books or newspapers.

4. EXPERIMENT 2

In the second study, we focused on how the presence of user reviews as an additional component to the medication website could influence younger and older adults' credibility judgments, and how the consistency of credibility cues in user reviews with the credibility cues in content message induced changes to their credibility judgment.

4.1 Participants

Twenty-two older adults (age between 58 and 80, Mean=68.45, SD=6.36, 59.1% are female) and twenty-two younger adults (age between 19 and 26, Mean=21.50, SD=1.95, 63.6% are female) participated in our study. All participants were recruited from the Champagin-Urbana area in Illinois, USA. There were no significant age difference in education levels, self reported frequency of health information seeking activities on Internet, or health domain knowledge.

4.2 Experiment Design and Material

A 2×2×2×2 mixed factor design was used in this study. There were two within-subjects variables: content cues (strong/weak) and contextual cues (strong/weak), and two between-subject variables: age (young/old), and user reviews (with/without user reviews). All participants were asked to finish the same task as in experiment 1. The 32 Web pages, which were combinations of weak/strong content and contextual cues, were the same as those

used in experiment 1(see Table 2). Then we added user reviews to each of the page to create the condition with user reviews. User reviews were randomly assigned to be consistent or inconsistent with the content cues, and evenly distributed across all content cue and contextual cue combinations.

User Reviews

We selected material of user reviews from the same website where we selected content articles (revolutionhealth.com) and modified them to fit the particular medicine. User reviews on the site were given anonymously together with ratings on a one star (not recommended) to five star (highly recommended) scale. The ratings were generally consistent with how negative or positive user reviews were used (see Table 2). These user reviews were primarily about users' experience with the medication, such as its efficacy and side effects, if any. Each entry of user review has less than 100 words. 4-6 entries of user reviews were given to each medication. We manipulate the consistency of user review with content cue strength by selecting positive (three to five stars) or negative (one to three stars) reviews.

4.3 Procedure

Participants were given the same instruction and tasks as used in experiment 1, and were randomly assigned to conditions with or without user reviews. For condition with user reviews, participants could click on a "Read Users' Review" link to read user reviews, which were presented on the same web page. After that participants clicked on the "Rate" button on the aggregator interface to submit their ratings for the medication.

4.4 Results

To understand how the presence of user reviews influenced younger and older adults' credibility judgment, we performed a four-way ANOVA with age and presence of user reviews as between subjects variables, and content cue and contextual cue strength as between subjects variables. The results showed that the main effects of content cue (F(1,40)=56.66, p<.01) and contextual cue (F(1,40)=23.05, p<.01) were significant. The interactions between content cue and age (F(1,40)=3.42, p=0.07), interaction between contextual and age (F(1,40)=3.42, p=0.07) and interaction between content cue and presence of user reviews (F(1,40)=3.79, p=0.06) were marginally significant. Interestingly, there was a significant three-way interaction between content cue, age and presence of user reviews (F(1,40)=4.45, p=0.04).

Figure 4 explained the three-way interaction between content cue, age and presence of user reviews: while the presence of user

reviews significantly reduced younger adults' ability to differentiate between strong and weak content cues, this effect was not observed in older adults. Three-way ANOVA with content cue, contextual cue and presence of user review was performed within each age group to further verify this conclusion. Results showed that while there was significant interaction between content cue and presence of user reviews among younger adults (F(1, 20)=5.903, p=0.02), this two-way interaction was not observed among older adults (F(1,20)=0.022, p=0.88). It indicated that while the presence of mixed user review significantly affected younger adults' credibility ratings after reading them, it did not show such effect on older adults' credibility ratings.

