

# Searching for the highest number

Piers D. L. Howe · Daniel R. Little

Published online: 27 November 2014  
© The Psychonomic Society, Inc. 2014

**Abstract** When viewing a collection of products how does a consumer decide which one to buy? To do this task, the consumer not only needs to evaluate the desirability of the products, taking into account factors such as quality and price, but also needs to search through the products to find the most desirable one. We studied the search process using an abstraction of a common consumer choice task. In our task, observers searched an array of numbers for the highest. Crucially, the observers did not know in advance what this number would be, which made it difficult to know when the search should be terminated. In this way, our search task mimicked a problem often faced by consumers in a supermarket setting where they also may not know in advance what the most desirable product will be. We compared several computational models. We found that our data was best described by a process that assumes that observers terminate their search when they find a number that exceeds an internal threshold. Depending on the observer and the circumstances, this threshold appeared either to be fixed or to decrease over the course of the trial. This threshold can explain why in some situations the observers terminate the search without inspecting all the numbers in the display, whereas in other situations observers act in a seemingly irrational manner, continuing the search even after inspecting all the numbers.

**Keywords** Visual search · Satisficing · Decision making · Heuristics

---

P. D. L. Howe (✉) · D. R. Little  
School of Psychological Sciences, University of Melbourne, 12 Floor  
Redmond Barry Building, Melbourne, VIC 3010, Australia  
e-mail: pdhowe@unimelb.edu.au

## Introduction

Consumer spending accounts for approximately 70 % of the U.S. gross domestic product (Chandra, 2012). Understanding how consumers decide which products to buy is therefore a matter of considerable importance. A typical store may sell thousands of different products, and even within a single product category (e.g., cookies, pain relievers, pasta sauces, etc.) there can be tens or even hundreds of different options to choose from (Botti & Iyengar, 2006). To solve this decision problem, the consumers need to perform two tasks: an evaluation task and a search task. They need to evaluate the desirability of the products, taking into account factors such as quality, price, and their own individual needs while simultaneously searching through these products for the most desirable one according to these assessments.

Previous investigations have studied situations where participants were required to perform both tasks simultaneously (Krajbich et al., 2010; Reutskaja et al., 2011). However, the desirability of a product is not a stable quantity and can be influenced by a variety of factors, including the current state of the consumer, for example, their mood or how hungry they are, as well as the environment and context in which the product is presented. For example, the presence of similar products can make the original product either more or less desirable depending on the perceived relationships between the products (Tversky, 1972; Huber et al., 1982; Simonson, 1989). There is therefore a considerable advantage to isolating the search processes from the evaluation process to allow the former to be studied independently.

In our study, we wished to investigate the visual search strategy employed by observers independently of any product evaluations that the observers might normally do. To do this, we used an abstraction of a common consumer choice task. Observers were presented with arrays of numbers where each number represented the desirability of a hypothetical product.

Instead of searching for the most desirable product, our observers searched for the highest number in the array. In this way, we largely removed the evaluation aspect of the task, allowing us to concentrate on the visual search aspect of the task.

Our search task is challenging, because the observers do not know in advance what the highest number will be on any given trial, which makes it difficult for them to know when the search should be terminated. In this way, our task mimics a problem often faced by consumers in a supermarket setting where they also may not know in advance what the most desirable product will be. The purpose of this paper is to understand how observers search in such situations and how in particular they decide when to stop searching.

How people terminate search has been previously studied in sequential decision-making problems, such as the Secretary Problem. In the Secretary Problem (Seale & Rapoport, 1997; Seale & Rapoport, 2000; Zwick et al., 2003; Bearden & Murphy, 2007), candidates for a secretarial post are interviewed one at a time and the task is to pick the best candidate for the job. The candidates are presented in a random order, and it is assumed that each candidate can be accurately ranked relative to previously interviewed candidates. With  $n$  candidates, the optimal solution is to reject the first  $n/e$  of the candidates, where  $e$  is the base of the natural logarithm, and then select the first candidate that is better than every candidate interviewed so far (Buss, 1984). A typical finding is that observers terminate their search sooner than expected under an optimal stopping policy and accept progressively worse candidates the longer the search progresses (Lee et al., 2004; Lee, 2006).

In this type of task, options are presented sequentially and are not revisited once rejected (Gilbert & Mosteller, 1966; Bearden & Murphy, 2007). Consequently, a subset of items must be sampled and rejected prior to setting an aspiration level, thereby encouraging observers to use a satisficing approach (Simon, 1955). That is, observers decide to terminate once some internal subjective threshold is met. Conversely, in a typical decision-making situation (and in our experiments) observers are allowed to revisit previously-viewed items. Caplin, Dean, and Martin (2011) have previously studied a similar task to ours. Like us, they were primarily interested in studying the search aspect of decision making. In their experiment, observers were shown a list of options with each option comprising a series of additions and subtractions, expressed in dollar amounts. By solving the calculation, the observer could determine the dollar worth of any given option. Observers were required to search for the option corresponding to the highest dollar amount. They found that their data could be explained in terms of satisficing (Simon, 1955). Observers would search the options sequentially, stopping when they found an option whose value exceeded an internal threshold.

The purpose of the current study was to investigate the computational processes used by consumers when faced with

a simpler task, one that substantially simplified the evaluation process. In our task, observers were explicitly told the desirability of each item so did not have to calculate it. In this way, much of the evaluation of a product's desirability was avoided. We sought to determine whether satisficing would still be found for this task.

In our first experiment, the observers were required to search the items using a computer mouse. The desirability rating of each item was revealed only when the cursor was placed above it. As such, the paradigm was similar to that popularized by MouseLab (Camerer et al., 1993; Johnson et al., 2002; Gabaix et al., 2006). The advantage of this paradigm is that it mimics those situations where an observer can evaluate the desirability of a product only by manually interacting with it.

In other situations, the desirability of a product can be estimated simply by looking at it. To study these situations in our second and third experiments, the desirability ratings were permanently on display, and we used an eye tracker to study how observers visually searched the display. We were especially interested in how the search process would differ in this situation compared with the previous situation where manual interaction was required.

We proposed and compared three competing models of search: (i) a predetermined model, (ii) a satisficing model, and (iii) a decreasing threshold model. All of these models are variations of dynamic search models used in economics (McCall, 1970; Jovanovic, 1979; Caplin et al., 2011; Reutskaja et al., 2011). The first two models were selected as they are very similar to those proposed by Reutskaja et al. (2011) for similar circumstances. The third model was selected because, as discussed later, our data imply that an observer's internal choice threshold does decrease over the course of the trial, which in turn suggests that a model based on a such a mechanism might be the most natural way to explain our data.

A common finding is that observers often reinspect previously inspected items. The three models account for this in the same way by assuming that the observers inspect the items one at a time but that there is a probability  $\beta$  that any given inspection may be directed to an item that has previously been inspected. The higher the  $\beta$  parameter, the more likely a given inspection will be directed to a previously inspected item.

The three models further assume that while performing this sequence of inspections the observer attempts to remember which of the inspected items has the greatest value (i.e., has the highest number). However, our data show that observers do not always choose this item. The models account for this by assuming that sometimes observers fail to recognize the value of an item. Thus, although they may have inspected a highly valuable item, they do not always realize that they have done so and therefore do not remember it. The parameter  $\alpha$  denotes the probability that observers will remember the most desirable item when they encounter it.

The predetermined model is so named, because it assumes that observers will inspect a predetermined number of items before making their decision. This number varies from trial to trial in a random fashion. This model assumes that observers will attempt to choose the best item from the set of inspected items but, as discussed above, will sometimes fail to do this because they did not remember the most valuable item.

The satisficing model is similar to the predetermined model except that it additionally assumes that observers will terminate the search early if they find an item whose desirability exceeds an internal threshold. This threshold is assumed to be constant and not to be affected either by the duration of the trial or the desirability of the items that had been viewed so far in that trial. If no item exceeds this threshold, then the search is terminated after a predetermined number of inspections and the observer attempts to select the best item from the previously viewed set as before.

The satisficing model is particularly interesting because it makes similar predictions to an ideal observer model under the assumption that observers can reliably remember which items they have previously inspected but that each inspection incurs an implicit cost. According to this ideal observer model, the observer will terminate the search when the cost of the next inspection exceeds the expected gain of making another inspection. Because the items are necessarily inspected at random with respect to desirability, this implies that the observer will act as if there is a fixed threshold. Terminating the search once an item has been found whose desirability exceeds this threshold, because then the implicit costs of continuing the search will outweigh the expected gain. Of course, once the observer has inspected all the items in the display, the ideal observer model predicts that the observer should immediately terminate the search, choosing the most desirable item seen so far. The ideal observer model is therefore similar to the satisficing model except that the latter allows for the possibility that the observer may opt for a different cutoff, allowing for the possibility that the observer may choose to inspect only some of the items. Thus, the satisficing model is somewhat more general than the ideal observer model, which is why we choose to test it.

Like the satisficing model, the decreasing threshold model also assumes that an internal threshold can determine when the search is terminated. However, unlike the satisficing model, it assumes that this threshold decreases as the observer inspects more items. Thus, even if none of the items in the display are particularly desirable eventually the internal threshold will drop enough for one of them to exceed the threshold, after which the trial will be terminated with the observer usually, but not always, selecting the most desirable item of those viewed in that trial.

One could argue that the decreasing threshold model makes similar predictions to an ideal observer model under the assumption that the observers cannot remember which items

they had previously inspected. In this case, as the trial progresses, the chance of the observer inadvertently reinspecting a previously inspected item increases, so the expected gain of each additional inspection decreases. Assuming a fixed implicit cost per inspection and assuming that the observer will terminate the trial when the expected gain of an additional inspection falls below the implicit cost of an inspection it follows that as the trial progresses the threshold for terminating the trial will decrease. In other words, the observer would be willing to terminate the trial for a less desirable item. While this version of the ideal observer model would indeed make similar predictions to the decreasing threshold model, it only does so if one assumes that observers cannot remember where they have previously inspected and, as will be discussed later, in all three experiments we found that this was not the case. Observers were able—to some extent—to remember which items they had previously inspected.

#### A functional description of the three models

To more precisely define these three models, in this section we present a functional description of them.

