MIS Quarterly

RESEARCH ARTICLE

THE DARK SIDE OF REVIEWS: THE SWAYING EFFECTS OF ONLINE PRODUCT REVIEWS ON ATTRIBUTE PREFERENCE CONSTRUCTION[1]

Qianqian Ben Liu
Department of Information Systems, College of Business, City University of Hong Kong,
Hong Kong, CHINA {ben.liu@cityu.edu.hk}

Elena Karahanna
MIS Department, Terry College of Business, University of Georgia,
Athens, GA 30602-6273 U.S.A. {ekarah@uga.edu}

Based on the "wisdom of the crowd" effect, consumer-generated online reviews are supposed to help consumers make more accurate product evaluations. However, the large amount of information in the reviews, coupled with conflicting opinions, can make it difficult for consumers to identify and consider those attributes relevant to their decision. Thus, while online product reviews are generally believed to empower consumers, we suggest that they may have "swaying" effects in that attribute preferences (i.e., the relative importance consumers place on various product attributes in product evaluation) are more heavily influenced by characteristics of the online reviews rather than by the relevance of the attributes to the consumers' decision context. We propose that three characteristics of online reviews affect the assessment of attribute preferences: (1) the amount of information about attribute-level performance, which is often unevenly distributed across attributes, (2) the degree of information conflict about attribute-level performance, and (3) the relationship between the overall numeric rating and the attribute-level performance information in the reviews. We test our hypotheses in two randomized experiments and a free simulation study. Results from the three studies show that the three review characteristics influence attribute preferences and that their effects are strong enough such that attribute preferences are influenced more by these online review characteristics than by the relevance of the attributes to the consumers' decision context. Our work, which illustrates a dark side to online reviews, has implications for both online word-of-mouth and preference construction research.

Keywords: Online reviews, word-of-mouth, preference construction, product evaluation, wisdom of the crowd, choice architecture, nudge, process tracing, verbal protocol analysis, decision bias

Introduction

Based on the "wisdom of the crowd" effect, consumer-generated online reviews are supposed to help consumers make more accurate product evaluations. However, the large amount of information in the reviews, coupled with conflicting opinions, can make it difficult for consumers to identify the product attributes that are most relevant to their decision. It has been shown that, in the context of advertising, when information about irrelevant attributes is available, consumers may treat irrelevant attributes as though they were critically important in evaluating the product (Brown and Carpenter 2000). Therefore, consumers' attribute preferences, that is, the relative importance consumers place on

[1]Soon Ang was the accepting senior editor for this paper. Andrew Burton-Jones served as the associate editor.

The appendices for this paper are located in the "Online Supplements" section of the *MIS Quarterly*'s website (http://www.misq.org).

various product attributes when evaluating a product (e.g., Noseworthy et al. 2012; Scholz et al. 2010), might not be constructed based on careful consideration of the relevance of the attributes to the consumers' decision context but rather may be primarily influenced by characteristics of the information environment. As such, while online product reviews are generally believed to empower consumers, they may also have "swaying" effects in that after reading reviews people may overweigh irrelevant attributes but underweigh relevant attributes when assessing a product.

We propose that three prominent *characteristics of online reviews* influence the construction of attribute preferences: (1) the amount of information about attribute-level performance, which is often unevenly distributed across attributes, (2) the degree of information conflict about attribute-level performance, and (3) the relationship between the overall numeric rating and the attribute-level performance information in the reviews. We further suggest that the construction of attribute preferences can potentially be *swayed* by reviews such that attribute preferences are influenced more by the proposed review characteristics than by the relevance of the attributes to the consumers' decision context.

The impact of online product reviews on *attribute* preferences is an overlooked, yet important, issue. Existing studies on online reviews have examined the impact of online product reviews from both the seller and consumer perspectives. From the seller perspective, the literature typically examines the impact of online product reviews on product sales (e.g., Chevalier and Mayzlin 2006, Duan et al. 2008, Liu 2006). More related to the current research are the studies that examine the impact of online product reviews from the consumer perspective. These typically look at the impact of online reviews on consumers' overall attitude toward the reviewed product and their purchase intention (for a review, see Cheung and Thadani 2012) or the factors that make a review perceived helpful in purchase decision-making (e.g., Mumdabi and Schuff 2010, Yin et al. 2014). This stream of work focuses on review message characteristics (e.g., argument quality, message sidedness, message length, emotions embedded in the message etc.), review source characteristics (e.g., reviewer's reputation and geographic location), and the moderating role of consumer involvement, prior knowledge, and product type (for a review, see King et al. 2014).

Given the rich attribute-level information in the reviews, it is unlikely that people only evaluate the product holistically when reading the reviews. Although consumers may use a variety of strategies in evaluating products and making purchase decisions, accurate and justifiable product evaluation requires people to identify attributes that should be used to evaluate the product and learn the performance of the product on these attributes (Bettman 1998). To the best of our knowledge, we are the first to theorize and test the impact of online product reviews on attribute preference construction. Our research takes an information-processing approach, which takes into account how people actually process information in an information-rich environment. This approach generates a more detailed understanding than prior approaches of how online product reviews influence consumers' product evaluation. There is a strong association between people's perceptions of attribute importance and their information acquisition and processing activities in product evaluation (MacKenzie et al. 1986). Therefore, from a practical perspective, understanding the impact of online product reviews on attribute preference construction can provide insights into the design of online review systems that can nudge consumers toward more informed decisions (Thaler and Sunstein 2008).

The paper is organized as follows. We first review related literature and develop our hypotheses. We then describe our research methods and results. Specifically, we test our hypotheses via three studies, two randomized experiments and a free simulation study, all of which focus on the evaluation of a single product that has multiple attributes.[2] Finally, we discuss our research findings and their theoretical and practical implications.

Related Literature

Preference is generally defined as people's disposition to choose one object or course of action over another (Simonson 2008). Researchers have shown that people may reverse their preferences when the same options are described differently (Tversky and Kahneman 1986), when options are presented with or without some extraneous options in the choice set (Shafir et al. 1993), and when different methods are used to measure preferences (Slovic and Lichtenstein 1983). These findings led to the general consensus that preferences are labile, inconsistent, and subject to a range of contextual factors (Payne et al. 1992). Payne and his colleagues were among the first few researchers to systematically investigate the constructive nature of preference. Defining preference construction as a process of arriving at a decision, they suggested that rather than relying on utility maximization, people

[2]Our study examines attribute preferences for a single product under the multi-attribute decision making framework (i.e., for products that consumers evaluate based on their attributes). People may use a holistic approach, rather than an attribute-based approach, in evaluating experience goods (e.g., food). Also, attribute preferences can be different when the product is evaluated in a comparative (i.e., considering multiple product alternatives) versus a noncomparative (i.e., single product) context.

use a wide variety of strategies and heuristics (i.e., simplifying methods) to construct preferences (Bettman et al. 1998). They concluded that the use of different strategies and heuristics, affected by the decision context (e.g., decision goals and the task environment), is why people's preferences over the same set of options may change across different contexts.

Similar to the general literature on preference construction, extant research on attribute-level preferences has examined the impact of attribute description and presentation, measurement methods, and the decision-making process on attribute importance weighting. For example, Weber et al. (1988) showed that people attach greater importance weight to attributes presented in more detail than to the attributes described more generally. Similarly, MacKenzie et al. (1986) found in the context of advertising that the concreteness of attribute description increases the importance weight attached to the attribute. Researchers have also shown that different measurement methods (e.g., free elicitation, direct rating, conjoint method, etc.) could lead to differences in attribute importance weighting (van Ittersum et al. 2007). Furthermore, the decision-making process also influences attribute importance weighting. The theory of reason-based choice views decision making as a process of generating reasons for and against the available options (Shafir et al. 1993). Therefore, people will attach greater importance weights to the attributes based on which clear reasons can be generated for preferring one option compared to its competitors. Based on this argument, Brown and Carpenter (2000) showed that people treat irrelevant attributes as though they were critically important when these attributes differentiate the available options.

There is also a literature specifically looking at the construction of attribute preferences in the *online* environment. Since information presentation is extremely flexible in the online environment and many information-processing aids are readily available, much attention has been given to the information presentation factors and ease-of-processing effects.

The *salience* and *partitioning* of attribute information are two major information presentation factors examined in this literature. Researchers have shown that product attributes receive greater weights than they normally would when these attributes are made more salient by explicitly displaying these attributes on the website (Häubl and Murray 2003) or by priming through subtle differences in the background of the website (Mandel and Johnson 2002). Partitioning an attribute into more detailed attributes also affects attribute preferences. An attribute presented as separate categories tends to receive greater importance weight than attributes presented under an umbrella category. For example, on an online dating website, the separate categories of "intelligence," "sense of human," "kindness," and "generosity" together receive a greater importance weight than the umbrella category of "personality" (Martin and Norton 2009). Clearly the effects of information presentation are not unique to the online environment since it is possible to manipulate attribute information similarly in offline settings. However, information structuring and restructuring is much more flexible in the online environment (e.g., it is difficult to rearrange the physical display of product information within a short amount of time in a physical store, but it is extremely easy in the online environment). Because of the ease and flexibility of information presentation, attribute preferences are likely more susceptible to information presentation in the online environment.

Furthermore, the online environment provides various functions to aid consumers' information processing. Researchers have consistently found ease-of-processing effects (i.e., when information about an attribute is made easier to process, people will attach a greater weight to that attribute). For example, an attribute on which product alternatives can be sorted receives a greater importance weight in the evaluation of the products (Quaschning et al. 2014). Similarly, when price comparison was made easy, people's price sensitivity increased, but when quality information was made easy to evaluate, people attached greater weight to quality and their price sensitivity deceased (Lynch and Ariely 2000).

Our reading of the literature is that the extant studies have a focus on attribute preference construction based on *seller-provided information* (e.g., product specifications, descriptions of various attributes, and the accompanying information-processing aids). Our research differs from the extant studies in that we examine the impact of a repository of *consumer-generated information* on attribute preference construction. The breadth, depth, and amount of information in consumer-generated reviews greatly exceed those of seller-provided information. For example, multiple pieces of information with varying levels of detail often exist for many product attributes and the amount of information for each attribute may vary greatly, with some attributes discussed extensively in the reviews and others considerably less so. Further, the information may be consistent or conflicting for the same product attribute (e.g., some consumers may evaluate the product positively on an attribute and others negatively). Therefore, consumer-generated reviews constitute an information-rich environment where multiple characteristics of the environment influence attribute preference construction and product evaluation. Our research is a step toward identifying specific characteristics of this information environment that have an impact on attribute preference construction.

Theory Development

Our theory development draws from the constructive preference perspective (Bettman et al. 1998; Payne et al. 1992). A major tenet of the theory is that preferences are shaped by the interaction between the properties of the *information environment* of the choice problem (Payne et al. 1992) and the properties of the *human information-processing system* (e.g., basic memory principles and cognitive biases). For example, presenting the outcomes of medical programs in terms of "lives saved" versus presenting them in terms of "lives lost" induces different preferences for these programs because of the interaction between information presentation and people's tendency to perceive losses as greater than the equivalent gains (Tversky and Kahneman 1986).

Attribute preference is defined as the relative importance consumers place on the various attributes of a product in product evaluation or choice (e.g., Noseworthy et al. 2012; Scholz et al. 2010). Attribute preference construction occurs as part of learning about the product from the reviews. The constructive preference perspective would suggest that the key to understanding attribute preference construction is to examine the characteristics of the reviews (i.e., the information environment) and how the human information-processing system operates in the presence of these review characteristics.

Since our focus is attribute-level preference construction, we examine the characteristics of the reviews that pertain to attribute-level information. Prior studies have examined two prominent review characteristics: the volume of the reviews (e.g., Liu 2006) and the variance in the numeric ratings (e.g., Sun 2012). Both of these describe information at the product level of analysis and provide a holistic assessment of the product. We examine information volume and variance at the attribute-level (i.e., the amount of attribute-information and the degree of information conflict about attribute-level performance). Furthermore, an underexplored but important review characteristic is the relationship between overall numeric rating and the individual reviews. Prior research showed that the conflict between the overall numeric rating and the individual review rating affects people's perceptions of the review (Qiu et al. 2012). We extend this characteristic to the relationship between the overall numeric rating and the attribute-level performance information in the reviews. To sum up, we examine three review characteristics: (1) the amount of information about attribute-level performance, which is often unevenly distributed across attributes, (2) the degree of information conflict about attribute-level performance, and (3) the relationship between the overall numeric rating and the attribute-level performance information in the reviews.