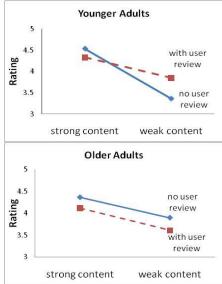


Figure 4. Credibility ratings for pages of strong/weak content cues with or without user reviews

We then investigated the younger group and older group with user reviews separately by introducing the consistency with content cue strength (consistent/ inconsistent) as a within-subject variable. Four-way ANOVA with content cue, contextual cue, age and consistency of user reviews showed that the main effects of content cues (F(1,20)=13.27,p<0.01) and contextual cues $(F(1,20)=8.03,\ p=0.01)$ were still significant. Two way interaction between content cue strength and consistency of user reviews was significant $(F(1,20)=43.86,\ p<0.01)$, and the two-way interaction between consistency and age was significant $(F(1,20)=4.30,\ p=0.05)$. However, the interaction between content cue and age became non-significant when there

Strong		Weak
Content cue	chosen from "high ranked" medicine, with research evidence, explanation of treating mechanism, comprehensive and accurate information, professional writing	chosen from "low ranked" medicine, lack of evidence, biased information, ill logic, commercial or unprofessional writing tone
Contextual cue	nice layout/color/information structure, with reference/contact information/third party endorsement	bad design, lack of reference/source, typo, advertising, commercial features (e.g. donate button)
User review	Five-star review: "Really effective product!. Combined with the right diet and exercise this meds is capable of producing rigid control of blood sugars in a very tight range." Four-star review: "It works for me. I have been on it for about 5 weeks now. My blood pressure is now high/normal. For the first 4 weeks, this medication caused moderate fatigue	One-star review: "This was a complete waste of time and money for me. I tried different brands one after the other and never even lost half a kg!" Two-star review: "Didn't see much difference. I used this remedy for a few months; it did help a little with sugar cravings but I probably loss only a couple of
	but this side effect has now disappeared."	pounds. It made my mouth dry, and I was irritable"

Table 2. Examples of content cue, contextual cue and user review manipulation

were user reviews (F(1,20)=0.03, p=0.87).

These results indicated that while older adults were less sensitive to the credibility cues in Web page contents, as shown in experiment 1, this age difference disappeared when user reviews were presented. The significant interaction between consistency and content cue, and that between consistency and age suggested that user reviews that were consistent and inconsistent with content cue strength had differential impact on older and younger users' credibility judgment. To further understand this effect, we divided all the 32 Web pages into two categories: pages with user reviews consistent with content cue credibility, and those inconsistent with content cue credibility, and analyzed them separately.

Effects of Consistent User Reviews

We analyzed the 16 Web pages with user reviews that were consistent with the content cue credibility. Four-way ANOVA with content cue, contextual cue, age and presence of consistent user reviews showed that the main effects of content cue $(F(1,40)=102.00,\ p<.01)$ and contextual cue $(F(1,40)=14.27,\ p<.01)$ were significant. The interactions between content cue and age $(F(1,40)=4.87,\ p=<.01)$, and interaction between content cue and presence of consistent user reviews $(F(1,40)=7.34,\ p=0.01)$ were also significant.

The results indicated that the presence of consistent user reviews had positive effects on users' ability to differentiate between Web content of high credibility and those of low credibility, possibly because they provided additional consistent cues to reinforce participants' initial attitude formed by reading the article. We then performed a three-way ANOVA within each age group. While there was a significant two-way interaction between content cue and consistent user reviews among older adults (F(1,20)=6.80, p=0.02), this two-way interaction was not significant among younger adults (F(1,20)=1.66, p=0.21). Figure 5 illustrated this difference: while consistent user reviews did not significantly benefit younger adults' credibility judgment with content cues, it enhanced older adults' ability to differentiate messages with strong and those with weak content cues.

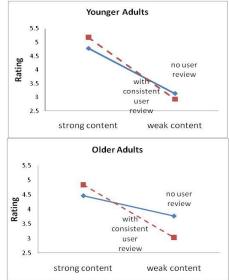


Figure 5. Credibility ratings for pages of strong/weak content cues with or without consistent user reviews

In summary, we found that consistent user reviews i.e., favorable user reviews given to credible medication information, and

unfavorable user reviews given to non-credible ones, enhanced older adults' differential reaction to credible medication information and non-credible one more significantly than younger adults. It implied that older adults could benefit from user reviews that may confirm their initial credibility judgment with the original contents.

