

## Search Strategies in Decision Making: The Success of “Success”

BEN R. NEWELL<sup>1\*</sup>, TIM RAKOW<sup>2</sup>, NICOLA J. WESTON<sup>1</sup>  
and DAVID R. SHANKS<sup>1</sup>

<sup>1</sup>*University College London, UK*

<sup>2</sup>*University of Essex, Colchester, UK*

### ABSTRACT

Examination of search strategies has tended to focus on choices determined by decision makers' personal preferences among relevant cues, and not on learning cue-criterion relationships. We present an empirical and rational analysis of cue search for environments with objective criteria. In such environments, cues can be evaluated on the basis of three properties: validity (the probability that a cue identifies the correct choice if cue values differ between alternatives); discrimination rate (the proportion of occasions on which a cue has differing values); and success (the expected proportion of correct choices when only that cue can be used). Our experiments show that though there is a high degree of individual variability, success is a key determinant of search. Furthermore, a rational analysis demonstrates why success-directed search is the most adaptive strategy in many circumstances. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS search strategies; heuristics; cue hierarchies, adaptive behaviors

When faced with a decision it is often impossible to consider all of the options, their attributes and their potential consequences simultaneously, so we must do so sequentially. The order that we adopt to dictate this sequential search may have profound effects on our decisions and the consequences of those decisions (Hastie & Dawes, 2001). In this paper we present a new conceptualization of search in decision making in which search is dictated by the “success” of information in predicting the correct outcomes of decisions. Success is a function of the likelihood that a piece of information is usable and its accuracy in leading to a correct inference. We report empirical evidence showing that people's search patterns are allied to those determined by success, and a rational analysis that demonstrates the variety of circumstances in which success-directed search achieves the best outcomes.

Decision making can be considered to have three component processes: information acquisition; evaluation/action; and feedback/learning (Einhorn & Hogarth, 1981; Payne, 1982). Arguably, information acquisition has received the least attention of these three aspects. Indeed, Anderson (1990) notes that the literature

---

\*Correspondence to: Ben R. Newell, School of Psychology, The University of New South Wales, Sydney 2052, Australia.  
E-mail: ben.newell@unsw.edu.au

Contract/grant sponsors: Economic and Social Research Council, UK; The Leverhulme Trust, UK.

on decision making under uncertainty focuses mainly on one-step problems that obviate the need for information search.

The literature on information purchase (for an overview, see Connolly & Gilani, 1982; or, Connolly & Thorn, 1987) shows that people generally respond in the appropriate direction to changes in task characteristics such as the cost or diagnosticity of information. However, the magnitude of response is typically less than normative principles specify (Hershman & Levine, 1970; Fried & Peterson, 1969; Lanzetta & Kanareff, 1962; Pitz, 1968; Van Wallendael & Guignard, 1992). Relative to the prescriptions of normative principles (e.g., Edwards, 1965; Marschak, 1954; Stigler, 1961; Wendt, 1969), findings regarding the scale of acquisition are equivocal: overpurchase; underpurchase; and near optimal purchase are all observed (Hershman & Levine, 1970; Kaplan & Newman, 1966; Pruitt, 1961; Tversky & Edwards, 1966).

There has been some focus on sequential processing in decision making (e.g., Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Russo & Rosen, 1975), and this research has implications concerning the order in which we search through a fixed amount of information. There is clear evidence that the order in which information is presented (and processed) influences judgment (Hogarth & Einhorn, 1992). Likewise, choice can be influenced by the order in which alternatives are presented (Plott & Levine, 1978), and by whether information relevant to the alternatives is presented sequentially or simultaneously (Russo, 1977). In this paper we examine information acquisition in detail, considering situations in which people have control of *both* the amount of information they acquire *and* the sequence in which they search through that information.

Gigerenzer (2001) suggests that search can be conceived as the exploration of two dimensions: a search for alternatives or “the choice set”; and a search for cues on which to evaluate the alternatives. Our concern here is the search for cues in environments with an objective criterion in which cues are probabilistically related to this criterion.

## SEARCHING FOR CUES

One approach to decision making that examines search in such environments is the fast-and-frugal heuristics framework (e.g., Gigerenzer, Todd and The ABC Research Group, 1999). The view of Gigerenzer and colleagues is that humans possess an adaptive toolbox comprising a variety of heuristics, and building blocks to construct heuristics, that can be applied to a wide range of decision problems.

In our previous work we have questioned the “toolbox” approach, suggesting that much of the empirical evidence for the use of heuristics could just as easily, and more parsimoniously, be incorporated into a general evidence-accumulation model of decision making (e.g., Newell & Shanks, 2003; Newell, Weston & Shanks, 2003—see also Lee & Cummins, 2004), for a formal instantiation of such a model). Such a general model has the potential to replace the “ultimate” free parameter—i.e., a blank page on which new heuristics can be invented to accommodate each and every variation in behavior—with a single, empirically determined, evidence-threshold parameter. The important consideration for the current work is that both approaches require a mechanism for determining the sequential search through information. Thus our examination of potential search mechanisms is consistent with both a general evidence-accumulation model and the “toolbox” approach. We will not reiterate here the arguments why we believe a general model may be more appealing but rather focus on potential mechanisms for search required both by heuristics and a general model.

The most prominent heuristic in the adaptive toolbox—“take-the-best”—specifies a process for serial search through binary cues. It uses computed *cue validities* to order this search, where validity is defined as the probability that the cue will identify the correct alternative on a random selection of alternatives that differ on this cue. The “take-the-best” heuristic (TTB) comprises three basic building blocks: the search rule (search cues in order of descending validity); the stopping rule (stop after the first discriminating cue is discovered); and the decision rule (choose the outcome pointed to by the first cue that discriminates). In many of the situations in which TTB is thought to be applied people are assumed to have access to a reference class of

cues ranked according to their validities. TTB is a special case of a lexicographic rule (e.g., Fishburn, 1974) formulated for inference rather than preference, where search is directed by the objective validity of cues rather than a subjective measure of cue importance.

A simple example of an inferential search problem is answering an almanac question like, “Which has the larger population, Hamburg or Leipzig?” (cf. Gigerenzer & Goldstein, 1996). The reference class accessed to answer such a question is assumed to include cues such as, “Is the city the capital?”, “Does it have an airport/university/football team?”, etc. Assuming both cities are recognized, as soon as a cue is discovered that has different values for the two cities (e.g., Hamburg has a soccer team but Leipzig does not) this single cue is used to infer (correctly in this case) that Hamburg has the larger population.

It is worth noting here that TTB focuses on the use of binary rather than continuous cues, as indeed do many empirical analyses of multiple-cue judgment (e.g., Castellan, 1974; Castellan & Edgell, 1973) and categorization (e.g., Smith & Minda, 2000; Nosofsky, Palmeri, & McKinley, 1994) (which can be considered as formally equivalent—see Juslin, Olsson, & Olsson, 2003). Although this focus may seem unrepresentative of many “real-world” situations, it is common for cues to take a binary form even when they could be treated as continuous. For instance, a glance through any medical journal will reveal many analyses where a continuous patient variable, such as age, is treated as a binary cue (e.g., age > 50 years). This is reflected in how doctors are encouraged to make decisions with these cues (see Choussat et al., 1978, for an example), and are observed to do so (Abernathy & Hamm, 1994).

The psychological plausibility and simplicity of the validity-based search rule used by TTB has been championed (e.g., Martignon & Hoffrage, 1999). Gigerenzer, Hoffrage, and Kleinbölting (1991) proposed that people construct the hierarchy of cue validities in their reference class through the operation of some frequency-encoding mechanism. This monitoring of the frequencies of co-occurrences in the environment appears plausible given the research into the effectiveness of frequency estimation (e.g., Hasher & Zacks, 1984), but empirical tests remain scarce. In order to aid theoretical development, we and others (Bröder, 2000; Bröder, 2003; Jones, Juslin, Olsson, & Winman, 2000; Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003; Rieskamp & Hoffrage, 1999) have used experimental environments in which the explicit building blocks of the fast-and-frugal heuristics can be examined.

Existing empirical evidence suggests that people have great difficulty in learning a hierarchy of cue validities, although they can employ a validity-based search rule if told this hierarchy (Bröder 2003; Newell & Shanks, 2003; Newell et al., 2003; Rakow, Hinvest, Jackson, & Palmer, 2004). In this paper we focus on the way in which participants search through cues in an attempt to discover why participants might find it difficult to learn about validities, and to examine what other properties of cues they might also be sensitive to.