##### *Predetermined model*

1. In the course of the trial, the model attempts to keep track of the most desirable item inspected so far in that trial and labels this quantity *max\_inspected*. This value is set to zero at the beginning of each trial.
2. At the start of the trial, the model decides on the maximum number of inspections that will occur in that trial. This number,  $N$ , is drawn from a Poisson distribution with mean  $M$ .
3. For each inspection, the model decides whether to inspect a new item or inspect a previously inspected item. The probability of inspecting a previously inspected item is given by  $\beta$ . Obviously, this rule is subject to the logical constraint that if all the items have already been inspected then the model will necessarily reinspect a previously inspected item and if no item has yet been inspected the model will necessarily inspect a new item.
4. Having decided what type of item to inspect (i.e., a previously inspected item or a new item), the model then randomly chooses an item to inspect from the appropriate subset of items.
5. The value of the inspected item is then compared to *max\_inspected*. If the currently inspected is more desirable than *max\_inspected* then *max\_inspected* is increased to the value of the currently inspected item with probability  $\alpha$ . This means that the value of *max\_inspected* is not always updated even when it is less than the value of the currently inspected item.

6. The model then determines if  $N$  inspections have occurred. If they have occurred and if the currently inspected item has desirability equal to  $max\_inspected$ , then the model chooses it and terminates the trial. Otherwise, the model reinspects the item corresponding to  $max\_inspected$  and then terminates the trial, choosing that item.
7. If the trial is not terminated in step 6 then steps 3–7 are repeated.

In summary, this model is defined by three parameters:  $\alpha$ ,  $\beta$ , and  $M$ .  $\alpha$  is the probability that if the value of a currently inspected item exceeds the value of  $max\_inspected$ , then the value of  $max\_inspected$  will be updated with the value of the currently inspected item.  $\beta$  the probability that any given inspection will be directed towards a previously inspected item.  $M$  is the mean of the Poisson distribution used to determine how many items will be inspected on any given trial.

#### Satisficing model

1. In the course of the trial, the model attempts to keep track of the most desirable item inspected so far in that trial and labels this quantity  $max\_inspected$ . This value is set to zero at the beginning of each trial.
2. At the start of the trial the model decides on the maximum number of inspections that will occur in that trial. This number,  $N$ , is drawn from a Poisson distribution with mean  $M$ .
3. For each inspection, the model decides whether to inspect a new item or inspect a previously inspected item. The probability of inspecting a previously inspected item is given by  $\beta$ . As before, this rule is subject to the logical constraint that that if all the items have already been inspected then the model will necessarily reinspect a previously inspected item and if no item has yet been inspected the model will necessarily inspect a new item.
4. Having decided what type of item to inspect (i.e., a previously inspected item or a new item), the model then randomly chooses an item to inspect from the appropriate subset of items.
5. The value of the inspected item is then compared to  $max\_inspected$ . If the currently inspected is more desirable than  $max\_inspected$ , then  $max\_inspected$  is increased to the value of the currently inspected item with probability  $\alpha$ .
6.  $Max\_inspected$  is then compared to an internal threshold,  $T$ . If  $max\_inspected$  exceeds this threshold, then the model terminates the trial, choosing that item.
7. If the trial has not been terminated on step 6, the model then determines if  $N$  inspections have occurred. If they have occurred, then the model takes one of two courses of action. If the currently inspected item has desirability equal to  $max\_inspected$ , then the model chooses it and terminates the trial. Otherwise, the model reinspects the item corresponding to  $max\_inspected$  and then terminates the trial, choosing that item.
8. If the trial is not terminated on step 7 then steps 3–8 are repeated.

In summary, this model is defined by the four parameters:  $\alpha$ ,  $\beta$ ,  $T$ , and  $M$ .  $\alpha$  and  $\beta$  are defined in the same way as previously.  $T$  represents the choice threshold.  $M$  represents the mean of the Poisson distribution used to determine how many items will be inspected on any given trial.

#### Decreasing threshold model

1. In the course of the trial, the model attempts to keep track of the most desirable item inspected so far in that trial and labels this quantity  $max\_inspected$ . This value is set to zero at the beginning of each trial.
2. For each inspection, the model decides whether to inspect a new item or inspect a previously inspected item. The probability of inspecting a previously inspected item is given by  $\beta$ . As before, this rule is subject to the logical constraint that that if all the items have already been inspected then the model will necessarily reinspect a previously inspected item and if no item has yet been inspected the model will necessarily inspect a new item.
3. Having decided what type of item to inspect (i.e., a previously inspected item or a new item), the model then randomly chooses an item to inspect from the appropriate subset of items.
4. The value of the inspected item is then compared to  $max\_inspected$ . If the currently inspected item is more desirable than  $max\_inspected$ , then  $max\_inspected$  is increased to the value of the currently inspected item with probability  $\alpha$ . This means that the value of  $max\_inspected$  is not always changed even when it is less than the value of the currently inspected item.
5.  $Max\_inspected$  is then compared to an internal threshold. This threshold starts at a fixed value ( $T$ ) and then decreases by a set amount  $\Delta T$  for each inspection carried out by the model. If  $max\_inspected$  exceeds this threshold, then the trial will be terminated. If the currently inspected item has desirability equal to  $max\_inspected$ , then the model chooses it and terminates. Otherwise, the model reinspects the item corresponding to  $max\_inspected$  and then terminates the trial, choosing that item.
6. If the trial is not terminated on this inspection then steps 2–6 are repeated.

In summary, this model is defined by the parameters:  $\alpha$ ,  $\beta$ ,  $T$ , and  $\Delta T$ .  $\alpha$  and  $\beta$  are defined in the same way as previously.  $T$  and  $\Delta T$  are used to define the choice threshold.

Each of the models builds on the previous model by adding an additional level of complexity. The predetermined model simply chooses a maximum number of inspections and continues searching until reaching that number. The highest inspected number is then usually chosen. The satisficing model builds upon this by adding a decision threshold. If this decision threshold is exceeded, the trial terminates early. Otherwise observers continue to search for a predetermined number of inspections before attempting to choose the highest inspected number. Finally, the decreasing threshold model refines the concept of the decision threshold by allowing it to decrease across trials. This means that the model no longer needs to define a priori a maximum number of potential inspections per trial, because it can be guaranteed that eventually an item will be found that will exceed the decision threshold regardless of the values of inspected items.

### Experiment 1: 24 partially hidden items

In many conventional settings, the observer might need to touch an item to be able to evaluate it. For example, in a shop a consumer may need to pick up a product to read its price tag. Glockner and Betsch (2008) found that with multi-attribute stimuli, forcing participants to manually search each attributes value by clicking on the attribute with the mouse led to an increased use of heuristic strategies, such as making a decision based on a single attribute. Likewise, in search tasks, forcing an observer to physically interact with a display caused observers to alter their search behaviour and terminate their search earlier than they would with a purely visual display due to the increase costs associated with the manual interactions (Sonnemans, 1998; Pedersini et al., 2010). As a prelude to considering displays where the items could be evaluated purely by inspection (Pieters, 2008), Experiment 1 investigated how observers would search a display that required manual interaction. Specifically, the desirability value of each item was revealed only when a cursor was positioned on that item. The number of items in the display was rather large ( $n=24$ ) to reflect the number of choices that an observer might be presented with in a real consumer choice situation.

#### Method

**Participants** There were eight participants (age range 26–36; 5 males, participant PH is an author). A near vision (40 cm) Good-Lite® eye chart was used to verify that all observers had normal or corrected-to-normal (20/20) near vision. As in all the experiments reported here, the observers provided

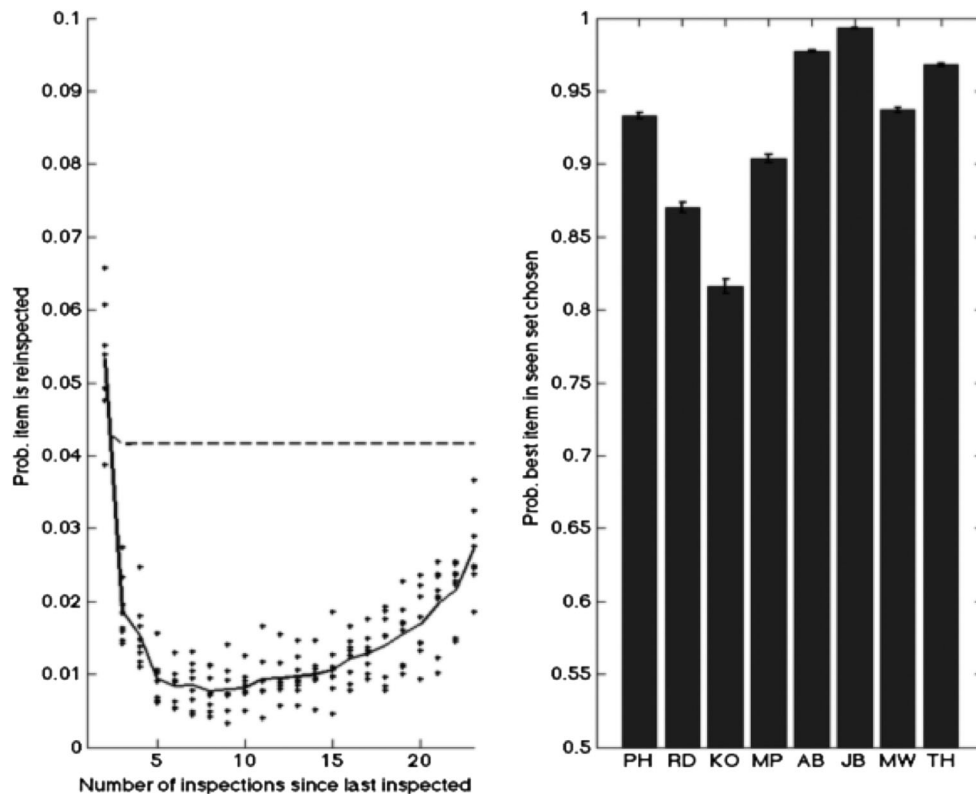
informed written consent, and the study was approved by the Department Human Ethics Advisory Group in the School of Psychological Sciences at the University of Melbourne.