The Amount of Information about Attribute-Level Performance

We posit that the amount of attribute-level performance information affects the importance weight assigned to that attribute via two different and non-mutually exclusive routes. The first route is an *automatic* route that operates outside people's conscious awareness. A large amount of information for an attribute increases the accessibility of that attribute in memory, which will automatically lead people to judge that attribute as important. The second route is a *conscious inference* route (i.e., people consciously infer the importance of an attribute based on the amount of information available for that attribute). The distinction between the automatic and conscious inference routes parallels the two systems of the mind where System 1 processes are automatic and intuitive and System 2 processes are deliberate and logical (Kahneman 2011).

First, the amount of information affects information-processing activities including attention, encoding, consolidation, and retrieval. When people pay attention to a piece of information, it registers in working memory (WM), making this information available for further processing, for more permanent storage in the long-term memory (LTM), or both. Given that WM has a small capacity, a large amount of information for an attribute draws attention to this attribute at the expense of other attributes. WM is a holding buffer for LTM. As long as a piece of information resides in WM, there is a tendency to transfer it to LTM, which is known as "encoding" (Lieberman 2012). Thus, repeated attention to an attribute leads to better encoding and consolidation of that attribute in LTM, making this attribute highly accessible in LTM (Lieberman 2012). When judging attribute importance, people need to first retrieve that attribute from LTM. When the attribute is retrieved from memory, memory for that attribute will be further strengthened and competing memories for other attributes will be less accessible afterwards (i.e., "retrieval-induced forgetting"; see Anderson et al. 1994; for an example, see Appan and Browne 2010). Therefore, a large amount of information for an attribute increases accessibility of that attribute in LTM. According to the mere-accessibility effect (Menon and Raghubir 2003), the accessibility of attribute and the resulting ease-of-retrieval will automatically lead people to judge that attribute as important. The mere-accessibility effect occurs involuntarily, effortlessly, and outside of people's awareness (Menon and Raghubir 2003).

Second, people also consciously use the amount of information for an attribute to infer the importance of that attribute. The social learning perspective suggests that people infer others' beliefs from their actions and incorporate these beliefs in their own judgment (for a review, see Chamley 2003).

When an attribute is mentioned in a review, the inference is that the attribute must be important to the reviewer, otherwise it would not have been mentioned. If an attribute is mentioned in many reviews, the inference is that the attribute must be important to many people. As a result, an information cascade occurs where the consumer also views the attribute as important because he or she "follows" the inferred belief irrespective of his or her decision context (Bikhchandani et al. 1992), a phenomenon also known as herd behavior (Banerjee 1992).

The automatic and conscious inference routes both predict that increasing the amount of attribute information will lead to a greater importance weight attached to that attribute. We thus posit the following:

> *H1: The greater the amount of attribute information read in the reviews, the greater the importance weight attached to the attribute.*

The Degree of Information Conflict

A prominent characteristic of consumer-generated reviews is the presence of conflicting information about product performance at both the attribute and product level. Conflicting information often elicits a deeper level of information processing, since resolving conflicting information requires a careful consideration of the positive and negative information in relation to each other (Maheswaran and Chaiken 1991).

The level of information processing ranges on a continuum from shallow *maintenance rehearsal* to deep *elaborative rehearsal* (Craik and Tulving 1975). Maintenance rehearsal is a superficial and passive level of processing. Simply seeing a comment from a review without thinking about its meaning is an example of maintenance rehearsal. In contrast, elaborative rehearsal is a deeper and active level of processing that involves thinking about the meaning of the attended information and connecting it to other information held in memory. People are, in general, spontaneously motivated to resolve conflicting information to form an integrated conclusion (Hastie 1980). Therefore, conflicting information will induce elaborative rehearsal. For example, when one sees conflicting comments on a camera's battery life, this person may read the review carefully to understand the context in which the comments were made so that he or she can decide if this comment is valid. The person may also seek corroborating evidence or converging opinions from additional reviews. In this example, resolving the conflict requires elaborative rehearsal, a more thoughtful processing of the reviews. Elaborative rehearsal is more effective in encoding information into LTM, which will make the information more accessible in memory (Craik and Tulving 1975). Again, based on the mere-accessibility effect, we argue that the increased accessibility, as a result of the deeper level of processing induced by conflicting information, will make people assign a greater importance weight to the attribute.

> *H2: The greater the degree of information conflict for an attribute, the greater the importance weight attached to the attribute.*

Presence of Both Numeric Overall Rating and Textual Attribute Information

As people read reviews, they will constantly judge if the product is good or bad (Hastie and Park 1986). This ongoing judgment will be anchored on the overall numeric rating for the product (typically represented by a star-rating) and adjusted as new information is processed. This anchoring effect will be strong for the following reasons. First, the overall numeric rating is visually salient and easily accessible. Second, the fact that the overall numeric rating is the average of all individual ratings may lead people to believe that the overall rating is more "representative" than any individual review. Third, people normally prefer numeric information to textual information because numeric information is more concrete and less effortful to process than textual information. Therefore, the overall numeric rating's visual salience, seeming representativeness, and concreteness will lead to a strong anchoring effect. The adjustments made based on reading the reviews will be insufficient as in other anchoring and adjustment contexts (e.g., Epley and Gilovich 2006) such that one's ongoing judgment will be primarily driven by the overall numeric rating. We call the ongoing judgment that precedes the final judgment an *emerging judgment*.

People's attribute-level evaluation based on their reading of the reviews (e.g., "the camera's battery life is short") can be coherent or incoherent with the emerging judgment. For example, a 4.5 (out of 5) overall numeric rating will lead to a positive emerging judgment on the camera. However, if the reviews say that the camera has a short battery life (a negative evaluation on the "battery life" attribute), the negative attribute-level evaluation is incoherent with the positive emerging judgment. The coherence or incoherence between the emerging judgment and attribute-level evaluation will affect the importance weight attached to the attribute.[3]

[3]It is worth noting that the factor hypothesized here is the coherence between the emerging judgment of the product (anchored on the overall numeric rating) and attribute-level evaluation. Coherence is also what we measured in all studies reported in the paper. In the randomized experiments, level of coherence was varied by manipulating the overall numeric rating.

Specifically, cognitive consistency theories suggest that judgment and decision-making processes tend to settle at a state of coherence where the favored option must be supported by the attributes that are deemed as important (i.e., the favored option must be positively evaluated on the "important" attributes) and the attributes that do not support the favored option must be deemed unimportant (Simon et al. 2004). With an emerging judgment in mind (e.g., "the camera looks like a good option"), people tend to bolster attributes that support the emerging judgment and suppress attributes that do not support the emerging judgment so as to arrive at the state of coherence (Nickerson 1998). In the earlier example, the negative attribute-level evaluation (short battery life) does not support the positive emerging judgment (the camera looks like a good option), which is anchored on the high overall numeric rating (4.5 out of 5). To reduce the incoherence, people will suppress the battery life attribute by assigning a low importance weight to it. The coherence-directed processing is a type of confirmation bias whereby people construct attribute preferences to support their emerging judgment.

> *H3: The greater the level of coherence between an individual's attribute-level performance evaluation and emerging judgment of the product, the greater the importance weight attached to the attribute.*

Can Online Reviews Sway Attribute Preference Construction?

The context of a specific purchase decision often makes some product attributes more relevant than others to the consumer. For example, a different set of attributes is relevant in assessing whether to purchase a camera for a 10-year-old child (e.g., ease of use, durability) versus for an expert photographer (e.g., image quality, manual control). Payne et al. (1999) suggest that the quality of constructed preferences can be assessed and that well-constructed preferences are based on careful consideration of the aspects of the decision that are most critical to the individual. This means that well-constructed attribute preferences should be driven by the relevance of the attributes to the decision context. In the preceding section, we argued that consumers' attribute preferences are influenced by three characteristics of the reviews. An important question arising from this argument is whether online reviews could sway preference construction such that the assessment of attribute importance is more heavily influenced by characteristics of the reviews than by the relevance of the attribute to the decision context. Consumer-generated reviews are supposed to help people make a more informed product evaluation. However, if the review characteristics overpower the relevance of the attributes in the assessment of attribute importance, design interventions are necessary to nudge people to more carefully consider the relevance of the attributes to their decision context.

We posit that although attribute preferences will be influenced by both the relevance of the attributes and the review characteristics, the review characteristics are likely to play a more dominant role in shaping attribute preferences. In the general domain of pre-decisional information acquisition, available evidence shows that people are substantially responsive to contextual factors that are irrelevant to decision quality and only weakly responsive to factors that are relevant to decision quality (Connolly and Thorn 1987). Such suboptimal information acquisition persists in simplified tasks where the relevant decision quality factors are made easy to assess as well as in incentivized tasks where real money is at stake (Connolly and Thorn 1987). In our context, for attribute relevance to play a dominant role in constructing attribute preferences, people need to keep in mind the decision context throughout the information acquisition task and process review information in light of the decision context. Given the limits of working memory, the large amount of information in the reviews, and the fact that humans are "cognitive misers," people are not likely to perform such a cognitively demanding task.

We suggest that the review characteristics will play a dominant role in constructing attribute preferences for the following reasons. First, as we have argued, both the amount of attribute information and the degree of information conflict increase the attribute importance weight partly through the mere-accessibility effect. The mere-accessibility effect, occurring outside people's conscious awareness, is difficult to discount (Menon and Raghubir 2003). Second, the amount of attribute information increases the attribute importance weight because people tend to infer attribute importance from the amount of attribute information. Such an inference, prompted by people's herd instinct, is less effortful than retrieving the decision context and examining the reviews against the decision context. Third, people only know if their product evaluation is accurate after purchasing and using the product. When accuracy of product evaluation cannot be assessed immediately, justifiability of product evaluation becomes a more salient goal of decision making (Bettman et al. 1998). The coherence-directed processing, whereby attributes are weighed to support the emerging judgment, is instrumental in generating a justifiable product evaluation. All of these reasons lead to the prediction that the review characteristics will dominate the influence of attribute relevance to the decision context in constructing attribute preferences.

H4: Online product reviews have a swaying effect, that is, the assessment of attribute importance will be more heavily influenced by characteristics of the reviews than by the relevance of the attribute to the decision context.

In summary, our theoretical development presented three characteristics of online reviews that influence how people judge the importance of various attributes in product evaluation. We also theorized that these review characteristics may sway attribute preferences such that attribute importance weights are not constructed based on careful consideration of the relevance of the attributes to the consumers' decision context but rather are primarily influenced by online review characteristics.

Research Methods, Design, and Results

We conducted three studies to test our hypotheses. The first study was a randomized experiment where we manipulated the three hypothesized factors and examined their effects on the constructed attribute preferences. Participants were paid a flat fee for participation in the first study. Therefore, a potential threat to validity was whether results were influenced by participants not being sufficiently motivated to process review information. To rule out lack of motivation as a potential threat, we ran a second study in which a monetary incentive was used to induce high motivation to process review information. These two studies were randomized experiments that enabled us to assess causality. The third study was a free simulation experiment to provide more realism and to allow for higher generalizability.[4]

Study 1

We ran a 2 × 2 × 2 factorial experiment where we varied the amount of attribute information, the degree of attribute information conflict, and the coherence between attribute-level performance evaluation and emerging judgment on product performance. Participants were given a scenario and were asked to evaluate a digital camera based on 10 reviews.

[4]Free simulation differs from controlled experiment in that in a controlled experiment the experimenter maximizes control over the nature and timing of experimental events whereas in a free simulation events and their timing are determined by both the researcher and the behavior of participants (Fromkin and Streufert 1976; Jenkins 1985).

Participants and Design

Fifty-two students participated in the study (31 females and 21 males, M_{age} = 21.92, SD_{age} = 1.74). Participants were recruited through a subject pool at a public university and were paid a flat fee of $5 for their participation. Seventy-eight percent of participants owned a digital camera at the time of the experiment, and 58% percent had at some point researched digital cameras on the Internet. The participants reported to have low to medium knowledge of digital cameras (M = 3.60, SD = 1.35 on a 1 to 7 scale).

A panel of five experienced photographers helped us create the following scenario:

> Your friend Bob is considering buying a digital camera for his grandfather. Bob's six-year-old sister is going to stay with his grandparents for the summer and the grandfather wants to take many pictures of Bob's sister. The grandfather is in his 60s and his vision is not what it used to be. Although Bob's grandfather is showing some interest in photography, he is not very tech-savvy. Bob asks you to help choose a digital camera for his grandfather.

The expert panel selected eight camera attributes: four attributes including autofocus, intuitiveness of operation, image stabilization, and LCD screen were deemed as relevant to novice photographers like Bob's grandfather and the other four attributes including optical viewfinder, manual mode, macro mode, and raw format are only relevant to more advanced photographers. To ensure that our selection of the relevant and irrelevant attributes was valid, we invited a similar group of 50 students who were not participants in the study to rate the importance of these attributes to Bob's grandfather. These students were presented with the scenario and brief explanations of the attributes and were asked to allocate 100 points among the attributes based on attribute importance. Results show that the relevant attributes received significantly higher points than the irrelevant attributes ($M_{difference}$ = 18.620, p-value < .001, also see panel (a) of Figure 1).