To illustrate the properties of cues consider the example mentioned above of determining which of two German cities has the larger population (e.g., Gigerenzer & Goldstein, 1996). Each of the pieces of information in memory (e.g., knowing if the city has a university, or knowing if it has an airport, or knowing if it has a football team, etc.) will have different validities. In addition to validity, a cue also has a discrimination rate—that is, the proportion of occasions on which the cue discriminates between the two (or more) alternatives. For example, a cue such as “Is the city the capital?” will have a high validity because the capital city has a very large population, but a low discrimination rate because only in the small proportion of comparisons involving the capital will knowing the answer to “Is the city the capital?” help in determining the population size. In contrast, knowing whether or not the city has a university will have a much higher discrimination rate (because there are many, but not all, cities with universities, but only one capital) but a lower validity because it is not always the case that universities are in the largest cities.

We argue that the overall usefulness of a cue is a function of both its discrimination rate and its validity. More useful cues are those that can frequently be used to make an inference, and, when used, usually point in the correct direction. We are not the first to assert this; Gigerenzer and Goldstein (1999) comment that, “A cue with a high ecological validity, however, is not very useful if its discrimination rate is small” (p. 85). However, in this study, we seek to develop this position further, and to consider people’s ability to search

through cues ordered according to both validity and discrimination. Note that the influence of discrimination rate on cue search has not been examined empirically before in the context of fast-and-frugal heuristics, because in previous studies the discrimination rate has been held constant across all cues (e.g., Bröder, 2000; Bröder, 2003; Newell & Shanks, 2003; Newell et al., 2003).

One previously described function of both discrimination rate and validity is the *success (rate)* of the cue (see Martignon & Hoffrage, 1999). The success of a cue is the expected proportion of correct inferences when *only that cue* can be used to make a choice. In the context of two-alternative forced choice (2AFC), this can be considered in two components. The first component is the expected proportion of correct inferences from choices on which the cue does discriminate. As a proportion of all choices in a set, this is the product of the cue's discrimination rate and its validity. The second component is the expected proportion of correct inferences from those choices on which the cue does not discriminate, forcing the decision maker to guess (because the cue has the same value for both objects). As a proportion of all choices in a set, this is the product of the complement of the discrimination rate and 0.5 (0.5 being the expected accuracy when randomly selecting one of the two alternatives). The success of the cue is hence the sum the of two component terms shown in Equation 1:

$$d^*v + (1 - d)^*0.5 \quad (1)$$

where  $d$  is the discrimination rate of the cue and  $v$  is the cue validity.  $d^*v$  is the expected proportion of correct inferences from occasions when the cue discriminates.  $(1 - d)^*0.5$  is the expected proportion of correct inferences from occasions when the cue does not discriminate, forcing a guess (with a 0.5 probability of a correct choice in a 2AFC task).

A highly successful cue is one with *both* a high discrimination rate *and* a high validity. Success is low when a cue has *both* a low discrimination rate *and* a low validity. Across all possible pairs of objects sampled from a large set of objects, validity ranges from 0.50 (non-predictive) to 1.0 (perfectly predictive), whereas discrimination rate for binary cues ranges from 0.0 (when all objects take the same cue value, either all positive or all negative) to 0.5 (when half of the objects take one cue value, and half the other). Consequently, for large sets of objects, the upper bound of success is at 0.75 and the lower bound is at 0.50. Intermediate levels of success may derive from many combinations of validity and discrimination rates: intermediate values for both cue parameters; or high validity combined with a low discrimination rate, or vice versa. For instance, three cues having the following combinations of cue parameters all have identical success rates (of 0.59): (1) validity of 0.68, discrimination rate of 0.5; (2) validity of 0.8, discrimination rate of 0.3; (3) validity of 1.0, discrimination rate of 0.18.

Using the earlier example of German cities, the success rate of the university cue (0.61) exceeds that of the national capital cue (0.51) (based on figures from Gigerenzer & Goldstein, 1999, p. 85). Despite the high validity of the national capital cue (1.0, indicating it *always* points in the correct direction), it has a very low discrimination rate (0.02, reflecting that, in almost every pair of cities, *neither* one is the capital). In contrast, the validity of the university cue is moderate (0.71), but it has a high discrimination rate (0.51, reflecting that, in about half of all possible pairs, one city will have a university and the other will not—because about half of the cities have a university).

Gigerenzer and Goldstein (1999) noted that discrimination and validity were negatively correlated in their German cities environment. Thus, more valid cues tended to have lower discrimination rates. This means that the first few cues looked at under a TTB strategy often fail to discriminate, necessitating the search for further cues. In such circumstances, TTB is less frugal (i.e., searches more information) than other one-reason algorithms such as Minimalist (which searches randomly through cues) or Take The Last (which starts with the last cue that discriminated, therefore increasing the likelihood of needing only one cue). Such a negative correlation between discrimination and validity is presumably not uncommon in real-world

environments, as the ABC Research Group found that the Minimalist strategy was more frugal than TTB for inferences made across 20 data sets (Czerlinski, Gigerenzer, & Goldstein, 1999).

If information search is costly in such an environment, then searching through cues in descending order of validity (as prescribed by TTB) may not be ideal. The discrimination rate of cues should have a significant bearing upon the order of cue search, as selecting cues with high discrimination rates tends to decrease the scale of information search. It is reasonable to assume that there is always some form of constraint on the amount of information that can or should be searched through. This follows if we agree that “information always costs something in terms of time, effort, or money” (Fried & Peterson, 1969, p. 525). Therefore, TTB, which searches following cue validity (with no regard to cue discrimination rates), is best adapted to environments where information is without cost. We doubt that such environments are common. Some (such as Fried and Peterson, above) would argue that they do not exist at all. In his rational analysis of memory, Anderson (1990) makes some similar points. He asserts that memory structures should be searched in (descending) order of the probability that they are relevant to the problem at hand. We take this to be equivalent to prescribing cue search in (“descending”) order of discrimination. However, Anderson’s discussion does not deal with the possibility that different pieces of relevant information may have differential value (“validity”). This is certainly the case in many situations of decision under uncertainty, where decisions are more likely to be correct when more valid cues are used. Therefore, if the accuracy of decisions or inferences is important, the validity (in addition to the discrimination rate) of cues should still have a bearing on the order of cue search.

Our hypothesis is that success is one of the determinants of cue search. This implies that people are sensitive to *both* the discrimination rate of cues (to reduce search) *and* the validity of cues (to enhance accuracy). To test this hypothesis we examine an environment where search is costly, accuracy is rewarded, and cues differ in both their discrimination rates and validities. To our knowledge, all previous tests of search behavior have been conducted in environments in which search by validity and success could not be distinguished because discrimination rates were held constant across cues, thus making validity and success orderings identical—see Equation 1 above (e.g., Bröder, 2003; Bröder, 2000, Newell & Shanks, 2003; Newell et al., 2003). The key test in the two experiments that follow is whether search behavior is more closely allied to the rank order predicted by success than to other plausible orders.

## EXPERIMENT 1

The purpose of Experiment 1 was to examine search strategies in a multiple-cue environment where participants’ choices could be assessed against an objective criterion. Our principal interest was to determine which of the three properties of the cues—*validity*, *discrimination rate (DR)*, or *success*—would drive participants’ information acquisition behavior. As can be seen from Table 1, the three properties predict three distinct acquisition orders. If participants buy information on the basis of *validity* we should observe an  $A > B > C > D$  order; in contrast, if *DR* drives behavior then the order would be  $C > D > B > A$ ; finally if *success* is most important then the information should be acquired in the order  $C > B > A > D$ .

Table 1. The validities, discrimination rates and success of Cues A–D and the search orders that they predict

	Cue properties				Predicted orders		
	A	B	C	D	V	DR	S
Validity	0.82	0.78	0.73	0.56	A > B > C > D		
Discrimination rate	0.22	0.37	0.50	0.47	C > D > B > A		
Success	0.57	0.60	0.62	0.53	C > B > A > D		

Note: V = Validity; DR = Discrimination Rate; S = Success.

*Participants*

Twenty members of the University College London community took part in the experiment. Fourteen were male and six were female. They had a mean age of 25.4 years (range 19–45,  $sd = 8.0$ ).

*Stimuli and design*

The experiment was programmed in Microsoft Visual Basic 6 and run on PCs. The experiment used a share profitability prediction task similar to the one used in a number of previous studies (Bröder, 2000; Newell & Shanks, 2003; Newell et al., 2003). Participants were presented with a series of two-alternative forced-choice decisions between the shares of two fictional companies (Share A and Share B). The shares were described by four binary cues with semantic labels concerning aspects of the company's financial status (see below). Each cue had an associated validity, discrimination rate and success rate.