**Stimuli and apparatus** The participants viewed a 21-inch CRT monitor at a resolution of 1280 by 1024 pixels with a frame rate of 85 Hz. Stimuli were presented in MATLAB (Mathworks, Natick, MA) using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). A combined head-and-chin rest maintained the viewing distance at 60 cm. The display comprised 24 black squares, arranged in a 4 by 6 grid. The grid subtended 33 by 24 degrees of visual angle (°) and each square subtended  $2.8^\circ \times 2.8^\circ$ . Every time an observer moved his cursor over a square, the square temporarily disappeared to reveal a two-digit number representing the value of that square. Each digit subtended  $0.15^\circ \times 0.1^\circ$ . Each number was randomly chosen in the range 10 to 99 inclusive, but once assigned to a square remained constant for the duration of the trial. The digits were black and the background was white.

**Procedure** The observer would start each trial by pressing the space bar. The 24 squares would then appear. The observer would move the cursor over the squares to reveal the numbers underneath, one at a time, until he was confident that he had located the highest number. At this point, he would click the mouse. Whichever number was uncovered when the mouse was clicked was deemed to be the observer's choice. The observer was then given feedback as to whether he had been successful in finding the highest number and invited to start the next trial. The purpose of the feedback was to encourage observers to rapidly learn a consistent strategy. Observers were allowed to take as long as they liked to complete a trial and there was no penalty for uncovering an item more than once. A single session comprised 200 trials. In total, each participant participated in six sessions, but the first session was treated as practice and not analyzed. It took each participant approximately 6 hours to complete these sessions.

#### Results and discussion

We found that observers typically did not view all the items before making a decision, terminating the search before they had inspected all the items on approximately 60 % of the trials. On average observers inspected 18.2 ( $\pm 3.0$ ) of the 24 items. Looking at the inspection traces we could see no pattern, a point that we will return to later. Because an item's desirability was distributed randomly and remained covered until the item was selected by the cursor, the initial search process was necessarily random with respect to item desirability. Figure 1a shows the probability that an item is reinspected as a function of the number of inspections since it was last inspected. The dashed line represents the probability that an



**Fig. 1** Data from Experiment 1. **a** The probability that an item is reinspected as a function of the number of inspections since it was last inspected. The dashed line represents the probability that an item would be reinspected if the process had been entirely random. The solid line

represents the average observer data and each cross represents a data point from a single observer. **b** Probability that the observer selects the most desirable of the viewed items

item would be reinspected if the process had been entirely random. The solid line represents the average observer. Most of the time the solid line is well below this dashed line, indicating that there is a strong tendency not to reinspect a previously inspected item. This was confirmed with a Pearson chi-square test  $\chi^2(21, N = 3967) = 5304, p < 0.001$ . Figure 1b shows the probability that when observers terminate the search they select the most desirable item out of all the items they have seen. While there is some interobserver variability, in general observers are quite good at doing this, selecting the best item on 92.5 % ( $\pm 5.9$  %) of the trials.

These findings justify the two assumptions common to all three models: observers have a tendency not to immediately refixate an item and will sometimes not choose the most desirable item out of the items viewed. However, this figure does not address the stopping rule employed by the observers, which is the principle difference between the three models. This issue is addressed in Fig. 2.

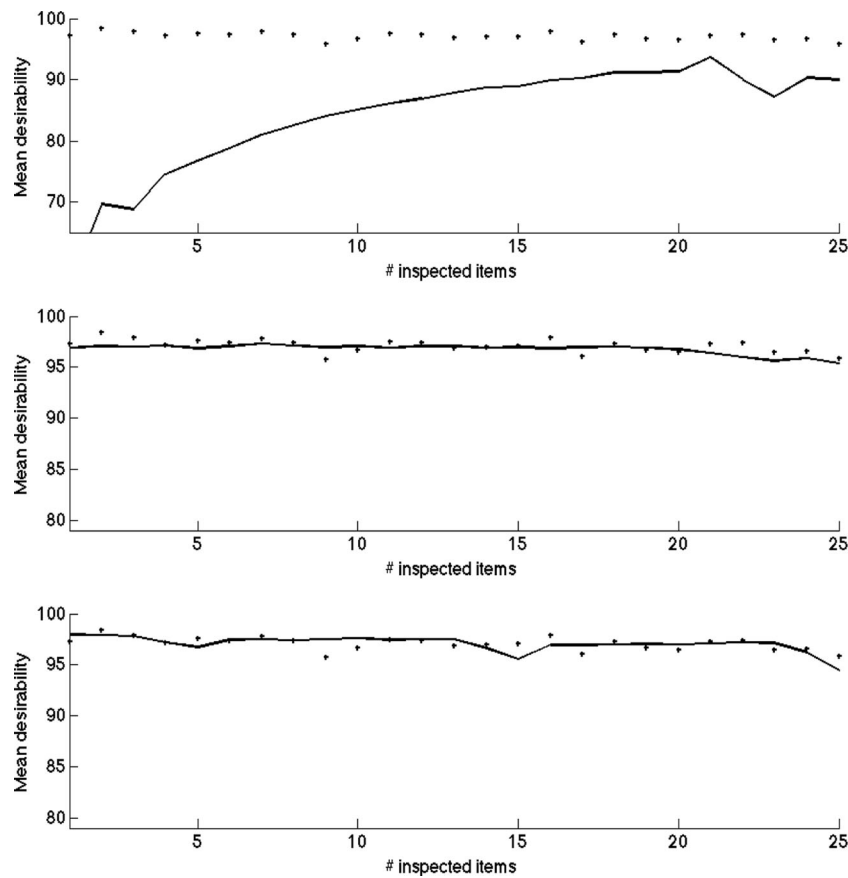
Figure 2 shows the value of the finally selected item in a trial as a function of the number of fixations in that trial. The crosses show the average data of the eight observers and the lines the model fits. Subplots a, b, and c show the model fit for the predetermined model, the satisficing model and the decreasing threshold model respectively. Considering the data

we see that the mean desirability of the selected item is inversely correlated with the number of items that are inspected:  $r(23) = -0.547, p = 0.0047$ .

Each of the three models was fit by first randomly selecting values for its parameters. Using these parameter values we then ran the model for 5,000 trials to calculate its predictions. By comparing these predictions to the actual measurements, we were able to calculate the mean square error (MSE). The MATLAB function *fminsearch* used the MSE to suggest new parameter values to reduce the MSE. This process was repeated until convergence was obtained in that it was not possible to further reduce the MSE by varying the parameter values.

The predetermined model predicts that for trials where observers happen to inspect more items they will on average tend to end up selecting a more valuable item, simply by virtue of the fact that they have more items to choose from. Thus, the model predicts that the mean value of the selected item should increase as a function of the number of inspections in a trial.

The satisficing model predicts that observers will terminate a trial when they inspect an item that exceeds an internal choice threshold. If none of the inspected items exceed this



**Fig. 2** The desirability of the finally selected item as a function of the number of inspections in that trial in Experiment 1. Model fits by (a) the predetermined model, (b) the satisficing model, and (c) the decreasing threshold model. Crosses represent data points; solid lines represent the model fits

threshold, the model will eventually attempt to choose the best item that it has inspected so far in the trial. So, the longer the trial the more likely that none of the items exceeded the internal threshold and the model had to just choose the best item it could find. The satisficing model therefore predicts that mean value of the selected item should decrease a function of the number of inspections in that trial.

The decreasing threshold model predicts that a trial is terminated when an inspected item exceeds an internal threshold, but this threshold continuously decreases across the duration of the trial. Thus, the longer the trial has continued for, the lower the choice threshold. So, this model predicts that the mean value of the finally selected item will also decrease as the number of inspections increases.

Increasing the number of parameters of a model is likely to reduce the MSE of the model fit but runs the risk of the model overfitting the data, i.e., fitting the noise in the data instead of fitting just the underlying relationship. This can result in poor predictive performance and choosing the wrong model to represent one’s data. One way to compensate for this problem is to introduce a penalty term for the number of parameters in the model. This ensures that we only choose models with higher number of parameters when the extra parameters result in a suitably substantial decrease in the MSE of the model fit.

There are different ways of penalizing models for the number of parameters they contain. In comparing our models, we utilized the Bayesian Information Criterion (BIC; Priestley, 1981). This allows us to compare models with different numbers parameters in a rigorous fashion as the one with the lowest BIC will be the one that fits the data the best.

$$BIC = n * \ln(MSE) + k * \ln(n) \tag{1}$$

where MSE is the mean square error (i.e., the error variance), n is the number of data points and k is the number of parameters.

As shown in Fig. 2, the predetermined model (subplot a) provides a very poor fit to the data (BIC =150, MSE =274). The satisficing model (subplot b) and the decreasing threshold model (subplot c) provide much better fits, with BICs of 31.0 and 73.90 respectively and MSEs of 2.07 and 11.08 respectively. Setting aside the predetermined model on the basis of its qualitative misprediction, we then chose to fit both the satisficing and decreasing threshold models to each observer’s individual data, as we suspected that there might be significant individual differences.

For each observer, we collected six measurements: the average number of inspections per trial, the average number

of reinspections per trial, the mean desirability (i.e., average value) of the chosen item, the probability that the observer chooses the maximum valued item of the items viewed, the mean inspection efficiency, and relative desirability of the chosen item. The inspection efficiency was calculated by averaging the relative desirability of all the inspected items within a trial. Relative desirability,  $r_d$ , was in turn calculated using the following formula adapted from (Reutskaja et al., 2011):

$$r_d = (d_{\text{sel}} - \text{mean}_d) / (\text{max}_d - \text{mean}_d) \quad (2)$$

where  $d_{\text{sel}}$  is the desirability of the selected item,  $\text{mean}_d$  and  $\text{max}_d$  are the mean and maximum desirability of all the items in the display, regardless of whether the items were viewed or not. The advantage of using relative desirability instead of absolute desirability is that on some trials, just by random chance, none of the items will be particularly valuable. This is more likely to occur where there are fewer choices in each trial. Using relative desirability allows us to avoid this issue.

In the model descriptions given above, we considered a version of both the satisficing model and the decreasing threshold model where  $\alpha$  and  $\beta$  were allowed to vary as free parameters. While this will almost certainly improve the model fits, it is not clear a priori whether the improvement is sufficient to justify including these model parameters. We therefore fitted to each subject's data four versions of each of these two models. In the first version, both  $\alpha$  and  $\beta$  were allowed to vary as free parameters. In the second version, it was assumed that observers would always choose the maximum inspected item, so  $\alpha$  was set to one but  $\beta$  was still allowed to vary freely, and the model had only three parameters. In the third version,  $\alpha$  was retained as a free parameter, but it was assumed that observers would never reinspect a previously inspected item (i.e.,  $\beta$  was set to 0). In the fourth version,  $\alpha$  was set to 1 and  $\beta$  was set to 0 so that there were only two free parameters.

Table 1 shows the BIC for each model for each observer (MSE are in Appendix 2). As can be seen, the decreasing threshold model with  $\alpha$  and  $\beta$  as free parameters is the best fit for five of the eight observers. The satisficing model with  $\alpha$  fixed was the best fit for two observers and the satisficing model with  $\beta$  fixed was the best fit for one observer. Thus, if we were to select just one model to represent all the observers we would do best to select the decreasing threshold model with both  $\alpha$  and  $\beta$  as free parameters. Figure 3 shows the model fits for this model for all 8 observers. Each dot represents a single observer. Appendix 1 lists the model parameters.

**Table 1** BICs for each the eight models for each observer in Experiment 1

	Sat $\alpha, \beta$	Sat $\sim\alpha, \beta$	Sat $\alpha, \sim\beta$	Sat $\sim\alpha, \sim\beta$	DT $\alpha, \beta$	DT $\sim\alpha, \beta$	DT $\alpha, \sim\beta$	DT $\sim\alpha, \sim\beta$
PH	-29.0	-30.9	-25.5	-25.6	<b>-38.5</b>	-33.4	-26.1	-25.7
RD	-27.8	-21.3	-32.0	-21.6	<b>-34.7</b>	-25.4	-21.1	-17.3
KO	-24.9	-12.4	-26.2	-14.1	<b>-28.4</b>	-19.9	-18.1	-12.6
MP	-24.4	-25.4	-25.5	-28.6	<b>-30.1</b>	-27.6	-30.0	-26.1
AB	-46.3	<b>-47.7</b>	-38.0	36.6	-35.2	-34.3	-35.6	-35.7
JB	-44.3	<b>-46.6</b>	-41.7	-43.1	-39.0	-38.7	-38.0	-40.2
MW	-33.9	-29.5	<b>-37.4</b>	-30.2	-32.2	-27.2	-29.5	-28.7
TH	-27.1	-26.5	-28.2	-31.2	<b>-34.5</b>	-32.0	-31.1	-31.0

DT  $\alpha, \beta$ =decreasing threshold model,  $\alpha$  and  $\beta$  allowed to vary as free parameters; Sat $\sim\alpha, \sim\beta$ =satisficing model, with neither  $\alpha$  and  $\beta$  allowed to vary as free parameters. Numbers in bold represent the best fits

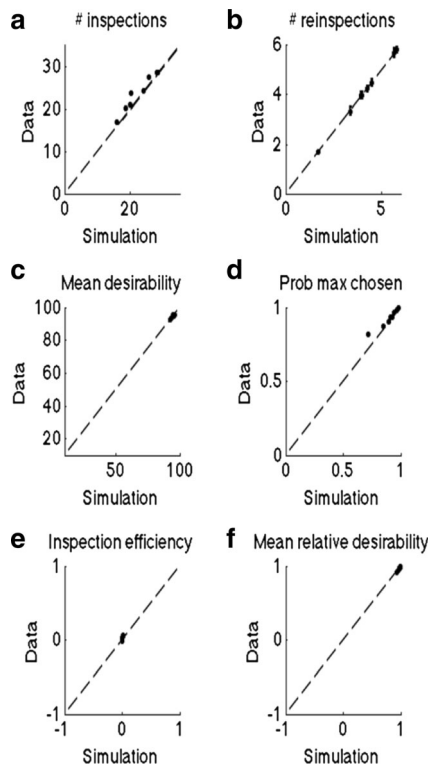
We also examined the data for possible locations biases in the inspections. Figure 4a shows the percentage of trials in which the first inspection occurs at each of the 24 locations. While the distribution is fairly uniform a Pearson chi-square test revealed some significant deviation from uniformity,  $\chi^2(23, N=3864)=59.6, p<0.001$ , although these deviations were not obviously systematic.

Figure 4b shows the percentage of the total number of inspections that occurred at each location. This display was more uniform and there was no evidence for any significant deviations from uniformity,  $\chi^2(23, N=80099)=16.3, p=0.80$ . There was therefore little evidence that there is any systematic bias to concentrate inspections on just one part of the display.

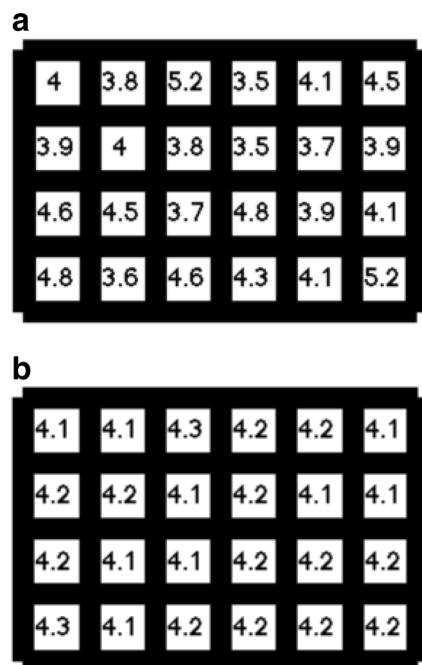
## Experiment 2: 12 uncovered items

Whereas the previous experiment was designed to investigate those situations where observers needed to manually interact with an item before they can evaluate it, this experiment was designed to investigate a situation that might be encountered in a shop where an observer visually searches a collection of shelves for the best product to purchase. As before, the display comprised an array of two digit numbers, each representing the desirability of an item, and the observer's task was to find the most desirable item (i.e., the highest number). Unlike Experiment 1, observers did not have to uncover the value of the items but instead could read an item simply by fixating on it. Consequently, any implicit costs associated with uncovering each item would be removed. From Glockner and Betsch (2008), we expected satisficing to no longer occur and instead for observers to simply inspect all items and then just select the most desirable one. Consequently, we expected our data to be best described by the predetermined model.





**Fig. 3** Decreasing Threshold Model fits for Experiment 1. Both  $\alpha$  and  $\beta$  were allowed to vary freely. Each dot represents one observer. Error bars represent the standard error of the mean and are often too small to be visible. **a** Mean number of inspections per trial. **b** Mean number of reinspections per trial. **c** Mean desirability of the selected item. **d** Probability that observer selects the most desirable of the viewed items. **e** Mean inspection efficiency. **f** Mean relative desirability of the selected item



**Fig. 4** Inspection biases for Experiment 1. **a** Percentage of initial inspections at each location. **b** Percentage of all inspections at each location

Method

*Participants* There were four participants. Ages ranged from 22 to 35 years; three were male. Two had participated in the previous experiment, and all had normal or corrected-to-normal visual acuity.

*Apparatus* The apparatus was the same as before except that an Arrington Research® 220Hz HeadLock™ eye tracker was used to monitor gaze positions.

*Stimuli* Participants viewed a display comprising twelve numbers, arranged in a 3×4 array. This was fewer items than in the previous experiment to ensure that the eye tracker could accurately identify which item was currently fixated. As before, each number comprised two digits and was randomly chosen in the range 10 to 99 inclusive. The digits were black and the background was white. Each digit subtended 0.15 by 0.1 degrees of visual angle (°). Neighboring digits were separated by 10°. Visual acuity falls off rapidly with eccentricity, decreasing by a factor of 5 at eccentricity of 10° (Westheimer, 1965; Brunner et al., 2010). Consequently, observers could read only the number directly foveated and could not read the neighboring numbers. Thus, the observers were forced to fixate each number in turn, as verified by eye tracking. This allowed us to use an eye tracker to measure which items were viewed before the decision was made.

*Procedure* The procedure was essentially the same as in Experiment 1 but modified to allow for eye tracking. At the start of the experiment, the eye tracker was calibrated using a 12-point calibration procedure, using the ViewPoint® computer program from Arrington Research. The 12 calibration points corresponded to the 12 points the numbers would occupy. This calibration procedure was repeated every 25 trials. The observer would start each trial by pressing the space bar. The 12 numbers would then appear, arranged in a 3×4 grid. The observer would read the numbers by fixating each one, one at a time, until he was confident that he had located the highest number. At this point, he would press the space bar while fixating that number. Whichever number was currently fixated when the space bar was pressed was deemed to be the observer’s choice. The observer was then given feedback as to whether he had been successful in finding the highest number and invited to start the next trial. As in Experiment 1, observers were allowed to take as long as they liked to complete a trial and there was no penalty for refixating a previously fixated item. A single session comprised 200 trials. In total, each participant participated in six sessions, but, as before, the first session was treated as practice and not analyzed. For each trial, we were able to determine which numbers had been fixated in which order, using custom analysis software written in MATLAB®. Fixations that lasted less than 75 ms were

assumed to be artifacts and discarded (Reichle et al., 1998). Choosing a different cutoff in the range of 50 ms to 100 ms did not substantially alter any of our findings.

