

Informing Decisions: How people use online rating information to make choices

Stelios Lelis, Andrew Howes
Manchester Business School
The University of Manchester
Booth Street West, Manchester, M15 6PB, UK
{Stylianos.Lelis, Andrew.Howes}@mbs.ac.uk

ABSTRACT

In this paper we investigate how people use online rating information to inform decision making. We examine whether a theory of searching for information to discriminate between alternative choices can explain behavior, and we contrast it to the normative theory. Partly in accord with the theory, findings from a controlled experiment suggest that in an environment dominated by positive reviews, such as the World-Wide Web, people gather more information for the best alternative under consideration, and they take more time to inspect reviews of lower rating. We discuss the theoretical and experimental implications, and propose a bounded optimal account of the way in which people acquire information in service of decision making.

Author Keywords

User modeling, information search, decision making, online consumer reviews, e-commerce.

ACM Classification Keywords

H.1.2 Models and Principles: User/Machine Systems – Human Information Processing; H.3.3 Information Storage and Retrieval: Information Search and Retrieval – Selection process.

General Terms

Human Factors, Experimentation, Theory.

INTRODUCTION

People sometimes search the World Wide Web so as to inform decision making. They use the web to help choose clothes, movies, vacations, and even homes. They choose which stocks to invest in, and which healthcare treatments to adopt. To make these decisions, people collect information about alternatives and engage in a process of comparison that is then terminated with a choice.

Choice information search presents an important Human-

Computer Interaction (HCI) design challenge. It has been known for many years, for example, that simply designing e-commerce sites so that they present a logical organization of products, with meaningful link labels, is insufficient to support critical aspects of the user's task. Rather, what is needed is functionality that allows user's to bring the details of alternative products into close temporal and/or spatial proximity so that they can be contrasted [18, 20]. It is also known that choice tasks can be supported through web sites that encourage the propagation of word-of-mouth, experiential, information [5]. Indeed, systems for providing reviews, ratings, and information about what others' also liked have become pervasive [17], while some studies of how people search reviews have been conducted [16].

In parallel with the proliferation of exciting design work, HCI research has made substantial advances in understanding the ways in which people navigate the web for information. Information Foraging Theory (IFT) [23], for example, has substantially advanced our understanding of how people search information using computer systems. IFT explains user behavior in general search tasks, for example how people locate articles relevant to a topic in a large collection of documents. IFT predicts that a patch of information (for example a set of documents) should be left, so as to exploit another, when the rate of within patch gains diminishes below the expected average rate of gain. However, Pirolli and Card noted [23] "information foraging is usually a task that is embedded in the context of some other task, and the value and cost structure of information foraging are consequently defined in relation to the embedding task". More generally, it might be the case that the embedding task has substantial consequences for the information search strategy.

The Bayesian Satisficing Model (BSM) [6] and SNIF-ACT [7] also offer substantial and rigorous explanations of information search in HCI. BSM [6] has been applied to explain information seeking in a map navigation task. BSM successfully explained when participants stopped gathering information and made a choice of a route. SNIF-ACT [7] has been applied to explain when people stop evaluating links in a web page and either follow a link or backup a page, and successfully predicted user navigation behavior. Through a deep mathematical analysis of optimal behavior, complemented with empirical investigation, BSM and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

SNIF-ACT offer significant insights into when computer users stop gathering information, and/or evaluating links on a web page.

In the current paper, we report models that are inspired by the rigor of IFT, SNIF-ACT and BSM but which aim to complement them with analysis of choice information search. We are concerned with how information search is shaped by expectations about the value of information in embedding tasks, particularly embedding tasks that involve discriminating between choices. We distinguish between expectations about the utility of alternatives and expectations about the value of information, and examine how the former guide the latter. We are particularly interested in the effect of prior expectations concerning the utility of the choices on which alternative people decide to find out more about. To make the contrast to existing work clear: when searching for information about which digital camera to purchase (a preferred choice A or an alternative B), we can (1) just search efficiently for more information about A and B and not worry about the choice between them (e.g. as modeled by IFT) (2) choose whether to get more information about ‘both A and B’ or stop and act (e.g. as modeled by BSM), (3) decide whether to look for information about the stronger, currently preferred choice A, or its weaker alternative. Here we focus on this latter situation. We refer to the effect of *prior utility* on choice information search.

In addition, we complement our interest in understanding whether people search for information about preferred products with understanding how they use product ratings. For example, a review site may contain two reviews for a particular camera, one with 5 stars and the other with 2 stars, and we might ask which review should a consumer read first? In the marketing literature there is evidence that people place more weight on negative information (the 2 star review here) than positive in their decisions [13, 15]. However, the evidence about whether people prefer to seek positive or negative information and the explanation of their preferences is limited [26]. In this literature the positivity or negativity of information is called *valence*, and in the current paper we adopt the same term.

The goal of this paper then is to examine the effects of (1) prior utility and (2) review valence on choice information search. We first review theoretical questions concerning the value of information in choice tasks. Subsequently we develop a new theory of searching for information to discriminate between alternatives, and contrast it to the normative theory of information search. The review and theory development take Optimal Experimental Design (OED) [21, 22] as a point of departure. Finally, we describe an empirical study designed to test both theories. The study uses an ecologically motivated product purchasing task that focuses on online consumer reviews search, and which is amenable to formal analysis.

RELATED WORK

The Value of Information in Decision Making

There are many explorations of theories of the value of information as explanations of choice information search [2, 8, 26]. These studies have largely focused on the normative definition of the value of information known as the Expected Value of Sample Information (EVSI) [8, 24]. As with all theories of the value of information, the starting point for EVSI is Bayesian inference. It is assumed that people have prior beliefs about the utility of alternatives in choice, and that these prior beliefs guide the estimations of the value of information about the alternative. Where EVSI is distinctive, is in the idea that the value is the utility that would be gained by improving the decision. EVSI is defined as the expected gain in choice utility obtained by a likely choice reversal.