Repeated choices were presented to participants in blocks of 64 trials. Each block was generated (independently) according to a fractional factorial design using SPSS ORTHOPLAN, to ensure that all four cues were independent (i.e., uncorrelated). Four binary cues were created such that 1/8, 1/4, 3/8 or 1/2 of the generated sets of cue-values were positive, resulting in cue discrimination rates of 0.22, 0.37, 0.47 and 0.50 respectively.

Table 1 displays the validities, discrimination rates and the success of the four cues A–D. The assignment of these properties to the four nominal (binary) cues was counterbalanced across participants, but the screen position of the cue labels was constant for all participants. (Note that this counterbalancing mitigates the possibility that participants' prior beliefs about the "usefulness" of particular cues systematically influenced search behavior. Each label (e.g., Share Trend Positive, see below) was assigned every combination of cue properties across participants.)

For each comparison there was an associated probability of Share A being the most profitable. After each choice the probability that the chosen share was most profitable was computed according to Bayes' Rule, assuming stochastic independence of the cues (details of the calculations can be found in Newell & Shanks, 2003). A random number generator then determined which share became most profitable according to this probability. Participants were motivated by the incentive of winning 3p (£0.03) for every correct prediction during the training trials and 6p (£0.06) for every correct prediction during test trials.

In the training phase cue information was supplied automatically and simultaneously to participants. At test, information was not provided automatically but participants were able to buy information about each cue in any order and were required to buy at least one cue. This feature was included to ensure that data on cue-search (our principal interest) was obtained for all participants. The cost of each piece of information was 1p (£0.01).

*Procedure* At the start of the experiment the participants were given a brief verbal description of the experiment telling them they would be making decisions about the profitability of company shares. Participants were then given full written instructions on the screen, which they read through with the experimenter. Participants were told that on each trial they would be asked to choose a share (one of two alternatives) that they thought would turn out to be the most profitable. To help them make this choice, participants were provided with four pieces of information about the two companies that the shares were from. These were:

1. Was the share trend positive over the last few months?
2. Does the company have financial reserves?
3. Does the company invest in new projects?
4. Is it an established company?

Participants were told that in making their predictions they should try to work out which pieces of information were most useful as not all the pieces were equally informative.

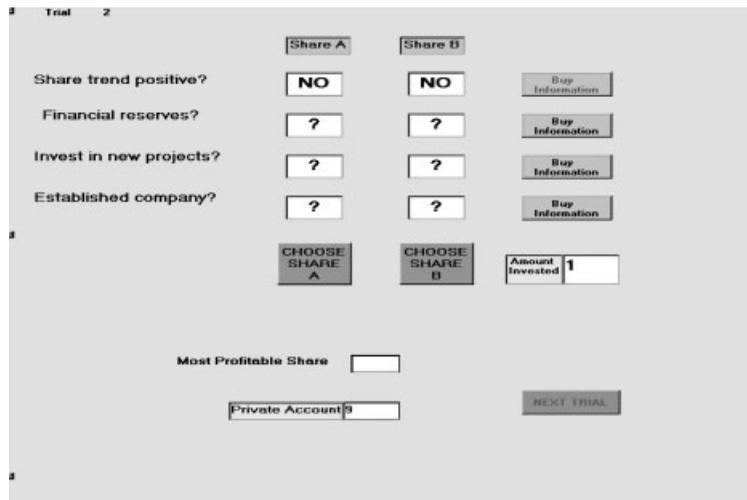


Figure 1. Screen layout during test trials in Experiments 1 and 2.

Participants then clicked a button to advance to the first trial. Participants worked through one 64-trial block as described above. The trials were presented in an order randomly selected for each participant. Participants were required to make a choice by clicking on the “CHOOSE SHARE A” or “CHOOSE SHARE B” buttons. On clicking either button, two windows below the choice buttons displayed the probability (expressed as a percentage) that the chosen share would be most profitable, and the most profitable share on that trial. If the “correct” share was chosen the Private Account window was incremented by 3p. On completion of the 64 training trials, a further set of instructions was displayed telling participants that for the next 128 trials (two more blocks of 64 trials) the information would no longer be automatically supplied to them but that instead they had to buy each piece. Participants were told that they could buy as many or as few pieces as they desired and in any order.

The screen layout for the test trials was slightly different to that of the training trials and is displayed in Figure 1. Clicking on the “BUY INFORMATION” buttons changed the contents of the pair of information windows from “?” to either “YES” or “NO”. Having bought at least one piece of information (minimum 1, maximum 4) participants made their choice. The most profitable share for that trial was then displayed in the window. If the “correct” share was chosen the private account was incremented by 6p minus any money that had been used to buy information. The 128 test trials were presented in a different random order for each participant.

On completion of the test trials, participants were presented with a new screen and asked to type in numbers between 0 and 100 that reflected the “usefulness” of each cue. The definition of usefulness was left intentionally rather open ended. Participants were simply told: “Give a score of 0 if you believe a piece of information was “no use whatsoever in making predictions”. Give a score of 100, if you believe that a piece of information was “as useful as any single piece of information could be” for this type of prediction task.” Participants were then thanked, debriefed and paid their participation fee plus their earnings from the experiment.

## RESULTS AND DISCUSSION

Our interest in Experiment 1 was to determine which property of the cues would drive participants’ search most strongly—*validity*, *discrimination rate (DR)*, or *success*. The three properties predict three

Table 2. Classification of search orders

	Search order classification				
	Success and DR	Validity	DR	Success	Incompatible
Experiment 1					
Mean percentage of trials	24.8	9.9	0.2	1.9	63.3
Percentage of participants	10.0	10.0	0.0	15.0	65.0
Experiment 2					
Mean percentage of trials	19.1	8.7	4.0	4.7	64.2
Percentage of participants	0.0	4.2	8.3	16.7	70.8

Note: DR = Discrimination rate.

distinct orders:  $A > B > C > D$ ,  $C > D > B > A$  and  $C > B > A > D$  for validity, discrimination rate and success, respectively (see Table 1).

To examine participants' search orders we classified each test trial for each participant as compatible/incompatible with the competing predicted orderings. If only cue C (most successful and most discriminating cue) was selected, this trial was classified as compatible with both success and DR. In addition, we classified each participant according to which search order he or she used most often.<sup>1</sup> The upper half of Table 2 summarizes these analyses.

The preponderance of incompatible classifications in these descriptive measures shows that only a minority of participants strictly followed any of the specified, deterministic orders. This rather disappointing finding highlights a potential problem with using the strict deterministic orders for classifying behavioral search patterns. The problem is that the measures displayed in Table 2 do not provide an indication of whether the "incompatible" orderings that were adopted were more akin to one specified order than to another. For example, although only three participants were classified as buying according to success, a further six participants purchased the most successful cue first but then deviated from the success ordering, resulting in being classified as "incompatible". Indeed if the analysis is restricted simply to the cue purchased *first* most often, then participants most commonly (45% of participants) bought the most successful/discriminating cue first and least commonly (15%) bought the most valid cue first. The remaining two cues were bought first most often by 20% of participants.

A further problem with using the strict deterministic classification of search patterns is the possible influence of the screen interface. As shown in Figure 1 the position of the "Buy Information" buttons on the screen, for which we counterbalanced, could result in deviations from the specified orders on individual trials and for individual participants. Consider a participant who decides always to buy the two most successful cues but because of the button position on the computer screen and the ease of using the mouse quite legitimately buys the second most successful followed by the most successful. Such a strategy would result in the participant being classified as "incompatible" despite his or her consistent use of the most successful cues.

To overcome the problems with these descriptive measures we considered the proportion of occasions—or *frequency*—with which each cue was used. We then aggregated this analysis across all participants in order to examine which of the three specified orders was most closely allied to participants' information acquisition behavior.

The first row of Section 3.1 of Table 3 displays the proportion of trials on which each cue was bought in the 128 test trials.<sup>2</sup> As can be seen from the mean values, the cue-purchase frequency follows the pattern

<sup>1</sup>In our previous work we have used an on-line ranking based on modal cue purchases to classify participants' search behavior. This measure is inappropriate for the current studies because three competing orders are considered (as opposed to one order—validity—in the previous studies) and no hint as to the correct ordering was provided (see Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003).

<sup>2</sup>Qualitative analysis of information purchase strategies across test trials revealed a high degree of consistency within subjects. We infer that learning was close to asymptote at the end of the 64 training trials—see General Discussion for further consideration of learning.