Effects of Inconsistent User Reviews

We then analyzed the 16 Web pages with user reviews that are inconsistent with the content cue strength. Four-way ANOVA showed that the main effect of contextual cue (F(1,40)=6.62, p=0.01) was still significant, the main effect of content cue was only marginally significant (F(1,40)=3.30, p=0.08). The interaction between age and contextual cue was marginally significant (F(1,40)=3.14, p=0.08), and interaction between content cue and presence of inconsistent user reviews was significant (F(1,40)=24.72, p<.01). We found that the interaction between content cue and age was not significant (F(1,40)=0.03, p=0.87) when inconsistent user reviews were presented. Interestingly, there was a marginally significant three-way interaction between content cue, age and presence of inconsistent user reviews (F(1,40)=2.43, p=0.10).

From Figure 6, one could understand the marginally significant three-way interaction among age, content cue strength, and presence of inconsistent user reviews could be attributed to younger adults' higher ratings when inconsistent user reviews were presented than those by older adults. It also led to the disappearance of the age difference in making credibility judgment based on content cues, as shown by the absence of the two-way interaction between age and content cue. The two-way interaction between inconsistent user review and central cue strength implied that inconsistent user reviews in general had a negative influence on users' ability to differentiate between high credibility and low credibility cues in Web page contents. Threeway ANOVA in each age group showed significant interactions between inconsistent user review and content cues for both adults(F(1,20)=16.17,p <.01) adults(F(1,20)=8.55, p<.01), which further confirmed this conclusion.

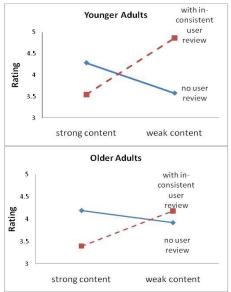


Figure 6. Credibility ratings for pages of strong or weak content cues with or without inconsistent user reviews

The effect of inconsistent user review is particularly interesting since negative reviews on well-argued articles and positive reviews on poorly argued articles could have distinctive implications to the use of user reviews for health information credibility control. While negative reviews on a well written (strong content cue) article could potentially help prevent users from mistrusting questionable sources, positive reviews on a poorly written (weak content cue) article, on the contrary, could imply deceptive manipulations or spamming related activities. Our results therefore implied that older and younger adults could have differential reaction to each of these cases of inconsistencies.

To further understand the effects of negatively inconsistent and positively inconsistent user reviews on older and younger adults' credibility judgment, we analyzed the main effects of the presence of negatively inconsistent user reviews on ratings for high credibility contents, and the main effects of positively inconsistent user review on ratings for low credibility content in each age group. It showed that, for older adults, the effect of negatively inconsistent user reviews on strong content cues was marginally significant (F(1,20)=3.17, p=0.09), but the effect of positively inconsistent user reviews on weak content cues was nonsignificant (F(1,20)=0.32, p=0.58). For younger adults the effect of positive user reviews on low credibility content was significant (F(1, 20) = 10.96, p < .01), and negative user reviews on high credibility content was marginally significant (F(1,20)=3.45,p=0.08). These results showed that older adults were less susceptible to the influence of inconsistent user reviews, and were especially less swayed by the combination of ill-made arguments and highly appraising user reviews. In that case, they tended to retain their initial negative attitude rather than change their ratings.

From experiment 2, we found that user reviews had strong impact on users' credibility judgment, and could influence older and younger adults' credibility judgment in different ways. When credibility cues in user reviews were consistent with the credibility cues in website contents, it could significantly facilitate older adults to make correct judgment about content cues, and thus help to make correct credibility judgment by supplementing the generally lower systematic/deliberative processing abilities of older adults. One of the most interesting finding in this study was the effect of user reviews that convey information inconsistent to the original contents. It showed that older adults were less subject to the influence of inconsistent user review than younger adults. Especially when appraising user reviews appeared on a flawed website, older adults tended to discount information from user reviews while younger adults would more likely be swayed by user reviews.