A number of studies of choice information search have examined whether people’s information gathering actions conform to the predictions of EVSI [8, 26]. Hagerty and Aaker [8] examined information gathering actions in alternative-by-attribute matrices. They found that in three-alternatives-by-seven-attributes choices EVSI correctly predicted what information participants collected in 19.3% of the cases. Although this is a three-fold increase compared to random selections, it also suggests that EVSI cannot offer a comprehensive account of information search. Simonson, Huber and Payne [26] also examined information search behavior in alternative-by-attribute matrices. They measured participants’ prior beliefs about alternatives and the certainty with which the beliefs were held, and they examined the acquisition order of information items. They found that, in accordance with EVSI, information concerning prior beliefs that are more uncertain is acquired earlier in the search process than information concerning more certain prior beliefs, and information about alternatives with high prior utility is acquired earlier compared to information about alternatives of low prior utility. Yet, they also observed an effect that cannot be explained by EVSI; that information concerning prior beliefs that are less favorable is acquired earlier in the search process than information concerning more favorable prior beliefs, especially for alternatives of lower utility. The access of information with negative-valence (i.e. less favorable) about inferior alternatives is not predicted by EVSI, or by an extension of EVSI with the value function of prospect theory [14], because this information cannot reverse the choice, while information with positive-valence about inferior alternatives can. In summary, these two studies [8, 26] offer partial support for EVSI as a model of information search in decision making.

In this article we explore an alternative theory of the value of information, the Discriminating Information Value (DIV). The theory is motivated by recent findings suggesting that people collect information to increase the discrimination between the alternative choices. Harvey and

Bolger [9] examined choice information search between apartments in alternative-by-attribute matrices. In these studies participants were remunerated according to a points system; the system attached a point score in each alternative and participants were awarded the points of the alternative they chosen. The studies were designed to examine whether, rather than using EVSI, participants collected information in an attempt to maximize the point score of the chosen apartment or, perhaps, to provide a basis on which alternatives could be eliminated. The authors found that neither premise offers a comprehensive account of information gathering actions, but rather that people collect information to increase the discrimination between the alternatives. In this article we explore this possibility. Below we describe the Discriminating Information Value (DIV) theory, and we also contrast it to EVSI - the normative theory. We contrast DIV to EVSI because we expect that, even if DIV cannot fully explain information search, the comparison will reveal the relative advantages of the discriminating search strategy.

THEORY DEVELOPMENT & PREDICTIONS

In this section we first describe the theory of the discriminating information value. Subsequently we deploy the theory to generate predictions about the preferences for which alternative to seek information, and the value of processing an information item (a review in our task) as a function of its valence. In addition we outline the predictions of EVSI, the normative theory.

Discriminating Information Value

In operationalising the value of information as the extent that it discriminates between alternatives we follow Optimal Experimental Design (OED) theories of the value of information [21, 22]. In OED theories the value of a question's usefulness is defined as the expected usefulness, given current knowledge of the possible answers. Each question is an information gathering action, and in our task the action is the one of obtaining and processing a review about an alternative in a choice scenario. We assume the rating of the review as the outcome of the information gathering action, or, in OED terminology, as the answer to the question. In a binary choice between alternatives i and j we define the Discriminating Information Value (DIV) of a rating r for alternative i , as the extent that it increases discrimination between the expected utilities of the two alternatives.

$$DIV_i(r) = \left| E(U_i | r) - E(U_j) \right| - \left| E(U_i) - E(U_j) \right| \quad (1)$$

Where $E(U_i)$ and $E(U_j)$ are the expected utilities of the two alternatives, and $E(U_i | r)$ is the expected utility of the alternative i after obtaining a review of rating r .

The value of processing one review for alternative i , the DIV of the alternative, is the sum of the values of processing a review of rating r across all ratings for this

alternative, weighted by the probability of obtaining the rating r for alternative i , $p_i(r)$ [21].

$$DIV_i = \sum_r p_i(r) \cdot DIV_i(r) \quad (2)$$

Now that we have stated the definition of DIV, we should see how it is estimated in order to generate the theory's predictions. To do that we should firstly define the prior beliefs about the proportions of alternatives' ratings, and derive alternatives' expected utility.

Let the decision maker hold beliefs about the proportion of alternative's i ratings, $\theta_{i,r}$, where r stands for rating. We assume that the beliefs follow a Dirichlet distribution with known parameter vector $(\alpha_{i,1}, \dots, \alpha_{i,R})$, where R is the number of possible ratings. The $\alpha_{i,r}$ terms can be seen as the number of times that a review of rating r has been obtained for alternative i and are called the prior strengths. $\nu = \sum_r \alpha_{i,r}$ is the total prior strength.

Given the prior beliefs, the utility of alternative i is given by the formula

$$U_i = \sum_r u_r \cdot \theta_{i,r} \quad (3)$$

Where u_r is the utility assigned to rating r , and is defined as the utility that the decision maker would obtain when experiencing an alternative of which all reviews are of rating r . The expected utility of alternative i is given by [1].

$$E(U_i) = \frac{\sum_r (\alpha_{i,r} \cdot u_r)}{\sum_r \alpha_{i,r}} \quad (4)$$

The DIV of each alternative and its ratings are estimated by projecting the processing of one more review for the alternative and are given by Equations (1) and (2). When a rating r is obtained the posterior beliefs of the proportion of ratings still follow a Dirichlet distribution with the r^{th} component of the parameter vector increased by one [4]. Therefore, the projected expected utility of the alternative after obtaining a review of rating r , $E(U_i | r)$, is given by Equation (4) with the updated beliefs. The probability of occurring a rating r , $p_i(r)$, is given by the Dirichlet distribution of prior beliefs. The calculation of the ratings' and alternative's DIV then follows.