Table 3. Section 3.1: The proportion of trials on which each cue was purchased—cue purchase frequency. Section 3.2: Participants' ratings of the usefulness, validity, discrimination rate and success of Cues AD. Right-hand columns: the significance of Page trend test for search orders predicted by validity (*V*), discrimination rate (*DR*), and success (*S*)

3.1	Cue-purchase frequency				Predicted orders		
	A	B	C	D	<i>V</i>	<i>DR</i>	<i>S</i>
Experiment 1	0.43	0.46	0.67	0.40	—	—	*
Experiment 2 (first test block)	0.35	0.60	0.60	0.39	—	—	*
Experiment 2 (VE Group)	0.24	0.59	0.53	0.28	—	—	*
Experiment 2 (DE Group)	0.54	0.53	0.67	0.34	—	—	—
3.2	Ratings						
Usefulness (Expt 1)	0.23	0.26	0.31	0.19	—	—	*
Usefulness (Expt 2)	0.21	0.26	0.27	0.22	—	—	*
Validity	0.55	0.61	0.61	0.47	*	—	*
Discrimination rate	0.38	0.50	0.54	0.51	—	*	*
Success	2.50	2.17	2.17	3.17	*	—	*

Note: VE stands for validities equal. DE stands for discrimination rates equal. Usefulness ratings are normalized to account for individual differences in interpretation of the measure. Validity and discrimination rate ratings are mean raw scores (unnormalized). Success rankings were calculated from the order in which participants selected cues; a lower rank indicates a higher estimate of success. \*Indicates significance at the  $p < 0.05$  level (one tailed).

predicted by cues acquired on the basis of success, that is  $C > B > A > D$ . To examine this pattern further we used the nonparametric Page Trend test for ordered alternatives (Siegel & Castellan, 1988). The Page Test evaluates the hypothesis that the proportions are ordered in a specific sequence versus the null hypothesis that the proportions are the same. A nonparametric test was appropriate because the nature of our predictions was ordinal. The experimental environment allows us to test whether the proportions are ordered in each of the sequences predicted by success, DR or validity.

The Page Trend test indicated a significant trend ( $z_L = 2.63, p = 0.004$ )<sup>3</sup> across the four values in the order predicted by success ( $C > B > A > D$ ). Page Trend tests for the cue-purchase frequencies predicted by validity and DR were not significant ( $ps = 0.323$  and  $0.115$ , respectively). The significant Page test for the success ordering indicates that there is an overall trend but it does not tell us whether there are significant differences in proportions between pairs of cues. In order to test whether pairs of cues differed from each other, we used a multiple-comparisons procedure for ranked scores (Siegel & Castellan, 1988, p. 180). Planned comparisons revealed that only the difference between the proportion of trials on which cue C and cue D were bought was statistically significant ( $p < 0.05$ ). This result demonstrates that participants were good at distinguishing between the most and least successful cues.

The first row of Section 3.2 of Table 3 displays the mean normalized estimated usefulness ratings of each cue. These were calculated by summing the ratings given to each cue and dividing each rating by this total for each participant. Consistent with the behavioral measure, the order of the means reflects that predicted by ratings on the basis of the *success* of cues. A Page Trend test confirmed this predicted ordering, indicating a significant trend in the order  $C > B > A > D$  ( $z_L = 2.90, p = 0.001$ ). Page Trend tests for the usefulness rating orders predicted by validity and DR were not significant ( $ps = 0.367$ , and  $0.071$  respectively). Planned comparisons again revealed that only the difference between the estimated usefulness of cue C and cue D was statistically significant ( $p < 0.05$ ).

<sup>3</sup>We report the  $z_L$  value for the Page trend tests following the conversion formula for an  $N$  exceeding 12 (Siegel & Castellan, 1988). For consistency, in Experiment 2 we report both the standard  $L$  value and the  $z_L$  value for the analyses involving an  $N$  of 12.

This consistency between the cue-purchase frequency and the rating of cue usefulness was confirmed by individual participant analyses that revealed that, for 80% of participants, the most highly rated cue was the cue most frequently purchased. For 12 out of 20 participants (60%), their frequency of purchase corresponded exactly to their stated order of usefulness (6 out of 20 participants if one does not permit ties in either sequence).

The aim of Experiment 1 was to examine which property of a set of cues would determine participants' search through those cues. Although the descriptive measures of search order indicated considerable individual variability in search patterns, the evidence from the purchase-frequency measure indicates strongly that search is influenced by the success of cues.

## EXPERIMENT 2

Experiment 1 indicated that participants' cue-purchase behavior was more consistent with search according to the success of cues (which is a function of a cue's validity *and* discrimination rate) rather than to the validity alone. In addition to their actual behavior, participants' post-test ratings of the usefulness of the cues also reflected the order predicted by success.

As we noted earlier, "useful" is a somewhat open-ended term, and though the ratings data from Experiment 1 suggest that participants interpret "useful" in terms that correspond to success, the ratings do not allow us to determine whether participants are sensitive to the individual properties of a cue that contribute to its success. In Experiment 2 we investigated this question by asking participants not only to rate the order of usefulness of the four cues, but also the orders of validity, discrimination rate, and success.

Furthermore, we introduced an additional test block after the ratings phase to examine whether participants' cue-purchase behavior would be affected by a change in the task environment. For one group the cue validities were all made equal in the additional test block (all set to 0.72), making purchase dictated by discrimination rate the optimal strategy. A second group was instead given a test block in which all the discrimination rates were equal (all set to 0.50), making purchase in the order of validity the optimal strategy. Note that when discrimination rates are set equal, success order is now identical to validity order (see Equation 1, above).

Thus, Experiment 2 aimed to answer two questions: (1) would participants' ratings of the orders of validity, discrimination rate, and success reflect the objective orderings; and (2) would participants' cue-purchase patterns adapt in accordance with changes in the task environment? In addition to answering these questions, we aimed to replicate the success ordering of cue purchase for the first test phase.

## METHOD

### *Participants*

Twenty-four members of the University College London community took part in the experiment. Twelve were male and 12 were female. They had a mean age of 23.8 years (range 18–29,  $sd = 3.52$ ). Participants were assigned randomly to two groups of 12—the Validities Equal group (VE) and the Discrimination Rates Equal group (DE).

### *Stimuli, design and procedure*

The training phase and the first test phase were identical to those of Experiment 1. At the end of the first test block of 128 trials, participants were given relevant definitions and asked to provide ratings of the cues in terms of their usefulness, discrimination, validity, and success. Ratings for usefulness, discrimination, and validity were given by entering numbers between 0 and 100 into text boxes on the screen. Success ratings

were obtained by asking participants to imagine that they could only buy one cue before making their choice and that they should click on the one they most wanted to buy. Once they had clicked on this cue it disappeared and a dialogue box appeared telling participants that the cue they had picked was no longer available and that they should pick the one they next most wanted to buy from the remaining cues. This procedure continued until all four cues had disappeared. The cue picked first was given a rank of 1, the cue picked second a rank of 2, etc.

Once all the ratings had been obtained participants were given one of two final test blocks consisting of a further 64 trials. Participants in the VE group were told that they would be given more trials in which the general nature of the task was the same as on the previous trials (i.e., predicting the more profitable share) but that now the validities of all the cues were equal (they were not told the actual value). Participants in the DE group were given similar instructions but told that all the discrimination rates of the cues were now equal (they were not told the actual value). On completion of this final block, participants were paid their earnings and debriefed.

## RESULTS AND DISCUSSION

The lower half of Table 2 displays the descriptive measures of search order. Consistent with Experiment 1, most search behavior was classified as incompatible with any of the three specified orders. However, for those that could be classified, the success ordering was again the most common classification. Analysis restricted to the *first cue* bought most often was more suggestive of the influence of the successful cues. Ten participants (42%) purchased cue C first, ten (42%) purchased cue B first, three (12%) purchased cue A first and one (4%) purchased cue D first. Thus a full 84% of subjects consistently bought one of the two most successful cues first most often.

The second row of Section 3.1 of Table 3 displays the frequency with which each cue was purchased across the 128 test trials. As can be seen from the mean values, the cue-purchase frequency follows the pattern  $C = B > D > A$ . Although this pattern is not exactly the same as that predicted by success ( $C > B > A > D$ ), a Page test for ordered alternatives indicated a significant trend in the order predicted by cues acquired on the basis of success ( $z_L = 2.58, p = 0.005$ ), but not on the basis of validity or discrimination rate ( $ps = 0.433$ , and  $0.074$ , respectively). The finding that cue purchase frequency was related to the order predicted by success and not by the other two properties of the cues replicates the finding of Experiment 1. Planned comparisons between individual cues revealed that no differences between pairs were significant.