## Results and discussion

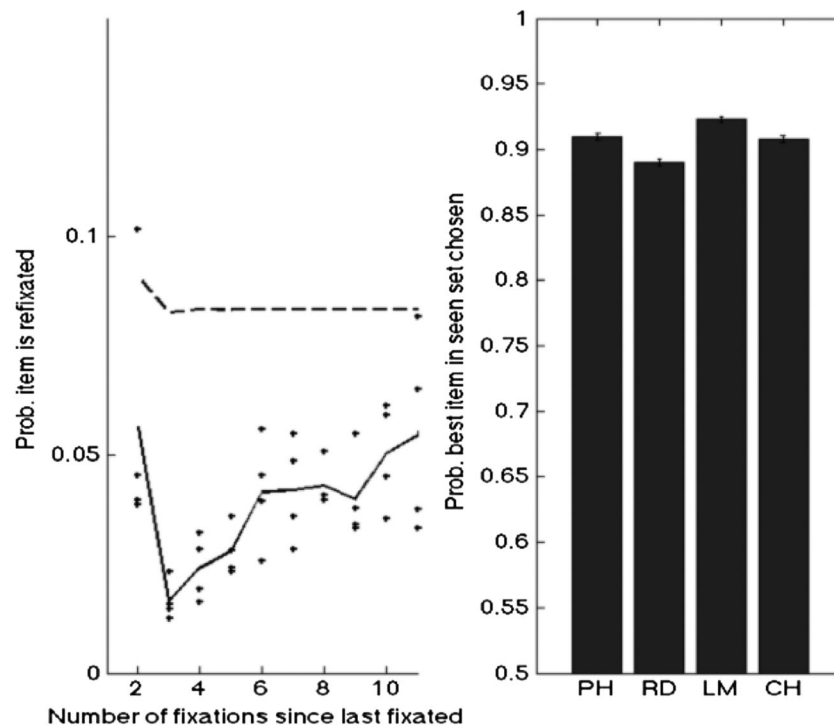
We found that on 75 % of the trials the observer would terminate the search before inspecting all the items, contradicting previous findings (Glockner & Betsch, 2008). On average observers would inspect 8.8 ( $\pm 1.4$ ) items before terminating the trial. Figure 5a shows the probability of an item being refixated as a function of when it was last fixated. The dashed line denotes what performance would be expected if the process had been entirely random, and the solid line denotes the average observer performance. A Pearson chi-square test revealed these two distributions to be significantly different,  $\chi^2(39, N = 1878) = 1444, p < 0.001$ , which demonstrates a tendency for observers not to refixate a recently fixated item. Figure 5b confirms that observers were good at selecting the most desirable item out of the set of viewed items. On average observers selected the highest number 91 % ( $\pm 1.4$  %) of the time.

Figure 6 shows how the average desirability of the finally selected item varies with number of fixations in that trial. Crosses represent the average data for the four observers, lines represent the model fits. Again, we can see a decrease as the number of fixations increases, but the resultant slope is not significant,  $r(13) = -0.286, p = 0.302$ . Subplot a shows the

model fit for the predetermined model. As before, this model provides the worst fit to the data (BIC = 79.5, MSE = 116), because the model predicts that mean desirability should increase with the number of fixations in the trial. Subplots b and c show the model fits for the satisficing and decreasing threshold models respectively. Both fits are considerably better than that of the predetermined model with BICs of 54.7 and 55.8 respectively and MSEs of 18.7 and 20.0 respectively.

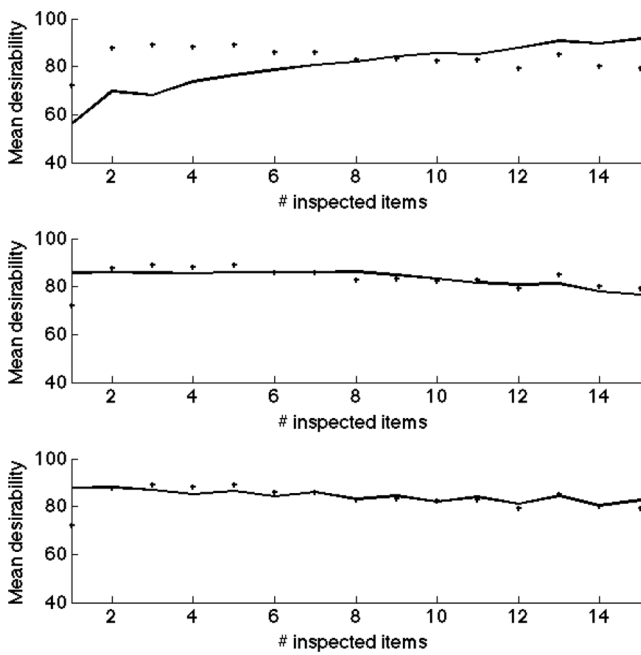
As before, we fitted four versions of the satisficing model and four versions of the decreasing threshold model to each observer's data. The BIC values of the resultant fits are shown in Table 2 (MSE values are in Appendix 2). For all of the observers, the best fit was achieved by the satisficing model threshold with  $\alpha$  as a free parameter but  $\beta$  fixed at zero. Figure 7 shows the fit for this model for the individual observers. As before, the fits are very good. Appendix 1 lists the model parameters.

Figure 8a shows the percentage of first inspections that occur at each of the 12 locations. The distribution is clearly not uniform,  $\chi^2(11, N = 4000) = 13,491, p < 0.001$ . It shows that fixations typically started on one of two locations: either the top left corner or at the number located just to the right of the centre of the display. Figure 8b shows the proportion of the time each location in the display was fixated. This distribution is much more uniform, although there are still significant biases,  $\chi^2(11, N = 47129) = 2,445, p < 0.001$ . In particular, there is still a slight bias to fixate the item just to the right of centre. However, that bias may be largely attributable to the



**Fig. 5** Data from Experiment 2. **a** The probability that an item is reinspected as a function of the number of inspections since it was last inspected. The dashed line represents the probability that an item would

be reinspected if the process had been entirely random. **b** Probability that the observer selects the most desirable of the viewed items



**Fig. 6** The desirability of the finally selected item as a function of the number of inspections in that trial in Experiment 2. Model fits by (a) the predetermined model, (b) the satisficing model, and (c) the decreasing threshold model. Crosses represent data points; solid lines represent the model fits

fact that observers typically started their fixations in that location.

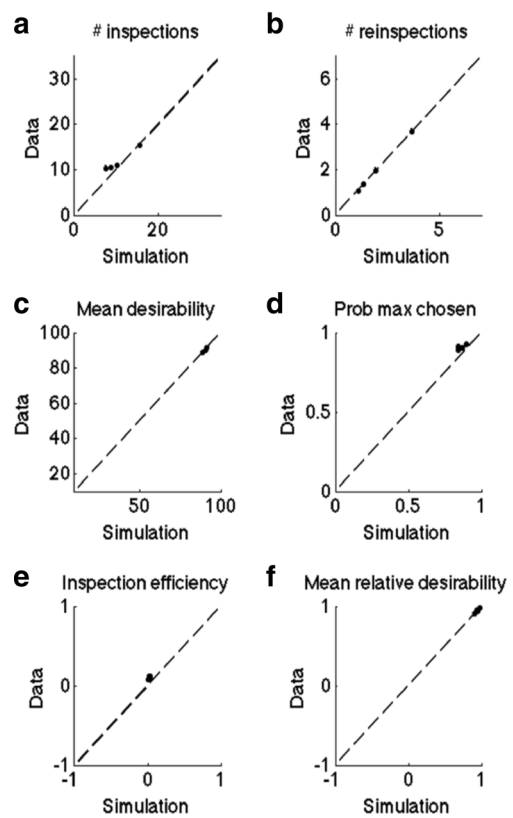
**Experiment 3: Four uncovered items**

In the previous experiment, we again found that people typically did not search all of the items and this occurred despite the items being presented uncovered. This finding contradicts the results of Glockner and Betsch (2008) who found that observers would search all the items before making a choice. Unlike Glockner and Betsch (2008), Experiments 1 and 2 used a large array of items. We wondered if observers would behave differently when confronted with only a few items in each display. In such situations, it would then be trivial to inspect each item and this

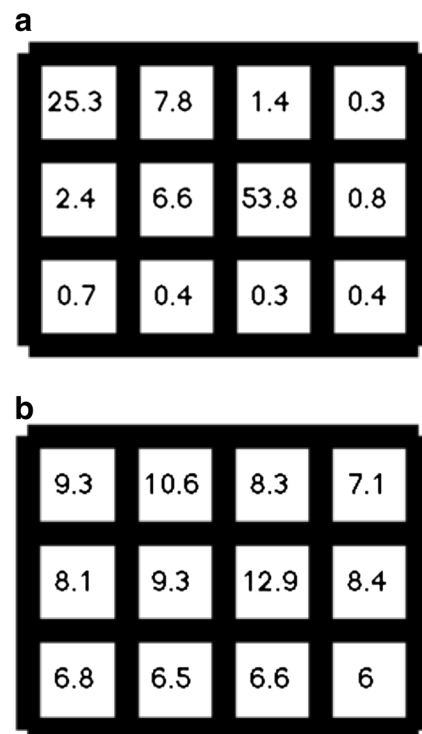
**Table 2** BICs for the each of the four models for each observer in Experiment 2

	Sat $\alpha, \beta$	Sat $\sim\alpha, \beta$	Sat $\alpha, \sim\beta$	Sat $\sim\alpha, \sim\beta$	DT $\alpha, \beta$	DT $\sim\alpha, \beta$	DT $\alpha, \sim\beta$	DT $\sim\alpha, \sim\beta$
PH	-27.0	-19.8	<b>-29.3</b>	-21.7	-27.7	-23.5	-19.0	-18.0
RD	-21.2	-17.4	<b>-26.8</b>	-20.1	-24.5	-17.4	-19.9	-16.2
LM	-28.4	-20.1	<b>-34.4</b>	-27.9	-29.5	-25.4	-23.6	-26.4
CH	-25.5	-23.4	<b>-30.1</b>	-23.1	-28.1	-24.2	-16.1	-21.3

DT  $\alpha, \beta$ =decreasing threshold model,  $\alpha$  and  $\beta$  allowed to vary as free parameters; Sat $\sim\alpha, \sim\beta$ =satisficing model, with neither  $\alpha$  and  $\beta$  allowed to vary as free parameters. Numbers in bold represent the best fits



**Fig. 7** Model fits for Experiment 2 for the satisficing model where  $\beta$  was fixed at zero but  $\alpha$  was allowed to vary freely. Same format as previously



**Fig. 8** Fixation biases for Experiment 2. a Percentage of initial inspections at each location. b Percentage of all inspections at each location

might encourage observers to compare all the items. We therefore reduced the number of items in the display to just four. We choose this number because it is reasonable to expect observers to easily remember (Cowan, 2001) and localise (Pylyshyn & Storm, 1988; Franconeri et al., 2007) such a small number of items. Also, previous decision studies have assumed that at least five items can be compared simultaneously (Hotaling et al., 2010; Tsetsos et al., 2010; Usher et al., 2010). In fact, under similar conditions, optimal decision strategies were observed when there were nine sources of information. Glockner and Betsch (2008) showed observers three alternatives, each of which had three attributes. Observers could inspect the attributes freely. Glockner and Betsch observed that observers would compare all the items despite the necessity of integration across three attributes for each alternative. In our task, no integration was required and there are only four items to search. Consequently, we expected to find a substantial decrease in heuristic processing and a switch to a more optimal search strategy.

