The adaptation of visual search to utility, ecology and design

Tseng, Yuan-chi; Howes, Andrew

DOI:
10.1016/j.ijhcs.2015.03.005

www.elsevier.com/locate/ijhcs

The Adaptation of Visual Search to Utility, Ecology and Design

Yuan-Chi Tseng, Andrew Howes

PII: S1071-5819(15)00058-0
DOI: http://dx.doi.org/10.1016/j.ijhcs.2015.03.005
Reference: YIJHC1950

To appear in: *Int. J. Human-Computer Studies*

Received date: 15 April 2014
Revised date: 20 March 2015
Accepted date: 30 March 2015

Cite this article as: Yuan-Chi Tseng, Andrew Howes, The Adaptation of Visual Search to Utility, Ecology and Design, *Int. J. Human-Computer Studies,* http://dx.doi.org/10.1016/j.ijhcs.2015.03.005

The Adaptation of Visual Search to Utility, Ecology and Design

Yuan-Chi Tseng[1], Andrew Howes[2]

[1]Department of Industrial Design, National Cheng Kung University, 1 University Road, Tainan, 701, Taiwan

(yctseng@mail.ncku.edu.tw)

[2]School of Computer Science, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

(HowesA@bham.ac.uk)

Running Head: *Utility Maximization and Scanning*

Corresponding Author: Yuan-Chi Tseng

yctseng@mail.ncku.edu.tw

Phone: 886-6-275-7575 ext 54335

ABSTRACT

An important question for Human-Computer Interaction is to understand the visual search strategies that people use to scan the results of a search engine and find the information relevant to their current task. Design proposals that support this task include space-filling thumbnails, faceted browsers, and textually enhanced thumbnails, amongst others. We argue that understanding the trade-offs in this space might be informed by a deep understanding of the visual search strategies that people choose given the constraints imposed by the natural ecology of images on the web, the human visual system, and the task demands. In the current paper we report, and empirically evaluate, a computational model of the strategies that people choose in response to these constraints. The model builds on previous insights concerning the human visual system and the adaptive nature of visual search. The results show that strategic parameters, including the number of features to look for, the evaluation-stopping rule, the gaze duration and the number of fixations are explained by the proposed computational model.

Keywords: Image Search, Strategic Adaptation, Utility Maximization, Ecology, Information Design, Eye movements

1. INTRODUCTION

The question of how to design interfaces so as to facilitate search for information on the web is an important challenge for Human-Computer Interaction researchers and designers (Cutrell & Guan, 2007; Klöckner, Wirschum, & Jameson, 2004; Rele & Duchowski, 2005; Russell-Rose & Tate, 2013; Tseng & Howes, 2008). For example,

many alternative designs for the standard list of search engine results have been proposed. They include Space-filling thumbnails (Cockburn, Gutwin, & Alexander, 2006), Tabular interface (Resnick, Maldonado, Santos, & Lergier, 2001), Faceted category interfaces (Yee, Swearingen, Li, & Hearst, 2003), and Textually-enhanced thumbnails (Woodruff, Faulring, Rosenholtz, Morrison, & Pirolli, 2001), amongst others (Bederson, 2001; de Bruijn & Spence, 2000 ; Fertig, Freeman, & Gelernter, 1996; Öquist & Goldstein; Snavely, Seitz, & Szeliski, 2006; Walter, Weßling, Essig, & Ritter, 2006). The number of proposals is, in part, a reflection of the scale of the design space, and in part, a reflection of how pervasively search engines are used for a range of everyday tasks.

Identifying which of the potentially hundreds of interesting points in this space is best might be informed by empirical usability testing that directly contrasts one design to another. However, while an empirical basis to any design work is essential, such an approach, used exclusively, carries the danger of leaving mysterious the underlying interactive processes that lead to one advantage, or the other, and may be unlikely to lead to rapid convergence on good designs. For example, some designs will be better or worse in different circumstances and explaining the differences demands theory. An alternative approach is to rely on design guidelines. For example, findings concerning the function of human mind, using conventional laboratory tasks, have implied user-centered design guidelines for more efficient use (Shneiderman, 1992). However, it is possible that guidelines may have similar problems to pairwise usability tests if they encourage a somewhat shallow understanding of the task and the constraints imposed by the design. This is one of a set of known problems with design guidelines (e.g. see Introduction in Johnson, 2010).

Empirical studies and guidelines can be complemented with cognitive modeling, with the potential advantage of developing a deeper understanding of how and why one design is better than another. Many cognitive models of visual search have been proposed. For example, integrated ACT-R and EPIC models of the cognitive, motor, and perceptual processing required to achieve visual search tasks provide one approach (Anderson et al., 2004; Halverson & Hornof, 2011; Kieras & Meyer, 1997; Meyer & Kieras, 1997a, 1997b). One key idea to emerge from this literature is that it is difficult to ascertain whether one design is better than another unless a detailed analysis is conducted of how the user's strategy changes with the task demands (Charman & Howes, 2003; Eng et al., 2006; Howes, Lewis, & Vera, 2009; Howes, Vera, & Lewis, 2007; Kieras & Meyer, 2000; S. J. Payne & Howes, 2013; S. J. Payne, Howes, & Reader, 2001; Pirolli, 2007). The purpose of the current article is therefore to report and test a computational model of visual search in which the search strategy is adjusted to the constraints imposed by (a) interface design, (b) the human visual system and (c) the priorities of the user, particularly priorities concerning time costs relative to the quality of the acquired results. The model explains how choices that people make about, for example, gaze duration, number of fixations, and which images to look at and which to select, are a consequence of ecological distributions of relevance, the diminishing acuity of the visual system with eccentricity from the fovea, and the priorities of the user.

In the following section we review the background to the problem addressed in the current article (Section 2). Subsequently, we described an experimental task environment that models the naturalistic task environment of images on the web (Section 3) and then a model of human behavior in this environment (Section 4),

followed by an empirical investigation (Section 5). The results are reported in Section 6 and discussed in Section 7.

2. BACKGROUND

2.1. The effect of design on strategy choice

In an effort to understand how to build better interfaces, researchers in HCI have suggested that the details of interface design affect visual search and overt attention strategies (e.g., Everett & Byrne, 2004; Halverson & Hornof, 2004, 2011; Pirolli, Card, & Van Der Wege, 2003; Tseng & Howes, 2008). Everett and Byrne (2004), for example, showed that a small difference of 1.6 degrees of visual angle between items can result in participants either fixating on an icon or not. Similarly, Halverson and Hornof (2004) provided evidence that low density, task-meaningless large font words could lead participants to use fewer and shorter fixations and so shorter overall search time than when given high density and small words. Presumably, when items are more closely packed together then more use can be made of peripheral vision. However, the pattern of findings is complex. In contrast to the previous results, Pirolli et al. (2003), for example, found that participants used more but shorter fixations when using a Hyperbolic browser (please see Lamping & Rao, 1996) than when using a standard browser, especially in areas of the Hyperbolic browser in which small size and low information scent items were grouped closely together.

There is further evidence showing that visual search strategy is adapted to the demands imposed by task environments, particularly the density of items on the

display (Bertera & Rayner, 2000; Näsänen, Ojanpää, & Kojo, 2001; Ojanpää, Näsänen, & Kojo, 2002; Vlaskamp & Hooge, 2006; Vlaskamp, Over, & Hooge, 2005). For example, Ojanpää et al. (2002) found that decreased spacing in a vertical list of words (common Finnish verbs, nouns and adjectives) resulted in longer but fewer fixations. Longer fixations enable more information to be gathered from fovea and peripheral vision, although longer fixations can only be effective if the information is available within the perceptual span. Vlaskamp et al. (2005) found that the fixation duration, number of fixations, and search time increased dramatically with decreasing item spacing, as the range of spacing became smaller than 1.5° visual angle. On the other hand, their data showed that at wide spacing range between 1.5° to 7.1° fixation duration, number of fixations, and search time increased slightly as the spacing increased. Bertera and Rayner (2000) found that as the item spacing increased, the number of fixations and fixation duration also increased. These results indicate that people need to manage the trade-off between the increased information gain of longer fixations and the effort and time cost of holding a fixation.

The spacing of fixations is also known to change during the course of a search (Over, Hooge, Vlaskamp, & Erkelens, 2007; Rao, Zelinsky, Hayhoe, & Ballard, 2002). Over et al. (2007) found that fixation duration increased and the amplitude of saccade decreased gradually as search progressed. They called this a coarse-to-fine strategy. Rao et al. (2002) used a coarse-to-fine matching mechanism to model the skipping saccades because it could increase the probability of an early match. In contrast, Brumby and Howes (2008) found a fine-to-coarse search strategy. People increased saccade amplitude once they had found, but not committed to, a highly relevant target.

Although allowing a high degree of experimental control, many tasks used in vision science lack ecological validity (Bertera & Rayner, 2000; Ojanpää et al., 2002; Vlaskamp et al., 2005). For example, Vlaskamp et al. (2005) used abstract shapes (e.g., squares) in their search task, and, Bertera and Rayner (2000) used an unstructured alphanumeric array. Ojanpää et al. (2002) used common-words, which reduced the task to a simple visual pattern match, rather than a match of information relevance but it is known that search behavior is contingent on label relevance (Brumby & Howes, 2008). The different materials may account for the different effects. Both tasks are far from an ecologically valid HCI task in which the stimuli are more heterogeneous and complicated.