### Ratings

The last four rows of Section 3.2 of Table 3 display the ratings for all four measures of cue properties. The order observed for the usefulness rankings ( $C > B > D > A$ ), although not exactly the same as that predicted by estimates made on the basis of success ( $C > B > A > D$ ), was shown by a Page test to manifest a significant trend in the order predicted by success ( $z_L = 2.09, p = 0.018$ ), but not in the orders predicted by the *validity* or *discrimination rate* ( $ps = 0.230$ , and  $0.352$ , respectively). Planned comparisons revealed that none of the differences between individual pairs of ratings were significant. The finding that estimates of cue usefulness are related to the order predicted by success and not to the other two properties of the cues is a replication of Experiment 1.

The Page test analyses indicated that participants were differentially sensitive to the validity, discrimination rate and success of the cues. For each cue property there was a significant trend in the order predicted by that property. However, for two of the three properties (validity and discrimination rate) there was also a significant trend for the order predicted by success. This “overlap” in orderings suggests that participants found it difficult to dissociate success from validity and discrimination rate. (For brevity only statistical significance is reported for these tests—see Section 3.2 of Table 3.)

*Final test block*

We considered adaptation to the new environment by examining cue purchase frequency—a measure that would be sensitive to changes in search regardless of whether those changes occurred on all, or only some, trials. Cue-purchase frequency for the final test block is shown in the lower two rows of Section 3.1 of Table 3 for Groups VE and DE respectively.

When validities are equal the optimal cue search is one dictated by the discrimination rate order  $C > D > B > A$ . The observed cue purchase frequency, as indicated in the third row of Section 3.1 of Table 3 was  $B > C > D > A$ . Page tests for ordered alternatives indicated a significant trend in the order predicted by the initial (i.e., first test block) success of cues ( $L = 324$ ,  $z_L = 2.40$ ,  $p = 0.008$ ), a nonsignificant trend for the order predicted by discrimination rate ( $L = 315$ ,  $z_L = 1.50$ ,  $p = 0.067$ ) and no trend for the order predicted by initial validity ( $L = 294$ ,  $z_L = -0.06$ ,  $p = 0.725$ ) (critical  $L = 317$ ,  $p = 0.05$ ). (Note that *initial* orders were tested to examine whether participants stayed with strategies adopted in the first test block or adapted to the new orders.) Planned comparisons between individual cues revealed that only the difference between Cue A and Cue B was significant ( $p < 0.05$ , one-tailed), indicating that participants were able to distinguish well between the *perceived* most and least discriminating cues.

When discrimination rates are equal the optimal cue-purchase order is one dictated by the order of validities  $A > B > C > D$ . The observed order, as indicated in the fourth row of Section 3.1 of Table 3 was  $C > A > B > D$ . Page tests indicated a nonsignificant trend for the order predicted by the validity of cues ( $L = 315$ ,  $z_L = 1.50$ ,  $p = 0.067$ ) and no trend for the orders predicted by initial discrimination rate ( $L = 304$ ,  $z_L = 0.4$ ,  $p = 0.345$ ) or initial success ( $L = 307$ ,  $z_L = 0.7$ ,  $p = 0.242$ ) (critical  $L = 317$ ,  $p = 0.05$ ).

Perhaps the most diagnostic result from the final test block is the pattern of use for Cue A. Cue A was the most valid and least discriminating cue and these two properties appear to have had a differential influence on the two groups. For the DE group, in which the optimal strategy was to buy in order of validities, there was an increase of almost 0.20 in the proportion of trials on which Cue A (the most valid cue) was bought in comparison to the first test block. In contrast, in the VE group in which the optimal strategy was to buy in the order of discrimination rates, there was an 0.11 decrease in the amount of trials on which Cue A (the least discriminating cue) was purchased. The difference between the purchase frequency of Cue A in the VE group (0.24) and in the DE group (0.54) was significant ( $t(22) = 2.05$ ,  $p < 0.05$ , one tailed). This difference indicates that participants did adapt their cue-purchase behavior in accordance with changes in the environment.

## A RATIONAL ANALYSIS OF CHOICE STRATEGIES INVOLVING SEARCH

In the introduction, we argued that the order of cues in a sequential search should have regard to *both* the validity *and* discrimination rates of cues. In our experiments, we did find that the frequency with which people acquired cues matched the order of cue success more closely than it matched other orders. To investigate the rationality of such a search strategy, we present an analysis of one-reason choice strategies that search through cues in sequence.

We consider three strategies that differ in the order in which they search through cues. Take The Best (TTB) searches cues in descending order of validity. Draw The Discriminator (DTD) searches cues in descending order of discrimination rate. Select The Successful (STS) searches cues in descending order of success. These strategies stop search and make a choice on the basis of the first discriminating cue that is found (i.e., they are “one-reason” decision strategies). In addition, we consider “families” of strategies, derived from each of these three basic strategies, which stop and guess if no discriminating cue is found by a particular point. Strategies that stop after a maximum of one, two, or three cues, we term “Truncated #1”, “Truncated #2” and “Truncated #3” strategies respectively (cf., Martignon & Hoffrage, 1999). Such strategies may have value in situations where cue search is costly. In order to distinguish them from these truncated strategies, we refer to the strategy within each family that potentially searches all available cues (four cues in our environment) as “Full” strategies.

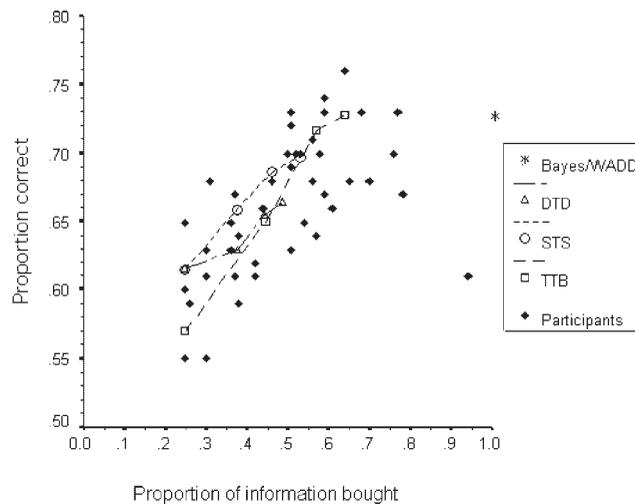


Figure 2. Proportion of information bought and proportion correct for participants and for three 'families' of decision strategy.

*Note:* For each family of strategies (joined by a line), the leftmost point represents the 'Truncated #1' strategy. Moving rightwards, the Truncated #2, #3, and Full strategies are shown. Values for all models are expected values computed from the 128 test trials, whereas those for participants are actual values observed from those same 128 test trials in Experiments 1 and 2.

For comparison, we also consider the performance of a weighted additive model (WADD) derived using multiple linear regression on the learning block of 64 trials (but applied to the test phase of 128 trials). The performance of every model we consider can be compared to that achieved if the option pointed to by the Bayesian calculation is selected.

Figure 2 shows the relationship between frugality and expected accuracy in our experimental environment for the TTb, DTD and STS families of strategies (open shapes). For a given level of frugality (proportion of information bought), Full-STs strategies are the most accurate. The Full-TTb strategy can achieve the highest level of accuracy (0.728), but only slightly more so than the Full-STs strategy (0.697), which is considerably more frugal. In order to beat the accuracy of Full-STs by 4.3%, Full-TTb uses 20.7% more information (both figures in relative terms). This effort/accuracy trade-off means that the two strategies would only give equivalent expected earnings were we to raise our reward for a correct choice from 6 to 14.4 times the cost of each cue. Full-TTb selects the option pointed to by the Bayesian calculation on 95% of choices, and is consequently only slightly less accurate. WADD *always* selects the option pointed to by the Bayesian calculation, and therefore matches its accuracy (0.736).

Six participants matched or beat the expected accuracy of the Full-TTb strategy, three of these did so with more frugal strategies.<sup>4</sup> Eight participants purchased more information than would be dictated by the Full-TTb stopping rule (i.e., > 63% of available information). With one (higher performing) exception, these participants had much the same levels of accuracy as participants with more frugal strategies (e.g., purchasing 50 to 60% of information). Simulations have demonstrated that compensatory strategies such as WADD only just outperform the more frugal (Full-) TTb strategy in data-fitting, on average (Czerlinski, Gigerenzer, & Goldstein, 1999). The pattern of performance here concurs with this finding. Levels of cue purchase sufficiently high to support some compensatory decision-making did not lead to levels of accuracy in excess of those achieved by participants with purchase patterns broadly consistent with a more frugal

<sup>4</sup>Due to the probabilistic nature of the task it is possible for individual participants to perform slightly above (or below) the level of accuracy that is expected given their choices.