Predictions

DIV predicts the value of processing a review for each alternative, and the value of processing a review as a function of the review's rating (valence). To develop DIV's predictions we assume ratings with values from 1 to 5, and take into consideration the distribution of review ratings in the WWW. Consumer reviews about products are highly positive, resulting in negatively skewed or J-shaped distributions of ratings per product [12], with average ratings greater than four [3]. Hu et al [12] analysis suggests that bimodal J-shaped distributions are the most common when all products are considered, and that negatively

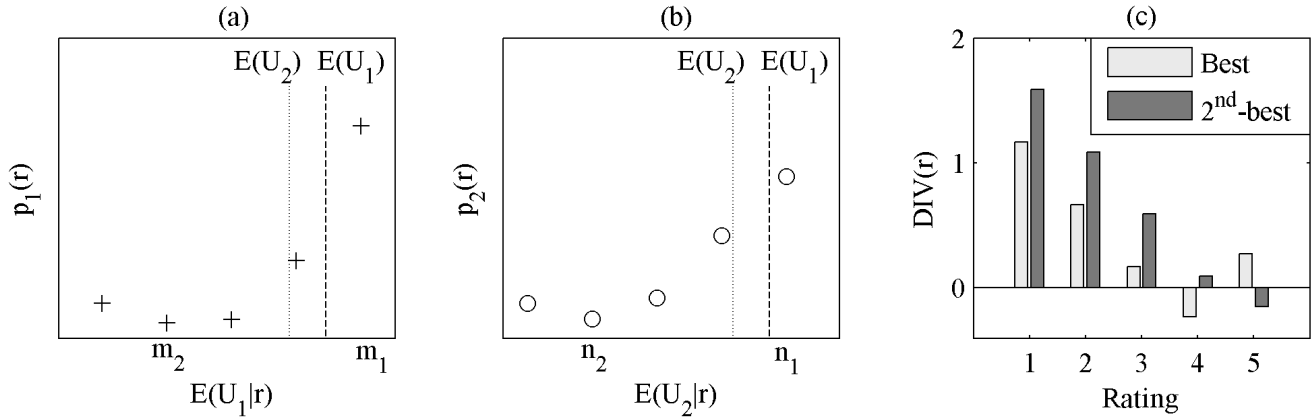


Figure 1 (a) Distribution of the posterior expected utility of the best alternative, $E(U_1|r)$. (b) Distribution of the posterior expected utility of the second-best alternative, $E(U_2|r)$. (c) Discriminating information value of each rating, $DIV(r)$ for both alternatives.

skewed distributions are equally common for products with high average rating. Consequently in the development of the predictions below we utilize J-shaped distributions with average rating larger than 4.

Preference for which alternative to seek reviews

Let a binary choice and one of the alternatives being better than the other, i.e. having a higher expected utility. Is there a preference for which alternative to seek information? To examine this we performed a simulation exercise to determine whether the information value of processing a review for the best alternative is higher than the information value of processing a review for the second-best. In the simulation we assumed that u_r , the utility assigned to each rating, is r . Alternatives' prior beliefs were initialized by firstly utilizing the hypergeometric distribution¹, and subsequently modifying ratings' occurrence probability so as to result in J-shaped distributions. The number of successes in the hypergeometric distribution was varied from 86 to 90 out of 100 for the best alternative and from 81 to 85 for the second-best. Subsequently 0.1 and 0.04 was subtracted from the probability of the 4- and 3-star ratings occurrence respectively, and 0.07 was added to both the probability of 1- and 5-star ratings occurrence. This procedure resulted in J-shaped beliefs with expected value larger than 4. Further we initialized the total prior strength of the beliefs to integer values from 1 to 10.

The simulation results showed that in 82.4% of the cases the DIV for information about the best alternative is higher than the DIV of the second-best. There are two things to

note about this percentage. First, obtaining a review for the best alternative is not always of higher value than obtaining a review for the second-best. Thus, depending on prior beliefs and total strength (i.e. the number of reviews that have been obtained), reviews of both alternatives can be of value. Second, the percentage is higher than 50%, demonstrating that in most of the cases obtaining a review for the best alternative is of higher value than obtaining a review for the second-best. We have verified that the percentage is similar for other J-shaped, and for negatively skewed distributions. Consequently, the DIV theory predicts that, when the prior beliefs are J-shaped or negatively skewed with expected value larger than 4, there is a higher preference to obtain reviews for the best alternative than the second-best.

To demonstrate why DIV predicts that people will have a higher preference to obtain reviews for the best alternative consider a choice scenario with J-shaped prior beliefs and average rating larger than 4. A positive 5-star rating for the best alternative increases discrimination, while the same rating (5-star) for the second-best alternative decreases discrimination. This is because a 5-star rating increases the expected utility of the best alternative, shifts it further away than the utility of the second-best and therefore increases discrimination. On the other hand, a 5-star rating for the second-best alternative decreases discrimination because it shifts its expected utility closer to the utility of the best alternative. Figures 1a and b illustrate this. They show the distributions of the posterior expected utility, $E(U|r)$ of the best and second-best alternative respectively (the leftmost point of the distribution is the expected utility of the alternative when obtaining a review of rating one, the next to the right a review of rating two, etc.). The prior expected utilities of the alternatives, $E(U_1)$ and $E(U_2)$ are also depicted. A 5-star rating for the best alternative increases the discrimination from $E(U_1) - E(U_2)$ to $m_1 - E(U_2)$ (Figure 1a), while a 5-star rating for the second-best alternative decreases it to $n_1 - E(U_1)$ (Figure 1b). On the other hand a moderate or negative rating for the second-best alternative

¹ The hypergeometric distribution describes the number of successes in a sequence of n draws without replacement from a sample of m objects, among which k are considered a success and $m-k$ a failure. For $n = 4$ the hypergeometric density function has 5 values (0, 1, ..., 4 successes), and for high k relatively to m the shape of the distribution is negatively skewed.

increases discrimination more than when it is obtained for the best alternative. Consider a 2-star rating for the second-best alternative. It considerably reduces its expected utility, shifts it further away than the utility of the best alternative, and increases discrimination from $E(U_1) - E(U_2)$ to $E(U_1) - n_2$ (Figure 1b). A 2-star rating for the best alternative also increases discrimination but only to $E(U_2) - m_2$ (Figure 1a). Note that for some of the moderate ratings of the best alternative the discrimination may be decreased, while it is increased when the same rating is obtained for the second-best alternative, for example the 4-star rating in Figure 1.

Consequently, as Figure 1c, which depicts the DIV of each rating for both alternatives, also illustrates, the DIV of negative and moderate ratings is higher for the second-best alternative, while the DIV of positive ratings is higher for the best alternative. However, positive ratings are very common, while negative and moderate ratings are rare. Thus, the DIV of the best alternative is generally higher, because the larger increases in discrimination for the best alternative realized by obtaining a positive rating, generally overshadow when taken in average, the larger increases in discrimination for the second-best alternative realized by obtaining a negative or moderate rating, because positive reviews are considerably more likely.

We also performed the simulation we described above for EVSI to examine the theory's prediction for which alternative to seek information. The simulation results showed that in 81.8% of the cases the EVSI for information about the best alternative is higher than the EVSI for information about the second-best. Consequently, and similarly to DIV, in a choice scenario EVSI predicts that people will obtain more reviews for the best alternative.

Value of processing a review according to rating

Here we examine the information value of processing a review as a function of a review's rating. We restrict the discussion, and therefore the predictions, to 4- and 5-star ratings, and distributions with expected value larger than 4.

DIV predictions: For the best alternative under consideration DIV predicts that reviews of 4-star rating are of lower information value than reviews of 5-star rating. This is because a 5-star rating increases the expected utility of the best alternative and therefore discrimination, while a 4-star rating decreases the utility of the best alternative, shifts it closer to the expected utility of the second-best, and therefore decreases discrimination (see also Figure 1a). For the second-best alternative DIV predicts that reviews of 4-star rating are of higher information value than reviews of 5-star rating. This is because a 4-star rating decreases the expected utility of the second-best alternative and therefore increases discrimination, while a 5-star rating increases the utility of the second-best alternative, shifts it closer to the expected utility of the best, and therefore decreases discrimination (see also Figure 1b).

EVSI predictions: Before developing the predictions of EVSI about the value of processing a review as a function of a review's rating let us recall the definition of EVSI. EVSI is the expected gain in choice utility obtained by a likely choice reversal. This would imply that if a rating does not reverse the choice it has zero EVSI. Nevertheless, as Chater et al. [2] noted, it is possible that the information may not immediately reverse the choice, but when combined with later information, it might do so. Consequently, a rating that brings the expected utilities of the two alternatives closer, thus making a future reversal more likely, has a higher value than a rating that increases the difference between the expected utilities of the two alternatives, thus making a future reversal less likely.

For the best alternative under consideration EVSI predicts that reviews of 4-star rating are of higher information value than reviews of 5-star rating. A 5-star rating for the best alternative increases its expected utility, does not reverse the choice, and makes a future reversal less likely. On the other hand a 4-star rating decreases the expected utility of the alternative and may reverse the choice. If it does reverse the choice it has a positive information value. Yet, even if it does not reverse the choice it makes a future reversal more likely than a 5-star review. This is because it shifts the expected utility of the best alternative closer to the one of the second-best (see also Figure 1a).

For the second-best alternative EVSI predicts that reviews of 4-star rating are of lower information value than reviews of 5-star rating. Note that a 4-star rating decreases the expected utility of the alternative, does not reverse the choice, and makes a future reversal less likely. On the other hand a 5-star review increases the expected utility of the second-best alternative and may reverse the choice. If it does reverse the choice – which is the case in Figure 1b – it has a positive information value. Yet, even if it does not reverse the choice, it makes a future reversal more likely than a 4-star review. This is because it shifts the expected utility of the second-best alternative closer to the one of the best.

To summarize, the two theories give contradicting predictions about the value of processing 4- and 5-star reviews. For the best alternative DIV predicts that 4-star reviews are of lower information value than 5-star, while EVSI predicts that 4-star reviews are of higher information value than 5-star. For the second-best alternative DIV predicts that 4-star reviews are of higher information value than 5-star, while EVSI predicts that 4-star reviews are of lower information value than 5-star. Finally, note that the predictions are the same for J-shaped and negatively skewed distributions with expected value larger than 4, as they are independent of ratings' occurrence probabilities.

EXPERIMENT

To test the predictions and contrast the two theories we conducted an experimental study of how human preferences

	Alt. 1	Alt. 2	Alt. 3
Price	75	81	87
LCD Size	2.5	2	2.25
Digital Zoom	4	2	3
Battery Life	275	225	175
Weight	170	200	230
Dimensions	8.5×2.0×5.3	9.4×2.4×6.2	9.0×2.2×5.8
<i>Note.</i> The remaining attributes and their values were: Resolution: 7 mega-pixel; Optical zoom: 3x; Battery type: Proprietary Lithium; Memory Type: Compact Flash Card; Supplied card: 32MB; Red eye reduction: Built-in; ISO Levels: Auto, 50, 100, 200, 400; Video: Yes;			

Table 1. An example of alternatives' attribute values.

are shaped by prior utility and valence. The experiment used a digital camera choice task with high ecological validity, in which participants were first exposed to attributes information and then reviews. We used multiple levels of differences between alternatives attribute values in order to ensure that the results are robust to these differences.

The Task

The task involved a choice between three alternative digital cameras². Participants were initially exposed to information about alternatives' attribute values. Subsequently, they were free to read consumer reviews for each camera, and the number and order of reviews obtained for each alternative, and the time they took to read each review, were measured.

Each participant performed a practice task, followed by two experimental tasks. The practice task involved a choice between three mega-pixel cameras, and the experimental tasks involved choices between five and seven mega-pixel cameras. In each task participants were first presented with attribute information in a single web page, and when they indicated that they had obtained as much information as they wanted, they were forwarded to another web page listing consumer reviews about the alternatives. While in this page participants could not go back to the page presenting alternatives' attributes and obtain more information.

Specific instructions were given to the participants. They were instructed to imagine that they needed to purchase a digital camera and select one that was suitable. They were told that they could take as much time as they wanted in order to inspect information and make their choice. Finally, and as a motive to engage with the task, they were informed

² Choices between three alternatives were utilized in the experiment although the reported theory describes opinion search in binary choices. The third, inferior in terms of ratings alternative, was introduced so as to increase the ecological validity of the task

that one of them, randomly selected, would be awarded one of the digital cameras s/he would choose.

Design and Materials

The attribute information was presented in a single web page, with layout very similar to the 'Compare Products' pages of real web sites like, for example, Epinions.com. To enhance the ecological validity of the study, choice alternatives were described in addition to features with a fictitious name and a camera picture. The pictures were randomly selected for each trial from a pool of ten camera images. The ten images were of similar looking compact digital cameras, downloaded from amazon.co.uk and processed to remove any brand information. Camera names, pictures, and order of presentation, were randomized for each choice.

Each alternative was described with fourteen attributes, and the values of six of them were randomized resulting in a large number of different choice sets. The randomization was performed so as to give multiple levels of differences between alternatives attribute values. As a result in some of the choices one alternative was better than the other two, while in others the three alternatives were more or less similar in terms of attribute values. This was done in order to make sure that people's preference for which alternative to obtain more reviews is robust to these differences. Table 1 shows an example choice scenario.

After inspecting the attributes information participants were free to read consumer reviews for each alternative. They were forwarded to another web page. In this page they were presented in a structure similar to the attributes web page, and in one column for each alternative, with its name, picture, and a numbered list of its reviews in the form 'Review 1', 'Review 2', etc. To access a review participants should place the mouse pointer over the text 'Review x' and the review opened automatically in a pop-up window. A delay of two seconds was introduced between placing the

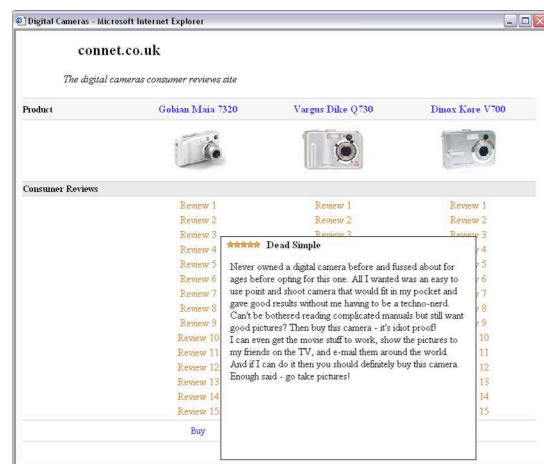


Figure 2. The reviews web page with one of the reviews being displayed.

	Ratings				
	Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
Best	5, 5, 4	5, 5, 4	5, 5, 4	5, 5, 4	5, 5, 4
Second-best	5, 4, 4	5, 4, 3	5, 5, 3	5, 5, 2	5, 4, 4
Worst	5, 4, 1	5, 3, 2	5, 4, 2	5, 4, 1	5, 4, 1

Table 2 The subsets that alternatives' ratings were grouped.

pointer over the review link, and the review being displayed. Figure 2 depicts the web page with one of the reviews opened.

There were 15 reviews for each alternative. The ratings of each alternative's reviews were the same in all choices. Alternatives were firstly assigned ratings and then reviews according to the assigned ratings. Three sets of review ratings were utilized, and each set was assigned to one of the alternatives. The means of the ratings sets ranked them as best, second-best and worst. The distribution of ratings was similar to the type of distributions obtained for real products in the World Wide Web; i.e. were negatively skewed or J-shaped. The best and second-best ratings' sets were negatively skewed, and the worst set was J-shaped – reflecting the statistics of the environment reported earlier. The sets were assigned to alternatives as follows. In the choices where one of the alternatives had better attribute values than the other two, the best set was assigned to this alternative, and the other two alternatives were randomly assigned one of the second-best and worst sets. In the choices where attributes were more or less similar for all attributes, the three sets were randomly assigned. Below I refer to the three alternatives as best, second-best and worst as prescribed by the rating sets that was assigned to it.