The task of Brumby & Howes (2008) had high ecological validity but the task, involving words, required limited graphical information processing. To explore graphical information processing in a task environment involving search for pictures, Tseng and Howes (2008) used real photo thumbnails and the real display to simulate the pages of thumbnails returned by a search engine and found that the number of alternatives in a search set and the density of the display influence how people make small but significant changes to eye-movement strategy. For example, as item density changed their participants adjusted the duration that they attended to each item. There was a negative correlation between the number of items and the gaze duration. Longer gazes were only used when they were efficient, i.e. when expected information gain was high. Also, longer visits to items were combined with skipping. Tseng and Howes (2008) found that participants were observed to reduce the number of items that they visited, i.e. they skipped, when there was a larger number of alternatives in a search set. These findings support the view that people adjusted their visual search strategy to their expectations of information gain, and that these expectations were

contingent on (a) the density of items, and (b) the prior likelihood that an item is the one that they will want to select.

2.2. Visual information processing constraints modulate strategy choice

There is a substantial body of literature showing how the constraints of the visual information processing system modulate strategy choice. Geisler (2011) reviews work that shows how visual search strategies are adapted to the biological constraints of the human visual system, such as the optics of the eye, the spatial and chromatic sampling by the photoreceptors, photon noise, and retinal spatial summation, all of which provide noisy neural representations of the stimulus.

Some of these constraints are embedded in an active vision model (Halverson & Hornof, 2011; Kieras & Hornof, 2014). By using the EPIC cognitive architecture (Kieras & Meyer, 1997) to model visual search behavior, they attempt to explain the relationship between strategies and the underlying architectural mechanisms. These models provide answers to the four questions of active vision: when do the eyes move? What can be perceived? Where do the eyes move to? What information is integrated between eye movements? In this model (Halverson & Hornof, 2011; Kieras & Hornof, 2014), they emphasized the importance of the constraints of the visual information processing system, particularly the spatial resolution of the retina which decreases as eccentricity increases (also see Geisler, 2011). Therefore, the saccade mechanism is required to bring the higher acuity visual receptors to where they can sample relevant information. In addition the model captures the fact that preparing and executing eye movements takes time, and that encoding the visual properties and

interpreting the visual information also take time. Time costs not only make performance slower, they motivate different strategy choices.

2.3. The effect of reward and user preference on visual search strategy

While there is some work that directly probes the ecological priorities of users (Toomim, Kriplean, Pörtner, & Landay, 2011), there is also work that tests the consequence of manipulating priorities by making use of different reward schemes. Selection bias toward previous rewarded items has been found in lots of recent studies using behavioral and neural approaches (see Awh, Belopolsky, & Theeuwes, 2012 for a review; Kiss, Driver, & Eimer, 2009; Libera & Chelazzi, 2009; Navalpakkam, Koch, Rangel, & Perona, 2010; Raymond & O'Brien, 2009). Furthermore, Tseng and Lleras (2013) manipulated the rewards of the spatial contexts and found that the rewards (either related to arousal or valence) associated with the spatial context can accelerate the implicit learning of the repeating contexts. In other words, a rewarding environment can have implications for the acquisition of efficient strategies.

The role of user preference and information gain in determining visual attention from stimuli with basic visual features has been modeled by accumulation models of perceptual and attentional decision tasks, such as two-alternative forced choice (2AFC) of the current directions of motion dots (Bogacz, Hu, Holmes, & Cohen, 2010; Britten, Shadlen, Newsome, & Movshon, 1993; Shadlen, Britten, Newsome, & Movshon, 1996; Shadlen & Newsome, 1996, 2001) and the 2AFC of the current oddball color (Tseng, Glaser, Caddigan, & Lleras, 2014). In spatial decision making, Najemnik and Geisler (2005) further show that fixations are better predicted by a

model that makes eye movements so as to maximize information gain about the location of the target than a model that makes eye movements so as to maximize the probability of fixating the target. The model exhibits phenomena such as center-of-gravity effects, where people fixate between two places each of which is associated with evidence that it contains the target. These findings support the view that search and selection is guided by the goal to maximize information gain (Najemnik & Geisler, 2005, 2008, 2009) given the natural texture and the acuity of the human visual information processing system, particularly the fall off in acuity with eccentricity.

More generally, information gain will be traded against other user priorities such as time cost. People can be thought of as seeking to maximize some subjective expected utility function (Cox & Young, 2004; Geisler, 2003; Howes et al., 2009; Lelis & Howes, 2011; Najemnik & Geisler, 2005; Sperling & Dosher, 1986; Trommershäuser, Maloney, & Landy, 2008; Tseng & Howes, 2008). Priorities can also include the internal costs of decision processing (Droll & Hayhoe, 2007; Duggan & Payne, 2009; O'Hara & Payne, 1998; J. W. Payne, Bettman, & Johnson, 1993; Smith & Walker, 1993) and external time costs (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Bogacz et al., 2010; Forstmann et al., 2008; Kocher & Sutter, 2006; Maule & Edland, 1997; J. W. Payne, Bettman, & Luce, 1996). For example, in two alternative forced-choice tasks, people try to adjust the threshold of evidence accumulation to balance the speed-accuracy tradeoff and thereby maximize the rate of rewards (i.e. reward rate or utility) (Bogacz, 2007; Bogacz et al., 2006; Bogacz et al., 2010; Gold & Shadlen, 2002). Moreover, gains and time constraints are known to play a role in perceptual-motor control tasks (Bogacz et al., 2006; Bogacz et al., 2010;

Howes et al., 2009; Hudson, Maloney, & Landy, 2008; Maloney & Mamassian, 2009).

3. THE TASK

In order to investigate the role of visual search in Human-Computer Interaction, we studied a scenario in which a person has the goal of finding an image with a particular set of features for some purpose such as illustrating a book, or contributing to a presentation. A goal may be, for example, to find an image of a church with blue sky and people. We assume that a set of keywords have been chosen and entered, and that the search engine has returned a 2 dimensional array of images (See Figure 1-search display). The user then scans the images to gather information about them using some pattern of fixations and saccades. Once the user identifies an image as meeting the criteria, i.e. matching to the features, then they select it by clicking on it and the task ends.

(It is sometimes the case that the scan of the results of a search leads a user to enter new search keywords and generate a different search page. Similarly, a user may expand an image but subsequently return to the search results to further consider alternatives. These activities are not addressed in the current paper.)

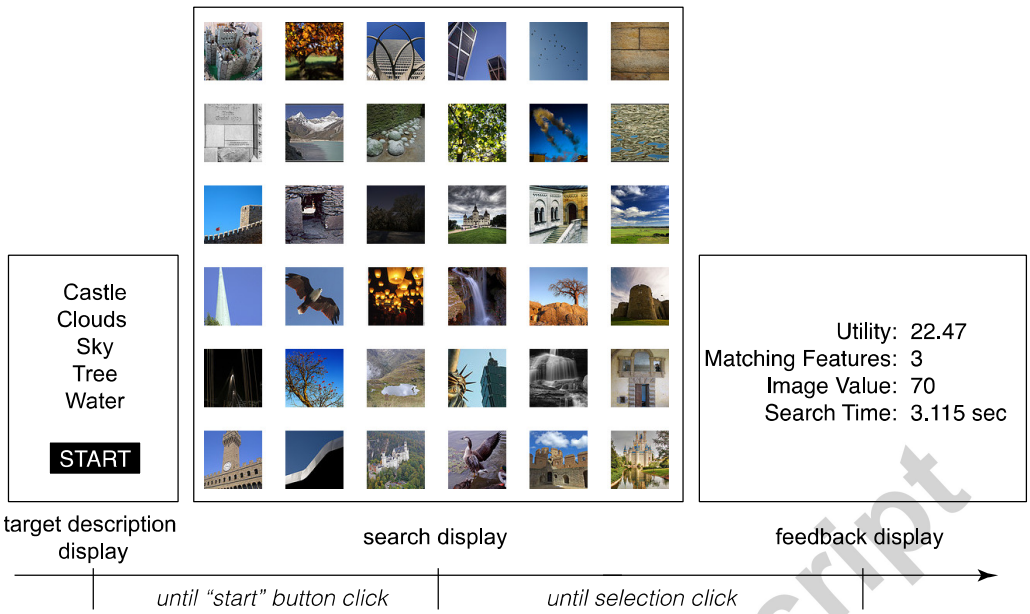

Figure 1. An example of one search trial. The gap between each thumbnail was 0.85° of visual angle. The task was to look for an image that included as many of the target features as possible and as quickly as possible. In this example, the target features are Castle, Clouds, Sky, Tree, and Water. There were 26 images with 1 target feature, 6 images with 2 target features, 2 images with 3 target feature, and 1 image with 4 target features and 1 image with 5 target features). The size of each thumbnail was 2.15° × 2.15° square.