5. CONCLUSION

The results of the two experiments showed that, when making credibility judgments of online health information, people integrate credibility cues from different sources including message content, contextual website features and user reviews, and we found age differences in all these three aspects: 1) for content cues, older adults were less able to differentiate contents with high credibility from those with low credibility. Protocol analysis revealed that older adults were more likely to read and accept what was stated in the article without adequate deliberation on the quality of information. This can be problematic as they may be easily misled by unreliable information. 2) For contextual cues, older adults were less sensitive to the credibility indicated by Website features. Protocol analysis showed that older adults were

more inclined to ignore the contextual Website features during credibility judgment. 3) For user reviews, older adults exhibited different reaction to user reviews that were consistent or inconsistent with the original message contents (content cues). Consistent user reviews seemed to better support older adults' credibility judgment of the medication by providing additional consistent cues to reinforce their initial judgment with the content. Thus, it helped to moderate the age difference in making correct credibility judgment. Inconsistent user reviews, however, had less impact on older adults' credibility judgment as compared to younger adults. Especially when positive user reviews were given to a low credibility Web page, older adults' credibility judgment was less swayed by the user reviews.

Our results showed that there are complex age differences in credibility judgment of online health information. Based on these results, implications for improving older adults' credibility judgment of online health information could be drawn. From a training perspective, we identified two credibility judgment strategies that older adults could be encouraged to adopt: 1) Older adults may make better credibility judgment when they actively deliberate on the credibility of the message, rather than passively reading facts stated on the Web page; 2) Older adults may benefit from starting the credibility judgment process by examining contextual cues first, as it takes less cognitive effort and can be more easily adopted by older adults.

From a design perspective, one could consider methods to compensate for older adults' lower sensitivity to credibility cues in message contents and website features. As older adults are less likely to spontaneously deliberate on the credibility dimension of information, it will be helpful to provide explicit, prominent guidance to encourage older adults to deliberate on the information. Similarly, designers should take into account older adults' generally limited Internet experience, as reflected by their tendency to pay limited attention to web-specific contextual cues in their credibility judgments. For example, while collaborative implementations to provide authority/ credibility indicators (e.g. HON stamp and other displayed awards) have been considered promising solutions for regulating health information website, our studies (as well as others) seem to suggest that users often fail to actively check these features. One way to ameliorate the problem is to utilize external tools such as the use of a pop-up checklist that "reminds" users to ask critical questions relevant to the credibility of information and to pay attention to Web-specific features, which may help older adults to naturally adopt more effective credibility judgment when consuming online health information. To successfully guide users' attention towards credibility related features and relieve users from excessively deliberating on the quality of information, visualization techniques seem to provide useful tools to guide users' credibility judgment in a systematic and user-friendly way [25][31].

Another focus of our study is the impact of user reviews on credibility judgment. As stories about how people benefit from other patients' experience shared online abound, the impact of "user reviews" on health information website is likely to be more than just another tool for people to socialize. Our study showed that user reviews could have complex influence on older and younger adults' credibility judgments. It suggested that knowing exactly how older adults would be influenced by user reviews is important in understanding the long-term effects of different designs of Web 2.0 technology for older adults.

As we found that older adults' lower sensitivity to content credibility could be moderated by user reviews consistent with the contents, it implied that a well designed user review system, which encourages users to give informative and accurate reviews, may facilitate older adults' credibility judgment. For example, customizable sorting function of user reviews may help users to more relevant reviews and make better credibility judgment. Also, highlighting the linkages between contents and user reviews (e.g., by providing hyperlinks between a statement on a page and user reviews that confirm or question it) may be helpful for users who look for evidence or counterevidence for their initial credibility judgment with the contents.