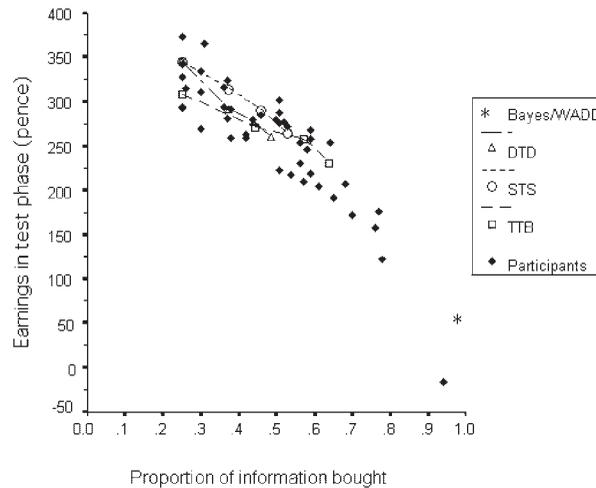


Figure 3. Proportion of information bought and earnings for participants and for all models

Note: See note accompanying Figure 2 for explanation.

one-reason decision strategy. Why this should be so, can be seen from the relatively similar levels of expected performance for the one-reason and compensatory strategies.

Figure 3 shows the relationship between the proportion of information purchased and the expected earnings for the three families of strategies. This scatter-plot illustrates the value of the STS family of strategies in environments (such as this) where information search is relatively expensive. In terms of earnings, Full-TTB is a relatively poor strategy in this environment, as its higher level of information purchase reduces expected earnings at a faster rate than its enhanced accuracy increases earnings. Seventy percent of participants earned more than would have been expected had they followed a Full-TTB strategy. They achieved this by adopting strategies that were more frugal than Full-TTB. In this environment, strategies of buying at most one or two items of information, or buying randomly until a cue discriminates, all lead to greater frugality and hence higher earnings than TTB. Consequently, observing that many people earn more than TTB would be uninteresting were it not for the fact that people do *not* buy randomly, but are influenced by the success of cues.

As was the case with purchase orders, there was considerable variation between individuals in the *amount* of information purchased. In general, our participants could be described as “frugal.” Eighty-two percent of them purchased less of the available information than Full-TTB would prescribe, and 59% of them purchased fewer cues than Full-STs would. However, for the “harsh” environment our participants encountered even this limited information search represented over-purchasing of (expensive) cues (to their financial detriment). This is consistent with prior research, where over-purchase is often observed when information is expensive though under-purchase occurs when it is cheap (Connolly & Gilani, 1982; Connolly & Serre, 1984). We cannot rule out the possibility that non-financial considerations influenced information search (and perhaps differentially for our participants). If participants value correct responses over and above earnings, or perceive their task, in part, as gaining knowledge or reducing uncertainty (see Lindley, 1956; Oaksford & Chater, 1998) it may not be quite so clear-cut that what we observe here is “gross” over purchase.

## GENERAL DISCUSSION

In the introduction we argued that search through cues should be sensitive to both the validity and the discrimination rate of those cues. We contended that success-directed search would often be adaptive because

validity and discrimination rate are frequently negatively correlated. To test the success of “success” we investigated search strategies when search was costly, decision outcomes could be objectively determined, and people had to learn an environment in which they had no prior hierarchy of the values of different cues. We set up a straightforward competition between three potential search orders: validity; discrimination rate; and success. Success was the “winner” in both of our experiments.

It is fair to say that a reader looking at Table 2 with its high proportion of “incompatible” classifications might find the conclusion that success was the winner rather unconvincing. Clearly, determining the winner of the competition is dependent on the method used to classify the responses. When the behavior of each individual on each trial is examined and compared with the strict deterministic orders, the ability of *any* of the proposed rules to capture the strategies used is rather poor. In contrast, a *frequency of purchase* measure, that aggregates across participants and takes into account possible idiosyncrasies resulting from the screen layout, or the possibility of errors in making selections, reveals search patterns much more akin to one order—success—than to the other two candidate orders.

However, even if one accepts that the frequency data provide good evidence for the success of success (as we do), it is still important to consider possible reasons for the failure of the strict deterministic orders to capture individual participants’ behavior. A good starting point is to ask whether we need better theories of how people search in these environments. The three search rules that we considered are deterministic in that they predict the same order of cue selection on *every* trial. In contrast to such deterministic rules, Tversky (1972) proposed a *probabilistic* search rule for the elimination-by-aspects model of choice (a model that is closely related to the lexicographic strategies under consideration here). In that model, cues (or “aspects”) are selected with a probability proportional to their weight or importance for the particular choice at a particular moment in time. As Tversky (1972) stated, “the probabilities merely reflect the fact that at different moments in time different states of mind (leading to different choices) may prevail” (p. 296).

It is possible that the momentary salience of a particular cue influenced search order in our experimental environment. For example, on some trials a cue might be selected because it predicted the correct outcome on the preceding trial (i.e., the search rule used in the “take-the-last” heuristic, Gigerenzer et al., 1999). The influence of such probabilistic rules may indeed give rise to the patterns of data we observed—especially if, as is likely, the momentary salience of particular cues was different for individual participants. Support for such an interpretation comes from Albert, Aschenbrenner, and Schmalhofer (1989) who proposed that “probabilistically”, important cues are processed earlier than less important ones, but that reading habits, accessibility (recency), salience and availability can also play a role.

An additional explanation of why participants’ search frequencies did not uniformly reflect cue success (or validity, or DR) could simply be that their experience was not uniform. Outcomes were determined probabilistically (in keeping with the nature of the environment) and, therefore, the cue orders according to success or validity that individual participants observed would sometimes differ (as a result of sampling variability) from those that were programmed. Note that sampling variability of this kind affects cue rankings by validity *and* success, and therefore, does *not* provide an explanation of why overall search frequencies reflected success more so than they did validity. Clearly, observed validity and success are far more likely to vary from their programmed values in the case of individuals (observing 64 learning trials) than across the experiment as a whole (with  $20 \times 64$ , or  $24 \times 64$  learning trials). For this reason, we are cautious of reading too much into the variability in search behavior revealed by our individual analysis shown in Table 2. That said, the possibility that a portion of this variability might reflect stable individual differences (e.g., preferences for speed over accuracy, or vice versa) is worthy of future investigation.

A related consideration is that the “definition” or “precision” with which validities and success rates can be learned is a function of the cues’ DRs. By examining Table 1, it can be seen that the most extreme example is for Cue A because it has the lowest DR. Given its DR of 0.22 there are only 14 discriminating trials for Cue A within a block of 64 trials ( $64 \times 0.22 = 14.08$ ). Its programmed validity is 0.82, but the closest validity to this that can be observed is 0.786 (11/14 choices correct) or 0.857 (12/14 correct). When cues have low

discrimination rates, quite large numbers of trials may be necessary before validities observed in samples of choice pairs can fall close to the corresponding validities for the population of all possible choice pairs. For instance, during 192 learning trials, (three times the sample used in the current experiments) Cue A would discriminate on 42 trials ( $192 \times 0.22 = 42.24$ ) allowing validities of 0.810 (34/42 discriminating trials) or 0.833 (35/42) to be observed for Cue A. Observed (sample) validities are most variable when DRs are low (yielding small numbers of discriminating trials). Readers may therefore wonder whether the chances of observing cue orders that are discordant from those programmed are greater for validity (which is computed from a subset of trials) than for success (which is computed from all trials). Using a binomial approximation to estimate the overlap of cue parameter distributions for each cue suggests that this is not the case. The lower variability in observed success is counteracted by the lower dispersion in success rates (as shown in Table 1 these rates differ by only 0.02, 0.03, and 0.04 compared with much greater dispersion for cue validities)—making the task of discriminating cues by success no easier than the task of discriminating cues according to their validity.

Recent evidence from our laboratory sheds some light on the issue of precision in learning cue properties. Rakow, Newell, Fayers and Hersby, (2003) employed an equivalent methodology to the current studies, with similar cue properties, but extended training. Analysis of the final test session (which followed 256 learning trials spaced across four sessions) indicated that only 15% of trials were incompatible with any of the search rules, while 43%, 33% and 9% were compatible with validity, success and discrimination rate respectively. Analysis of the dominant strategies among participants (i.e., those used on the majority of trials) revealed five (out of twelve) adopting validity, four adopting success, one adopting discrimination rate and two using “incompatible” strategies. The picture that emerges from the descriptive measures is that both success *and* validity vie for the dominant strategy, but with near equal numbers of participants opting for each one. Analysis of the frequency of purchase also suggested the persistence of individual variability, but, consistent with the current experiments, a strong influence of successful cues. For example, even in the final testing session there were five distinct rank orders of frequency with which the twelve participants purchased cues. However, for every testing session the most successful cue was the most frequently bought cue, and, aggregating across testing sessions, the frequency of cue purchase was much more akin to that prescribed by success than by validity. Furthermore, there was no evidence of dependable relative changes in cues use with increasing practice (nonsignificant cue by session interaction, ( $F(9, 99) = 0.73, p = 0.677$ )). This evidence suggests that individual differences in cue purchase persist even with extensive training.