A critical element of this natural task environment is the prior probabilities that an image will meet the criteria. Modeling these prior probabilities is an essential requirement of achieving ecologically valid study materials. It is possible to imagine a world in which the precision of searches is so high that they tend to return a very high proportion of relevant images (images that match the features specified in the user's

goal). It is also possible to imagine a world in which the precision of searches is very low. The natural distributions are likely somewhere between these extremes. We expect that the scanning strategy adopted by users will partly reflect the actual distribution of relevance in the environment and for this reason we gathered data concerning this distribution.

In order to investigate the real world distribution we conducted a pilot study with Google Images (https://images.google.com/). The goal was to determine the relationship between the number of keywords entered and the number of images returned that matched all of the keywords. The study did not use participants, rather the idea was to infer the distribution of images with different numbers of features by inspecting the number of images in Google search results for different queries. The study results are shown in Figure 2a. For example, when "water" and "bridge" were entered, 644,000,000 images were returned. When "water", "bridge" and "sky" were entered, 53,700,000 images were returned. The distribution follows a power-law reflecting that the more keywords are entered, the fewer the number of results that match *all* keywords entered. These results are consistent with statistics from natural environments and other man-made environments (Adamic & Huberman, 2000; Stewart, Chater, & Brown, 2006)

Given the above finding, we designed the materials for an experiment in which the number of images with each number of features (1 to 5) followed a power-law (Figure 2b). In each display of 36 images the number of images given the number of features was equal to $26 \times n^{-2.2}$ where n is the number of features.

In general, the benefits of obtaining a higher value image had to be offset against the cost of longer visual search times in a digital image repository. A search engine

will usually return more images than can be individually evaluated by a user. One problem for the user is to decide how to allocate time, one of their most scarce resources, to the evaluation of some subset of images. Another problem is to decide when to stop the search. It may be the case that the ideal is to find an image that matches 5 goal features but that an image that matches only 4 will do. This problem is known as a speed/accuracy trade-off problem. It is a type of problem that has been extensively studied in psychology (e.g., J. W. Payne et al., 1996). In order to model the speed/accuracy trade-off in the experimental task environment, images containing more target features had a higher value (e.g. 200 points for a *5-feature image*) but, of course, they were rarer in the environment (e.g. there is only one *5-feature image* in 36 alternatives in Figure 2b) and were therefore likely to require more time to find.

4. THE MODEL

The model makes the following assumptions:

a. Each image is scanned without repetition. This assumption follows from work suggesting that people seldom revisit the same items during visual search (Peterson, Kramer, Wang, Irwin, & McCarley, 2001) but no assumptions are made about the scan order.

b. The input to the scan function consists only of the features of the foveated image and the function generates a noisy estimate of the number of matching features. The noise is Gaussian distributed with variance σ^2 and centered on the true value. This assumption is motivated by evidence that visual search limited by signal noise that negatively impacts acuity (Geisler, 2011). The

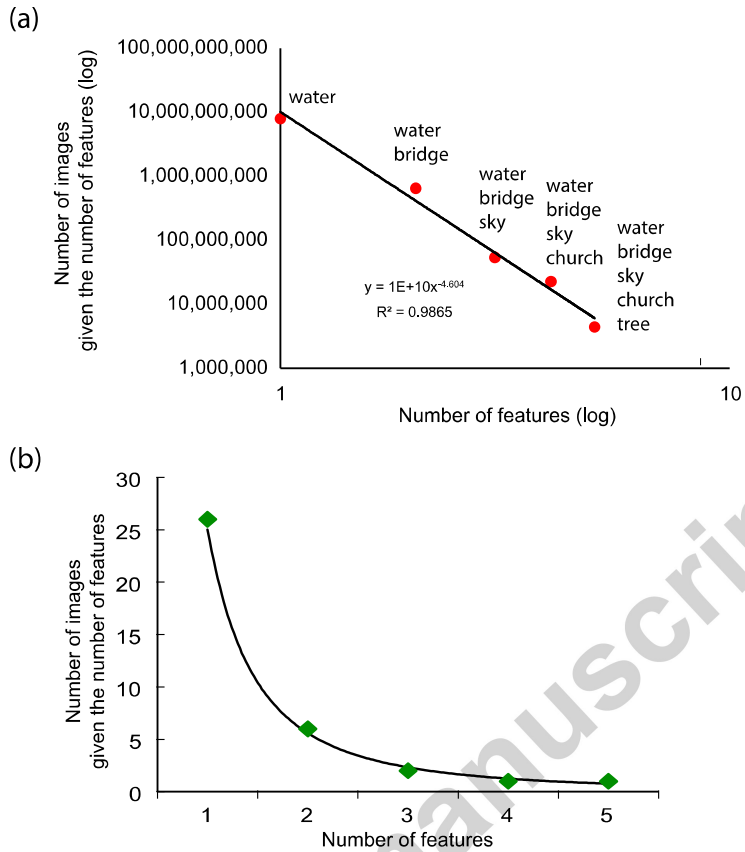

Figure 2. (a) The frequency of n-feature images (matching n keywords) in the real world. The log-log plot shows the power-law relationship between the number of returned images and their corresponding number of input keywords in a natural image search engine (data resource: Google image search 10/Aug/2013); (b). The Number of images with n features = 26 x n-2.2 in the search task of our experiment

level of noise is an invariant property of an individual's visual system and is not under strategic control.

c. The strategy space is defined by two thresholds. If the number of detected features in an image exceeds the target criterion *TC* then the image is selected. This assumption is motivated by signal detection theory (Green & Swets, 1966). If the number of gazes exceeds the evaluation-stopping rule, *ST* then an image with the highest number of detected features is selected. The strategy is defined by the pair *(TC, ST)*.

d. We assume that participants select a strategy *(TC, ST)* that maximizes a subjective expected utility function *U* which defines a speed/accuracy trade-off. This assumption follows works suggesting that during simple choices, the brain tries to maximize the reward rate (Bogacz, 2007; Bogacz et al., 2006; Bogacz et al., 2010; Gold & Shadlen, 2002).

The model incorporates the assumption that participants will adapt *(TC, ST)* to the error σ^2 in the perception of features, the utility function *U*, and the ecological distribution *E*. The model predicts the strategy that participants should select. This prediction is tested in the following experiment.

5. EXPERIMENT

Participants were given a series of tasks and asked to search and select images. On repeated trials they were presented with a display of 36 images that were arranged so as to look like the results page of an internet search engine, such as Flickr. They were free to select one image. After each trial they were given feedback that consisted of a reward signal that was calculated by dividing the image *value* (explained below) by

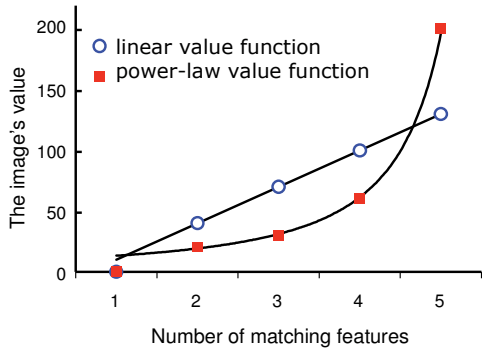

Figure 3. Image value as a function of its number of target features for each of the two value functions used in our experiment. Red filled squares denote the power-law value condition; open blue circles denote the linear value condition.

the trial search time (the duration between the time when the images first appeared to the time when a selection was made).

In the experiment, the value of an image was a function of the number of target features that it contained and there were two value conditions (Figure 3): a power-law condition and a linear condition. In the power-law condition a *5-feature image*, for example, had a value of 200 points whereas in the linear value condition it had a value of 120 points (Figure 3).

5.1 Participants

Sixteen naïve participants from the University of Manchester, between 18 and 28 years old, with normal color vision or corrected-to-normal vision, who provided informed consent, participated in the experiment for monetary compensation. Participants were aware that they would be paid a £5 voucher and entered a

competition for three more £5 vouchers based on overall points (utility) scored from performance on the task. The local ethics committee approved the protocol.

5.2 Apparatus

The experiment used a Tobii 1750 eyetracker system which was integrated with a 17" TFT display monitor. The monitor's response time was 8 msec. The eye tracker tolerated fairly large head movements. The freedom of head-movement (W × H × D) is 30 × 16 × 20 cm at a distance of 60 cm from the tracker. The Tobii tracker also provides long-lasting calibrations which allowed for a natural user environment. The eye tracker has a tracking rate or frame rate of 50 Hz. The position and duration of eye movements, the input of the participant's mouse click, and every screen event and display are recorded during the experiment. The temporal resolution of 20 msec is sufficient to monitor long fixations and eye movements in our task. Although low temporal resolution could cause noise in eye signal sampling, noise was reduced by averaging several gazes per trial. The Tobii 1750 is a completely unobtrusive system, which removes the need for chin-rests and other restraints. Moreover, long studies can be performed without fatigue for the participant or reduced quality of data. Visual Basic, using the default set of Tobii 1750 functions, was used to present the stimuli and collect eye movement data. Before the experiment started, the eyetracker system was calibrated.

