User reviews have become a standard feature on most online shopping websites. Despite the availability of a large number of user reviews for a product, studies have shown that users typically read only the first few reviews. It is unclear what aspects of user reviews impact the decision of whether or not to purchase a product or service. This study investigated which of the visual cues presented by user reviews are mostly likely to attract users’ attention and subsequently make the reviews have more impact on the purchasing decisions. A taxonomy was proposed for categorizing visual cues and applied in the design of a human subject experiment. Results from one online study conducted using the visual cue taxonomy were presented and the implications and benefits of considering it in the design of user review visualizations, from both a retailer and consumer’s perspective were discussed.

Visual Cues: What Combinations Make for a Better User Review?
Julian Archer¹, Jing Chen², and Nsikak Udo-Imeh¹

¹School of Industrial Engineering
Purdue University
West Lafayette, IN

²Department of Psychological Sciences
Purdue University
West Lafayette, IN

User reviews of products have become an important tool in guiding the decision making process of consumers (Chatterjee, 2001; Chen et al., 2003; Duan, Gu, & Whinston, 2005; Godes & Mayzlin, 2004; Zhang et al., 2004). User review systems are available for both the purchase of physical products and services. Whether shopping for physical products or services, potential customers are encouraged to use these user reviews considering the fact that consumers perceive consumer-provided information to be of higher credibility than marketer-provided information (Kim & Gupta, 2012). Clearly, there is some value in utilizing user review systems. However, how the system actually works to influence or aid the customer’s decision process is not well understood. Studies have also shown that people rarely read Web pages word by word; instead, they scan the page, picking out individual words and sentences (Nielsen, 1997). An identification of the visual cues that attract users’ attention can help us present reviews that help users make better purchase decisions. Therefore, the aim of this study is to investigate what types of visual cues contained in online user reviews attract attention and possibly make the corresponding reviews have more impact on the purchase decision making process.

BACKGROUND
An online user review is a written assessment based on personal experience or opinion of a product or service posted on a dedicated review webpage or the retailer's/service provider’s webpage. These product opinions which may be regarded as a new form of electronic word-of-mouth (eWOM) have become an indispensable tool for consumers and thus for retailers who want to attract and retain consumers (Pan and Zhang, 2011). Evidence from the literature indicates users often finalize the decisions to buy or not to buy after consulting these online reviews (Dellarocas 2003, Godes et al. 2005). In traditional shopping, the customer can see, touch and feel the product, or ask questions to clear any doubts before making the purchase decision.

However, in online shopping, the customer has to make a decision based only on the description of the product and/or pictures. Sometimes websites provide online customer service in the form of “chats” but they may not give the same effect as an in-store, person-to-person customer service experience. Therefore, user reviews, serve as additional information to resolve any lingering questions shoppers may have about the product. Since user reviews present the opinions of fellow customers, they appear to be trusted because of shared experiences and the perception of shared values or shared needs (Clemmons, Barnett, & Appadurai, 2007). Despite the increasingly important role that online user reviews play in the purchase decision, little is known about what causes consumers to find an online peer review helpful to their shopping tasks (Pan and Zhang, 2011). However, only a few have touched the visual cues (forms) of user reviews.

There are a number of studies investigating what characteristics of users’ reviews influence users’ perception of the helpfulness of reviews (e.g., word counts: Gan, Cao, & Jones, 2012; abstractness: Huang, Tan, & Ke, 2011; emotion: Kim & Gupta, 2012; review extremity, depth, and product type: Mudambi & Schuff, 2010; review valence and length: Pan & Zhang, 2011). However, only a few have touched the visual cues (forms) of user reviews.

For example, some studies examine whether the number of words or characters influence the usefulness of user reviews. Pan and Zhang (2011) showed that length (the number of typed characters) have positive effects on user review helpfulness, which was moderated by product type (i.e., experiential vs. utilitarian product). Similarly, but more theoretically, Gan, Cao, and Jones (2012) developed a cognitive model illustrating the cognitive process of evaluating user reviews. They found an inverse “U-shaped” effect of word counts on review helpfulness.

Also, some researchers found the emotions induced by the presence of user reviews have an effect on review usefulness. Kim and Gupta (2012) studied the effect of emotions in user review on helpfulness perception. Part of the emotions in their study were conveyed through emoticons (e.g., JJJ) and bold capital letters (e.g., WORKS WELL!!), which can be classified as visual cues. They found that these...
emotional cues do not influence users’ perception unless there are multiple expressions in multiple user reviews.

A more direct study was done by Cao, Duan and Gan (2011), which used the number of helpfulness votes a review actually received as the measurement of helpfulness perception. They used a text mining methodology to analyze online reviews, in terms of the basic (e.g., “pros” and “cons”), stylistic (e.g., writing style), and semantic characteristics (e.g., make sense or not) of user reviews. What they found is that the semantic characteristics have the most impact.

Most of these studies only dealt with one or several aspects of the visual cues without a good verification for why they anchored those specific characteristics. Thus, systematic investigation is needed to fully understand how reviews play a role in product evaluation to ultimately develop better online-review mechanisms.

Hypothesis

We hypothesized that some visual cues contained in online user reviews attract and focus the attention of online shoppers, and these visual cues have higher user preferences than others. Therefore, by utilizing particular combinations of these visual cues, online-review mechanisms can be product developed to facilitate the online purchase decision process.

Taxonomy

Studies of the way online users search for information on a web page suggest people skim and scan through the content presented, searching for keywords to enable quickly decide on the relevance of the website to their needs (Weinreich, Obendorf, Herder, & Mayer, 2008). Such keywords act as visual cues which attract and retain the attention of users. The attention is retained until the information that the cue provides is processed, then it moves on in search of the next cue. This process is repeated until the person has acquired and processed enough information to make a decision on the usefulness of the site in addressing his/her concerns or needs. For online product reviews, we assume that users adopt the same scan and skim strategy, noticing cues, processing the information they present, and using the information provided to decide whether to purchase the product or not. This assumption is based on the fact that user reviews are presented on a webpage and therefore can be considered information on that webpage. The skim and scan strategy makes the manner in which the information is provided important. In addition to confirming the skim and scan strategy for online reading, eye-tracking studies of reading online also indicate that the chances a reader will fixate on a section of what is presented to read it depends among other things, on whether the word is a content word or function word (Richardson & Spivy, 2004). Here also, the words serve as cues, determining whether users focus on the content or not. Therefore, it appears that the structure and content of online user reviews are cues that can attract, and retain the attention of shoppers who use them, which can in turn influence the decisions users make about the products. Figure 1 shows our preliminary taxonomy of visual cues.

Figure 1: Preliminary taxonomy of Visual Cues

METHOD

Participants

A total of 114 participants were recruited for this study on a voluntary basis. Participants were randomly assigned to one of four experimental conditions based on a first-come-first-serve basis. There were two independent variables: 1) review layout and 2) ordering of the review valence. Review layout had two levels: paragraph and bulleted. Ordering of the review valence had two levels: Positive reviews presented first and negative reviews presented first. Note that although the taxonomy of possible conditions was proposed in our study, we initially focused on the two independent variables in the current study, for an exploratory reason.

There were four experimental conditions: 1) positive paragraph reviews followed by negative bulleted reviews (PP + NB), 2) negative paragraph reviews followed positive bulleted reviews (NP + PB), 3) negative bulleted reviews followed by positive paragraph reviews (NB + PP), and 4) positive bulleted reviews followed by negative paragraph reviews (PB + NP).

Each participant was presented with one of the four experimental conditions previously mentioned, which constituted a between-subjects experimental design. Four reviews were presented to each participant. Of the four reviews presented, two were positive and two were negative. The same (in terms of content) reviews used for one participant were presented across all participants, regardless of whether they were in displayed in bulleted or paragraph format, or whether positive versus negative reviews were presented first.

Apparatus and Procedure

The apparatus of this study utilized a medium fidelity mockup of the Amazon website interface. In order to present this interface to the participants for them to interact with, Qualtrics was utilized. Qualtrics is an interface for designing and administering surveys. The Amazon mockup was a seamless combination of modvied print screen shots from the actual website, and was integrated into Qualtrics as though it was a question element in the survey.

In order to recruit participants for our study, Amazon Mechanical Turk (MTurk) was utilized. MTurk is an example of a crowdsorcing interface. We used MTurk to direct participants to our surveys on Qualtrics. Participants were compensated $0.10 to take part in our study. All the payments were automatically handled by MTurk once we set up an active account and launched the surveys via Qualtrics. The
restrictions on MTurk were set to “90% approval rate” to allow for quick recruitment of participants.

Once the participants were directed to Qualtrics, a consent form was presented. Participants were instructed to carefully read the consent form and decide whether they “agree” or “disagree” to continue with the study. If they agreed, further instructions were provided on the second page and the “simulated” Amazon website was presented in a typical scroll format. If they disagreed, they were directed to the end of the survey and the study was terminated. Once in the “simulated” Amazon section of the survey on Qualtrics, participants were presented with a GPS product to evaluate and determine whether they would buy it or not. The participants were shown a GPS product webpage, on which consisted of: the product picture, customer reviews, user ratings, price, etc., as on the real Amazon website. The only things that were manipulated in our study as mentioned before were the layout (paragraph versus bulleted) of user reviews and the ordering of the review valence (positive versus negative reviews were presented first). Figure 2 illustrates an example of what the participant would see for the “simulated” Amazon website. Note that the words in the figure may be impossible to read and therefore it is to be used just to get a sense for how the “simulated” Amazon website compares to the real one that you may be familiar with.

After completing the evaluation task of the GPS product, the participants would then scroll down in Qualtrics and answer a series of questions, both relevant to the experimental task and not relevant (demographic information). The questions were intended to help us evaluate three dependent variables: 1) extent to which the participant would buy the product 2), decision to buy the product, and (3) layout preference. The extent to which the participant would buy the product was evaluated on a 5-point likert scale, where 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree. The decision to buy the product was evaluated based on a yes or no response. Finally, layout preference was evaluated based on a participant’s subjective preference for paragraphs or bulleted, utilizing the 5-point rating scale previously mentioned and a few multiple choice questions. See appendix for the full list of questions used in our survey.

All in all, the task developed was intended to mimic an everyday online shopping decision making situation. At the same time, we did not want participants’ performance to be influenced by individual differences or preference, hence the reason we picked a GPS device as the product to be evaluated. A GPS typically isn’t something people shop for on a day to day basis, when compared to other products such as clothes and personal accessories to wear.

RESULTS

To ensure the quality of the data, we asked two questions: “I prefer the paragraph format over the bullet format” and “I prefer the bullet format over the paragraph format”. Forty-three participants were excluded because of their inconsistent answers in the two questions. The following analyses were based on the remaining 71 participants, among whom 24 did condition 1, 17 did condition 2, 15 did condition 3, and 15 did condition 4.

We collected basic demographic information from the users. In terms of age, 29 of them were 18-24, 30 of them were 25-34, 6 of them were 35-44, 3 of them were 45-54, three of them was 55-64, and two of them were 65 and over. Twenty-seven of them were female and 43 were male. Ten of them were white/Caucasian, five were Black/African American, 49 were Asian, one was Native American, and one was Pacific Islander. We also asked how often they shopped online. Most of them had online shopping experience: 48 shopped online less than once a month; nine shopped online less than once a week; and 12 shopped online more than once a week.

Overall Trend

When asked about which format they preferred, most of the participants (46 out of 71) preferred paragraph to bulleted. Contrary to that, there were more people who tended to think the reviews presented in bulleted format were most helpful than those presented in paragraph format. People thought the reviews presented in bullet format were easier to understand than those in paragraph format. At least based on subjective self-reports people are more likely to “read all” of the reviews in bulleted format than those in paragraph format. The participants “only scanned through” the reviews in bulleted format more than those in paragraph format, which contradicted the results from the “read all” questions. See Figure 3 for a summary of the results.

Analysis Based on Conditions

We then separated the answers by conditions |Condition 1 = (PP+PB), Condition 2 = (NP+PB), Condition 3 = (NB+PP), Condition 4 = (PB+NP). The analyses based on conditions were mainly focused on the “purchase decision” question: “Based on the simulated user reviews, would you purchase the GPS, not purchase the GPS, or Not sure?” These analyses were based on two hypotheses: 1) If the
bullet format attract more attention, then the positive reviews in bullet format would lead to more purchase decisions than the negative reviews in paragraph format; 2) If users tend to read the first few reviews more than the later ones, then the positive reviews first would lead to more purchase decisions than the negative reviews first. Unfortunately, the results did not confirm the above mentioned two hypotheses. All the purchase/not purchase decisions in the four conditions were very similar (see Figure 4).

**DISCUSSION**

**Overall Trend**

It was interesting to find that most of the participants preferred reviews in the paragraph layout as opposed to the bulleted layout. This didn’t make sense, considering that most of the data collected from those very same participants were supportive of the bulleted format, notably: participants said the bulleted format was more helpful, easier to understand, and quicker to go through when evaluating whether or not to buy the GPS product. This inconsistency may be attributed to the fact that those participants probably rushed through the survey in an attempt to “just complete it”, while not fully paying attention to the details of the question. Additionally, this may also be the reason why there are a fair amount of neutral responses across the results.

The next interesting discussion point is in comparing the “read all” data with the “only scanned through” data. Participants agreed to have scanned through bulleted reviews more often than they scanned through paragraph reviews. This result points to the fact that something about the bulleted format allowed the reader to evaluate the user review information more quickly. Maybe the answer is in the ability to identify more keywords, but that would be difficulty to determine at this stage.