## Method

**Participants** As before, there were four participants (age range 22–36, 2 males). Two had participated in the previous experiment. All were verified to have normal or corrected-to-normal visual acuity as before.

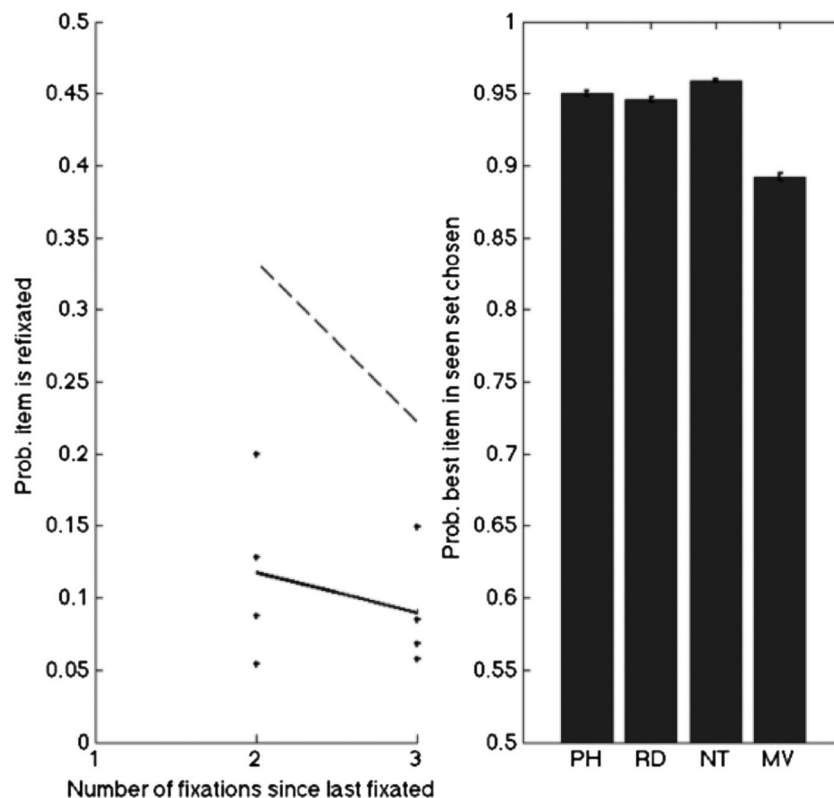
**Stimuli, apparatus, procedure, and model fitting** Other than reducing the number of items in a display to four, the apparatus and procedure were identical to that of Experiment 2. The numbers were the same size as before and appeared at the corners of imaginary rectangle with dimensions 16.8° by 13.2°.

## Results and discussion

In 32 % of the trials, the observers would still terminate the search before viewing all the items. However, this does not mean that on all the remaining trials observers viewed all four items exactly once and then chose the highest one, which would have been the optimal strategy. That strategy was in fact implemented on only 46 % of the trials.

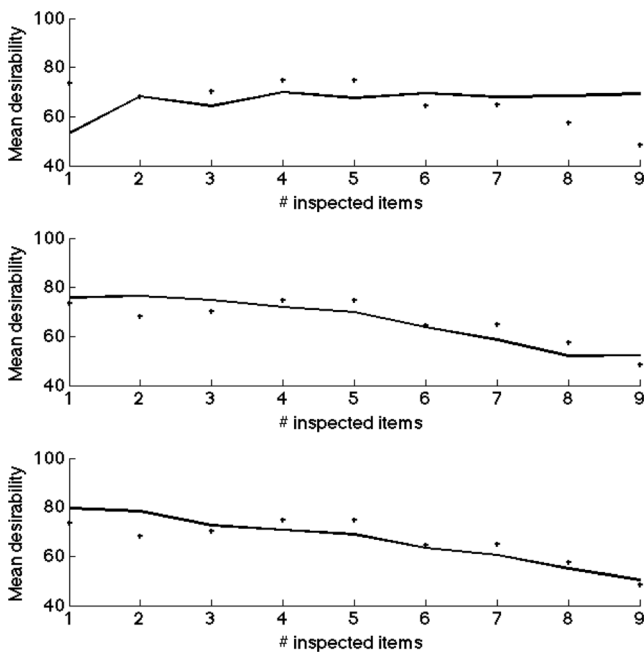
Figure 9a shows the probability of an observer fixating an item as a function of when the item was last fixated. The solid line represents the average observer and the dashed line the expected performance if the fixations were random. These two distributions are not equal,  $\chi^2(1, N=390)=425, p<0.001$ , again demonstrating a tendency to avoid refixating a recently fixated item. Figure 9b shows were good at choosing the most desirable items out of those that they had fixated. On average the most desirable item was chosen on 93.7 % ( $\pm 3.0\%$ ) of the trials.

Figure 10 shows that the average desirability of the finally selected item decreases a function of the number of fixations



**Fig. 9** Data from Experiment 3. **a** The probability that an item is reinspected as a function of the number of inspections since it was last inspected. The dashed line represents the probability that an item would

be reinspected if the process had been entirely random. **b** Probability that the observer selects the most desirable of the viewed items



**Fig. 10** The desirability of the finally selected item as a function of the number of inspections in that trial in Experiment 3. Model fits by (a) the predetermined model, (b) the satisficing model, and (c) the decreasing threshold model. Crosses represent data points; solid lines represent the model fits

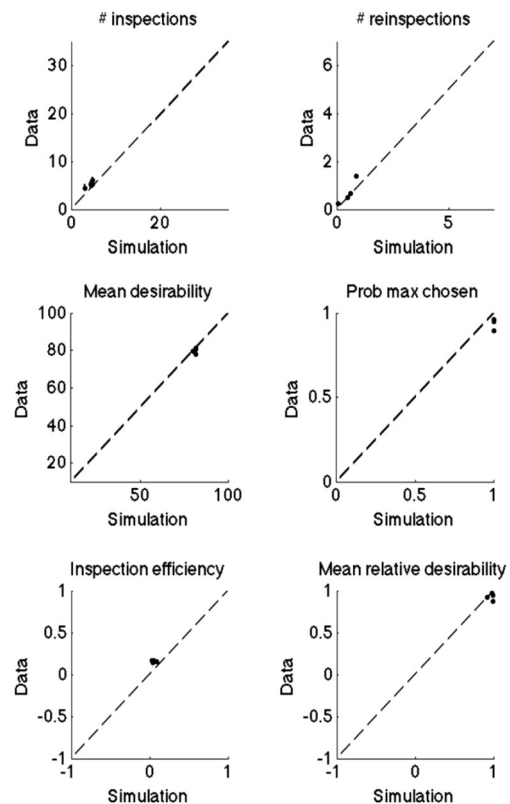
in that trial leading to a significant negative correlation,  $r(7) = -0.798, p = 0.010$ . As before, subplots a, b, and c show the model fits for the predetermined model, the satisficing model, and the decreasing threshold model respectively. The negative correlation ensures that the predetermined model cannot provide a good account of the data (BIC = 50.8, MSE = 136). The model fits of the other two models are better with BICs of 36.3 and 40.4 respectively and MSEs of 21.3 and 33.5 respectively.

As before, we fitted four versions of the satisficing model and four versions of the decreasing threshold model to each observer’s data. The BIC values of the resultant fits are shown in Table 3 (MSE values are in Appendix 2). For three of the four observers, the best fit (Table 3) was achieved by a decreasing threshold model. For two of these observers, the best fit was achieved with the version of the decreasing

**Table 3** BICs for each of the four models for each observer in Experiment 3

	Sat $\alpha, \beta$	Sat $\sim\alpha, \beta$	Sat $\alpha, \sim\beta$	Sat $\sim\alpha, \sim\beta$	DT $\alpha, \beta$	DT $\sim\alpha, \beta$	DT $\alpha, \sim\beta$	DT $\sim\alpha, \sim\beta$
PH	-17.3	-17.2	-17.3	<b>-29.1</b>	-26.7	-24.2	-28.9	-18.0
RD	-17.3	-17.4	-17.7	-19.5	-19.3	-21.2	-20.6	<b>-22.4</b>
NT	-11.4	-10.6	-11.4	-13.3	-24.1	-25.7	-26.7	<b>-27.4</b>
MV	-20.7	-17.5	-21.9	-19.3	-25.6	-15.1	<b>-27.5</b>	-18.3

DT  $\alpha, \beta$ =decreasing threshold model,  $\alpha$  and  $\beta$  allowed to vary as free parameters; Sat $\sim\alpha, \sim\beta$ =satisficing model, with neither  $\alpha$  and  $\beta$  allowed to vary as free parameters. Numbers in bold represent the best fits



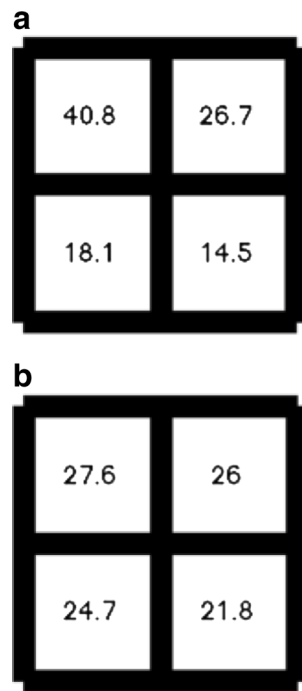
**Fig. 11** Model fits for Experiment 3 for the decrease threshold model where neither  $\alpha$  or  $\beta$  was allowed to vary. Same format as previously

threshold model where  $\alpha$  was constrained to equal 1 and  $\beta$  was constrained to 0. This makes sense, because with four items, reinspection is less useful than with more than four items. Figure 11 shows the fits for the decreasing threshold model with  $\alpha$  and  $\beta$  constrained. Most of the model predictions are highly accurate; however, the model under predicts the number of fixations (top left panel).

Figure 12a shows the probability that the first inspection occurs at each of the four locations. This distribution was not uniform:  $\chi^2(3, N = 4000) = 656, p < 0.001$ . There was a strong bias for fixations to start with the number at the top left. For this stimulus arrangement, there was no number located at the centre of the screen. Figure 12b shows the corresponding data when all fixations are considered. While the distribution is not uniform,  $\chi^2(3, N = 2086) = 150, p < 0.001$ , the bias is less pronounced.