5.3 Task

The participants were instructed to find an image based on a target description which was given at the beginning of the trial. The target description contained five

unique target features, e.g. Castle, Clouds, Sky, Tree, Water (see Figure 1-target description display). The task was similar to a visual search paradigm in which people are asked to search for a target among distractors. However, the task in this experiment was different in three respects: (i) the description of the five target features comprised higher level features such as 'cloud' and 'sky' rather than lower level features such as a letter shape; (ii) participants were allowed to select any images in the search set; and (iii) the result of the search was not dichotomous; it was not either correct nor incorrect. Rather than being judged as correct or incorrect, participants were given points according to a utility function defined in terms of the image value and their search time. Feedback was given after each trial (Figure 1-feedback display).

5.4 Stimuli and display

5.4.1 Sets of target descriptions

There were 5 sets of target descriptions in this experiment. The 5 sets were (a) Sun, Water, Sky, Boat, People; (b) Castle, Clouds, Sky, Tree, Water; (c) Bridge, Water, Sky, Church, Tree; (d) Animal, Patterned-Fur, Grass, Tree, Sky; and (e) Night, Building, Street, Car, People. One of these 5 sets was used in each trial of the experiment. For example, in the first trial a participant might be asked search for an image matching the set of target features in set 'a'; in the second trial they might search for matches to set 'b'; and so on. After all five sets of target descriptions were used, they were repeated (Although the target descriptions repeated, the images in the search display never repeated). An additional set "Boat, People, Water, Sky, Sun" was used in the practice trials. There were 20 repetitions of each set giving a total of 100

trials. The task order was generated for each participant in a counter-balanced random sequence.

5.4.2 Stimuli

Scene image thumbnails were used as search stimuli in this experiment. These images were carefully selected from a photo sharing website, Flickr. The standard square images in the website were used. Poor resolution pictures and those with special effects, such as High Dynamic Range (HDR) images were removed. The size of each square image was 75 × 75 pixels square, which subtends a visual angle of 2.15°. No image was used twice so as to prevent a possible memory effect on search performance.

5.4.3 Search display

On each trial, 36 images which were randomly positioned in a 6 x 6 display. Some images had more target relevant features and some had fewer. The distribution of images with each number of features was fixed. In each set of 36 images, there were 26 images with 1 target feature, 6 images with 2 target features, 2 images with 3 target feature, 1 image with 4 target features, and 1 image with 5 target features (Figure 2b). We define an *n-feature image* as an image with n number of target features. For example, a 3-feature image is an image with 3 target features; all of its other features were non-target features.

5.5 Trial procedure

In each trial, the goal for a participant was to maximize the utility feedback by selecting a thumbnail from the display 6 × 6 array of images. At the beginning of the

experiment they were told that on each trial utility would be calculated as the value of the image that they selected divided by the time that they had spent (the search time). Each trial had three steps (see figure 1). (i) In the first step, the target description and a start button were presented in the top-left corner of the display (see figure 1-target description display). Participants were informed that they had unlimited time to familiarize themselves with this list of the five target features; they were aware that the time spent on the target description screen was not included in the calculation of utility. (ii) Participants clicked on the start button below the target description to show the search display which contained the 6x6 array of images (see figure 1-search display). Participants were allowed to select any image in the search set using the mouse and a mouse button click. Participants were aware that the more target features in the selected image, the higher the potential utility; but they were also aware that longer search times would reduce their utility. (iii) After they had selected an image, the search display disappeared and was replaced by the utility feedback display (right panel-feedback display in figure 1). The feedback provided was an estimate of the utility, which was calculated by dividing the image value by the trial search time (from the search display onset to the selection click). The feedback screen consisted of the utility, the search time, and value of the selected image. Participants had unlimited time to study the feedback; they were encouraged to try to attain the highest utility by balancing the trade-off between image value and search time.

5.6 Experiment Design

The experiment was a repeated-measures design. There were 4 conditions (2 density conditions × 2 value function conditions). The presentation of the 4 conditions was counterbalanced across participants. In the high-density condition, the edge-to-

edge item spacings were 3 pixels (visual angle=0.085°). In the low-density condition, the edge-to-edge item distances were 30 pixels (visual angle=0.85°). In the linear value condition, an image's value decreased linearly with the decrease in the number of target features (Figure 3). The images that had 5, 4, 3 or 2 target features had an image value of 130, 100, 70 and 40, respectively (Figure 3). In the power-law value condition, the relationship between value of an image and its number of target features was a power-law function (Figure 3). If an image had 5, 4, 3 or 2 target features, the image's value would be 200, 60, 30 and 20, respectively (Figure 3). A short introduction, which described the relation between an n-feature image and its value (i.e. value function), was shown before each block. Participants were thereby made aware of the expected image value.

Participants performed an initial practice session of 8 trials followed by 100 test trials. The 100 trials were separated into 4 blocks. Block 1 had 10 trials, block 2 had 10 trials, block 3 had 40 trials and block 4 also had 40 trials. The utility function was manipulated within participant. Participants performed one 10 trial block with power-function utility and one 40 trial block. The other 10 and 40 trial blocks used the linear utility function. The order of the utility conditions was counter-balanced between participants.

The two display densities (high and low density image arrays) were mixed in each block and the presentations of two density conditions were counterbalanced across participants.

6. RESULTS AND DISCUSSION

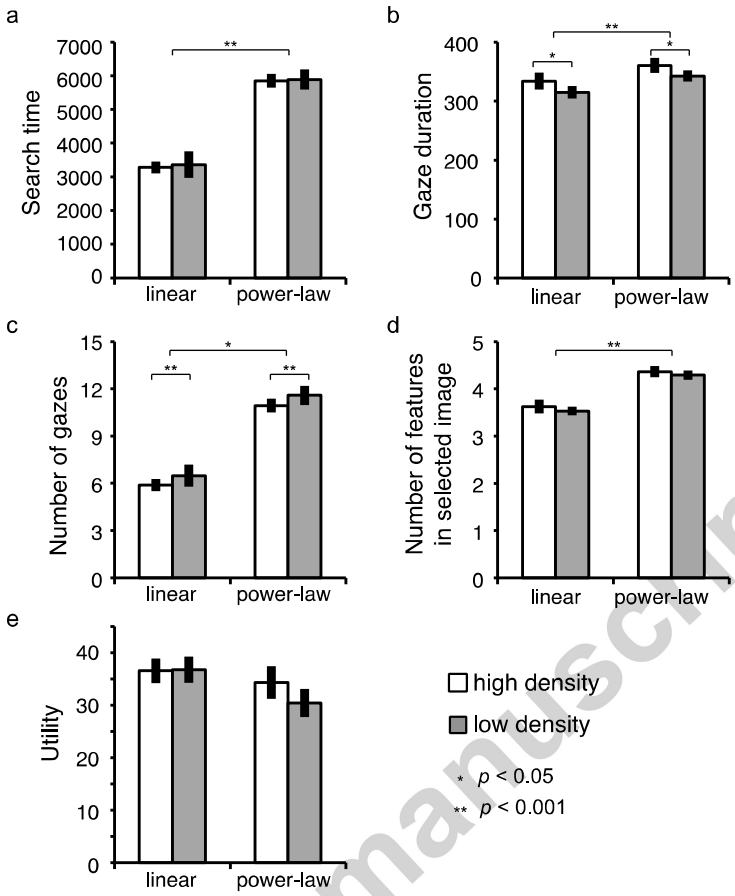

Figure 4. Mean search performance as a function of four conditions (16 participants). a. Search time (msec); b. Gaze duration (msec); c. Number of gazes; d. Average number of target features in the selected images; e. Utility=image value/search time (unit: sec). Error bars represent standard error (SEM).

We first inspected the data for evidence of practice effects. Participants performed better on the last two blocks (each of 40 trials) than on the first two blocks (each of 10 trials). As we were interested in performance once it is adapted we discarded the data from the first two blocks. All analyses presented in this section are therefore for the last two blocks.

We inspected the effect of the manipulation of utility on measures of strategic adaptation. Several measures of search performance (Figure 4) provided evidence of adaptation to the temporally constrained utility functions. Most importantly, and as predicted by the need for a higher threshold in the power-law condition, participants found higher value images (with more goal features) in the power-law condition than in the linear condition ($F[1,15]=56.98$, $p<0.001$; Figure 4d).

In addition, there was evidence that participants made a number of other adaptations to strategy. Gaze durations were longer in the power-law condition than in the linear condition ($F[1,15]=4.6$, $p<0.05$; Figure 4b). There were longer gaze durations ($F[1,15]=12.68$, $p<0.05$; Figure 4b) in the high-density display than in the low density display, suggesting, perhaps, that participants increased gaze durations to compensate for crowding effects -- though crowding of foveated images only occurs over very small distances (Levi, 2008).

The payoff function had a main effect on the number of gazes ($F[1,15]=64.414$, $p<0.001$; Figure 4c). Participants examined more thumbnails in the power-law condition. This is consistent with the fact that they found images with more matching features in this condition and with the greater reward associated with these high value images. Lastly, there were fewer gazes in the high density display ($F[1,15]=5,121$, $p<0.05$; Figure 4c), presumably because peripheral vision was more effective when the images were closer together.