Six sets of reviews according to alternatives' ratings were constructed; two for each alternative. Each set contained 15 reviews, which were selected so as not to contradict each other. Each alternative's sets of reviews were randomly, and separately for each participant, assigned between the two tasks that s/he made. So as to maintain ecological validity we sampled reviews from amazon.co.uk. The reviews were selected according to specific criteria. The criteria were that the review (i) had a length between 100 and 150 words, (ii) did not mention attributes different to the ones of our fictitious products, (iii) did not mention specific characteristics of other reviews, (iv) did not mention Amazon, (v) did not mention that the camera comes with extras such as case, camera dock etc., (vi) did not compare the reviewed camera to competitors, or describe the choice between two cameras, and (vii) did not extensively review the video capability. No further control was exercised on the sample. Further, a total of 90 reviews was used (44 5-star and 28 4-star), so as to minimize any effects of idiosyncratic reviews on the dependent variables. After sampling, some reviews were slightly edited for spelling, and brand and model names were replaced by the fictitious ones.

The order of review ratings was semi-randomized, and the order of reviews was fully randomized. This was done so as to (i) avoid any effects of sequential, highly positive, or negative ratings, (ii) equally position reviews of all ratings throughout the search process in order to control for the effect of fatigue on the dependent variables, and (iii) position full reviews in various stages of the search process so as to further control for the effects of fatigue and idiosyncratic reviews. The set of ratings for each alternative were grouped into subsets of mean rating similar to the average of the whole set (see Table 2). The order of subsets was held constant, yet the order of ratings within each subset was randomized. This procedure resulted in a large number of different ratings orders for each alternative. After determining the order of ratings, the full reviews were randomly selected according to their rating from each alternative's set of reviews.

Finally choice (five or seven mega-pixels), and order of differences between alternatives' attribute information were counterbalanced across participants.

Participants

Twelve members of a large UK university's community volunteered to take part in the experiment and were paid £5 for their participation. The participants were recruited by a posting an add to the university's web site. After the completion of the study, the incentive for engaging with the task was administered, and a randomly selected participant was awarded one of the digital cameras s/he chose. Since the digital cameras used in the experiment were fake, we awarded a camera with features similar to the one the randomly selected participant chose.

Results

Below we first provide an analysis of participants' preference for which alternative to seek reviews, followed by an analysis of the average time they took to inspect reviews of 4- and 5-star rating, which is an estimate of the value of processing a review according to rating.

Preference for which alternative to seek reviews

To examine participants' preference for which alternative to seek reviews we consider two dependent variables; the number of reviews participants acquired for each alternative, and the acquisition priority of each alternative's reviews. Figure 3a shows the number of reviews participants acquired per alternative. An analysis of variance (ANOVA) was conducted using alternative (best, second-best, and worst) as the within-subjects variable. The

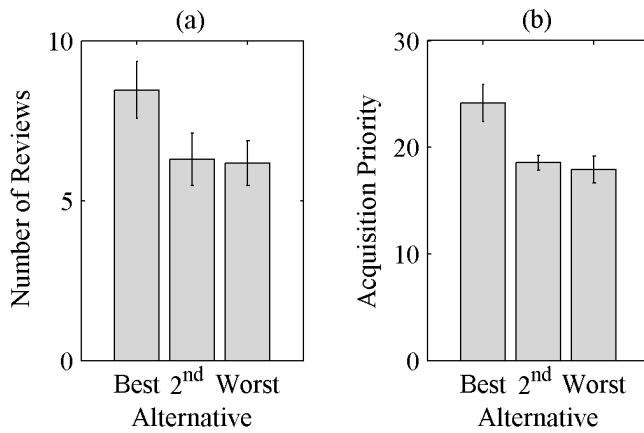


Figure 3. (a) The number of reviews participants acquired per alternative. (b) Alternatives' acquisition priority. Error bars represent standard errors of the means.

analysis revealed a significant effect of alternative, $F(2,36) = 9.7$, $MSE = 4.1$, $p < .001$. As Figure 3a illustrates participants obtained more reviews for the best alternative than for the second-best and the worst. The planned comparisons between the number of reviews participants obtained for the best and the other two alternatives were tested by performing pair-wise two-tailed t -tests. The t -tests show that the differences between the best, and the second-best and worst alternatives are significant ($t(23) = 3.78$, $p = .001$; and $t(23) = 4.41$, $p < .001$ respectively, two-tailed).

The acquisition priority (AP) is a measure of the order of reviews processing. It summarizes how early or late in the search process reviews about each alternative are acquired. AP is the measure described in [26], updated with the total number of information items (reviews) in our task (45), the number of alternatives (3), and the number of reviews that each participant acquired. The larger the value of AP the earlier the reviews about this alternative are acquired. Differences between the value of AP were tested with a one-way ANOVA using alternative (best, second-best, and worst) as the within-subjects variable. The analysis revealed a significant effect of alternative, $F(2,22) = 5.56$, $MSE = 25.6$, $p < .05$. As Figure 3b illustrates, the best alternative has the largest AP while the second-best and the worst alternatives have relatively equal. That is, participants obtained reviews for the best alternative earlier than for the second-best and the worst. The planned comparisons between the AP of the best and the other two alternatives were tested by performing pair-wise two-tailed t -tests. The t -tests show that the differences between the best and second-best, and the best and worst alternatives are significant ($t(11) = 3.17$, $p < .01$; and $t(11) = 2.32$, $p < .05$ respectively).

Time to inspect reviews per rating

To examine the value of processing a review according to its rating, we consider the average time participants took to inspect reviews of each rating. The analysis was done

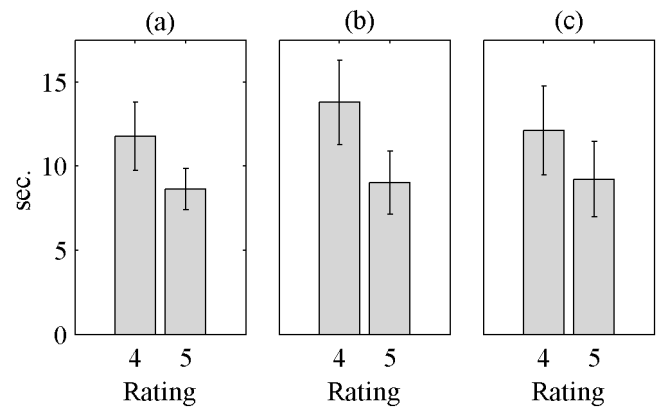


Figure 4. The average time participants took to inspect each review per rating for (a) best, (b) second-best, and (c) worst alternative. Error bars represent standard errors of the means.

separately for each alternative. Only 4- and 5-star reviews were examined.

For the best alternative a one way ANOVA was conducted using rating (4, 5) as within-subjects variable. The analysis revealed a significant main effect of rating, $F(1,11) = 5.05$, $MSE = 11.8$, $p < .05$. As Figure 4a illustrates participants took more time to inspect 4-star than 5-star reviews.

For the second-best alternative a one way ANOVA was conducted using rating (4, 5) as within-subjects variable. One participant who did not obtain a 5-star review was removed from the analysis. The effect of rating is significant, $F(1,10) = 17.2$, $MSE = 7.3$, $p < .01$. Participants took on average more time to inspect 4-star than 5-star reviews, Figure 4b.

Finally, a one way ANOVA was conducted for the worst alternative using rating (4, 5) as within-subjects variable. Two participants who did not obtain a 4- or 5-star review were removed from the analysis. The analysis revealed a significant effect of rating, $F(1,9) = 26.8$, $MSE = 1.57$, $p = .001$. As Figure 4c illustrates participants took more time to inspect 4-star than 5-star reviews.

GENERAL DISCUSSION

In this paper we have addressed the question of how people use rating information, found on many e-commerce web sites, in order to support choice information search tasks. We described how the Discriminating Information Value (DIV) theory made a number of predictions about user' behavior and we showed how results of an empirical study provide some evidence in favor of DIV. First, in an environment dominated by positive reviews, such as the World-Wide Web, participants gathered more information for the best alternative under consideration. Second, participants took more time to inspect reviews with lower ratings for inferior alternatives.

However, we also found evidence that is counter to DIV's predictions: when evaluating the best alternative,

participants took more time to inspect reviews with 4-star ratings than reviews with 5-star ratings. In fact, the experimental results suggest that people have a tendency to take more time to inspect a review with lower rating irrespective of the ranking of the alternative. This, in turn, suggests that for participants, if not for the DIV theory, the lower the rating of a review the higher its information value.

Further analyses of the implications of the data for the normative theory of information search (EVSI) were conducted. In contrast to DIV, we found that EVSI could explain why participants took more time to inspect 4-star than 5-star reviews for the best alternative – because it increases the chance of a response reversal. However, EVSI did not explain why participants had the same preference, for 4-star over 5-star reviews, when evaluating an inferior alternative. Therefore, neither EVSI nor DIV offers an explanation of participants' information gathering preferences in the experimental choice task.

In sum then, while the results show unequivocal effects of beliefs about prior utility and of rating value on choice information search, the direction of the effects does not support existing theoretical accounts. The findings therefore present a considerable challenge to explanation, a challenge that we start to address in the following section.

Alternative Explanation

In order to introduce a new approach to understanding choice information search we take a step back and reframe the problem as a problem of bounded utility maximization [10, 11, 19, 25, 27]. Bounded utility maximization assumes that individuals adapt rationally to a utility function given constraints imposed by their cognitive abilities and their experience. Bounded utility maximization radically differs from normative accounts of behavior, by taking into account cognitive and task constraints. For example, the normative account of choice information search, EVSI, is only normative if temporal information constraints are ignored. We need a theory that is optimal given the constraints experienced by users.

The constraints experienced by users include time. Users have limited time to process information, and they trade-off time for choice utility. A means to reduce the time effort of choice is to simplify the decision by increasing the discrimination between alternatives (DIV). On the other hand a means to increase choice utility is to search for information that may reverse the choice (EVSI). From this perspective we can see that EVSI and DIV are not in competition, but rather that they can each contribute toward solving the utility maximization problem in different ways. The balance between them may be quantified and studied by examining the maximization of the difference between the choice utility and the cost (as expressed in time or cognitive effort) of making the choice. Of course the balance selected by a user will depend on cognitive

constraints (e.g. limitations of short-term memory), and on a range of constraints that include aspects of the web site design such as the cost of retrieving reviews from different products and with different ratings.

Here we briefly sketch how DIV and EVSI might work together to explain the results of the reported experiment. DIV predicts that reviews of lower rating are of higher value for inferior alternatives because they largely increase discrimination. In effect these lower ratings decrease the utility of the alternative considerable, and allow the decision maker to eliminate it from further consideration and save information processing costs and time. On the other hand, EVSI predicts that reviews of lower rating are of higher value for the best-alternative because they significantly reduce its utility and have the potential to reverse the choice, therefore saving utility. That is, processing less favorable information for all alternatives results in gains; it reduces the costs of processing information and time effort (as exemplified by DIV), and increases the choice utility (as exemplified by EVSI). Therefore, it appears that both theories have advantages in explaining review processing, and together contribute to the understanding of why people take more time to inspect less favorable reviews.