Participants were presented with 10 reviews that jointly contained information about the eight attributes. We employed a 2 (attribute-information amount: high or low) × 2 (attribute-information conflict: high or low) × 2 (overall numeric rating: high or low) mixed factorial design with attribute-information amount and attribute-information conflict being the within-subjects factors, and the overall numeric rating being a between-subjects factor.

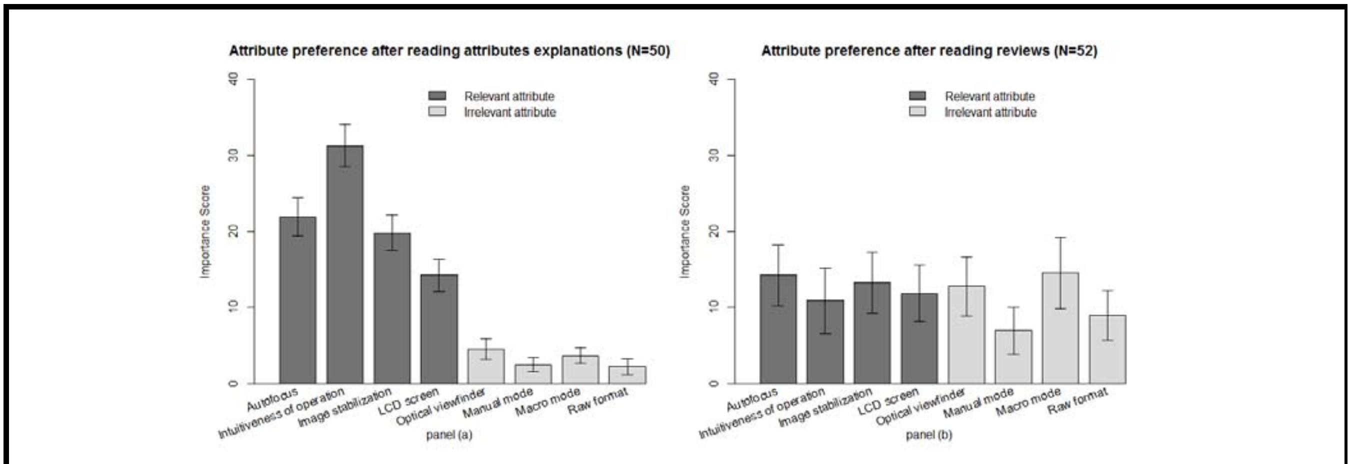


Figure 1. Participants' Constructed Attribute Preferences

The amount of information available for an attribute (A): In the high attribute-information amount conditions (denoted as H_A), all 10 reviews contained information about the attribute. In the low attribute-information amount conditions (denoted as L_A), only 2 out of the 10 reviews contained information about the attribute.[5]

The degree of information conflict for an attribute (C): In the high attribute-information conflict condition (denoted as H_C), half of the reviews that contained the attribute evaluated the attribute positively and the other half evaluated the attribute negatively. In the low attribute-information conflict condition (denoted as L_C), all reviews that contained the attribute consistently evaluated the attribute positively or negatively.

The overall numeric rating (R): We varied the numeric overall rating to elicit different levels of emerging judgment. Participants were told that the digital camera had more than 100 reviews but we only selected 10 reviews submitted by verified buyers. The numeric overall rating was displayed as either 4.5 (high, denoted as H_R) or 2.5 (low, denoted as L_R) on a 1 to 5 scale.

Each level of each within-subjects factor (i.e., H_A, H_C, L_A, and L_C) can be implemented on any of the eight attributes. For example, if a high amount of attribute information and a high degree of information conflict (i.e., H_AH_C) is implemented on the attribute of autofocus, participants will see information about autofocus in all 10 reviews with 5 of them being positive about autofocus and the other five being negative about it. We counterbalanced the implementations of the within-subjects factors so that each level of each within-subjects factor was implemented on each of the eight attributes equal times. The counterbalancing required us to create four sets of reviews (see Table 1). Each review set contains information about eight attributes (denoted as "Attr1" through "Attr8" in Table 1) across 10 reviews.[6] Each participant only read one randomly selected review set. For example, participants who were assigned to "review set 1" saw information about "Attr1" in all 10 reviews with 5 reviews evaluating the attribute positively and the other 5 evaluating the attribute negatively (H_AH_C), but participants who were assigned to "review set 3" saw information about "Attr1" only in 2 reviews with both reviews consistently evaluating "Attr1" positively or negatively (L_AL_C). In the four review sets, each level of each within-subjects factor (i.e., H_A, H_C, L_A, and L_C) was implemented on each of the eight attributes twice.

The coherence between attribute-level evaluations and the emerging judgment: We employed a matched-pair design to manipulate the coherence between attribute-level evaluations and the emerging judgment on the product. Specifically, each review set was further assigned a high and a low overall numeric rating, leading to a pair of matched groups for each review set (i.e., 1&1', 2&2', 3&3', and 4&4' in the "Experimental Group" column of Table 1). Participants in the matched

[5]In the low information amount condition, it was necessary to have two reviews (instead of just one) contain information about the attribute because we also needed to manipulate information conflict. If only one review contained information about that attribute, there would be no information conflict for the attribute.

[6]Table 1 does not show which review contains which attribute in each of the review sets. This information can be found in Table A1 in Appendix A.

Table 1. Implementation of the Experimental Factors

Review Set	Within-Subjects Factors (A: attribute-information amount, C: attribute-information conflict)								Between-Subjects Factor (R: overall numeric rating)	Experimental Group
	Attr1	Attr2	Attr3	Attr4	Attr5	Attr6	Attr7	Attr8		
1	H_AH_C	L_AL_C	H_AL_C	L_AH_C	H_AH_C	L_AL_C	H_AL_C	L_AH_C	H_R	1
									L_R	1'
2	H_AL_C	L_AH_C	H_AH_C	L_AL_C	H_AL_C	L_AH_C	H_AH_C	L_AL_C	H_R	2
									L_R	2'
3	L_AL_C	H_AL_C	L_AH_C	H_AH_C	L_AL_C	H_AL_C	L_AH_C	H_AH_C	H_R	3
									L_R	3'
4	L_AH_C	H_AH_C	L_AL_C	H_AL_C	L_AH_C	H_AH_C	L_AL_C	H_AL_C	H_R	4
									L_R	4'

Note: H and L represent high and low levels of experimental factors respectively.

groups saw the same review set but different levels of numeric overall rating. For example, both groups 1 and 1' saw "review set 1," but group1 saw a 4.5 overall rating (H_R) and group 1' saw a 2.5 overall rating (L_R). Therefore, in the matched groups, participants' attribute-level performance evaluations, informed by the same review set, should be about the same, but their emerging judgment, anchored on different levels of overall numeric rating, should be different. For example, suppose a review set is overall positive about autofocus. The group of participants who see this review set in the high overall numeric rating condition should (1) form a positive emerging judgment (anchored on the high overall numeric rating), and (2) evaluate autofocus positively (informed by the positive comments on autofocus in the review set). This will lead to coherence between attribute-level evaluation and the emerging judgment in the group. Participants in the matched group who see the same review set but see a low overall numeric rating will also evaluate autofocus positively but their emerging judgment on the camera will be negative. This will lead to incoherence between attribute-level evaluation and the emerging judgment in the matched group. This design led to eight experimental groups to which participants were randomly assigned (see Table 1).

The reviews: The 40 reviews (4 review sets × 10 reviews in each set) were created set by set. Before creating the reviews, we first *randomly* determined the placement of the attributes in the reviews (i.e., what attributes are discussed in each review of that set), and the valence and extremity of each attribute discussed in each review. The placement, valence, and extremity of the attributes in the reviews were represented numerically in a review design table (see Table A1 in Appendix A). We gathered comments on the eight attributes from real camera reviews on Amazon.com. The reviews were then assembled from these comments. The reviews started with a paragraph describing the background of the reviewer (all of the reviewers were described as being inexperienced with photography). The other paragraphs explained the attributes by discussing the implications of having a high or low value on the attributes (e.g., "a solid macro mode allows you to take good close-up pictures of small objects") and described the reviewer's evaluations of the camera on the attributes and the reviewer's experiences that support his or her evaluation.[7] Each of these paragraphs only contained information about one attribute and an attribute was either discussed only in one paragraph or not discussed at all in the review. The descriptions of reviewer background and explanations of the attributes were the same in all conditions but reviewers' evaluations of the camera on the attributes were manipulated differently in different conditions based on the review design table (see Table A1 in Appendix A). The length of the reviews ranged from 299 words to 422 words (M = 343.90, SD = 59.15).

Procedure

Participants performed the experimental task in a computer lab. After logging into the experiment website, they were randomly assigned to one of the eight experimental groups. They were first asked to complete a questionnaire that included

[7]The reviews did not contain any statement of attribute importance (e.g., "the attribute is important," "people should take into account the attribute in purchase decision," etc.) because statements of attribute importance will also affect attribute preferences so that they may confound the effects of the hypothesized review characteristics. Also, in reality it is uncommon that people directly state what attributes are important to them. For example, a text analysis we performed on the DSLR camera that has the most reviews on Amazon.com reveals that less than 2% of the reviews state what attributes are important.

questions about their age, gender, and knowledge of digital cameras. Next they were presented with the scenario and asked to list camera attributes that would be important to Bob's grandfather and indicate the relative importance of these attributes by allocating 100 points among these attributes. Participants were then presented with the 10 reviews with the designated overall rating on the digital camera.

The reviews were displayed to the participants one by one. We randomized the order of the 10 reviews and the order of the paragraphs that contained attribute information within each review. This allowed us to rule out order effects of information processing as a confound. The participants were told that the reviews were written by verified buyers of the camera. The website also displayed the name of the attributes discussed in each review and a star rating representing the valence and the extremity of each attribute (see Appendix B for an example).[8] After they finished reading each review, the participants could proceed to the next review by clicking a button. To ensure that they read all 10 reviews, the button was shown 30 seconds after the presentation of each review. Participants were told that the task is self-paced and they should read the reviews as they normally would. After they finished reading the third, the sixth, and the ninth reviews, the website asked them to evaluate the camera on a 1 to 5 scale based on the reviews they had read thus far. This allowed us to measure their emerging judgment on the camera.

After participants finished reading all 10 reviews, they were again asked to list the attributes they considered to be important to Bob's grandfather and indicate the relative importance of these attributes by allocating 100 points among these. They were then asked to make an overall evaluation of the camera on a 1 to 5 scale and indicate if they would recommend the camera to Bob's grandfather. At the end of the experiment, and to enable us to perform manipulation checks, we showed participants the eight attributes mentioned in the reviews and asked them to indicate the amount of information available for each attribute, the degree of information conflict for each attribute, and the performance of the camera on each attribute.

Results

Manipulation checks: We first checked if the participants perceived the valence and extremity of the attributes as intended. We invited 30 coders from the same subject pool to code the attributes in the 40 reviews. Each coder coded all 40 reviews (without being shown the attribute name and the corresponding attribute-level star ratings). For each review, they were required to code all of the attributes discussed in that review on a -2 to +2 scale, with -2 being extremely negative and +2 being extremely positive. We ran a multi-level ANOVA to check if the coders perceived the valence and extremity of the attributes as we intended.[9] The analysis showed that they did not perceive the valence and extremity of the attributes significantly different from what we intended ($M_{difference}$ = .116, p-value = .350).

Next we ran T-tests to see if our manipulations on the within-subjects factors of attribute-information amount and attribute-information conflict were successful. The analyses showed that participants did perceive a significantly higher amount of attribute information ($M_{difference}$ = 2.269, p-value < .001) and higher degree of conflict ($M_{difference}$ = .657, p-value < .001) when these factors were manipulated to be higher.

We manipulated the overall rating to induce different levels of emerging judgment on the camera. A MANOVA shows that different levels of overall rating led to differences in the three waves of emerging judgment collected while participants were reading the reviews (Wilks' lambda = .340, p-value < .001). A T-test further shows that the average of the three judgments is significantly higher when participants saw a higher overall rating ($M_{difference}$ = .442, p-value = .002). Moreover, in the high overall rating conditions, participants gave a significantly higher evaluation to the camera after reading all reviews ($M_{difference}$ = 1.018, p-value < .001) and they were more likely to recommend the camera (66.7% recommended in high overall rating conditions versus 16% in low overall rating conditions). From this analysis we conclude that manipulating the overall rating leads to different levels of emerging judgment.

Further, given that participants in the matched groups (e.g., groups 1 and 1' in Table 1) saw the same reviews, we expect that they would make similar attribute-level performance evaluations. T-tests support that participants' attribute-level performance evaluations in the high numeric overall rating conditions were not significantly different from those in the low overall rating conditions ($M_{difference}$ = .109, p-value = .275).

Hypotheses testing: Before the experiment, we invited a separate group of 50 students from the subject pool to rate the importance of the 8 attributes based on brief explanations of

[8]The display of attribute-level information, as commonly found in many e-commerce websites (e.g., bestbuy.com), was to reduce any misunderstanding of the attributes discussed in the reviews. Two of our pilot studies did not display the attribute-level information. Analysis of the pilot data showed that the display of attribute-level information did not materially change the results.

[9]The unit of analysis is the comment on an attribute. We performed a multi-level ANOVA to account for the fact that these comments on attributes are nested in different reviews.

these attributes. Unlike the reviews, the brief explanations have equal amounts of information for each attribute (one short paragraph for each attribute), no information conflict, and no overall rating that would cause confirmation bias. Therefore, we expect that, after reading the explanations, people assign greater weights to relevant attributes than to irrelevant attributes. This is supported by a t-test ($M_{difference}$ = 18.620, p-value < .001). However, when people were presented with the reviews, there was no significant difference in the importance weights attached to relevant and irrelevant attributes ($M_{difference}$ = 1.745, p-value = .244). We plotted the 95% confidence intervals of the importance weights from reading the brief explanations (Panel (a), Figure 1) as well as from the experiment in study 1 (Panel (b), Figure 1). The plots shows that while attribute preferences constructed by reading the brief explanations were clearly influenced by the relevance of the attributes, attribute preferences constructed by reading the reviews do not appear to be driven by the relevance of the attributes.

To test if participants' constructed attribute preferences are affected by the hypothesized review characteristics, we ran the following model:

$$\boldsymbol{y}_{ij} = \beta_{0i} + \beta_1 \boldsymbol{relevance}_j + \beta_2 \boldsymbol{pre}_{ij} + \beta_3 \boldsymbol{amount}_{ij} + \beta_4 \boldsymbol{conflict}_{ij} + \beta_5 \boldsymbol{coherence}_{ij} + \varepsilon_{ij} \qquad \text{(Model 1)}$$

where $\boldsymbol{y}_{ij}$ is the importance weight that participant i placed on attribute j (i = 1,…, 52; j = 1,…, 8) after reading the reviews, β_{0i} is the participant fixed effects that capture individual differences of these participants (e.g., motivation to process information, cognitive styles, etc.), $\boldsymbol{relevance}_j$ is a dummy variable that indicates if attribute j is a relevant attribute (attribute j is relevant when $relevance_j = 1$), $\boldsymbol{pre}_{ij}$ is the importance weight that participant i placed on attribute j before reading the reviews, $\boldsymbol{amount}_{ij}$ is a dummy variable that indicates the amount of information on attribute j presented to participant i (it is high when $amount_{ij} = 1$), $\boldsymbol{conflict}_{ij}$ is a dummy variable that indicates the degree of information conflict in attribute j for participant i (it is high when $conflict_{ij} = 1$), $\boldsymbol{coherence}_{ij}$ is a continuous variable that captures the extent to which participant i's evaluation of attribute j is coherent with his or her emerging judgment on the digital camera. This variable was constructed based on participant i's emerging judgment on the camera and his or her attribute-level performance evaluation on attribute j. Recall that we asked participants to judge the camera after they finished reading the third, sixth, and ninth reviews. We used the third judgment as a measure of participant i's emerging judgment of the camera.[10] Participant i's attribute-level performance evaluation on attribute j was reported after he or she finished reading the reviews. We rescaled both variables from a 1 to 5 scale to -2 to +2 scale by subtracting 3 from the data. The variable $\boldsymbol{coherence}_{ij}$ was then calculated as the product of the two rescaled variables. This operationalization is based on the intuition that when the attribute-level performance evaluation and the emerging judgment are of the same sign (both are positive or negative) and are both high in magnitude, the value of $\boldsymbol{coherence}_{ij}$ will be high.[11]

The results (see Table 2) show significant effects of the amount of attribute information (β = 16.990, p-value < .001), the degree of attribute information conflict (β = 2.279, p-value = .032), and the coherence between emerging judgment and attribute-level evaluation (β = 2.788, p-value < .001) on attribute preferences, supporting H1, H2, and H3. The relevance of an attribute to the decision context (β = 1.106, p-value = .260) and the importance of an attribute reported before reading the reviews (β = .062, p-value = .401) do not have significant effects on attribute preferences.

To assess if the reviews have a swaying effect (i.e., attribute importance is more heavily influenced by characteristics of the reviews than by the relevance of the attribute), we compared the relative importance of the review characteristics and the relevance of the attributes in predicting the constructed attribute preferences. Since the independent variables are measured on different scales (e.g., *coherence* is a product of two variables measured on -2 to 2 scale, but the other variables are binary), it is misleading to directly compare the coefficients to assess relative importance. We therefore ran a dominance analysis, which defines the relative importance as the contribution made by the individual independent variables to the prediction of the dependent variable (Azen and Budescu 2003). Our analysis decomposes the R-squared of a model into contributions from the individual regressors using the proportional marginal variance decomposition algorithm (Grömping 2007).[12] As shown in Table 2, attribute relevance only contributes .234% of the variations in the constructed attribute preferences. In contrast, the three hypothesized factors together contribute 41.060% of the vari-

[10]We use the third judgment to measure emerging judgment because it incorporates more information from the reviews than the other two judgments. Using an alternative measure that consists of the average of the three judgments does not generate materially different results.

[11]This measure does not differentiate different types of coherence or incoherence (e.g., a high overall evaluation but a low attribute score versus a low overall evaluation but a high attribute score). A *post hoc* analysis showed that different coherence or incoherence types have no significant impact on attribute preference above and beyond the coherence measure.

[12]The proportional marginal variance decomposition is generally recommended over other algorithms (Grömping 2007). Using other algorithms, including the squared standardized coefficients and the standardized coefficients adjusted by marginal correlation, did not generate materially different results.

Table 2. Model Estimation Results of Study 1

Variable	Coefficient	P-value	Relative Importance
Relevance	1.106 (1.191)	.260	.234%
Pre	.062 (.052)	.401	.318%
Amount	**16.990*** (1.161)**	**.000**	36.200%
Conflict	**2.279* (1.153)**	**.032**	.723%
Coherence	**2.788*** (.702)**	**.000**	4.137%
F(56, 359)	**4.291***	**.000**	/
Adj. R-squared	.346	/	/

Notes: (a) *p < 0.05; **p < 0.01; ***p < 0.001; Std. errors in parentheses; (b) Sample size: 416 (52 participants × 8 attributes).

ations in the constructed attribute preferences. Therefore, we conclude that attribute preference construction is swayed by the reviews in study 1, supporting H4.

Discussion

The results of study 1 support our hypotheses that the amount of attribute information, the degree of attribute information conflict, and the coherence between participants' emerging judgment on the product and attribute-level evaluation affect attribute preference construction. Moreover, when participants were presented with a brief description for each attribute and were only required to assess the importance of these attributes, participants did place more importance weights on the relevant attributes than the irrelevant attributes. Surprisingly when they were asked to evaluate the camera based on reviews that contained an uneven amount of information across different attributes, varying degrees of information conflict, and a numeric overall rating, relevance of the attributes did not have a significant impact on attribute preferences. A dominance analysis shows that attribute preferences that result from reading the reviews are primarily driven by the review characteristics not by attribute relevance. Taken together, this suggests that the hypothesized characteristics of reviews can sway attribute preference construction such that relevance of the attributes to the decision context has no significant impact on the constructed attribute preferences.

It is worth noting that attribute preferences prior to reading the reviews are not significant in predicting the final attribute preferences. Prior studies on the impact of online reviews often assume that consumers process the reviews according to Bayesian information updating (e.g., Archak et al. 2011), which requires both the prior beliefs and information in the reviews to be significant in forming their final beliefs. Our results suggest that attribute preference construction does not necessarily follow Bayesian information updating.

In study 1, participants were paid a flat fee, regardless of their performance in the experiment. It is possible that the non-significant effect of attribute relevance is due to participants not having sufficient motivation to process information in the reviews. As such, a threat to validity is that the participants' constructed attribute preferences are due to their lack of motivation to process the information rather than the hypothesized review characteristics. Although we controlled for motivation to process information using a fixed effects model, the effects of the review characteristics might be substantively different when people have a high motivation to process information. Therefore, we ran study 2 to investigate if our hypotheses hold when people have high motivation to process information.

Study 2

We extended the design of study 1 by adding motivation to process review information as a between-subjects factor. We used a monetary incentive to induce different levels of motivation to process review information.

Participants and Design

Ninety-nine students (77 females and 22 males, $M_{age} = 22.73$, $SD_{age} = 2.63$) participated in the study. Participants were recruited from the same subject pool used in study 1. About

80% of the participants owned a digital camera at the time of the experiment and 51% had researched at some point digital cameras on the Internet. The participants reported to have low to medium knowledge of a digital camera (M = 3.70, SD = 1.21 on a 1 to 7 scale).

We implemented a 2 (attribute-information amount: high or low) × 2 (attribute-information conflict: high or low) × 2 (numeric overall rating: high or low) × 2 (motivation to process review information: high or low) mixed factorial design by adding motivation to process review information as a new between-subjects factor to the design of study 1. Each of the 8 experimental groups in Table 1 was implemented in both high and low motivation conditions, leading to 16 experimental groups in study 2. All participants performed two tasks in this study. Task 1, the main task in study 2, was the same task as in study 1. Task 2 was a filler task, which was only used to induce different levels of motivation to process review information in task 1. In task 2, participants read 24 reviews for a hotel and evaluated the hotel based on the reviews. Participants in the high motivation conditions were led to believe that their monetary payoff was solely determined by the accuracy of their evaluation of the camera (task 1) and had absolutely nothing to do with the accuracy of their evaluation of the hotel (task 2). In contrast, participants in the low motivation conditions were led to believe that their monetary payoff was solely determined by the accuracy of their evaluation of the hotel (task 2) and had absolutely nothing to do with the accuracy of their evaluation of the camera (task 1). It was expected that participants would have higher motivation to process review information only for the task that determined their monetary payoff.

Procedure

After arriving at the lab, participants were paid a $5 show-up fee. The experimenter then described the two tasks that the participants would perform and told them that they had an opportunity to earn a monetary reward up to $15 depending on their performance in one of the two tasks specified on the experiment website. Participants logged into the experiment website and were randomly assigned to one of the 16 experimental groups. The website showed a message indicating which task would determine their monetary reward.

All participants performed task 1 (digital camera evaluation) first. The procedure was identical to the procedure in study 1. Following the completion of task 1, participants were presented with the scenario for task 2, which asked them to evaluate a hotel for a person who plans to travel to New York City. The website then presented 24 reviews to the participants one by one. After they finished reading the reviews, they were required to evaluate the hotel on location, room size, cleanliness, and staff. As discussed earlier, task 2 was a filler task in that we only used this task to induce different levels of motivation to process review information in task 1. After the completion of task 2, we measured their motivation to process review information in task 1 and task 2 using two questions adapted from the scale of purchase-decision involvement (Mittal 1989).[13] After participants completed both tasks, we calculated their monetary reward based on the Euclidean distance between their attribute-level performance evaluations and the actual ratings of the product on the attributes.[14] The monetary reward participants received ranged from $0 to $15 (M = $9.89, SD = 4.26).

Results

Manipulation checks: To check the manipulation on motivation, we ran a T-test to see if the manipulation led to differences in participants' self-reported motivation to process review information. The results show that participants in the high motivation condition reported significantly higher motivation to process review information ($M_{difference} = 1.14$, $p < .001$). As in study 1, T-tests show that participants perceived significantly higher amount of attribute information ($M_{difference} = 2.271$, p-value < .001) and higher degree of conflict ($M_{difference} = .642$, p-value < .001) when these factors were manipulated to be higher. A MANOVA shows that different levels of overall rating led to differences in the three waves of emerging judgment collected while they were reading the reviews (Wilks' lambda = .342, p-value < .001). A T-test further shows that the average of the three judgments is significantly higher when participants saw a higher overall rating ($M_{difference} = .609$, p-value < .001). Therefore, manipulating the overall rating led to different levels of emerging judgment as we expected. Further, T-tests support that participants' *attribute-level* performance evaluations in the high numeric overall rating conditions were not significantly different from those in the low overall rating conditions ($M_{difference} = .250$, p-value = .651). From the analysis, we conclude that our manipulations were successful.

Hypothesis testing: We ran the model from study 1 (model 1) separately on the high and low motivation conditions and contrasted the results (see Table 3). The results for the high

---

[13]The two questions are: (1) How important was it to you to read the reviews carefully during the task? (2) When you performed the task, to what extent were you concerned about reading the reviews carefully?

[14]The actual rating on an attribute was calculated by taking the average of the attribute ratings across all reviews in the review design table (Table A1) in Appendix A.