6.1 Modeling the experimental results

While the experimental results reported in the previous section are encouraging, suggesting effects of the utility and density manipulations that are in the predicted

directions, they only offer evidence of adaptation and do not tell us whether the amount of adaptation was predicted by the utility maximization theory. The following analyses focuses on answering this question. We report two models, a best fit model and a utility maximization model. We examine whether the utility maximization model predicts the best fit model.

6.1.2 The best fit model

The model's parameters were fitted to two measures of each participant's performance. The first measure was the observed percentage of target features that matched the selected images in each condition. We called this measure *Percent-match*. The second measure was the number of gazes. However, rather than using the raw number of gazes we used the quantile gaze probability, called *Quantile-gaze* (Hu, Tseng, Winkler, & Li, 2014; Ratcliff & Tuerlinckx, 2002; Tseng et al., 2014). We used this measure because fit to quantile performance takes the shape of the performance distributions into account and not just the mean values (Ratcliff, 1979). We used this measure to fit each participant's distribution of gazes across conditions and thereby capture that on some trials they used very few gazes and on others many more. The quantile-gaze method makes a chi-square assumption that was met by our data.

Three of the model parameters were fitted to these measures: (i) target criterion, *TC*; (ii) evaluation-stopping rule, *ST*; and (iii) the sensitivity to the stimulus, *d'*. These parameters were calibrated to each individual participant's data. *TC* had a range from 1 to 5 which corresponded to the space of possible targeting strategies. The range of evaluation-stopping rule thresholds, *ST* was [1, 3, 6, 9, 12, 15, 18, 21, 24, 36, 48, 60, 72]. *d'* was set to $1/\sigma$ where σ had a range from 0.3 to 1.3. The size of the search space

was therefore *TC* x *ST* x *d'* is 5 x 13 x 12. The model was calibrated separately to each individual participant's data using Monte Carlo simulation (10,000 simulations for each parameter set). The calibration parameters were determined using the minimum sum of Pearson's chi-square (χ^2) value of the model fitting of *Percent-match* and *Quantile-gaze*. *Percent-match* was similar to the four types of responses in STD so it was sensitive to *TC* and *d'*. E.g., when *TC*=3 all 3-, 4-, and *5-feature images* were accepted. When *d'* was higher, an image with fewer matching features was less likely to be selected (i.e. fewer false alarms). (Note: the source code of these simulations can be found in the Open Science Framework https://osf.io/x2rt6/).

The mean parameter values for the individual best fit models are shown in Table 1. The higher target criterion (*TC*) values for the Power condition are consistent with the fact that on average participants selected images with more matching features in this condition. In the next section, we examine whether these parameter values correspond to the parameter values when utility is maximized.

6.2 The utility maximization model

In addition to finding the best fit model, we also derived the utility maximization model. Rather than fitting to the data, the utility maximizing model was found by determining the utility maximizing strategy given the constraints on each individual's performance. We tested the correspondence between the utility maximization model and the best fit model and therefore whether the former explained the latter. In order to do this we set *d'* to the best fitting value and then found the values of *TC* and *ST* that maximized utility. We found that the utility maximizing values were *TC*=51 and *ST*=2.94 for the linear-high density condition, *TC*=46.13 and *ST*=2.94 for the linear-

Search condition	ST	*d'*	TC
Linear-High	49.50(4.98)	0.87(0.05)	2.94(0.11)
Linear -Low	35.44(5.17)	0.84(0.06)	3.00(0.09)
Power-High	48.38(4.60)	0.91(0.05)	3.81(0.10)
Power-Low	46.13(4.22)	0.87(0.04)	3.81(0.10)

Table 1. the mean parameter values for the individual best fit models (SEM in Parentheses)

low density condition, *TC*=54 and *ST*=3.94 for the power-law-high density condition and *TC*=56.25 and *ST*=3.94 for the power-law-low density condition. These correspond closely to the best fitting values reported in Table 1.

The correspondence between best fit and utility maximization requires further explanation. We pursue this in the following section by expanding the utility of all plausible points in the strategy space. We first consider variation in the evaluation-stopping rule *ST*, then consider variation in the target criterion *TC*, and lastly the sensitivity of the results to variation in *d'*.

6.2.1 Evaluation-stopping rule

Figure 5 shows the predicted utility as a function of each level of evaluation-stopping rule, *ST*, given the best fitting *TC* and *d'* for each individual. It shows low values of *ST* offered lower utility. It also shows the utility reaches the asymptote after 24 evaluations and 36 evaluations in the Linear and Power-law condition, respectively. This finding is consistent with the intuition that the utility maximization theory predicts that participants should make more effort to look for the higher value items in the Power-law condition. The shape of the evaluation-stopping rule curve in

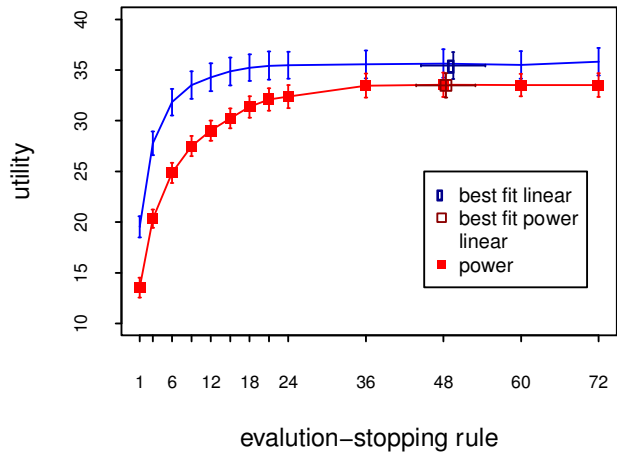

Figure 5. Predicted utility for each evaluation-stopping rule (blue solid circle: linear value condition; red solid square: power value condition) given the best fitting target criterion, *TC* and *d'* of each individual (N=16, data are from high-density display). The dark blue open circle is the best fitting utility in linear condition. The dark red open square is the best fitting utility in power-law condition. Error bars represent SEM.

Figure 5 is not hard to understand. First, low values of *ST* stop search before an item is above *TC* and therefore offer limited utility. Second, high values of *ST* tend not to have an effect because an item will usually be found that is above *TC* before *ST* is reached and, hence, the utility plateau.

6.2.2 Target criterion

The *d'* and *ST* from the best fitting model was used to test the utility predicted by each of the plausible strategies, *TC*. Figure 6 provides a plot of mean utility against mean number of features in the selected image in each level of *TC* for each value condition (blue solid circle: linear value condition and red solid square: power-law

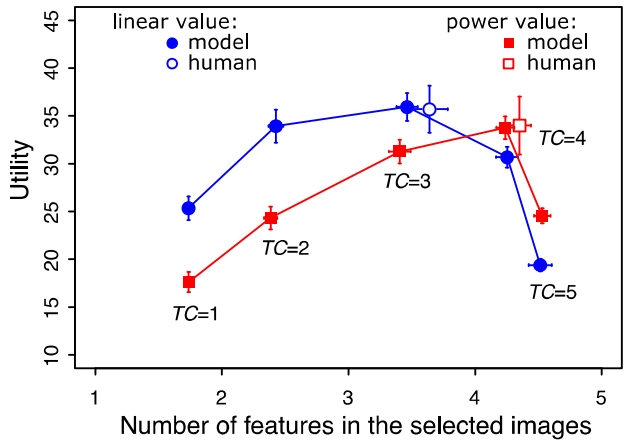

Figure 6. Model utility with each possible targeting strategy (Target Criterion, TC=1 to 5) and the best fitting d' of each participant (N=16, The data are from high-density display) and empirical utility by human subjects. Model utility for each targeting strategy plotted as curves for linear value (blue line and blue solid circle) and power-law value (red line and red solid square) condition. Human utility in the linear value condition (blue open circle) and power-law value (red open square) are shown (N=16). Error bars represent standard error (SEM) for each utility (vertical axis) and image feature gain (horizontal axis).

value condition) and the mean human observations (blue open circle: linear value condition and red open square: power-law value condition). For the linear condition the highest utility model corresponded to a strategy with *TC*=3 and for the power-law condition it was *TC*=4. It can be seen that the highest utility model (the target criterion yielding the greatest utility) in both conditions, corresponds to the human data. The fact that, on average, participants selected images with 3 features in the

Subject	Best fitting *TC*				Predicted bounded optimal *TC*			
	Linear-High	Linear-Low	Power-High	Power-Low	Linear-High	Linear-Low	Power-High	Power-Low
1	3	3	4	4	3	3	4	4
2	3	3	4	4	3	3	4	4
3	3	3	4	4	3	3	4	4
4	2	2	3	3	3*	3*	4*	4*
5	3	3	4	4	3	3	4	4
6	3	3	4	4	3	3	4	4
7	3	3	4	4	3	3	4	4
8	3	3	4	4	3	3	4	4
9	3	3	4	4	3	3	4	4
10	2	3	3	3	2	3	4*	4*
11	4	4	4	4	3†	3†	4	4
12	3	3	4	4	3	3	4	4
13	3	3	4	4	3	3	4	4
14	3	3	4	4	3	3	4	4
15	3	3	4	4	3	3	4	4
16	3	3	3	3	3	3	4*	4*

* The predicted bounded optimal is higher than the best fitting (observed) *TC*.
† The predicted bounded optimal is lower than the best fitting (observed) *TC*.