Note that more participants “read all” of the bulleted reviews than the paragraphed reviews. This shouldn’t be so, since we expected that the “only scanned through” data would be the reverse, to some extent, of the “read all” data. One logical explanation for this unmet expectation is the fact that some participants were probably confused with what the two questions we were asking. For instance, some of them probably though “scanned through” meant “read” rather than “filter” or “ignore some details”. Because of this discrepancy in the results, we took a further look at the demographic data. What we noticed was that the majority of people taking the survey were Asian. We cannot assume or infer that English was not their first language, hence why the difficulty in interpreting the questions may have occurred. However, it made us aware that we should have probably collected demographic data on language proficiency.

Another point to note when comparing read all” data with the “only scanned through” data is the fact that the major amount of participants “read all” for both the paragraph and bulleted reviews. This is interesting and goes against previous literature that participants don’t “read all” user review. Nevertheless, this could be an erroneous finding considering that only four reviews were presented, and therefore cannot be considered a burdening amount for individuals to mentally process.

**Analysis Based on Conditions**

In terms of the four different conditions we assessed, purchasing decisions to buy or not buy the product were not noticeably influenced by the layout and/or valence of the user reviews. We expected that if bulleted reviews attracted more attention, positive bulleted reviews would increase the likelihood that the GPS product would be purchased. However, this was not the case. Additionally, if users tend to read the first few reviews more than the later ones, then the positive reviews first would lead to more purchase decisions than the negative reviews first. Again, this was not the case. Therefore, if we carefully look at the results for purchasing decisions, the decision to buy the GPS product was quite high across all conditions. This suggests that the participants in the study were probably not money conscious, and regardless of what they thought of the product, they would buy it just because that option was available to them without any real penalty.

**CONCLUSION**

There was no conclusive evidence or validation that the user review layout and/or order of valence influences a
person’s decision to buy a product. However, we did notice a strong preference for the user reviews presented in bulleted format. This bit of information leads us to speculate that visual cues may not actually affect purchasing decisions directly, per se, but may aid in capturing a person’s attention to guide them through a series of user reviews. Therefore, by further understanding how other visual cues affect the interaction process with user review systems, we may be able to adapt future findings for the purpose of improving online shopping experience with particular websites.

FUTURE RESEARCH

The current study used a mock-up of the Amazon website, which had some drawbacks and did not allow users to interact the website. To increase the ecological validity of the study, future research should use a more realistic simulated “Amazon” website or other online shopping websites. The new website should be improved in terms of the following aspects: 1) There will be multiple reviews, more than four that were used in the current study, due to the reason that the usual products online have many reviews; 2) The reviews will be written by multiple reviewers, and some neutral names of the reviewers will be presented, due to whether the pros and cons are from the same person may influence the persuasion of the reviews; 3) The five-star rating of the single reviewers will be kept and be consistent with the contents of the reviews, due to the reason that the star ratings may provide a good summary of the reviewers’ perspective; 4) Participants will be able to click around and scroll the webpage, if they want to look at more reviews in the next few pages, due to the reason that this modification will make the webpage more like the real one.

The current study was based on the participants’ subjective evaluations of the user reviews. Future research can be conducted to combine techniques like eye-tracking to get more objective measurements. In the new design, each user review will cover a large enough region on the screen, so that it can be defined as a single region of interest to reduce the error caused by the eye-tracking device and get the corresponding fixation duration and frequency data. Hypotheses, such as 1) whether certain combination of visual cues would attract more attention (e.g., reflected by more fixation counts and longer fixation durations); 2) whether users look first at those reviews with certain combination of visual cues then jump to (or skim over) other reviews; 3) whether users mainly read the first few reviews on the first page as self-reported, will be tested in the future study.

Last but not least, the current study only proposed an exploratory taxonomy of visual cues. More efforts should be put to work on a more thorough and systematic taxonomy. Also, the current study tested only two types of visual cues among many, future study can be conducted to test out other visual cues and their combinations. After testing all the possible visual cues, those are of help should be picked out and encouraged in online reviews systems. The next step is, then, to establish how to promote users to write reviews by utilizing these different types of visual cues to help the later online shoppers to make better decisions. Related research could be done to modify the current interface, for example, to include examples of good/helpful reviews in the website. Studies can be conducted to test how these examples should be provided (e.g., text or pictures illustrating the helpful reviews) and where (e.g., on top of all the available reviews on the review page or above the input box on the page for inputting reviews).

References