Table 3. Model Estimation Results of Study 2

Variable	Low Motivation	p-value	Relative Importance	High Motivation	p-value	Relative Importance
Relevance	1.981 (1.318)	.134	.644%	**2.599* (1.079)**	**.018**	1.131%
Pre	.058 (.058)	.321	.309%	.064 (.065)	.324	.228%
Amount	**16.548*** (1.276)**	**.000**	34.306%	**17.600*** (1.047)**	**.000**	39.231%
Conflict	2.055 (1.173)	.081	.546%	**1.922* (1.052)**	**.039**	.590%
Coherence	**3.378*** (.824)**	**.000**	3.639%	**3.740*** (.596)**	**.000**	4.991%
F-value	**4.498***	**.000**	/	**6.083***	**.000**	**/**
Adj. R-squared	.323	/	/	.407	/	/

Notes: (a) *p < 0.05; **p < 0.01; **p < 0.001; Std. errors in parentheses; (b) Sample size: 368 in the low motivation condition (46 participants * 8 attributes), 424 in the high motivation condition (53 participants x 8 attributes).

motivation condition are consistent with those of study 1 with one exception: the relevance of an attribute significantly affects the importance weight assigned to that attribute (β = .062, p-value = .018). Therefore, we can conclude that, with the exception of relevance of an attribute, lack of motivation to process information is not a threat to the results of study 1 and that people's constructed attribute preference is affected by both the relevance of an attribute and the characteristics of reviews. We also ran a dominance analysis to assess the relative importance of the independent variables. It can be seen that even when people had high motivation to process review information, attribute relevance still made a much smaller contribution to the variations in the constructed attribute preferences than the review characteristics (1.131% versus 44.811%).

Discussion

Study 2 shows that our hypotheses are robust under the high motivation to process review information condition. Although attribute relevance has a significant effect on the constructed attribute preferences in the high motivation condition, the dominance analysis shows that attribute preferences are still primarily driven by the review characteristics.

The results show that the degree of attribute information conflict only affects attribute preferences when people have high motivation to process information. The manipulation check shows that, in the low motivation condition, participants could assess accurately if an attribute had a high degree of conflicting information. However, in the low motivation condition, participants did not have the motivation to resolve the conflicting information by carefully processing information in the reviews. Consequently, the attributes with a high degree of conflicting information did not become more accessible in memory. This finding is consistent with our argument for H2 that conflicting attribute information induces a deeper level of processing when people are motivated to process information to resolve this conflict.

Through randomized experiments, studies 1 and 2 provide causal evidence for the impact of review characteristics on constructed attribute preferences. In both studies, participants read 10 researcher-manipulated reviews that contained information about 8 attributes. In reality, many products have a large number of reviews that contain information about a wide range of attributes, and people can freely choose to read all or only a subset of the available reviews. To investigate if the results from the randomized experiments hold in less restrictive settings, we designed study 3 that allows participants to freely choose which reviews to read from a larger set of reviews that contain a large number of attributes.

Study 3

The objective of study 3 is to enhance the external validity of our results. We ran a free simulation study to allow participants to determine which reviews to read and in what sequence, as they would in real life. Moreover, instead of directly manipulating the three factors that affect attribute preferences, we used the verbal protocol analysis (VPA) to capture and measure them. VPA requires participants to continuously report their thoughts when performing a task (Ericsson and Simon 1993). Collected verbal protocols allow researchers to measure participants' cognitive activities while they are performing their tasks.

Participants and Design

Thirty-four undergraduate students participated in this study (12 females and 22 males).[15] Participants were recruited from two undergraduate MIS courses. Sixty-eight percent of the participants owned a digital camera at the time of the experiment, and 65% of the participants had researched digital cameras on the Internet at some point. The participants reported a medium level of knowledge about digital cameras prior to reading the product reviews (Mean = 4.05, SD = 1.43 on a 1 to 7 scale).

We presented the participants with 60 reviews about a digital camera, spanning 10 pages on the study website. The 60 reviews were real customer reviews for the Canon A590 IS camera randomly selected from Amazon.com (see Appendix C). The brand and model name were removed from reviews so that participants' evaluation would not be biased by the brand name. The free simulation study differs from the randomized experiments in that participants can freely determine how many reviews to read, when to read which reviews, and in which sequence. Therefore, we did not directly manipulate the amount of attribute information, the degree of attribute information conflict, and the coherence between emerging judgment and attribute-level performance evaluations. These factors were jointly determined by the randomization of the review presentation order and participants' review selection.

Procedure

We ran the participants individually through the experimental session. Participants received course credit for participation. Each session consisted of the following procedures.

Pre-task survey: Upon arrival at the lab, participants were asked to complete the pre-task survey, which contained questions about their prior knowledge of and level of interest in digital cameras, their level of motivation to process information, and cognitive style.

Think-aloud training exercises: Because thinking aloud (required by verbal protocol analysis) is not natural to participants, we showed participants a think-aloud demonstration video. We then gave them two exercises that were similar to the primary task. Participants were required to complete the exercises while verbalizing their thoughts. The exercises were not time limited. The participants were instructed to stop when they were able to continuously verbalize their thoughts.

Primary task introduction and pre-task interview: After the training exercises, participants were introduced to the primary task, which required them to evaluate a digital camera based on 60 reviews. They were presented with a scenario in which they were asked to imagine that they were considering buying a digital camera as a birthday gift for a friend who occasionally takes long trips and has started to show interest in photography. Participants were then asked in a pre-task interview to list the attributes that were important to them in evaluating the digital camera.

Primary task: The primary task was to evaluate, based on the 60 reviews, whether the digital camera was a viable option they would consider buying for their best friend. The 60 reviews were presented across 10 pages, each page showing 6 reviews in 3 rows and 2 columns. The order of the reviews was randomized when they were presented to the participants. Participants were asked to perform the task as they normally do. They were required to continuously verbalize their thoughts as they did in the training exercises. To avoid distracting the participants, the experimenter was not visually accessible to participants. All sessions were video-recorded and clickstream data was captured for each session.

Post-task interview and survey: In the post-task interview, the experimenter asked participants to describe the steps they went through to make the evaluation of the camera. Following the interview, participants took a survey asking them to list the attributes they considered in evaluating the camera, rate the importance of each attribute by allocating 100 points among them, evaluate the camera on each attribute (on a 1 to 5 scale with 1 being extremely bad and 5 being extremely good), and make an overall evaluation on the camera.

Data Analysis and Results

Following Ericsson and Simon's (1993) guidance, our data analysis included coding the verbal protocols and developing a statistical model using data from the verbal protocols, clickstream data, surveys, and interview.

Coding the Verbal Protocol Data: The verbal protocols were first transcribed from the videos. Complete protocol data from 31 participants were available for analysis. Al-

[15]Given that VPA is labor intensive, requiring considerable efforts for transcription, coding, and analysis, VPA studies are typically conducted with small samples of less than 20 participants (Eveland and Dunwoody 2000) making our sample size of 34 more than adequate.

though 34 students participated in the study, we excluded three protocols because they were not properly recorded due to a technical problem. After transcription, we extracted from the verbal protocols all attribute-related thoughts. We coded (1) the times a participant paid attention to an attribute (as indicated by their verbalization), (2) every mention of whether an attribute is relevant, (3) every assessment of the overall performance of the camera, (4) every assessment of the attribute-level performance (i.e., whether the camera is good or bad on the attribute), and (5) whether information on an attribute is consistent across the attended reviews.

We took an iterative coding approach to ensure that the verbal protocols were coded reliably. Two researchers, who were aware of the research objectives, developed and refined the coding scheme. The two researchers first coded two randomly selected protocols independently. They agreed on 332 out of 492 segments, or 67.5 percent for these two protocols. They then carefully examined the coded protocols, discussed disagreements, and revised the coding scheme. Another two protocols were then randomly selected and coded by the same two researchers. The two researchers achieved a high inter-rater agreement (211 out of 244 segments or 86.3 percent) for these two protocols. Subsequently, one researcher coded all protocols based on the revised coding scheme.

To further validate the reliability of the researcher's coding, we invited two additional coders who were blind to the objective of the study to code a randomly selected sample of eight verbal protocols (25% of all protocols). The inter-rater reliability (Cohen's Kappa) between the researcher and first coder was .800 and between the researcher and the second coder was .779, indicating substantial agreement (Landis and Koch 1977). As a result, the verbal protocols coded by the researcher were then used together with the clickstream, interview, and survey data for statistical analysis.

Operationalization of variables: To measure the constructed attribute preferences, we used the self-reported importance score for each attribute (on a 1 to 100 scale) collected in the post-task survey. We measured the amount of information for an attribute using the number of times a participant saw an attribute in the reviews (as indicated in their verbal protocol). To measure information conflict, we coded the valence of information attended by each participant for each attribute as a sequence of +1 (positive information about an attribute) and -1 (negative information about an attribute). We then calculated the variance of the sequence of +1 and -1 for each attribute as p*(1-p), with p being the proportion of +1 in the sequence. We used the variance as the measure of degree of information conflict (the bigger the variance, the greater the degree of information conflict for an attribute).

As in studies 1 and 2, we operationalized the coherence between product-level emerging judgment and attribute-level performance evaluations as a product of a participant's self-reported likelihood to buy the digital camera (capturing their emerging judgment) and the participant's attribute-level performance evaluations. We first rescaled the self-reported likelihood to buy from a 1 to 7 scale to a -3 to +3 scale. Then we created a categorical variable to capture the participant's attribute-level performance evaluations.[16] The variable is 1 if the participant's evaluation on the attribute is positive (e.g., the picture quality is good); it is -1 if the evaluation is negative (e.g., long lag time between shots). The variable is zero if the evaluation is neutral (e.g., the camera uses AA batteries). As in studies 1 and 2, the coherence variable was calculated as the product of these two variables.

We created a dummy variable to indicate whether an attribute was initially reported as important before participants read the reviews. The variable is 1 if the participant reported the attribute as important before reading the reviews, otherwise it is zero. We also created a categorical variable to indicate whether a participant stated that an attribute was relevant to the decision in the verbal protocols. The variable is 1 if the participant indicated that the attribute was relevant. The variable is -1 if the participant indicated that the attribute was irrelevant. The variable is zero if the participant did not make any comment about the relevance of the attribute. These two variables were included as controls. Appendix D provides the descriptive statistics on these variables for each attribute.

Model estimation: Data collected from this study contains 31 participants (N = 31), and there are 24 attributes (J = 24, see Appendix C for these attributes) that may be considered by each participant, resulting in a sample size of 744 (i.e., N*J = 31*24). Before the estimation, all independent variables except the dummy variables were standardized by subtracting the mean and then dividing by the standard deviation. The model estimation results are shown in Table 4.

The freedom of choosing reviews to read in the free simulation study may lead to an endogeneity issue in the above estimation. Specifically, we attempt to establish the causal impact of information amount for an attribute on the importance weight attached to that attribute (i.e., attribute-information amount → attribute preference). However, in the free simulation study, a participant may choose to ignore information about attributes that he or she thinks are not

[16]Attribute-level performance evaluations are reported in the verbal protocols. Participants often reported multiple evaluations on the same attribute over time (they updated their evaluations as they processed new information from the reviews). We used the last evaluation if multiple evaluations were reported for an attribute.

Table 4. Model Estimation Results of Study 3

Variable	Model 1 (Fixed Effects)	p-value	Relative Importance	Model 2 (Fixed Effects with IV)	p-value
Relevance	**5.925*** (.987)**	**.000**	5.667%	**4.934*** (1.079)**	**.000**
Pre	**7.429*** (.992)**	**.000**	8.691%	**7.015*** (.065)**	**.000**
Amount	**2.443*** (.424)**	**.000**	13.389%	**4.167** (1.491)**	**.005**
Conflict	**2.259*** (.392)**	**.000**	3.855%	1.369 (.838)	.103
Coherence	**1.969*** (.460)**	**.000**	2.344%	**1.851*** (.476)**	**.000**
F-value	**84.456***	**.000**	/	**79.300**	**.000**
Adj. R-squared	.356	/	/	.344	/

Notes: (a) *p < 0.05; **p < 0.01; ***p < 0.001; Std. errors in parentheses; (b) Sample size: 744 (31 participants × 24 attributes).

important, leading to less information processed for those attributes (i.e., attribute preference → attribute-information amount). An OLS regression will only show correlation but not the direction of causality. To verify the direction of causality, an instrument variable is needed to isolate the part of variation in the amount of information processed for an attribute that is not due to the perceived importance of that attribute. A valid instrument is the amount of attribute information available on the pages visited by participants. This is because the amount of information on the visited pages affects the amount of information processed (information cannot be processed if it is not on the visited pages) and it is uncorrelated with participants' attribute preferences (the display of the reviews on the pages was randomized). The results of estimating the specification of Model 1 with the instrument variable are summarized in Table 4 (Model 2).

Model 2 (the fixed effects model with instrumental variable) differs from Model 1 (the fixed effects model) in that the degree of attribute-information conflict becomes nonsignificant when the instrument is used. To decide which model to use, we ran a Hausman test to assess if endogeneity is a serious concern. If endogeneity is significant, the results of the fixed effects model without instrument (Model 1) are biased, but if endogeneity is not significant, the fixed effects model with instrument (Model 2) may generate inflated standard errors. The test does not reject the null hypothesis that the amount of information processed is endogenous (χ^2 (1) = 1.485, p = .156). Therefore the fixed effects model (Model 1) is appropriate, preferred, and provides support for H1, H2, and H3.

We ran the dominance analysis on Model 1 to assess the relative importance of the independent variables. It can be seen that, in this study, attribute relevance made a larger contribution to the variations of constructed attribute preferences than in the randomized experiments (5.667% versus .234% in study 1 and 1.131% in study 2). However, compared to the review characteristics, attribute relevance still makes a smaller contribution to the variations of attribute preferences in study 3 (5.667% versus 19.588% for review characteristics), providing support for H4.

Discussion

In this study, participants were able to perform the task in a less restrictive setting and more similar to what they would normally do when reading reviews. The variation in the amount of attribute information, the degree of attribute information conflict, and the coherence between emerging judgment and attribute-level evaluation comes from participants' selective processing of the reviews (the endogenous source) and the randomization of the review display (the exogenous source). Exploiting the exogenous source of the variation, we were able to provide arguably causal evidence for the impact of the review characteristics on attribute preferences. The results of this study are consistent with those of study 1 and study 2, providing evidence for generalizability of our findings beyond a controlled experimental setting.

A minor difference between the results of study 3 and those of studies 1 and 2 is that the initial attribute importance (i.e., importance of attributes reported before participants read the reviews) has a significant impact on the constructed attribute preferences (β = 7.429, p-value < .001) in study 3 but not in studies 1 and 2. An explanation is that reviews in study 1 and 2 only contained information about 8 attributes but reviews in study 3 contained information about 24 attributes. The eight attributes in studies 1 and 2 may not necessarily be the "important" attributes participants had in mind before they read the reviews. Thus, in studies 1 and 2 the important attributes reported before reading the reviews are less likely to affect the importance weights placed on the eight attributes discussed in the reviews.

General Discussion

To summarize, study 1 shows that when people were simply asked to determine the importance of attributes based on brief descriptions of the attributes, they did weigh the relevant attributes more than the irrelevant ones. However, when information about the attributes was presented in the form of reviews, attribute importance weights were primarily influenced by the manipulated review characteristics and considerably less so by the relevance of the attributes to the decision context. Study 2 further shows that these review characteristics affect attribute preferences even when people are highly motivated to process review information. Although the relevance of attributes has significant effects on attribute preferences when people have high motivation to process information, constructed attribute preferences remain primarily driven by the review characteristics. Finally, study 3 supports the effects of the hypothesized review characteristics and the swaying effect of the reviews in a more realistic setting. Taken together, the three studies consistently support the effects of the hypothesized review characteristics on attribute preference construction and the swaying effect of online product reviews.

In all three studies, the amount of attribute information in the reviews had the greatest impact on constructed attribute preferences. These studies show that this relationship is *causal* and not merely correlational. When an attribute is relevant to many people, the attribute will likely be mentioned frequently in the reviews. Further, since the attribute is relevant to many people, there is also a greater chance that a consumer will assign a high importance weight to that attribute before reading the reviews. Therefore, it could be that the amount of attribute information in reviews does not cause people to assign a higher importance weight to that attribute. Rather, the relationship could be merely correlational because they are both driven by the fact that the attribute is relevant to many people (a third variable explanation). When the link is merely correlational, had the attribute not been frequently mentioned in the reviews, the consumer would have still assigned a high importance weight to that attribute as long as the attribute is relevant to the decision at hand. Clearly, given the large effect size of the amount of attribute information, differentiating between correlation and causation is important to both researchers and practitioners. Through the randomized and quasi experimentation, we ruled out the possibility that the link between the amount of attribute information and attribute preference is merely correlational.

Theoretical Implications

Our work has implications for both the online word-of-mouth and preference construction literatures. Although existing studies of online word-of-mouth have generated important insights into the effects of online product reviews, many are based on simplified assumptions of human information processing. Examples of these assumptions include consumers having inherent and fixed preferences over the course of product evaluation and choice, and consumers processing reviews using Bayesian information updating (e.g., Archak et al. 2011). In the extant literature, the effects of online product reviews boil down to the "informative effect" (i.e., the more reviews posted for a product, the more likely it is that people become aware of the product) and the "persuasive effect" (i.e., the higher the average rating, the more likely it is that people will buy the product) (for reviews, see Etzion and Awad 2007; Liu 2006). Using an information-processing approach, our research shows that the effects of online product reviews are richer than the informative effect and persuasive effect. Contrary to the research that treats attribute preferences as predetermined factors affecting information acquisition and processing in product evaluation, our research shows that review characteristics affect how information is processed, which in turn affects attribute preferences. There is an emerging literature on "choice architecture" that aims at designing the choice environment to nudge people toward personally or socially desirable outcomes (Thaler and Sunstein 2008). Clearly, designing a choice architecture involves an in-depth understanding of how online product reviews influence product evaluation. We believe that incorporating psychological theories about how people actually process information into this line of research will greatly advance our understanding of the impact of online product reviews.

Our research raises questions about the wisdom of the crowd effect. The utility of online reviews is premised on the wisdom of the crowd effect, that is, that the aggregation of many people's opinions is a better approximation of the truth than an individual's opinion (Surowiecki 2005). By aggregating opinions from many customers, online review websites intend to provide more accurate information about the reviewed product. However, simply displaying a few quantitative metrics (e.g., overall rating) together with the individual reviews may not fully realize the promise of the wisdom of the crowd. Our research shows that the current information environment of online product reviews may sway people's attribute preferences. Therefore, although useful knowledge is present in the reviews, it can be difficult to distill. Furthermore, studies 1 and 2 show that people's prior attribute preferences (i.e., the attribute preferences before reading the reviews) have no significant impact on attribute preference construction. Using Bayes' rule as a benchmark, this suggests that people could put too little weight on their prior beliefs and too much weight on the reviews in constructing attribute preferences. Existing evidence shows that online product reviews are subject to a self-selection bias with reviewers not being representative of the general population (Li and Hitt 2008).[17] Therefore, overweighing the reviews and underweighing one's prior belief may lead to biased product evaluation. It is important that future research investigates how the reviews should be organized and presented to help people better distill the wisdom embedded in the reviews and properly weigh the information from the reviews in product evaluation. It would also be beneficial to examine whether these effects differ based on the consumer's level of expertise with the product category (e.g., cameras).

Our research also contributes to the literature on preference construction in the online environment. First, as discussed in our literature review, the current literature on online preference construction has a focus on the presentation of seller-provided information. Many of the existing studies only examine a single feature of the information environment (e.g., the salience of information, ease of information processing, or partitioning of product attributes). Because of user participation and content generation, social media have enabled an increasingly more complex online information environment where multiple features of the information environment can affect preference construction through multiple mechanisms. Our research highlights the importance of extending the literature of online preference construction to the context of social media. Second, although the current literature on online preference construction has identified many features of the information environment that influence people's preference construction, the existing studies do not assess if these effects are sufficiently large to sway preference construction away from the desired judgment and decision making. The current study assesses this swaying effect by defining the factor that is critical to well-constructed preferences (i.e., relevance of attributes to the decision context) and comparing the effects of the review characteristics to the effect of this factor. Through assessing the swaying effect, we show that the effects of review characteristics are large enough to matter and should be taken into account by future studies of online preference construction. Third, a criticism on the current literature on preference construction is that many studies exaggerate preference construction by using a well-controlled experimental task and carefully designed stimuli (Simonson 2008). Therefore it becomes important to investigate the robustness of preference construction in multiple studies using mixed methods. Through combining well-controlled experiments with the less restrictive free simulation, our research demonstrates a useful approach to strengthen the realism and generalizability of preference construction research.

Finally, the current research focuses on attribute preference construction in the evaluation of a single option. In such contexts, people may not evaluate the different options independently. Therefore, attribute preference construction might be more complicated in the context of choosing from multiple options. Future research may extend the current study to the context of choosing among multiple options by identifying additional online review characteristics or other relevant design aspects of the online information environment.

Practical Implications

There is a heated discussion in the popular press on winning the "zero moment of truth," which is a name given to consumer's online research before they try or buy a product (Lecinski 2011). Online product reviews play an important role in the zero moment of truth as consumers rely on these to evaluate products of interest. Our research raises several issues regarding how online reviews can sway people's attribute preferences in product evaluation. Below, we discuss these issues and make suggestions for remedies.

First, our findings suggest that seeing a large amount of attribute information may lead to a greater importance weight attached to the attribute even though this attribute may be irrelevant. To make an informed decision, it is important to carefully consider the attributes that are critical to the decision at hand. A possible solution is to show brief descriptions of the reviewers' background and the context of the reviewer's purchase decision, and allow consumers to choose reviews from people who have similar background or similar purchase decision contexts. For example, reviews on newegg.com describe the reviewer's background in terms of "tech level" and time of "ownership." This is based on the assumption that reviews from people with similar backgrounds are more likely to contain information about attributes relevant to the consumer.

Second, our findings suggest that high degree of attribute information conflict can also lead people to assign a high importance weight to an attribute. A potential remedy for the effect of conflicting attribute information is to visualize the

[17]We thank an anonymous reviewer for this insight.

degree of information conflict for each attribute and to present this visualization up front. An example is shown in Liu et al. (2011), in which the amount of positive and negative comments on an attribute are displayed as two bars of different colors positioned side by side and the relative length of the two bars represents the degree of conflict. With this presentation, consumers can identify both the overall degree of conflict and the majority opinion for each attribute at a glance. Since making sense of conflicts becomes less cognitively demanding, this may reduce the impact of conflicting attribute information on the constructed attribute preferences.

Finally, our findings show that the numeric overall rating has a strong impact on people's emerging judgment, which further leads to a confirmation bias in the construction of attribute preferences. The overall numeric rating serves as a reference point in people's judgment. However, due to the differences in taste or decision context, the numeric overall rating might not always be an appropriate reference point. For example, if most reviews were submitted by experienced photographers, the numeric overall rating might not be very informative to a novice. A potential remedy is to allow people to see the overall rating from reviewers who have similar decision contexts.

We call for future design studies to test the effectiveness of these suggestions and investigate other possible interventions to nudge consumers toward better decision making. The literature on choice architecture shows that even a subtle change in the information environment may have a substantial impact on consumer behavior (Thaler and Sunstein 2008). Theoretically informed and empirically tested studies are necessary to better understand the design of different information organizations and presentations in the context of online reviews.

Acknowledgments

We would like to thank the senior editor, Soon Ang, the associate editor, Andrew Burton-Jones, and the anonymous reviewers whose insightful comments enhanced the development of this paper throughout the review process. We would also like to thank Arun Rai, Rick Watson, Dale Goodhue, and the participants at research colloquia at the University of Georgia and the Georgia Institute of Technology for their helpful comments. The research has been partially supported by a New Faculty Start-up grant at City University of Hong Kong.

References

Anderson, M. C., Bjork, R. A., and Bjork, E. L. 1994. "Remembering Can Cause Forgetting: Retrieval Dynamics in Long-Term Memory," *Journal of Experimental Psychology: Learning, Memory, and Cognition* (20:5), pp. 1063-1087.

Appan, R., and Browne, G. J. 2010. "Investigating Retrieval-Induced Forgetting During Information Requirements Determination," *Journal of the Association for Information Systems* (11:5), pp. 250-275.

Archak, N., Ghose, A., and Ipeirotis, P. G. 2011. "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management Science* (57:8), pp. 1485-1509.

Azen, R., and Budescu, D. V. 2003. "The Dominance Analysis Approach for Comparing Predictors in Multiple Regression," *Psychological Methods* (8:2), pp. 129-148.

Banerjee, A. V. 1992. "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics* (107:3), pp. 797-817.

Bettman, J. R., Luce, M. F., and Payne, J. W. 1998. "Constructive Consumer Choice Processes," *Journal of Consumer Research* (25:3), pp. 187-217.

Bikhchandani, S., Hirshleifer, D., and Welch, I. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy* (100:5), pp. 992-1026.

Brown, C. L., and Carpenter, G. S. 2000. "Why Is the Trivial Important? A Reasons-based Account for the Effects of Trivial Attributes on Choice," *Journal of Consumer Research* (26:4), pp. 372-385.

Chamley, C. P. 2003. *Rational Herds: Economic Models of Social Learning*, Cambridge, UK: Cambridge University Press.

Cheung, C. M. K., and Thadani, D. R. 2012. "The Impact of Electronic Word-of-mouth Communication: A Literature Analysis and Integrative Model," *Decision Support Systems* (54:1), pp. 461-470.

Chevalier, J. A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* (43:3), pp. 345-354.

Connolly, T., and Thorn, B. K. 1987. "Predecisional Information Acquisition: Effects of Task Variables on Suboptimal Search Strategies," *Organizational Behavior and Human Decision Processes* (39:3), pp. 397-416.

Craik, F. I. M., and Tulving, E. 1975. "Depth of Processing and the Retention of Words in Episodic Memory," *Journal of Experimental Psychology: General* (104:3), pp. 268-294.

Duan, W., Gu, B., and Whinston, A. B. 2008. "Do Online Reviews Matter? An Empirical Investigation of Panel Data," *Decision Support Systems* (45:4), pp. 1007-1016.

Epley, N., and Gilovich, T. 2006. "The Anchoring-and-Adjustment Heuristic: Why the Adjustments Are Insufficient," *Psychological Science* (17:4), pp. 311-318.

Ericsson, K. A., and Simon, H. A. 1993. *Protocol Analysis*: *Verbal Reports as Data* (rev. ed.), Cambridge, MA: The MIT Press.

Etzion, H., and Awad, N. 2007. "Pump Up the Volume? Examining the Relationship between Number of Online Reviews and Sales: Is More Necessarily Better?," in *Proceedings of the 28th International Conference on Information Systems*, Montreal, Canada, paper 120.

Eveland, W. P., and Dunwoody, S. 2000. "Examining Information Processing on the World Wide Web Using Think Aloud Protocols," *Media Psychology* (2:3), pp. 219-244.

Fromkin, H. L., and Streufer S. 1976. "Laboratory Experimentation," in M. D. Dunnette (ed.), *Handbook of Industrial and Organizational Psychology*, Chicago: Rand McNally, pp. 415-465.

Grömping, U. 2007. "Estimators of Relative Importance in Linear Regression Based on Variance Decomposition," *The American Statistician* (61:2), pp. 139-147.

Häubl, G., and Murray, K. B. 2003. "Preference Construction and Persistence in Digital Marketplaces: The Role of Electronic Recommendation Agents," *Journal of Consumer Psychology* (13:1/2), pp. 75-91.

Hastie, R. 1980. "Memory for Information Which Confirms or Contradicts a General Impression," in *Person Memory: The Cognitive Basis of Social Perceptions*, R. Hastie, T. M. Ostrom. E. B. Ebbesen, R. S. Wyer, Jr., D. L. Hamilton, and D. E. Carlston (eds),Hillsdale, NJ: Erlbaum, pp. 155-177.

Hastie, R., and Park, B. 1986. "The Relationship Between Memory and Judgment Depends on Whether the Judgment Task Is Memory-Based or On-Line," *Psychological Review* (93:3), pp. 258-268.

Jenkins, A. M. 1985. "Research Methodologies in MIS Research," in *Research Methods in Information Systems*, E. Mumford, R. Hirschheim, G. Fitzgerald, and A. T. Wood-Harper (eds.), Amsterdam: Elsevier, pp. 103-118.

Kahneman, D. 2011. *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux.

King, R. A., Racherla, P., and Bush, V. D. 2014. "What We Know and Don't Know About Online Word-of-Mouth: A Review and Synthesis of the Literature," *Journal of Interactive Marketing* (28:3), pp.167-183.

Landis, J. R., and Koch, G. G. 1977. "The Measurement of Observer Agreement for Categorical Data," *Biometrics* (33), pp. 159-174.

Lecinski, J. 2011. *Winning the Zero Moment of Truth–ZMOT*, Google.

Li, X., and Hitt, L. M. 2008. "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research* (19:4), pp. 456-474.

Lieberman, D. A. 2012. *Human Learning and Memory*, Cambridge, UK: Cambridge University Press.

Liu, Q. B., Karahanna, E., and Watson, R. T. 2011. "Unveiling User-Generated Content: Designing Websites to Best Present Customer Reviews," *Business Horizons* (54:3), pp. 231-240.

Liu, Y. 2006. "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing* (70:3), pp. 74-89.

Lynch Jr., J. G., and Ariely, D. 2000. "Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution," *Marketing Science* (19:1), pp. 83-103.

MacKenzie, S. B., Lutz, R. J., and Belch, G. E. 1986. "The Role of Attitude Toward the Ad as a Mediator of Advertising Effectiveness: A Test of Competing Explanations," *Journal of Marketing Research* (23:2), pp.130-143.

Maheswaran, D., and Chaiken, S. 1991. "Promoting Systematic Processing in Low-Motivation Settings: Effect of Incongruent Information Pn processing and Judgment," *Journal of Personality and Social Psychology* (61:1), pp. 13-25.

Mandel, N., and Johnson, E. J. 2002. "When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices," *Journal of Consumer Research* (29:2), pp. 235-245.

Martin, J. M., and Norton, M. I. 2009. "Shaping Online Consumer Choice by Partitioning the Web," *Psychology & Marketing* (26:10), pp. 908-926.

Menon, G., and Raghubir, P. 2003. "Ease-of-Retrieval as an Automatic Input in Judgments: A Mere-Accessibility Framework?," *Journal of Consumer Research* (30:2), pp. 230-243.

Mittal, B. 1989. "Measuring Purchase-Decision Involvement," *Psychology & Marketing* (6:2), pp.147-162.

Mudambi, S. M. and Schuff, D. 2010. "What Makes a Helpful Review? A Study of Customer Reviews on Amazon.com," *MIS Quarterly* (34:1), pp.185-200.

Nickerson, R. S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," *Review of General Psychology* (2:2), pp. 175-220.

Noseworthy, T. J., Wang, J., and Islam, T. 2012. "How Context Shapes Category Inferences and Attribute Preference for New Ambiguous Products," *Journal of Consumer Psychology* (22:4), pp. 529-544.

Payne, J. W., Bettman, J. R., and Johnson, E. J. 1992. "Behavioral Decision Research: A Constructive Processing Perspective," *Annual Review of Psychology* (43:1), pp. 87-131.

Payne, J. W., Bettman, J. R., and Schkade, D. A. 1999. "Measuring Constructed Preferences: Towards a Building Code," *Journal of Risk and Uncertainty* (19:1), pp. 243-270.

Qiu, L., Pang, J., and Lim, K. H. 2012. "Effects of Conflicting Aggregated Rating on eWOM Review Credibility and Diagnosticity: The Moderating Role of Review Valence," *Decision Support Systems* (54:1), pp. 631-643.

Quaschning, S., Pandelaere, M., and Vermeir, I. 2014. "When and Why Attribute Sorting Affects Attribute Weights in Decision-Making," *Journal of Business Research* (67:7), pp. 1530-1536.

Scholz, S. W., Meissner, M., and Decker, R. 2010. "Measuring Consumer Preferences for Complex Products: A Compositional Approach Based on Paired Comparisons," *Journal of Marketing Research* (47:4), pp. 685-698.

Shafir, E., Simonson, I., and Tversky, A. 1993. "Reason-Based Choice," *Cognition* (49:1-2), pp. 11-36.

Simon, D., Krawczyk, D. C., and Holyoak, K. J. 2004. "Construction of Preferences by Constraint Satisfaction," *Psychological Science* (15:5), pp. 331-336.

Simonson, I. 2008. "Regarding Inherent Preferences," *Journal of Consumer Psychology* (18:3), pp. 191-196.

Slovic, P., and Lichtenstein, S. 1983. "Preference Reversals: A Broader Perspective," *The American Economic Review* (73:4), pp. 596-605.

Sun, M. 2012. "How Does the Variance of Product Ratings Matter?," *Management Science* (58:4), pp. 696-707.

Surowiecki, J. 2005. *The Wisdom of Crowds: Why the Many Are Smarter than the Few*, New York: Anchor Books.

Thaler, R. H., and Sunstein, C. R. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*, New Haven, CT: Yale University Press.

Tversky, A., and Kahneman, D. 1986. "Rational Choice and the Framing of Decisions," *Journal of Business* (59:S4), pp. 251-278.

Van Ittersum, K., Pennings, J. M., Wansink, B., and Van Trijp, H. C. 2007. "The Validity of Attribute-Importance Measurement: A Review," *Journal of Business Research* (60:11), pp. 1177-1190.

Weber, M., Eisenführ, F., and Von Winterfeldt, D. 1988. "The Effects of Splitting Attributes on Weights in Multiattribute Utility Measurement," *Management Science* (34:4), pp.431-445.

Yin, D., Bond, S., and Zhang, H. 2014. "Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews," *MIS Quarterly* (38:2), pp.539-560.

About the Authors

Qianqian Ben Liu is an assistant professor in the Department of Information Systems at City University of Hong Kong. His research focuses on judgment and decision-making in information-intensive environments and the use and impact of IT in healthcare. His work has been published in *Information Systems Research* and other scholarly and practitioner journals.

Elena Karahanna is Distinguished Research Professor and the L. Edmund Rast Professor of Business in the MIS Department at the Terry College of Business, University of Georgia. Her research focuses on the implementation and use of information systems, health IT, IS leadership, social media, and cross-cultural issues. Her work has been published in leading scholarly journals, including *MIS Quarterly, Information Systems Research, Management Science,* and *Organization Science*, as well as in practitioner journals (*MIS Quarterly Executive, Cutter Benchmark Review, Business Horizons*). She has served as senior editor for *MIS Quarterly, Information Systems Research,* and *Journal of the Association for Information Systems* and serves as an associate editor for *Management Science*. She was named an AIS Fellow in 2012.

MIS Quarterly

RESEARCH ARTICLE

THE DARK SIDE OF REVIEWS: THE SWAYING EFFECTS OF ONLINE PRODUCT REVIEWS ON ATTRIBUTE PREFERENCE CONSTRUCTION

Qianqian Ben Liu
Department of Information Systems, College of Business, City University of Hong Kong,
Hong Kong, CHINA {ben.liu@cityu.edu.hk}

Elena Karahanna
MIS Department, Terry College of Business, University of Georgia,
Athens, GA 30602-6273 U.S.A. {ekarah@uga.edu}

Appendix A

The Review Design Table

We created four sets of reviews that implemented the within-subjects factor combinations assigned to each attribute (i.e., one of H_AH_C, H_AL_C, L_AH_C, or L_AL_C). Each set of reviews contained 10 reviews. Attribute information contained across the 10 reviews determines the amount of information on an attribute (high if the attribute is discussed in all 10 reviews; low if it is discussed in only two reviews), and the amount of attribute-information conflict (high if half the reviews that discussed the attribute were positive on the attribute and the other half negative; low if all reviews that discussed the attribute discussed it either consistently positively or consistently negatively).

Given this, before creating the reviews, we first had to randomly determine what attributes will be discussed in each review, and the valence and extremity of each attribute discussed in each review. Consider review set 2 in Table 1 as an example: all the 10 reviews in review set 2 will discuss the attribute "Attr1" with the same valence (i.e., H_AL_C on "Attr1"). We first flipped a coin to decide the valence of "Attr1" in review set 2 (e.g., head is positive and tail is negative). Once the valence was determined, we flipped a coin 10 times to decide the extremity of "Attr1" in each of the 10 reviews (e.g., head is extremely positive or negative and tail is positive or negative). The attribute "Attr2" in review set 2 (L_AH_C) will be discussed in only two reviews (i.e., low amount of attribute information) with a different valence (i.e., high conflict of attribute information). We first determined which two reviews would discuss "Attr2" by randomly sampling two whole numbers from 1 to 10 without replacement (e.g., if the numbers 2 and 5 are sampled, then only reviews 2 and 5 will discuss "Attr2"). Next we flipped a coin to determine the valence and extremity of "Attr2" in each of the two reviews. Using this randomization, we determined the placement of all the attributes in the reviews for all the review sets (i.e., what attributes are discussed in each review of that set), and the valence and extremity of each attribute discussed in each review. Based on the results of the randomization, we created a "review design table" (see Table A1) to numerically represent the placement, the valence, and the extremely of the attributes in the 40 reviews. The texts of the reviews were written according to the review design table.

The numbers in the attribute columns of Table A1 represent the extremity and valence of that attribute in reviews (the valence and extremity are represented on a 1 to 5 scale with 1 being extremely negative and 5 being extremely positive). An empty cell means that the attribute is not discussed in that review.

Table A1. The Review Design Table

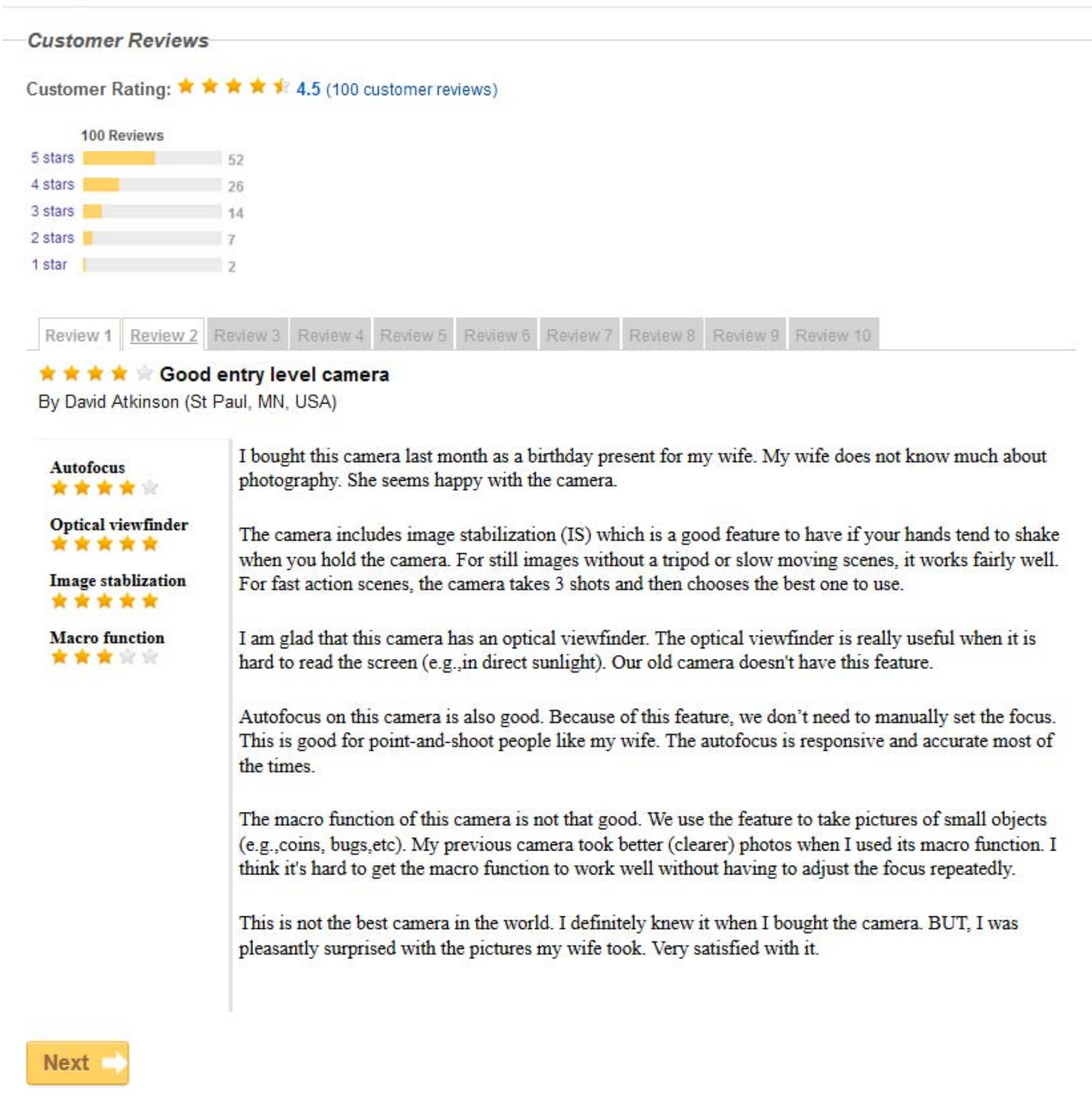
Review Set 1	Attr1: Autofocus (H_AH_C)	Attr2: Ease of Use (L_AL_C)	Attr3: Image Stabilization (H_AL_C)	Attr4: LCD Screen (L_AH_C)	Attr5: Optical Viewfinder (H_AH_C)	Attr6: Manual Mode (L_AL_C)	Attr7: Macro Mode (H_AL_C)	Attr8: Raw Format (L_AH_C)	Review Rating
Review1	4		4		4		2		3.5
Review2	2	4	5		1		1		2.5
Review3	4	5	4		5		2		4
Review4	1		5	4	1		1		2.5
Review5	3		4	1	4	4	2		3
Review6	2		5		2	5	1		3
Review7	4		4		4		2		3.5
Review8	1		5		1		1		2
Review9	4		4		5		2	5	4
Review10	2		5		2		1	2	2.5
Review Set 2	**Attr1: Autofocus (H_AH_C)**	**Attr2: Ease of Use (L_AL_C)**	**Attr3: Image Stabilization (H_AL_C)**	**Attr4: LCD Screen (L_AH_C)**	**Attr5: Optical Viewfinder (H_AH_C)**	**Attr6: Manual Mode (L_AL_C)**	**Attr7: Macro Mode (H_AL_C)**	**Attr8: Raw Format (L_AH_C)**	**Review Rating**
Review1	1		2	5	5		5		4
Review 2	2		4		4	2	3		3
Review 3	1		2		5	3	5		3
Review 4	2	3	4		4		2		3
Review 5	1		1		5		5	1	2.5
Review 6	1	5	1		5		4		3
Review 7	2		4	4	4		2		3
Review 8	2		4		4		2		3
Review 9	1		1		5		5		3
Review10	2		4		4		3	2	3
Review Set 3	**Attr1: Autofocus (H_AH_C)**	**Attr2: Ease of Use (L_AL_C)**	**Attr3: Image Stabilization (H_AL_C)**	**Attr4: LCD Screen (L_AH_C)**	**Attr5: Optical Viewfinder (H_AH_C)**	**Attr6: Manual Mode (L_AL_C)**	**Attr7: Macro Mode (H_AL_C)**	**Attr8: Raw Format (L_AH_C)**	**Review Rating**
Review1	1	4		4		2		5	3
Review 2	2	5		1		1		2	2
Review 3		4	5	5		2		4	4
Review 4		5	2	2		1		1	2
Review 5		4		5	2	2		4	2.5
Review 6		5		2	1	1		2	2
Review 7		4		4		2	4	5	4
Review 8		5		1		1	1	1	2
Review 9		4		5		2		5	4
Review10		5		2		1		2	2.5

Table A1. The Review Design Table (Continued)

Review Set 4	Attr1: Autofocus (H_AH_C)	Attr2: Ease of Use (L_AL_C)	Attr3: Image Stabilization (H_AL_C)	Attr4: LCD Screen (L_AH_C)	Attr5: Optical Viewfinder (H_AH_C)	Attr6: Manual Mode (L_AL_C)	Attr7: Macro Mode (H_AL_C)	Attr8: Raw Format (L_AH_C)	Review Rating
Review1		5	2	2		4		5	4
Review 2		2	1	1		2		4	2
Review 3	5	1		2		5		5	4
Review 4	1	4		1		1		4	2
Review 5		5		2		4		5	4
Review 6		1		1	2	2		4	2
Review 7		4		2	4	5	5	5	4
Review 8		2		1		1	4	4	2.5
Review 9		2		2		4		5	3.5
Review10		4		1		2		4	3

Appendix B

Review Website Used in Studies 1 and 2



Appendix C

Experimental Materials for Study 3

Experimental product. The digital camera was chosen based on a survey of a similar group of students that were not participants of the study. The students were asked to rate a large number of products on (1) their interest in the product and (2) whether they purchased these online. The digital camera emerged as one of the top products on both interest and purchase. The final selection of the product for the study also took into account the number of attributes that might be considered before a purchase decision.

Reviews. The 60 reviews were real customer reviews for Canon A590 IS randomly selected from Amazon.com. At the time of data collection, this camera had about 600 reviews. We randomly selected 60 reviews because 60 was approximately the average number of reviews digital cameras had on Amazon.com (among all the digital cameras that had reviews) at the time of the data collection. The brand and model name were removed from reviews so that participants' evaluation would not be biased by the brand name.

Attributes in the reviews. When we created the experiment materials for study 3, Amazon.com showed the attributes discussed in the reviews for the best-selling digital cameras (this feature is no long available on Amazon.com). We created a list of attributes discussed in the camera reviews from Amazon.com. One author of the paper and a coder who was blind to the objectives of the research read the 60 reviews used in the protocol study and removed the attributes that were not discussed in the 60 randomly selected reviews. This left us with 24 attributes discussed at least once in the 60 reviews (see Table C1).

Table C1. Digital Camera Attributes

Attribute	Description
Image quality	The quality of pictures produced by the camera
Battery	Whether the battery life is satisfactory
Portability	Whether the camera is easy to carry around
Ease of use	Whether the camera is easy to operate
Value for the money	Whether the camera offers good value
Manual mode	The availability and performance of manual mode
Lag time between shots	The delay between two consecutive shots
Viewfinder	The availability and usefulness of viewfinder
Feature	The usefulness of features provided by the camera
Video	The quality of video produced by the camera
Construction quality	Whether the camera is sturdy
Zoom	The performance of zoom
Look & feel	Whether the camera looks good and feels good in hand
LCD screen	The performance of LCD screen
Image stabilization	The availability and usefulness of image stabilization
Auto mode	The availability and performance of auto mode
Movement shooting	The quality of movement shooting
Low light performance	The performance of the camera under low light condition
Flash	The performance of the flash
Accessory	Whether necessary accessories (e.g. memory card, case) are provided
Lens	The quality of the lens
Face recognition	The performance of face recognition
Red eye reduction	The availability and performance of red eye reduction function
Documentation	Whether the manual is well organized

Appendix D

Descriptive Statistics for Study 3

Table D1. Mean and Standard Deviation for Each Variable per Attribute

Attribute	Importance weight	Amount of information	Degree of conflict	Coherence	Initial criterion?	Relevance
Image quality	26.065 (18.690)	4.161 (3.579)	.297 (.226)	.581 (.992)	.548 (.506)	.258 (.445)
Battery	23.339 (15.637)	5.000 (3.975)	.323 (.214)	.516 (1.458)	.290 (.461)	.532 (.499)
Portability	9.710 (9.353)	1.903 (1.491)	.053 (.142)	.258 (.930)	.387 (.495)	.065 (.359)
Ease of use	3.871 (7.079)	1.226 (1.746)	.028 (.109)	.290 (.902)	.065 (.250)	.032 (.180)
Value for the money	11.839 (17.506)	2.129 (2.202)	.030 (.117)	.194 (1.046)	.484 (.508)	.226 (.425)
Manual mode	3.613 (7.256)	2.065 (1.413)	.089 (.184)	.161 (.779)	.129 (.341)	-.048 (.373)
Lag time between shots	1.774 (4.573)	2.839 (2.464)	.132 (.212)	.097 (.944)	.032 (.180)	-.097 (.700)
Viewfinder	.000 (.000)	.484 (.626)	.000 (.000)	.065 (.359)	.000 (.000)	.000 (.000)
Feature	2.419 (5.458)	.903 (1.012)	.000 (.000)	.129 (.718)	.258 (.445)	.032 (.180)
Video	.000 (.000)	.903 (1.044)	.047 (.147)	.032 (.547)	.032 (.180)	-.097 (.396)
Construction quality	7.323 (11.441)	.548 (1.060)	.028 (.109)	-.129 (1.024)	.129 (.341)	.032 (.315)
Zoom	.645 (2.497)	.419 (.564)	.000 (.000)	.000 (.516)	.065 (.250)	-.065 (.250)
Look & feel	1.613 (6.375)	.839 (1.003)	.016 (.090)	.129 (.670)	.097 (.301)	-.097 (.301)
LCD screen	.903 (2.937)	1.355 (1.427)	.062 (.164)	.129 (.499)	.000 (.000)	-.113 (.442)
Image stabilization	.323 (1.796)	.935 (1.031)	.000 (.000)	.065 (.629)	.000 (.000)	.032 (.315)
Auto mode	.968 (5.388)	.613 (1.086)	.013 (.072)	.065 (.250)	.000 (.000)	.000 (.000)
Movement shooting	.000 (.000)	.129 (.341)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)
Low light performance	.000 (.000)	.161 (.454)	.000 (.000)	.065 (.359)	.000 (.000)	.000 (.000)
Flash	.806 (3.188)	.161 (.374)	.000 (.000)	.000 (.000)	.032 (.180)	.032 (.180)
Accessory	.323 (1.796)	.194 (.477)	.016 (.090)	.000 (.000)	.226 (.425)	-.032 (.180)
Lens	0.710 (3.598)	.742 (.773)	.016 (.090)	.065 (.359)	.065 (.250)	.016 (.273)
Face recognition	.000 (.000)	.258 (.575)	.000 (.000)	.097 (.396)	.000 (.000)	.000 (.000)
Red eye reduction	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)	.032 (.180)	.000 (.000)
Documentation	.000 (.000)	.097 (.301)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)

Notation: Mean (standard deviation)