Table 2. the utility maximization model's predicted thresholds in both Linear and Power-law condition for 12 out of 16 of the individual participants.

linear condition and 4 features in the power-law condition is precisely explained by the utility maximization models.

Further, the utility maximization model accurately predicted thresholds in both the Linear and Power-law conditions for 12 out of 16 of the individual participants (Table 2). However, there is very little between participant variation in the utility maximization models predictions and much more variation between best fitting thresholds. Only the utility maximization prediction for participant 10 Linear High-density is different to the predictions for other participants (a threshold of 2 versus 3 for the rest). In addition, participants 4, 10, and 16 had best fitting thresholds that were lower than utility maximization in at least one of the conditions, and participant

11 had best fitting thresholds that were above utility maximization in the linear conditions.

6.2.3 Sensitivity analysis

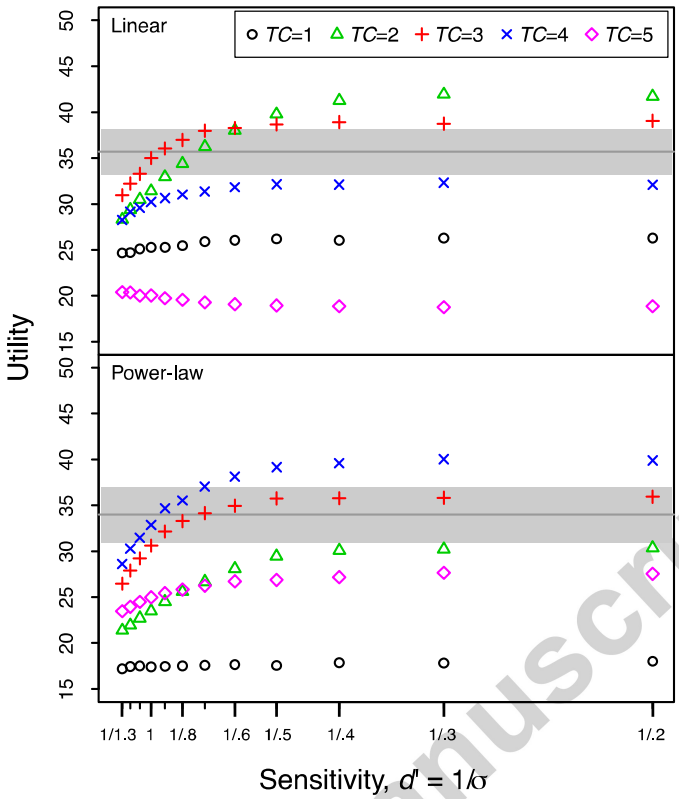

Figure 7. Predicted utility as function of target criterion, *TC*, for each level of sensitivity *d'*=1/σ (σ = 0.2 to 1.3). Upper panel is for the linear value condition and lower panel is for the power-law value condition. The predicted data are from high-density display, black circle: *TC*=1; green triangle: *TC*=2; red plus: *TC*=3; blue cross: *TC*=4; magenta diamond: *TC*=5, *ST*=36 was used in the model. Human utility plotted as the horizontal grey line (N=16). Error bars (the gray bar) represent SEM.

It may seem that the excellent correspondence between the prediction derived from the utility maximization model and the best fit to the observed human behavior was inevitable and that, therefore, the test only offered a weak test of the model.

However, this was not the case. Here we look at two sources of evidence for this argument.

First, recall that the utility maximization model and the best fitting model only corresponded exactly for 12 of the 16 participants. In the remaining 4 cases there was a difference between utility maximization and best fit. The fact that not all participants were predicted by utility maximization supports the argument that correspondence was not inevitable and that instead utility maximization offers an explanation of 12 of the 16 observed behaviors.

Second, we tested whether there were differences between the utility maximization target criterion *TC* and the observed value *TC* for the plausible range of *d'* (Figure 7). Our aim was to investigate whether different values of *TC* would be required given different values of *d'*. Variation of *TC* with *d'* provides further evidence that the utility maximization prediction was not an inevitable consequence of an overly constrained theory. Examine the Power-law panel (bottom panel) of Figure 7 showing that the maximum utility *TC* (i.e. *TC*=4, blue crosses) corresponds precisely to the best fitting *TC*=4 (blue cross) at *d'*=1, as previously observed, but that at values of *d'* > 1/0.8 the best fitting strategy is *TC*=3 (red plus) whereas the utility maximization strategy remains *TC*=4. Now examine the linear panel (top panel). Here the utility maximization *TC*=3 (red plus when *d'*<1/0.5) corresponds precisely to best fitting strategy at *d'* = 1, but that at values of *d'* > 1/0.5 while the best fitting strategy remains *TC*=3, the utility maximization strategy is *TC*=2.

As an aside, note that it is somewhat counterintuitive that for values of *d'* > 1/0.5 in the Linear value condition the maximum utility strategy is lower (*TC*=2) when for lower values of *d'* a higher *TC* (*TC*=3) gives maximum utility. Higher values of *d'* are

associated with greater discrimination and therefore should, one might think, allow higher threshold strategies. However, further investigations of the model revealed that the greatest advantage of higher *d'* is that it allows low value items, with fewer than 2 features, to be avoided, even with low *TC*. We take from these findings that it is only in a subset of the possible discrimination values that the performance of the best fitting model corresponds to the predictions of the utility maximizing model.

These results could be found in the url. http://myweb.ncku.edu.tw/~yctseng/expt5_model(individuals)_2015.zip

7. GENERAL DISCUSSION

The goal of the paper was to report and test a model of the choices that people make about strategies for scanning search engine results. The empirical evidence and computational model presented in this paper show that people adapt visual search strategies to the statistics of the ecology of images on the web (the skewed distribution of probability of finding an image that has multiple features), to the human visual system (longer fixations enable more information to be gathered from foveal and peripheral vision, although longer fixations can only be effective if the information is available within the perceptual span), and to the shape of the reward function (in the power-law condition the reward was much higher for the rarer targets). Strategy choice, including adjustments to gaze duration and number of fixations can be explained as an optimal adaptation to these constraints.

The findings are important for understanding human computer interaction because they expose how the strategy choices that people make are affected by interface

design and user priorities. Small changes in the visual angle between items and changes in the reward structure *cause* qualitative changes in strategy. These findings support an adaptive view of interaction (S. J. Payne & Howes, 2013; Pirolli, 2007) in which people are modeled as agents that are sensitive to the statistical properties of the ecological context as well as to the costs and benefits of action. This framework for understanding human interaction with technology holds promise as a means of explaining, rather than merely describing, interaction. It thereby has the potential to predict how people will use new devices. The model reported above, for example, might form a basis for an approach to predicting performance in the various design proposals that have been made over recent years including Space-filling thumbnails (Cockburn et al., 2006), Tabular interface (Resnick et al., 2001) and Faceted category interfaces (amongst others).

In addition, the empirical results, and the reported model, add to the growing body of evidence that computational models of adaptation to cognitive constraints offers a viable means of explaining behavior in HCI (Brumby, Salvucci, & Howes, 2009; Fu & Pirolli, 2007; Janssen, Brumby, Dowell, Chater, & Howes, 2011; S. J. Payne et al., 2001; Vera, Howes, McCurdy, & Lewis, 2004). These models have been used to help explain not only visual search, but also web search, multi-tasking at the desktop, and multi-tasking in complex dynamic tasks such as driving. By framing the problem of explaining Human-Computer Interaction as the problem of calculating the implications of cognitive and ecological constraints for strategies, this work has embraced the extraordinary flexibility of human cognition while also enhancing the power of theory to predict rather than just describe behavior.

In addition, to a contribution to the science of understanding how people adapt to interface design our work also adds evidence in support of the argument that some aspects of design might be automated. Pirolli (2007), Bailly, Oulasvirta, Kötzing, and Hoppe (2013) and Bailly and Oulasvirta (2014) have argued that it might be possible to discover 'optimal designs'. For example, Bailly et al. define an optimal menu system as the menu system that best meets all design goals and relevant constraints and they thereby define menu design as an optimization problem. Bailly et al. pose five challenges that must be met if optimal designs are to be found. One of the challenges is to determine predictive models of human behavior – much like we have described above. A key property of a predictive model is that, given a proposed design, the strategy prediction be determined automatically, much as *TC* and *ST* were in our utility maximization model. The performance implications of design variations that have not been investigated empirically, e.g. levels of icon spacing other than those explored above, may then be determined by running the model.

However, there are also limitations of our specific study. While our model explained the behavior of three-quarters of the participants, the quarter who failed to find a utility maximizing strategy remained unexplained. One possible reason is that while we ensured that our study materials were sampled according to the ecological distribution of images returned by a web browser, we did not attempt to discover the ecological distribution of subjective utility or preferences (Toomim et al., 2011). It is known that people vary in their preferences concerning speed and accuracy (Bogacz et al., 2010; Hu et al., 2014). They may vary in the extent to which they choose to adopt the tradeoff imposed by an experiment and/or they will vary in their everyday preferences. More careful study of the ecology of preferences, such as is advocated by Toomim et al. (2011) would allow more careful study of user adaptation to design.

In conclusion, we have reported and empirically evaluated, a computational model showing that user choice of interactive strategy, including adjustments of gaze duration and number of fixations, can be explained as an optimal adaptation to a combination of properties derived from the individual, from the statistics of images on the web, and from speed/accuracy trade-off functions.

ACKNOWLEDGMENTS.

We thank Glyn Humphreys, Alejandro Lleras, Alex Kirlik, Chiang-Shan Ray Li, Geoffrey B. Duggan, for suggestions and comments on earlier drafts. Yuan-Chi Tseng was supported partially by OSS Awards and MBS scholarship from University of Manchester and Grant No. MOST 102-2410-H-006-019 & MOST 103-2221-E-006 -210 from the Ministry of Science and Technology, Taiwan and the Ministry of Education, Taiwan - The Aim for the Top University Project to the National Cheng Kung University (NCKU). Andrew Howes was supported by a grant from Office of Naval Research.

VITAE

Yuan-Chi Tseng is a design researcher with interests in human cognition and attentional decision, and their applications in the designs of experience, information, interactive media and persuasive technology. He was awarded his PhD in 2009 from the University of Manchester and has held posts at the Human Factors Division, Psychology and Beckman Institute, University of Illinois at Urbana-Champaign, and the Smith-Kettlewell Eye Research Institute. He is currently Associate Professor in the Department of Industrial Design and the Director of the Cognition & Experience Design Lab at the National Cheng Kung University.

Andrew Howes is a computer scientist with an interest in computational models of human behavior. He was awarded his PhD in 1992 and has held posts at the MRC Applied Psychology Unit, Cambridge, School of Psychology, Cardiff University, and Manchester Business School. He is currently Professor of Human-Computer Interaction in the School of Computer Science at the University of Birmingham.

REFERENCE

Adamic, L. A., & Huberman, B. A. (2000). Power-law distribution of the world wide web. *Science, 287*, 2115.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in cognitive sciences*.

Bailly, G., & Oulasvirta, A. (2014). Toward optimal menu design. *interactions, 21*(4), 40-45.

Bailly, G., Oulasvirta, A., Kötzing, T., & Hoppe, S. (2013). *Menuoptimizer: Interactive optimization of menu systems.* Paper presented at the Proceedings of the 26th annual ACM symposium on User interface software and technology.

Bederson, B. (2001). *PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps*. Paper presented at the 14th annual ACM symposium on User interface software and technology, Orlando, Florida

Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of the effective visual stimulus in visual search. *Perception & Psychophysics, 62*, 576-585.

Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. *Trends in cognitive sciences, 11*(3), 118-125.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review, 113*(4), 700-765.

Bogacz, R., Hu, P. T., Holmes, P., & Cohen, J. D. (2010). Do humans produce the speed-accuracy trade-off that maximizes reward rate? *The Quarterly Journal of Experimental Psychology, 63*(5), 863-891.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Vis Neurosci, 10*(6), 1157-1169.

Brumby, D. P., & Howes, A. (2008). Strategies for Guiding Interactive Search: An Empirical Investigation Into the Consequences of Label Relevance for Assessment and Selection. *Human-Computer Interaction, 23*(1), 1-46.

Brumby, D. P., Salvucci, D. D., & Howes, A. (2009). *Focus on driving: How cognitive constraints shape the adaptation of strategy when dialing while driving.* Paper presented at the Proceedings of the SIGCHI conference on human factors in computing systems.

Charman, S. C., & Howes, A. (2003). The adaptive user: an investigation into the cognitive and task constraints on the generation of new methods. *J Exp Psychol Appl, 9*(4), 236-248. doi: 10.1037/1076-898X.9.4.236

Cockburn, A., Gutwin, C., & Alexander, J. (2006). *Faster document navigation with space-filling thumbnails.* Paper presented at the SIGCHI conference on Human factors in computing systems (CHI '06).

Cox, A. L., & Young, R. M. (2004). *A Rational Model of the Effect of Information Scent on the Exploration of Menus.* Paper presented at the Sixth International Conference on Cognitive Modeling.

Cutrell, E., & Guan, Z. (2007). *What are you looking for? An eye-tracking study of information usage in Web search.* Paper presented at the SIGCHI conference on Human factors in computing systems (CHI '07).

de Bruijn, O., & Spence, R. (2000). *Rapid serial visual presentation: a space-time trade-off in information presentation*. Paper presented at the working conference on Advanced visual interfaces.

Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of experimental psychology: Human perception and performance, 33*(6), 1352.

Duggan, G. B., & Payne, S. J. (2009). Text skimming: The process and effectiveness of foraging through text under time pressure. *Journal of Experimental Psychology: Applied, 15*(3), 228-242.

Eng, K., Lewis, R. L., Tollinger, I., Chu, A., Howes, A., & Vera, A. (2006). *Generating automated predictions of behavior strategically adapted to specific performance objectives.* Paper presented at the SIGCHI conference on Human factors in computing systems (CHI '06).

Everett, S. P., & Byrne, M. D. (2004). *Unintended effects: Varying icon spacing changes users' visual search strategy.* Paper presented at the SIGCHI conference on Human factors in computing systems (CHI '04).

Fertig, S., Freeman, E., & Gelernter, D. (1996). Lifestreams: an alternative to the desktop metaphor. *SIGCHI conference on Human factors in computing systems (CHI '96)*, 410-411.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E. J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences, 105*(45), 17538.

Fu, W.-T., & Pirolli, P. L. (2007). SNIF-ACT: A Cognitive Model of User Navigation on the World Wide Web. *Human-Computer Interaction, 22*.

Geisler, W. S. (2003). Ideal observer analysis. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 825-837). Boston: MIT press.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research, 51*(7), 771-781.

Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the Brain:: Decoding the Relationship between Sensory Stimuli, Decisions, and Reward. *Neuron, 36*(2), 299-308.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: John Wiley and Sons.

Halverson, T., & Hornof, A. J. (2004). *Local Density Guides Visual Search: Sparse Groups are First and Faster*. Paper presented at the Human Factors and Ergonomics Society Annual Meeting.

Halverson, T., & Hornof, A. J. (2011). A Computational Model of "Active Vision" for Visual Search in Human–Computer Interaction. *Human–Computer Interaction, 26*(4), 285-314.

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review, 116*(4), 717-751.

Howes, A., Vera, A., & Lewis, R. L. (2007). Bounding rational analysis: Constraints on asymptotic performance. In W. D. Gray (Ed.), *Integrated models of cognitive systems*. New York: Oxford University Press.

Hu, S., Tseng, Y. C., Winkler, A. D., & Li, C. S. R. (2014). Neural bases of individual variation in decision time. *Human Brain Mapping, 35*(6), 2531-2542.

Hudson, T. E., Maloney, L. T., & Landy, M. S. (2008). Optimal compensation for temporal uncertainty in movement planning. *PLoS computational biology, 4*(7).

Janssen, C. P., Brumby, D. P., Dowell, J., Chater, N., & Howes, A. (2011). Identifying Optimum Performance Trade‐Offs Using a Cognitively Bounded Rational Analysis Model of Discretionary Task Interleaving. *Topics in Cognitive Science, 3*(1), 123-139.

Johnson, J. (2010). *Designing with the mind in mind: simple guide to understanding user interface design rules*: Morgan Kaufmann.

Kieras, D. E., & Hornof, A. J. (2014). *Towards accurate and practical predictive models of active-vision-based visual search.* Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Canada.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction, 12* 391-438.

Kieras, D. E., & Meyer, D. E. (2000). The role of cognitive task analysis in the application of predictive models of human performance. In J. M. Schraagen & S. F. Chipman (Eds.), *Cognitive task analysis* (pp. 237-260).

Kiss, M., Driver, J., & Eimer, M. (2009). Reward priority of visual target singletons modulates event-related potential signatures of attentional selection. *Psychological Science, 20*(2), 245-251.

Klöckner, K., Wirschum, N., & Jameson, A. (2004). *Depth and breadth-first processing of search result lists.* Paper presented at the Ext. Abstracts of SIGCHI conference on Human factors in computing systems (CHI '04).

Kocher, M. G., & Sutter, M. (2006). Time is money--Time pressure, incentives, and the quality of decision-making. *Journal of Economic Behavior & Organization, 61*(3), 375-392.

Lamping, J., & Rao, R. (1996). The hyperbolic browser: A focus+ context technique for visualizing large hierarchies. *Journal of Visual Languages & Computing, 7*(1), 33-55.

Lelis, S., & Howes, A. (2011). *Informing decisions: how people use online rating information to make choices.* Paper presented at the Proceedings of the 2011 annual conference on Human factors in computing systems.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research, 48*(5), 635-654.

Libera, C. D., & Chelazzi, L. (2009). Learning to attend and to ignore is a matter of gains and losses. *Psychological Science, 20*(6), 778-784.

Maloney, L. T., & Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. *Visual Neuroscience, 26*(01), 147-155.

Maule, A. J., & Edland, A. C. (1997). The effects of time pressure on human judgement and decision making. In R. Ranyard, W. R. Crozier, & O. Svenson (Eds.), *Decision making: Cognitive models and explanations* (pp. 189-204). London: Routledge.

Meyer, D. E., & Kieras, D. E. (1997a). A computational theory of executive control processes and human multiple-task performance: Part 1. Basic Mechanisms. *Psychological Review, 104*, 3-65.

Meyer, D. E., & Kieras, D. E. (1997b). A computational theory of executive control processes and human multiple-task performance: Part 2. Accounts of Psychological Refractory-Period Phenomena. *Psychological Review, 104* 749-791.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategy in visual search. *nature, 434*, 387-391.

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision, 8*(3), 1-14.

Najemnik, J., & Geisler, W. S. (2009). Simple summation rule for optimal fixation selection in visual search. *Vision Research, 49*(10), 1286-1294.

Näsänen, R., Ojanpää, H., & Kojo, I. (2001). Effect of stimulus contrast on performance and eye movements in visual search. *Vision Research, 41*(14), 1817-1824.

Navalpakkam, V., Koch, C., Rangel, A., & Perona, P. (2010). Optimal reward harvesting in complex perceptual environments. *Proceedings of the National Academy of Sciences, 107*(11), 5232-5237.

O'Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive Psychology, 35*(1), 34-70.

Ojanpää, H., Näsänen, R., & Kojo, I. (2002). Eye movements in the visual search of word lists. *Vision Research, 42*(12), 1499-1512.

Öquist, G., & Goldstein, M. Towards an improved readability on mobile devices: evaluating adaptive rapid serial visual presentation. *Interacting with Computers, 15*(4), 539-558.

Over, E. A. B., Hooge, I. T. C., Vlaskamp, B. N. S., & Erkelens, C. J. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Research, 47* 2272-2280.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*: Cambridge Univ Pr.

Payne, J. W., Bettman, J. R., & Luce, M. F. (1996). When time is money: Decision behavior under opportunity-cost time pressure. *Organizational Behavior and Human Decision Processes, 66*, 131-152.

Payne, S. J., & Howes, A. (2013). Adaptive Interaction: A Utility Maximization Approach to Understanding Human Interaction with Technology. *Synthesis Lectures on Human-Centered Informatics, 6*(1), 1-111.

Payne, S. J., Howes, A., & Reader, W. R. (2001). Adaptively distributing cognition: a decision-making perspective on human-computer interaction. *Behaviour & information technology, 20*(5), 339-346.

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science, 12*(4), 287-292.

Pirolli, P. L. (2007). *Information foraging theory: Adaptive interaction with information*: Oxford University Press.

Pirolli, P. L., Card, S. K., & Van Der Wege, M. (2003). The effects of information scent on visual search in the hyperbolic tree browser. *ACM Transactions on Computer Human Interaction (TOCHI), 10*(1), 20-53.

Rao, R. P. N., Zelinsky, G., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research, 42*(11), 1447-1463.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86*(3), 446.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9*(3), 438-481.

Raymond, J. E., & O'Brien, J. L. (2009). Selective visual attention and motivation: the consequences of value learning in an attentional blink task. *Psychological Science, 20*(8), 981-988.

Rele, R. S., & Duchowski, A. T. (2005). *Using Eye Tracking to Evaluate Alternative Search Results Interfaces*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society, September, Orlando, FL.

Resnick, M. L., Maldonado, C. A., Santos, J. M., & Lergier, R. (2001). *Modeling On-line Search Behavior Using Alternative Output Structures.* Paper presented at the Human Factors and Ergonomics Society 45th Annual Conference.

Russell-Rose, T., & Tate, T. (2013). *Designing the Search Experience: The Information Architecture of Discovery*: Morgan Kaufmann.

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci, 16*(4), 1486-1510.

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proceedings of the National Academy of Sciences of the United States of America, 93*(2), 628-633.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J Neurophysiol, 86*(4), 1916-1936.

Shneiderman, B. (1992). *Designing the user interface: strategies for effective human-computer interaction* (Vol. 2): Addison-Wesley Reading, MA.

Smith, V. L., & Walker, J. M. (1993). Monetary rewards and decision cost in experimental economics. *Economic Inquiry, 31*(2), 245-261.

Snavely, N., Seitz, S. M., & Szeliski, R. (2006). Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics (TOG), 25*(3), 835-846.

Sperling, G., & Dosher, B. A. (1986). Strategy and optimization in human information processing. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. I, Sensory processes and perception*. New York: Wiley.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology, 53*(1), 1-26.

Toomim, M., Kriplean, T., Pörtner, C., & Landay, J. (2011). *Utility of human-computer interactions: Toward a science of preference measurement.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends in cognitive sciences, 12*(8), 291-297.

Tseng, Y.-C., Glaser, J., Caddigan, E., & Lleras, A. (2014). Modeling the Effect of Selection History on Pop-out Visual Search. *PLoS ONE, 9*(3), e89996. doi: 10.1371/journal.pone.0089996

Tseng, Y.-C., & Howes, A. (2008). *The adaptation of visual search strategy to expected information gain.* Paper presented at the SIGCHI conference on Human factors in computing systems (CHI '08), Florence, Italy.

Tseng, Y.-C., & Lleras, A. (2013). Rewarding context accelerates implicit guidance in visual search. *Attention, perception & psychophysics, 75*(2), 287-298.

Vera, A., Howes, A., McCurdy, M., & Lewis, R. L. (2004). *A constraint satisfaction approach to predicting skilled interactive cognition.* Paper presented at the SIGCHI conference on Human factors in computing systems (CHI '04).

Vlaskamp, B. N. S., & Hooge, I. T. C. (2006). Crowding degrades saccadic search performance. *Vision Research, 46*(3), 417-425.

Vlaskamp, B. N. S., Over, E. A. B., & Hooge, I. T. C. (2005). Saccadic search performance: the effect of element spacing. *Experimental Brain Research, 3*, 1-14.

Walter, J. A., Weßling, D., Essig, K., & Ritter, H. (2006). *Interactive hyperbolic image browsing-towards an integrated multimedia navigator*. Paper presented at the SIGKDD '06, Philadelphia, USA.

Woodruff, A., Faulring, A., Rosenholtz, R., Morrison, J., & Pirolli, P. (2001). *Using Thumbnails to Search the Web.* Paper presented at the SIGCHI conference on Human factors in computing systems (CHI '01).

Yee, K. P., Swearingen, K., Li, K., & Hearst, M. (2003). *Faceted metadata for image search and browsing.* Paper presented at the SIGCHI conference on Human factors in computing systems (CHI '03).

Search condition	ST	*d'*	TC
Linear-High	49.50(4.98)	0.87(0.05)	2.94(0.11)
Linear -Low	35.44(5.17)	0.84(0.06)	3.00(0.09)
Power-High	48.38(4.60)	0.91(0.05)	3.81(0.10)
Power-Low	46.13(4.22)	0.87(0.04)	3.81(0.10)

Subject	Best fitting *TC*				Predicted bounded optimal *TC*			
	Linear-High	Linear-Low	Power-High	Power-Low	Linear-High	Linear-Low	Power-High	Power-Low
1	3	3	4	4	3	3	4	4
2	3	3	4	4	3	3	4	4
3	3	3	4	4	3	3	4	4
4	2	2	3	3	3*	3*	4*	4*
5	3	3	4	4	3	3	4	4
6	3	3	4	4	3	3	4	4
7	3	3	4	4	3	3	4	4
8	3	3	4	4	3	3	4	4
9	3	3	4	4	3	3	4	4
10	2	3	3	3	2	3	4*	4*
11	4	4	4	4	3†	3†	4	4
12	3	3	4	4	3	3	4	4
13	3	3	4	4	3	3	4	4
14	3	3	4	4	3	3	4	4
15	3	3	4	4	3	3	4	4
16	3	3	3	3	3	3	4*	4*

* The predicted bounded optimal is higher than the best fitting (observed) *TC*.
† The predicted bounded optimal is lower than the best fitting (observed) *TC*.

Highlights (maximum of 85 characters, including spaces)

- Reports a computational model of eye movements over image search engine results.
- Reports an empirical study of visual attention for image search engine results.
- Shows that eye movements are rational given interface design and user priorities.
- Provides evidence that strategies are adapted to the ecology search engine results.

ABSTRACT

An important question for Human-Computer Interaction is to understand the visual search strategies that people use to scan the results of a search engine and find the information relevant to their current task. Design proposals that support this task include space-filling thumbnails, faceted browsers, and textually enhanced thumbnails, amongst others. We argue that understanding the trade-offs in this space might be informed by a deep understanding of the visual search strategies that people choose given the constraints imposed by the natural ecology of images on the web, the human visual system, and the task demands. In the current paper we report, and empirically evaluate, a computational model of the strategies that people choose in response to these constraints. The model builds on previous insights concerning the human visual system and the adaptive nature of visual search. The results show that strategic parameters, including the number of features to look for, the evaluation-stopping rule, the gaze duration and the number of fixations are explained by the proposed computational model.