### General discussion

The goal of our study was to investigate visual search performance in an abstract version of the supermarket choice problem. Observers were presented with an array of numbers and asked to find the largest one. Crucially, our task greatly simplified the evaluation of the desirabilities of the items; the desirability of an item was



**Fig. 12** Fixation biases for Experiment 3. **a** Percentage of initial inspections at each location. **b** Percentage of all inspections at each location

entirely determined by its numerical value. This avoided possible instabilities in product evaluations. Also, there was no explicit time pressure in that the observers were never penalized for taking too long to make a choice. Instead, any time pressure was implicit and due solely to the requirement for the observers to make a large number of these choices within a single session. A similar time pressure is presumably experienced by shoppers in a supermarket setting.

Our purpose was to investigate the computational processes used by observers in performing this task and how these processes are affected both by the number of items in the display and by how the observers are required to interact with the displays, specifically whether the observers can evaluate the desirability of the products simply by fixating them or whether the evaluation requires a manual interaction. Across all three experiments, we found that observers were typically very good at selecting the most desirable item from the set of items that has been inspected and this item was typically either the most desirable or close to the most desirable item in the display. This occurred despite the items being inspected at random with respect to desirability and even though the observers would sometimes waste time reinspecting previously inspected items.

To analyse our data, we initially considered three models, all variations of dynamic search models previously used in economics (McCall, 1970; Jovanovic, 1979; Caplin et al., 2011; Reutskaja et al., 2011). Our

predetermined model assumed that observers would always inspect a predetermined number of items in a trial and choose the most desirable item from the inspected set. It predicted that on those trials where more items were inspected the desirability of the chosen item should be on average higher since there was then a greater chance of an observer inspecting a highly desirable item. This prediction was not found to be correct for any of our experiments. Conversely, our satisficing model and the decreasing threshold model did predict that the desirability of the chosen item should be less on those trials where there were more fixations. Both models were able to fit the data much better than the predetermined model.

Because we were interested in potential individual differences, we then fitted four nested versions of both the satisficing model and the decreasing threshold model to the individual observer data. In all three experiments, both sets of models could provide good fits to the data. In Experiment 1, the best fitting model overall was the decreasing threshold model that allowed both  $\alpha$  and  $\beta$  to vary as free parameters. In Experiment 2, the best fitting model was the satisficing model that allowed  $\alpha$  to vary as a free parameter but fixed  $\beta$  to zero. In Experiment 3, the best fitting model was the decreasing threshold model that fixed  $\alpha$  to one and  $\beta$  to zero.

#### Comparison to previous decision-making studies

The decreasing threshold model is not consistent with the findings of Caplin et al. (2011). In their study, observers were shown a list of options, with each option expressed as a series of additions and subtractions in dollar amounts. Observers were instructed to find the option corresponding to the highest dollar amount. Caplin et al. reported that their data was best described by a satisficing model where observers would continue searching until they found an option that exceeded a fixed threshold. They explicitly considered a model with a decreasing threshold and found that it did not fit their data as well. However, their task was somewhat different from ours. In our task, the options were arranged in an array and the observers searched the array randomly. Conversely, in their task the options were presented as a vertical list, thereby encouraging the observers to consider the options one at a time in a systematic fashion. In addition, while in our experiments the value of each option was explicitly shown, in their experiments the observer had to calculate the value of each option, a purposely mentally intensive process. For these two reasons, it is likely that the observers in the Caplin et al. study adopted a more deliberate strategy than the observers in our study.

Reutskaja et al. (2011) considered a similar situation to the one investigated here. In their study, observers were simultaneously presented with multiple product images. Using an eye tracker, they studied the order in which the images were viewed. They attempted to model search based on the subjective desirability of each product, which had been measured previously for each item using a liking-rating task. Like us, they considered three models: (i) an optimal search model with zero search costs, (ii) a satisficing search model, and (iii) a hybrid search model.

Unlike in our experiments, in their experiments the observers had to select a product under strict time constraints. Their optimal search model assumed that observers would fixate as many items within this time limit as possible and then chose the item that they had seen that had the greatest desirability (although they were not always successful in this regard). Their satisficing model assumed that search would continue until either the observer had fixated on an item whose desirability exceeded that of an internal threshold or until a fixed time had elapsed, at which point the observer would attempt to select the most desirable item among the fixated items. Their optimal search model and their satisficing search model were therefore very similar to our predetermined and satisficing models respectively. Their hybrid model represented a combination of both these models. The model assumed that at the end of each fixation the observer would decide to stop with probability  $p$ . This probability would increase with the elapsed time within the trial and desirability of the most desirable of the items fixated in that trial, what we would refer to as *max inspected*. They concluded that a hybrid model provided the best account of their data.

Their hybrid model makes a prediction that is incompatible with our data. They explicitly predict that in those situations where the observer had fixated all the items he would then immediately terminate the trial. In our experiments, this did not always happen. For example, in Experiment 3 the observers would view all four items on 68 % of the trials. Only in 67 % of these trials, i.e., those trials where all four items were viewed, would the observer immediately terminate the trial either by selecting the currently fixated item or, if that item was not the most desirable one, by refixating the most desirable item before terminating the trial.

Our satisficing and decreasing threshold models make at least two counter-intuitive predictions. The first is that observers will sometimes continue searching even after viewing all the items in the display. This occurred in all three experiments, but especially in the third. As implied above, for those trials where the observer

viewed all four items, 33 % of the time the observer would still not immediately terminate the trial. According to the satisficing model, the rationale for this is that in displays where none of the items exceed the internal threshold, the model is forced to keep searching until it reaches the predetermined maximum number of inspections allowed in that trial. Conversely, according to the decreasing threshold model the rationale for this behaviour is that in displays where none of the items happen to be particularly desirable, the observer needs to make a large number of fixations to allow the internal threshold to decrease to a point where one of the items can be accepted.

The second counter-intuitive prediction made by both models is that observers often terminate their search without inspecting all the items in the display. This should occur whenever the observer happens upon an item whose desirability exceeds the internal choice threshold. Again, this also was observed in all three experiments. As noted above, even in Experiment 3 with just four items in each display, in 32 % of the trials the observer terminated the trial without viewing all the items. Note that it would be problematic to attempt to attribute this finding to lack of motivation on the part of observers. If observers were unmotivated, it would be hard to understand why they would often keep searching after inspecting all the items in the display.

#### Relevance to visual search

The present work also has implications for research on visual search. Because the items in Experiment 2 were arranged in a grid, we expected the observers to view them in a systematic fashion, inspecting the items one row at a time, left to right, top to bottom, just as they would read a page of printed text. In fact, on examining the eye tracker traces for this experiment we could not find any trials where the observers actually did this. Indeed, we could not discern any systematic patterns. This finding is consistent with previous work showing the observers search displays in a quasi-random fashion. It seems that observers do this, because it is the quickest way to search a display. Planning saccades takes a non-negligible amount of processing time, making it much quicker to execute a saccade to a random object than to execute a saccade to a predetermined object. Consequently forcing the observers to search the display in a preordained fashion significantly reduces the search rate (Wolfe et al., 2000).

There is currently a debate about to what extent the seemingly random selection is biased to items that the observer has not recently fixated (Scinto et al., 1986;

Horowitz & Wolfe, 1998; Kristjansson, 2000; Peterson et al., 2001; Horowitz & Wolfe, 2003). It has been claimed that negative priming, often referred to as inhibition of return, causes observers to tend not to revisit an item that they fixated a few items earlier (Kristjansson, 2000; Peterson et al., 2001). In Experiment 2, we examined the probability that a given item would be fixated as a function of when it was last fixated. The results are shown in Fig. 5a. There is evidence that observers are attempting not to refixate previously fixated items. The refixation rates are below what one would expect if the items were being fixated at random, as denote by the dashed line.

There is a high degree of similarity between performance in our task, in which every item is a potential target, and performance in the target absent trials of a visual search task in which no item is a target. In both cases, the observer can only be sure of performing the task correctly by inspecting every item in the display. However, in neither task does this typically occur in practice. According to the multiple-decision model (Wolfe & Van Wert, 2010), every time the observer inspects an item that is revealed not to be the target, an internal variable is incremented. When this internal variable exceeds the quitting threshold, the observer terminates the search. Thus, in the multiple-decision model and in both our satisficing and decreasing threshold models, search is terminated when an internal threshold is exceeded.

## Conclusions

Because visual acuity is highest near the point of fixation, to evaluate an item an observer often needs to fixate it. To be realistic, decision theories need to acknowledge this and account for the influence of fixation patterns on decision making (Krajbich et al., 2010). In particular, they need to acknowledge that not all items can be considered simultaneously and an observer may not even consider all the items before making a decision, contrary to what many decision theories currently assume (Berger, 1985; Kahneman et al., 1992; Roe et al., 2001; Usher & McClelland, 2004; Padoa-Schioppa & Assad, 2006; Wallis, 2007; Rangel et al., 2008; Hotaling et al., 2010; Tsetsos et al., 2010; Usher et al., 2010). These previous theories were designed for situations in which there were only a small number of options to choose from. They all assume that the observers aim to find the most desirable option from those presented to them, an approach known as desirability maximisation. Contrary to this, our data show that observers often do not consider all the available options

and instead try to find an item whose desirability exceeds an internal threshold, a form of satisficing (Simon, 1956), albeit one where the internal threshold may decrease over the duration of the trial. Models that assume desirability maximisation are therefore unlikely to be accurate descriptions of human behaviour when there are a large number of choice alternatives. Curiously, even when there were just four alternatives, our data were still best described by a model that assumed an internal threshold. This suggests that desirability maximisation may be less widespread than previously assumed.

## Appendix 1. Model parameters

**Table 4** Estimated model parameters for the decreasing threshold model,  $\alpha$  and  $\beta$  allowed to vary as free parameters, for each observer in Experiment 1

Observer	T	$\Delta T$	$\alpha$	$\beta$
PH	103.1	0.499	0.883	0.168
RD	104.2	0.839	0.772	0.123
KO	103.4	0.730	0.555	0.243
MP	101.7	0.422	0.812	0.207
AB	105.6	0.387	0.927	0.098
JB	105.8	0.393	0.964	0.104
MW	105.6	0.445	0.839	0.094
TH	102.9	0.350	0.896	0.059

**Table 5** Estimated model parameters for the satisficing model,  $\alpha$  allowed to vary freely but  $\beta$  fixed at 0

Observer	T	M	$\alpha$
PH	91.0	16.3	0.711
RD	83.5	17.5	0.676
LM	98.2	15.0	0.844
CH	88.1	15.9	0.761

**Table 6** Model parameters for the decreasing threshold model with  $\alpha$  fixed at 1 and  $\beta$  fixed at 0

Observer	T	$\Delta T$
PH	102.0	8.34
RD	86.6	10.8
NT	96.0	7.52
MV	94.8	5.62



**Appendix 2. Mean square error (MSE) values**

**Table 7** Mean square error for each the 8 models for each observer in Experiment 1

	Sat $\alpha, \beta$	Sat $\sim\alpha, \beta$	Sat $\alpha, \sim\beta$	Sat $\sim\alpha, \sim\beta$	DT $\alpha, \beta$	DT $\sim\alpha, \beta$	DT $\alpha, \sim\beta$	DT $\sim\alpha, \sim\beta$
PH	2.43e-3	2.37e-3	5.82e-3	7.72e-03	4.97e-4	1.56e-03	5.27e-03	7.59e-03
RD	2.97e-3	1.17e-2	1.97e-3	1.50e-02	9.26e-4	5.92e-03	1.21e-02	3.08e-02
KO	4.78e-3	5.17e-2	5.18e-3	5.25e-02	2.68e-3	1.48e-02	2.00e-02	6.74e-02
MP	5.17e-3	5.92e-3	5.82e-3	4.68e-03	2.02e-3	4.10e-03	2.75e-03	7.10e-03
AB	1.34e-4	1.44e-4	7.25e-4	2.45e+02	8.62e-4	1.34e-03	1.08e-03	1.43e-03
JB	1.87e-4	1.73e-4	3.91e-4	4.18e-04	4.58e-4	6.45e-04	7.25e-04	6.77e-04
MW	1.06e-3	2.99e-3	8.01e-4	3.59e-03	1.41e-3	4.39e-03	2.99e-03	4.60e-03
TH	3.28e-3	4.93e-3	3.71e-3	3.04e-03	9.67e-4	1.97e-03	2.29e-03	7.59e-03

DT  $\alpha, \beta$ =decreasing threshold model,  $\alpha$  and  $\beta$  allowed to vary as free parameters; Sat $\sim\alpha, \sim\beta$ =satisficing model, with neither  $\alpha$  and  $\beta$  allowed to vary as free parameters

**Table 8** Mean square error for the each of the four models for each observer in Experiment 2

	Sat $\alpha, \beta$	Sat $\sim\alpha, \beta$	Sat $\alpha, \sim\beta$	Sat $\sim\alpha, \sim\beta$	DT $\alpha, \beta$	DT $\sim\alpha, \beta$	DT $\alpha, \sim\beta$	DT $\sim\alpha, \sim\beta$
PH	3.36e-3	1.51e-2	3.09e-3	1.49e-2	2.98e-3	8.18e-3	1.71e-2	2.72e-2
RD	8.79e-3	2.25e-2	4.71e-3	1.76e-2	5.12e-3	2.24e-2	1.47e-2	3.72e-2
LM	2.66e-3	1.43e-2	1.32e-3	5.30e-3	2.23e-3	5.90e-3	7.94e-3	6.80e-3
CH	4.30e-3	8.28e-3	2.68e-3	1.18e-2	2.80e-3	7.27e-3	2.80e-2	1.59e-2

DT  $\alpha, \beta$ =decreasing threshold model,  $\alpha$  and  $\beta$  allowed to vary as free parameters; Sat $\sim\alpha, \sim\beta$ =satisficing model, with neither  $\alpha$  and  $\beta$  allowed to vary as free parameters

**Table 9** Mean square error for each of the four models for each observer in Experiment 3

	Sat $\alpha, \beta$	Sat $\sim\alpha, \beta$	Sat $\alpha, \sim\beta$	Sat $\sim\alpha, \sim\beta$	DT $\alpha, \beta$	DT $\sim\alpha, \beta$	DT $\alpha, \sim\beta$	DT $\sim\alpha, \sim\beta$
PH	2.28e-2	2.31e-2	2.28e-2	2.28e-2	3.56e-3	7.24e-3	3.31e-3	2.73e-2
RD	2.27e-2	2.24e-2	2.14e-2	2.14e-2	1.21e-2	1.20e-2	1.31e-2	1.32e-2
NT	6.12e-2	7.00e-2	6.12e-2	6.03e-2	5.47e-3	5.48e-3	4.74e-3	5.71e-3
MV	1.31e-2	2.21e-2	1.05e-2	2.21e-2	4.28e-3	3.31e-2	4.19e-3	4.79e-3

DT  $\alpha, \beta$ =decreasing threshold model,  $\alpha$  and  $\beta$  allowed to vary as free parameters; Sat $\sim\alpha, \sim\beta$ =satisficing model, with neither  $\alpha$  and  $\beta$  allowed to vary as free parameters

**References**

Bearden, J. N., & Murphy R. O. (2007). On generalized secretary problem. *Uncertainty and Risk – Mental, Formal and Experimental Representations*. R. D. L. M. Abdellaoui, M. J. Machina and B. Munier New York, Springer.

Berger, J. O. (1985). Chapter 2: Utility and Loss. *Statistical Decision Theory and Bayesian Analysis*, Springer Series.

Botti, S., & Iyengar, S. S. (2006). The dark side of choice: When choice impairs social welfare. *Journal of Public Policy and Decision-Making*, 25(1), 24–38.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.

Brunner, P., Joshi, S., Briskin, S., Wolpaw, J. R., Bischof, H., & Schalk, G. (2010). Does the ‘P300’ speller depend on eye gaze? *Journal of Neural Engineering*, 7(5), 056013.

Buss, T. F. (1984). A unified approach to a class of best choice problems with an unknown number of options. *Annals of Probability*, 12(3), 882–891.

Camerer, C., Colin, F., Johnson, E. J., Rymon, T., & Sen, S. (Eds.) (1993). Cognition and framing in sequential bargaining for gains and losses. *Frontiers of Game Theory*.

- Caplin, A., Dean, M., & Martin, D. (2011). Search and Satisficing. *American Economic Review*, *101*(7), 2899–2922.
- Chandra, S. (2012). “Consumer Spending in U.S. Increases 0.8% as Incomes Climb.” Retrieved 24 December, 2012, 2012, from <http://www.bloomberg.com/news/2012-10-29/consumer-spending-in-u-s-increases-0-8-as-incomes-climb-0-4-.html>
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. discussion 114–185.
- Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected at once? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1003–1012.
- Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. (2006). Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model. *American Economic Review*, *96*(4), 1043–1068.
- Gilbert, J. P., & Mosteller, F. (1966). Recognizing the maximum of a sequence. *American Statistical Association Journal*, *61*, 35–73.
- Glockner, A., & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1055–1075.
- Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394*(6693), 575–577.
- Horowitz, T. S., & Wolfe, J. M. (2003). Memory for rejected distractors in visual search? *Visual Cognition*, *10*(3), 257–298.
- Hotaling, J. M., Busemeyer, J. R., & Li, J. (2010). Theoretical developments in decision field theory: comment on Tsetsos, Usher, and Chater (2010). *Psychological Review*, *117*(4), 1294–1298.
- Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*, 90–98.
- Johnson, E. J., Camerer, C., Sen, S., & Rymon, T. (2002). Detecting failures of backward induction: Monitoring information search in sequential bargaining. *Journal of Economic Theory*, *104*(1), 16–47.
- Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy*, *87*(5), 972–990.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219.
- Krajibich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.
- Kristjansson, A. (2000). In search of remembrance: evidence for memory in visual search. *Psychological Science*, *11*(4), 328–332.
- Lee, M. D. (2006). A hierarchical Bayesian model of human decision-making on an optimal stopping problem. *Cognitive Science*, *30*, 1–26.
- Lee, M. D., O’Conner, T. A., & Welsh, M. B. (2004). *Human decision making on the full-information secretary problem*. 26th Annual Conference of the Cognitive Science Society, Mahwah, NJ, Erlbaum.
- McCall, J. J. (1970). Economics of information and job search. *Quarterly Journal of Economics*, *84*(1), 113–126.
- Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223–226.
- Pedersini, R., Morvan, C., Maloney, L. T., Horowitz, T. S., & Wolfe, J. M. (2010). An abstract equivalent of visual search: Gain maximization fails in the absence of visual judgments. *Journal of Vision*, *10*(7), 1311.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.
- 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.
- Pieters, R. (2008). A review of eye-tracking research in marketing. *Review of Marketing Research*, *4*.
- Polyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197.
- Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125–157.
- Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *Economic Review*, *101*, 900–926.
- Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: a dynamic connectionist model of decision making. *Psychological Review*, *108*(2), 370–392.
- Scinto, L. F., Pillalamarri, R., & Karsh, R. (1986). Cognitive strategies for visual search. *Acta Psychologica*, *62*(3), 263–292.
- Seale, D. A., & Rapoport, A. (1997). Sequential decision making with relative ranks: An experimental investigation of the “Secretary Problem”. *Organizational Behavior and Human Decision Processes*, *69*, 221–236.
- Seale, D. A., & Rapoport, A. (2000). Optimal stopping behavior with relative ranks. *Journal of Behavioral Decision Making*, *13*, 391–411.
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*, 99–118.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138.
- Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*, 158–174.
- Sonnemans, J. (1998). Strategies of search. *Journal of Economic Behavior and Organization*, *35*, 309–332.
- Tsetsos, K., Usher, M., & Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological Review*, *117*(4), 1275–1293.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281–299.
- Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*(3), 757–769.
- Usher, M., Tsetsos, K., & Chater, N. (2010). Postscript: Contrasting Predictions for Preference Reversal. *Psychological Review*, *117*(4), 1291–1293.
- Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annual Review of Neuroscience*, *30*, 31–56.
- Westheimer, G. (1965). Visual Acuity. *Annual Review of Psychology*, *16*, 359–380.
- Wolfe, J. M., Alvarez, G. A., & Horowitz, T. S. (2000). Attention is fast but volition is slow. *Nature*, *406*(6797), 691.
- Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, *20*, 121–124.
- Zwick, R., Rapoport, A., Lo, A. K. C., & Muthukrishnan, A. V. (2003). Consumer sequential search: Not enough or too much? *Marketing Science*, *22*, 503–519.