However, one key aspect of the design—specifically the cost of information—makes detailed comparisons between the data sets problematic. Simply put, when information was very expensive—as it was in one of the conditions of the experiment in Rakow et al. (2003)—search was curtailed and therefore participants’ strategies were more likely to be compatible with one of the search orders. This in turn reduced the over-all proportion of incompatible classifications. Hence there are interactions between length of training and cost of information, which make it difficult to draw strong conclusions about the isolated effect of increased numbers of trials.

Given these considerations of sampling variability and the limits on precision, it is in many ways all the more remarkable that the present experiments demonstrated the appropriate learning of validities and success rates. It is likely that participants benefited from the summary feedback (the probability that they chose the more profitable share), which was not subject to such variability or imprecision. Nonetheless, we feel that our participants’ performance in the far from trivial task of learning validity, DR, and success can be seen as an example of successful intuitive statistics on their part. Here we consider two classes of explanation for *how* people are able to achieve such success.

#### *Learning the cue hierarchy*

Gigerenzer et al. (1991) proposed that people can learn validities by logging the frequency of correct responses within a given reference class of events. This ability is a central feature of the Probabilistic Mental

Model (PMM) of probability judgment (Gigerenzer et al., 1991), and of TTB (Gigerenzer & Goldstein, 1999). However, in order to determine the validity of a cue, the required reference class of events is the subset of occasions on which the cue discriminates. For every cue, this is a different subset of events. In environments where cues have differing discrimination rates, each reference class (for each cue) will be of a different size. This strikes us as a difficult task—to construct separate reference classes, of different size, and establish a hierarchy of validity based on differing frequencies of correct inferences. This difficulty is most acute in cases where cue discrimination rates and validities are inversely related. In such cases, accurate learning of validities requires that people assess a cue that leads to a smaller number of correct inferences as more valid than one leading to a greater number of correct inferences. In other words, establishing a correct hierarchy of validities requires learning a series of *relative frequencies*, it cannot be done by logging frequencies alone. Participants' difficulty in dissociating success and validity in the second experiment may reflect precisely these problems.

In contrast, learning the success rates of cues strikes us as a far more robust process. The reference class is the same for all cues, and is the full set of observed events rather than a subset of events. The proportion of correct inferences when only that cue is used (i.e., success) is, in principle, directly observable. As all cues have the same reference class, logging frequencies is sufficient to establish a hierarchy of cues by success.

Further to this, we noted that, for the environment we considered, the order of the unstandardized regression weights for the WADD model derived using MLR followed the order of validity, whereas the order of the standardized regression weights ( $\beta$ s) followed that of success. It is known that an interesting class of connectionist networks—single layer networks employing the delta rule—asymptotically compute weights equivalent to *unstandardized* regression weights (Stone, 1986). This suggests that such networks, which are known to capture many important aspects of human learning (McClelland & Rumelhart, 1985), do so by learning about cue validity rather than success. Our results and rational analysis suggest that learning about success would be more adaptive. The success ordering of the  $\beta$ s might also be taken to imply further support for the notion that the informativeness of cues is better reflected by success than by validity alone.

#### *Strategic advantages of success-based search*

There is an eminent tradition within psychology proposing that behavior should be understood as adaptation to the environment, and therefore modeling behavior requires understanding the environment (Tolman & Brunswik, 1935; Simon, 1956; Gibson, 1979; Anderson, 1990; Gigerenzer & Todd, 1999). Guided by this approach, we consider under what conditions success-based search is better than validity-based search, by examining two one-reason strategies that adopt these orders (STS and TTB).

The TTB strategy is more accurate than STS in our environment, and we suspect that this would be the case in all environments. However, the advantage is small (3% in absolute terms), and similar to the advantage that Martignon and Hoffrage (1999) report for the German cities environment (2%). Nonetheless, if accuracy is the sole consideration, there is no doubt that TTB is the better of the two strategies. However, there are many situations where aspects other than accuracy come into play.

We have illustrated that, when there is some cost to search, the more frugal the strategy the better. STS is more frugal than TTB (because discrimination rates play some part in the selection of cues), yet is so without severely damaging accuracy (because success partially reflects validity). If we concur that there is always some cost to search (Fried & Peterson, 1969), it follows that situations where information is sufficiently cheap to favor TTB are relatively rare. In addition, if time and information are sufficiently inexpensive to favor the less frugal TTB, it may be more appropriate to use a compensatory strategy (which is even less frugal). Czerlinski, Gigerenzer, Goldstein (1999) report that the advantage of WADD over TTB in fitting data is small (approximately 2% higher accuracy over 20 environments). Though small, this advantage is similar to the accuracy advantage we noted for TTB over STS. Thus, it is precisely those situations where low information costs favor TTB over STS, where an information-heavy compensatory strategy may be

preferable to TTB. Of course, with a compensatory strategy one also needs to consider the additional cost or effort of computing the solution (Payne, Bettman, & Johnson, 1993). Notwithstanding, when the cost of search is taken into account, it is unclear whether there are many environments where TTB is preferable to the more frugal (though less accurate) STS, and preferable to the typically more accurate (though more effortful) compensatory strategies. Indeed, empirical evidence demonstrates that people often depart from the TTB stopping rule by purchasing additional cues when information costs are low (Bröder, 2000; Newell & Shanks, 2003).

In addition to the cost of information search, there may also be constraints upon the length of search. Time pressure may be such that there is a maximum length for any one search, or for the average length of search. Competitive choice situations may be such that, the longer you search the more likely it is that one choice option will disappear (because someone else has chosen it first). In these situations, search can be truncated before a discriminating cue is found. This acts as a disincentive for strategies with greater search lengths: either the maximum search length is reached, forcing a guess; or the choice set is smaller by the time a decision is made. Our rational analysis shows why STS is preferable to TTB in such situations. If lower levels of frugality are required, the STS family of strategies is more accurate. Only when it is permissible to undertake a lengthier search is the accuracy of TTB greater. STS is more likely to reach a quick decision, and therefore less likely to lose a race to select an option.<sup>5</sup>

A corollary follows in that we may not always know exactly how long we have to make a decision. This uncertainty about the decision situation also works against TTB. We can be confident that TTB will be a more accurate strategy than STS *only* if we know that time permits an exhaustive (or near exhaustive) search through the available information. Otherwise, the possibility of truncating search (due to some external constraint) risks lowering the accuracy of TTB (due to increased guessing.)

#### *The ecological rationality of success-based search*

Gigerenzer and colleagues suggest that heuristics develop to exploit the structure of environments, and have demonstrated that there is a class of environments where a one-reason heuristic that searches according to cue validity (TTB) is surprisingly accurate (Czerlinski et al., 1999). However, our analyses suggest that when the costs of search are considered, success-based search is likely to exploit these environments more effectively. In addition, we argue that learning a success-based cue hierarchy is more robust than learning a hierarchy of validities. Thus, it is hard for us to see how, and why, people might develop and use validity-based search strategies when both validities and discrimination rates vary. Importantly, the behavior of participants in our experiments accords with this reasoning.

We acknowledge that an empirical question remains as to whether the costs of search in external environments as we have characterized them (i.e., explicit monetary costs for uncovering information on a computer screen) are synonymous with the “cognitive costs” of accessing information in memory. Validity-based search may still prove to be a better description of what occurs in memory. We hope that new techniques for investigating the use of heuristics in memory-based tasks (cf., Bröder & Schiffer, 2003a; 2003b) will help to answer this question.

Our focus in the rational analysis was on the use of one-reason heuristics in order to demonstrate that the validity search rule employed by one member of the “adaptive toolbox” (TTB) could be bettered by a different search rule (success) in our environment. However, our principal finding regarding the importance of success for determining the sequential search through information is equally applicable to a general evidence

---

<sup>5</sup>We note that, in general, TTB can appropriately be described as a frugal strategy. In their analysis of 20 environments, Czerlinski, Gigerenzer, & Goldstein (1999) found that TTB used 31% of available information. The point here is that STS is more frugal than TTB, and this does confer benefits.

accumulation model of forced-choice decision making, of the kind alluded to in the introduction and discussed in more detail in Lee and Cummins (2004) and in our previous work (Newell & Shanks, 2003; Newell et al., 2003). Such a model requires a mechanism for search and we argue that using the success of cues provides a good candidate.