Design Implications

As we have said, it is critical that web site designs are sufficiently flexible to allow people to easily compare alternative choices, whether the choices are between health plans ,holidays, or financial investments [20]. Supporting comparison involves supporting low cost of access to the most useful information when it is needed. To the extent that the reported study reveals the information that participants find most useful in choice information search, it could form the basis of research aimed at informing design. For example, the results of the study suggest that in an environment dominated by positive information, such as the World-Wide Web, people will prefer to process more information for the best alternative, and that they will prefer to process negative information about all alternatives under consideration. These findings suggest that any interface that reduces the cost of access to information about higher utility products and to negative reviews will make it easier for people to make choices. There may be ways to take advantage of these ideas in single vendor sites, as well as on product comparison sites, and product forums.

Limitations

Further experimentation is required to ascertain whether the results are replicated with more participants and to ascertain whether they generalize beyond the sample student population.

CONCLUSION

The research reported in this paper extends previous work in the following ways: It developed an OED theory of the

discriminating value of information in decision making, and examined whether people gather information to discriminate between alternatives. Contrary to earlier findings [9], it suggests that people do not collect information merely to discriminate between alternative choices. In addition, by contrasting the theory to the normative account of information search (EVSI), it offers a descriptive, yet formally justified, explanation of users' preference to obtain less favorable information in environments dominated by positive information, such as the World-Wide-Web. Finally, building on its theoretical and experimental insights, it paves the way for the development of a promising theory of information search in choice tasks.

ACKNOWLEDGMENTS

This work was supported by a PhD Fellowship Award to the first author from the School of Informatics and Manchester Business School at the University of Manchester.

REFERENCES

- Bernard, J.-M. Non-parametric Inference about an Unknown Mean using the Imprecise Dirichlet Model. *ISIPTA 2001* (2001).
- Chater, N., Crocker, M.J. and Pickering, M.J. The rational analysis of inquiry: the case of parsing. In Oaksford, M. and Chater, N. eds. *Rational Models of Cognition*, OUP (1998), 441-568.
- Chevalier, J. and Mayzlin, D. The effect of Word of Mouth on sales: Online Book Reviews. *Journal of Marketing Research* 43, 3 (2006), 345-354.
- DeGroot, M. *Optimal Statistical Decisions*. McGraw-Hill (1970).
- Dellarocas, C. The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science* 49, 10 (2003), 1407-1424.
- Fu, W.-T. and Gray, W. Suboptimal tradeoffs in information seeking. *Cognitive Psychology* 52 (2006), 195-242.
- Fu, W.-T. and Pirolli, P. A Cognitive Model of User Navigation on the World Wide Web. *Human-Computer Interaction* 22, 4 (2007), 355-412.
- Hagerty, M. and Aaker, D. A Normative Model of Consumer Information Processing. *Marketing Science* 3, 3 (1984), 227-246.
- Harvey, N. and Bolger, F. Collecting information: Optimizing outcomes, screening options, or facilitating discrimination? *The Quarterly Journal of Experimental Psychology* 54A, 1 (2001), 269-301.
- Howes, A., Lewis, R. and Vera, A. Rational Adaptation Under Task and Processing Constraints: Implications for Testing Theories of Cognition and Action. *Psychological Review* 116, 4 (2009), 717-751.
- Howes, A., Vera, A. and Lewis, R. Cognitive constraint modeling: A formal approach to reasoning about behavior. In *Proc. CogSci 2004*, Erlbaum (2004), 595-600.
- Hu, N., Pavlou, P. and Zhang, J. Can Online Reviews Reveal a Product's True Quality? Empirical Findings and Analytical Modeling of Online Word-of-Mouth Communication. In *Proc. EC 2006*, ACM Press (2006), 324 - 330
- Huang, J.-H. and Chen, Y.-F. Herding in Online Product Choice. *Psychology & Marketing* 23, 5 (2006), 413-428.
- Kahneman, D. and Tversky, A. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica* 47, 2 (1979), 263-291.
- Lee, J., Park, D.-H. and Han, I. The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications* 7 (2008), 341-352.
- Lelis, S. and Howes, A. A Bayesian Model of How People Search Online Consumer Reviews. In *Proc. CogSci 2008*, Cognitive Science Society (2008), 553-558.
- Li, X. and Hitt, L. Self-Selection and Information Role of Online Product Reviews. *Information Systems Research* 19, 4 (2008), 456-474.
- Lohse, G. and Johnson, E. A Comparison of Two Process Tracing Methods for Choice Tasks. *Organizational Behavior and Human Decision Processes* 68, 1 (1996), 28-43.
- Malloney, L. and Mamassian, P. Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. *Visual Neuroscience* 26 (2009), 147-155.
- Miles, G., Howes, A. and Davies, A. A framework for understanding human factors in web-based electronic commerce. *Int. J. Human-Computer Studies* 52 (2000), 131-163.
- Nelson, J. Towards a rational theory of human information acquisition. In Oaksford, M. and Chater, N. eds. *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* OUP (2008), 143-163.
- Oaksford, M. and Chater, N. Optimal data selection: Revision, review, and reevaluation. *Psychonomic Bulletin & Review* 10, 2 (2003), 289-318.
- Pirolli, P. and Card, S. Information Foraging. *Psychological Review* 106, 4 (1999), 643-675.
- Raiffa, H. and Schlaifer, R. *Applied Statistical Decision Theory*. Division of Research, HBS (1961).
- Russel, S.J. and Subramanian, D. Provably bounded-optimal agents. *Journal of Artificial Intelligence Research* 2 (1995), 575-609.
- Simonson, I., Huber, J. and Payne, J. The Relationship between Prior Brand Knowledge and Information Acquisition Order. *Journal of Consumer Research* 14, 4 (1988), 566-578.
- Vera, A. and Howes, A. A constraint satisfaction approach to predicting skilled interactive performance. In *Proc. CHI 2004*, ACM Press (2004), 121-128.