One most interesting finding in this study was older adults' lower sensitivity to inconsistent user reviews, especially when positive user reviews appeared on a flawed website. This phenomenon could be interpreted from two perspectives: First, older adults' lower susceptibility to attitude change may be a possible explanation. As suggested by previous research, older adults are more resistant to attitude change when facing persuasive influence, in which detailed argument is provided to convince message recipient to adopt the position advocated. Also, research on age difference in dual processing of external cues indicates that motivational and emotional variables may vary older adults' deliberative processing level and affect the outcome of influence [21]. We infer that these age-related differences may cause older adults to be less sensitive to the influence of user reviews, especially in the situation of negative initial attitude. It was possible that older adults initial negative attitude towards lowcredibility content made them to selectively stop further deliberative processing when reading and comprehending user reviews.

The second possible reason for the lower sensitivity to inconsistent user reviews could be attributed to the generally lower Internet experience of older adults. Previous research provided substantial evidence that older adults in general have poorer Internet experience as compared to younger adults[3][4]. Also, research shows that frequent users of Internet tend to have more certainty and more confidence in online information. We may further extend this view to social networking applications. As there is research showing that younger adults often use Internet for entertainment and social networking, while older adults tend to use Internet as a tool for research, shopping and banking [29], it may imply that, compared to general Internet experience, older adults may have an even lower experience with user reviews and other social networking features (both in terms of actively contributing and passively reading). This may lead to a higher tendency for older adults to distrust and discount cues associated with user reviews.

We should point out that older adults' relatively insensitive reaction to positive user reviews on a low credibility site, and to negative user reviews on a seemingly credible site could be interpreted as a positive sign. One could argue that younger adults are more susceptible to attitudinal change by biased information than older adults. While in our study we did not deliberately test this difference, future research could be directed to how easily younger and older adults can detect biased or even wrong information in user reviews.

On the hand, to develop a user review system that helps quality control of online health information, designers should give more thoughts on how to motivate older adults to utilize such system. Good user experience, both in terms of the interaction with the system and the quality of user reviews it provides, is the key to encourage older adults to deliberately process and have higher trust in user reviews. Further research is needed to understand how different policies adopted by user review systems, such as those that limit the minimum length of review, and designs that allow collaborative evaluation of reviews, such as the "like" button, will provide better solutions to harness the potential of user reviews for information quality control, and to benefit older adults.

6. REFERENCES

- [1] Alsudani, F. and M. Casey., The effect of aesthetics on web credibility. In Proc. 23rd British HCI Group Annual Conference on People and Computers (2009), 512-519.
- [2] Cheung, M., et al., Credibility of Electronic Word-of-Mouth: Informational and Normative Determinants of On-line Consumer Recommendations. Int. J. Electron. Commerce 13, 4 (2009), 9-38.
- [3] Chin, J. and W.-T. Fu., Interactive effects of age and interface differences on search strategies and performance. In Pro. CHI2010. ACM(2010), 403-412.
- [4] Czaja, S.J., et al., Older Adults and Internet Health Information Seeking.In Proc. HFES2009. 126-130.
- [5] Dutton, W.H. and A. Shepherd, Trust in the Internet as an experience technology. Information, Communication & Society 9, 4 (2006), 433 - 451.
- [6] Ericsson, K. A., Simon, H., Protocol Analysis(1984), Cambridge, MA: MIT Press
- [7] Eysenbach, G., Credibility of Health Information and Digital Media: New Perspectives and Implications for Youth. The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning (2007), 123-154.
- [8] Fogg, B.J., Prominence-Interpretation Theory: Explaining How People Assess Credibility Online. In Proc. CHI 2003. ACM Press(2003),
- [9] Fogg, B.J., et al., What makes Web sites credible?: a report on a large quantitative study. In Proc. CHI2001. ACM(2001), 61-68
- [10] Gefen, D. and D.W. Straub, Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services 32, 6 (2004), 407-424.
- [11] Hamilton, M.A., Message Variables That Mediate and Moderate the Effect of Equivocal Language on Source Credibility. Journal of Language and Social Psychology 17,1 (1998), 109-143.
- [12] Hanson, V.L., Cognition, Age, and Web Browsing. In Proc. UAHCI2009. Springer-Verlag(2009), 245-250
- [13] Kunst, D.G., Pallavi M Latthe, Manish Latthe, Khalid S Khan, Accuracy of information on apparently credible websites: survey of five common health topics. British Medical Journal 324 (2002), 581-582.
- [14] Hilligoss, B. and S.Y. Rieh, Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. Information Processing & Management 44. 4 (2008), 1467-1484.

- [15] Hong, T., The influence of structural and message features on Web site credibility. Journal of the American Society for Information Science and Technology 57. 1 (2006), 114-127.
- [16] Lynn W.Phillips, B.S., Age Differences in Information Processing: A Perspective on the Aged Consumer. Journal of Marketing Research Journal of Marketing Research 14, (1977), 444-457.
- [17] Man Yee Cheung, C.L., Choon, Ling Sia, and Huaping Chen, Credibility of Electronic Word-of-Mouth: Informational and Normative Determinants of On-line Consumer Recommendations. International Journal of Electronic Commerce 13, 4 (2009), 9-38.
- [18] Meola, M., Chucking the checklist: A contextual approach to teaching undergraduates Web-site evaluation. Libraries and the Academy, 4,3(2004) 331–344.
- [19] Metzger, M.J., Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. Journal of the American Society for Information Science and Technology 58. 13 (2007), 2078-2091.
- [20] O'Grady, L., Future directions for depicting credibility in health care web sites. International Journal of Medical Informatics (2006), 58-65.
- [21] Peters, E., et al., Adult Age Differences in Dual Information Processes: Implications for the Role of Affective and Deliberative Processes in Older Adults' Decision Making. Perspectives on Psychological Science 2. 1 (2007), 1-23.
- [22] Petty, R.E., Cacioppo, J.T, The Elaboration Likelyihood Model. Advances in Experimental Social Psychology 19. 123-205 (1986).
- [23] Rieh, S.Y., Credibility: A Multidisciplinary Framework. Annual Review of Information Science and Technology 41 (2007), 307-364.

- [24] Sakamoto, Y. (2010). The impact of collective opinion on online judgment. Proceedings of the 32nd Annual Conference of the Cognitive Science Society.
- [25] Schwarz, J. & Morris, M., Augmenting Web Pages and Search Results to Support Credibility Assessment. In Proc. CHI2011. ACM(2011), 1245-1254.
- [26] Shon, J., J. Marshall, and M.A. Musen. The Impact of Displayed Awards on the Credibility and Retention of Web Site Information. In Proc. AMIA Symp. 2000.794-798
- [27] Sharit, J., et al., Investigating the Roles of Knowledge and Cognitive Abilities in Older Adult Information Seeking on the Web. ACM Trans. Comput.-Hum. Interact 15. 1 (2008), 1-25.
- [28] Sillence, E., et al., A framework for understanding trust factors in web-based health advice. International Journal of Human-Computer Studies 64. 8 (2006), 697-713.
- [29] Sydney Jones, S.F., Generations Online in 2009, Pew Internet & American Life Project. 2009.
- [30] Walther, J.B., Z. Wang, and T. Loh, The Effect of Top-Level Domains and Advertisements on Health Web Site Credibility. Journal of Medical Internet Research 6. 3 (2004), e24.
- [31] Yamamoto, Y. & Tanaka, K., Enhancing Credibility Judgment of Web Search Results. In Proc. CHI2011. ACM(2011), 1235-1244.
- [32] Yoon, C., Cole, C. & Lee. M., Consumer Decision Making and Aging: Current Knowledge and Future Directions. Journal of Consumer Psychology 19(2009) 2-16

APPENDIX

- The track: human factors
- The preferred allocation of reviewing expertise: primary- social/behavioral science, secondary-information science
- Topics: Interactions with health information technologies