#### ACKNOWLEDGEMENTS

The first and second authors contributed equally to this project and order of authorship was determined arbitrarily. The support of the Economic and Social Research Council (ESRC) and The Leverhulme Trust is gratefully acknowledged. The work was part of the programme of the ESRC Centre for Economic Learning and Social Evolution, University College London.

#### REFERENCES

- Abernathy, C. M., & Hamm, R. M. (1994). *Surgical scripts: Master surgeons think aloud about 43 common surgical problems*. Philadelphia: Hanley & Belfus.
- Albert, D., Aschenbrenner, K. M., & Schmalhofer, F. (1989). Cognitive choice processes and the attitude-behavior relation. In A. Upmeyer (Ed.), *Attitudes and behavioral decisions*. New York: Springer.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bröder, A. (2000). Assessing the empirical validity of the “Take-The-Best” heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1332–1346.
- Bröder, A. (2003). Decision making with the adaptive toolbox: Influence of environmental structure, personality, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 611–625.
- Bröder, A., & Schiffer, S. (2003a). Bayesian strategy assessment in multi-attribute decision-making. *Journal of Behavioral Decision Making*, 16, 193–213.
- Bröder, A., & Schiffer, S. (2003b). “Take-the-Best” versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, 132, 277–293.
- Castellan, N. J. (1974). The effect of different types of feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance*, 11, 44–64.
- Castellan, N. J., & Edgell, S. E. (1973). An hypothesis generation model for judgment in nonmetric multiple-cue probability learning. *Journal of Mathematical Psychology*, 10, 204–222.
- Choussat, A., Fontan, F., Besse, P., Vollot, F., Chauve, A., & Bricaud, H. (1978). Selection criteria for Fontan’s procedure. In R. H. Anderson, & E. Shinebourne (Eds.), *Paediatric Cardiology* (pp. 559–566). Edinburgh: Churchill Livingstone.
- Connolly, T., & Gilani, N. (1982). Information search in judgment tasks: A regression model and some preliminary findings. *Organizational Behavior and Human Decision Processes*, 30, 330–350.
- Connolly, T., & Serre, P. (1984). Information search in judgment tasks: The effects of unequal validity and cost. *Organizational Behavior and Human Performance*, 34, 387–401.
- Connolly, T., & Thorn, B. K. (1987). Predecisional information acquisition: Effects of task variables on suboptimal search strategies. *Organizational Behavior and Human Decision Processes*, 39, 397–416.
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.
- Edwards, W. (1965). Optimal strategies for seeking information. *Journal of Mathematical Psychology*, 2, 312–329.
- Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, 32, 53–88.
- Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, 86, 465–485.
- Fishburn, P. C. (1974). Lexicographic orders, utilities and decision rules. *Management Science*, 20, 1442–1471.
- Fried, L. S., & Peterson, C. R. (1969). Information seeking: Optional versus fixed stopping. *Journal of Experimental Psychology*, 80, 525–529.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.
- Gigerenzer, G. (2001). The adaptive toolbox. In G. Gigerenzer, & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 37–51). Cambridge, Mass: MIT Press.

- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.
- Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 75–97). Oxford: Oxford University Press.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.
- Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple Heuristics that Make us Smart* (pp. 3–34). Oxford: Oxford University Press.
- Gigerenzer, G., Todd, P. M., & The ABC Research Group (Eds.) (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.
- Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information. *American Psychologist*, *39*, 1327–1388.
- Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world*. Thousand Oaks, CA: Sage.
- Hershman, R. L., & Levine, J. R. (1970). Deviations from optimal purchase strategies in human decision making. *Organizational Behavior and Human Performance*, *5*, 313–329.
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief adjustment model. *Cognitive Psychology*, *24*, 1–55.
- Jones, S., Juslin, P., Olsson, H., & Winman, A. (2000). Algorithm, heuristic or exemplar: Process and representation in multiple-cue judgment. In L. R. Gleitman, & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 244–249). Hillsdale, NJ: Lawrence Erlbaum.
- Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*, 133–156.
- Kaplan, R. J., & Newman, J. R. (1966). Studies in probabilistic information processing. *IEEE Transactions on Human Factors in Electronics, HFE-7*, 49–63.
- Lanzetta, J. T., & Kanareff, V. T. (1962). Information cost, amount of payoff, and level of aspiration as determinants of information seeking in decision making. *Behavioral Science*, *7*, 459–473.
- Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying “take the best” and “rational” models. *Psychonomic Bulletin & Review*, in press.
- Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, *27*, 986–1005.
- Marschak, J. (1954). Towards an economic theory of organization and information. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes* (pp. 187–220). New York: Wiley.
- Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). New York: Oxford University Press.
- McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–188.
- Newell, B. R., & Shanks, D. R. (2003). Take-the-best or look at the rest? Factors influencing “one-reason” decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 53–65.
- Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast and frugal heuristic: not everyone “takes-the-best”. *Organizational Behavior and Human Decision Processes*, *91*, 82–96.
- Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule plus exception model of classification learning. *Psychological Review*, *101*, 53–79.
- Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world: Essays on the cognitive science of human reasoning*. Hove, UK: Psychology Press.
- Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin*, *92*, 382–402.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, UK: Cambridge University Press.
- Pitz, G. F. (1968). Information seeking when available information is limited. *Journal of Experimental Psychology*, *76*, 25–34.
- Plott, C. R., & Levine, M. E. (1978). A model of agenda influence on committee decision. *American Economic Review*, *68*, 146–160.
- Pruitt, D. G. (1961). Informational requirements in decision making. *American Journal of Psychology*, *74*, 433–439.
- Rakow, T., Hinvest, N., Jackson, E., & Palmer, M. (2004). Simple heuristics from the adaptive toolbox: Can we perform the requisite learning? *Thinking & Reasoning*, in press.
- Rakow, T., Newell, B. R., Fayers, K., & Hersby, M. (2003). *Information search with variable cost, time pressure and uncertainty*. Manuscript in preparation.

- Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics and how can we tell? In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 141–167). Oxford: Oxford University Press.
- Russo, J. E., & Rosen, L. D. (1975). An eye movement analysis of multi-alternative choice. *Memory and Cognition*, 3, 267–276.
- Russo, J. E. (1977). The value of unit price information. *Journal of Marketing Research*, 14, 193–201.
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review*, 63, 129–138.
- Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 3–27.
- Stigler, G. J. (1961). The economics of information. *The Journal of Political Economy*, 69(3), 213–225.
- Stone, G. O. (1986). An analysis of the delta rule and the learning of statistical associations. In D. E. Rumelhart, J. L. McClelland, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations of the microstructure of cognition*. (vol. 1: Foundations, pp. 444–459). Cambridge: MIT Press.
- Tolman, E. C., & Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological Review*, 42, 43–77.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281–299.
- Tversky, A., & Edwards, W. (1966). Information versus reward in binary choices. *Journal of Experimental Psychology*, 71, 680–683.
- Van Wallendael, L. R., & Guignard, Y. (1992). Diagnosticity, confidence, and the need for information. *Journal of Behavioral Decision Making*, 5, 25–37.
- Wendt, D. (1969). Value of information for decisions. *Journal of Mathematical Psychology*, 6, 430–443.

*Authors' biographies:*

**Ben Newell** (PhD 2001, University of New South Wales) is a Lecturer in Cognitive Psychology at the University of New South Wales and a Research Fellow of the ESRC Centre for Economic Learning and Social Evolution. His interests include decision heuristics and the implicit/explicit distinction in human learning and judgment.

**Tim Rakow** (PhD 2001, University College London) is a Teaching Fellow in Psychology at the University of Essex. His research interests include probability elicitation, risk assessment and decision making in medicine, information search, and inter-temporal choice.

**Nicola J. Weston** is a PhD student at the University of Plymouth. She was a research assistant at University College London where she completed this work. Her research interests include decision making, face recognition and eyewitness memory.

**David Shanks** (PhD 1985, University of Cambridge) is Professor of Psychology and Head of the Psychology Department at University College London, and Scientific Director of the ESRC Centre for Economic Learning and Social Evolution. His research interests include human learning, computational modelling (especially with neural network models), and decision making.

*Authors' addresses:*

**Ben R. Newell, Nicola J. Weston** and **David R. Shanks**, Department of Psychology, University College London, Gower Street, London WC1E 6BT, UK.

**Tim Rakow**, Department of Psychology, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK.