

Evaluating Three Criteria for Establishing Cue-Search Hierarchies in Inferential Judgment

Tim Rakow
University of Essex

Ben R. Newell
University of New South Wales

Kathryn Fayers and Mette Hersby
University of Essex

The authors identify and provide an integration of 3 criteria for establishing cue-search hierarchies in inferential judgment. Cues can be ranked by information value according to expected information gain (Bayesian criterion), cue-outcome correlation (correlational criterion), or ecological validity (accuracy criterion). All criteria significantly predicted information acquisition behavior; however, in 3 experiments, the most successful predictor was the correlational criterion (followed by the Bayesian). Although participants showed sensitivity to task constraints, searching for less information when it was more expensive (Experiment 1) and when under time constraints (Experiment 2), concomitant changes in the relative frequency of acquisition of cues with different information values were not observed. A rational analysis illustrates why such changes in the frequency of acquisition would be beneficial, and reasons for the failure to observe such behavior are discussed.

Keywords: cue search, fast-and-frugal, heuristics, inference, information acquisition

Consider an employee of a motor insurance company who is charged with the task of predicting which of two clients will make larger value claims in future years. Such predictions will often be wrong, but there are cues (age, sex, etc.) that may allow him to perform better than chance at this task (see Table 1). However, none of these cues perfectly predicts which client will have the greater claim value. Performing well in this task might require one to learn several probabilistic cue–criterion relations—hence, tasks of this kind are often described as a multiple-cue probability-learning (MCPL) tasks (Smedslund, 1955).

The MCPL paradigm has fostered hundreds of experimental studies (Holzworth, 2001). Typically, the focus is on judgments for individual objects rather than inferences involving pairs of objects or the choice between multiple objects. Many everyday choice tasks share properties of the MCPL choice task outlined above. When choosing between applicants, an employer needs to identify

who will perform best in the job on the basis of cues that predict only imperfectly (prior experience, qualifications, etc.). When identifying which of two patients in the emergency room is more seriously ill, the doctor may consider whether each of the patients is conscious, short of breath, complaining of pain, and so on. Often the relevant cues are not immediately apparent—in which case, information search is presumably neither random nor exhaustive. Predecisional information search is the focus of this article. Specifically, what determines which cues are considered and the frequency with which they are examined over repeated choices?

To address these questions, we focus on two facets of the MCPL domain: primarily, the cue-outcome relationships for the probabilistic environment in which the decision maker operates, but also the constraints or demands placed on the decision maker by task characteristics (e.g., information cost, time limitations). Our main interest is whether what people learn about cue-outcome relationships can usefully guide information search. By examining different task characteristics, we can also investigate whether people use what they have learned about cue-outcome relationships to counter the effects of various environmental constraints.

Learning Features of the Probabilistic Environment: Search Patterns and the Value of Information

The standard prescription for predecisional information search when multiple cues (or *attributes*) are relevant to the choice at hand is to search cues in descending order of importance (Albert, Aschenbrenner, & Schmalhofer, 1989) or at least to do so probabilistically (Tversky, 1972). When choice is not purely a matter of satisfying idiosyncratic personal preferences but demands maximizing value according to an objective criterion, how should cue importance or the value of information be defined? We have been

Tim Rakow, Kathryn Fayers, and Mette Hersby, Department of Psychology, University of Essex, Colchester, United Kingdom; Ben R. Newell, School of Psychology, University of New South Wales, Sydney, Australia.

Tim Rakow and Ben R. Newell contributed equally to this project, and order of authorship was determined arbitrarily. The work was part of the program of the Economic and Social Research Council (ESRC) Research Centre for Economic Learning and Social Evolution. We thank Roger Grace, who programmed Experiments 2 and 3, for his assistance and the ESRC and the Leverhulme Trust for their financial support.

Correspondence concerning this article should be addressed to Tim Rakow, Department of Psychology, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ United Kingdom, or Ben R. Newell, School of Psychology, University of New South Wales, Sydney 2052, Australia. E-mail: timrakow@essex.ac.uk or ben.newell@unsw.edu.au

Table 1
Multiple-Cue Probability-Learning Choice Task: Assessing Insurance Risk From Imperfect Predictors

Cue	Is the client . .	Client A	Client B
Sex	Male?	Yes	No
Age	Under 25 years of age?	—	—
Previous convictions	A convicted felon?	No	No
Marital status	Married?	—	—

Note. Dashes indicate that information is not yet known.

able to identify three different criteria that provide slightly differing answers to this question—we term these the *Bayesian*, the *correlational*, and the *accuracy* criteria. Each criterion defines the magnitude or strength of the cue-outcome relationship in a particular way—thus, these different criteria sometimes differ in how they rank order cues by importance.

The Bayesian Criterion

The Bayesian framework lends itself to the consideration of information search and the relative value of different pieces of information. Bayes's theorem can be used to update the subjective probability of an event or the confidence in some proposition on the receipt of additional information. It is therefore quite natural to conceive of the Bayesian decision maker as one who revises belief step-by-step on the sequential acquisition of cue information and chooses between options when sufficiently certain of the best option. Information is then valued according to the magnitude of the reduction of uncertainty or the revision of belief that it yields. Because we cannot know with certainty exactly how to revise belief until we have acquired the relevant information, the appropriate Bayesian measure for determining the order of information acquisition is the *expected change in belief* (ECB; Lindley, 1956). ECB, also known as *expected information gain* (Oaksford & Chater, 1998), has been used to account for human information search strategies consistent with a positive test strategy that have been observed in several cognitive tasks that require information acquisition (Klayman & Ha, 1987). In the context of MCPL choice tasks, the Bayesian criterion would imply that, at a given stage in the search process, the next cue to be acquired should be the cue that yields the greatest ECB (if all cues are equally costly).

The Correlational Criterion

Natural, behavioral, and social scientists are quite used to measuring the strength of a relationship according to correlation coefficients and to ranking the strength of different relationships accordingly (e.g., as typified by the use of Pearson's *r* as a measure of effect size; Cohen, 1969). In the context of judgment and decision making under uncertainty, Egon Brunswik proposed that successful judgment of an external criterion variable in probabilistic environments depended on weighting cues according to their strength of relationship with the criterion variable (Brunswik, 1955). Thus, judgment and choice involving objective criteria should mimic cue-criterion relations that are observable in the environment. Social judgment theorists who take their lead from Brunswik have usually (though not exclusively) assessed these

relations using correlational measures such as Pearson's *r* or beta weights derived using multiple linear regression (Hammond & Stewart, 2001; Slovic & Lichtenstein, 1971). Although what Brunswik provided was more a theory of information integration and less a theory of information search, he was explicit that not all cues would be examined on all occasions that a particular multiple-cue judgment is made, as, often, cues can effectively and efficiently be substituted for one another (*vicarious functioning*). Thus, it is possible to conceive of cue weights as the value ascribed to a cue in the integration process, and/or the propensity to attend to (or search for) a cue. Empirical investigations of information purchase indicate more frequent acquisition of cues that are more strongly correlated with an outcome variable, consistent with partial (though imperfect) sensitivity to variation in cue-outcome correlation (Connolly & Serre, 1984).

The Accuracy Criterion

Gigerenzer and colleagues have forwarded a vision of human judgment and choice built on an "adaptive toolbox" consisting of simple heuristics (Gigerenzer, Todd, & ABC Research Group, 1999; Todd & Gigerenzer, 2000). These relatively uncomplicated strategies search through information in an order determined by the cue-outcome environment, then stop the search and choose (or estimate) according to clearly defined rules. One of these simple heuristics, "take the best" (TTB), is applicable to the kind of MCPL choice tasks that we consider in this article (Gigerenzer & Goldstein, 1996). TTB searches through cues in descending order of their *ecological validity* (EV)—in which EV is defined as the probability of a correct choice given that the cue points to one object.¹ The reference class for this probability is the set of all possible objects pairs in the consideration set. In addition to its EV, a cue's *discrimination rate* (DR) is also of relevance to the value of the information carried by the cue. The DR is the probability that a cue will point to a unique choice (again defined for all possible object pairs). For instance, in the case shown in Table 1, the first cue (sex) does discriminate between the objects and is likely to do so relatively frequently for this task, probably yielding a DR close to the maximum possible value of 0.5. The third cue (criminal record) does not discriminate in the case shown and, in general, rarely will (as for most pairs of clients neither will be a convicted felon), resulting in a DR close to the minimum possible value of zero. However, of these two cues, criminal record may well have a higher EV. If so, when it can be used to make a choice, it is the more predictive of the two cues. It will be seen that searching through cues by descending EV places a premium on accuracy with the potential drawback that many cues must be searched through before a discriminating cue is examined.

¹ The terms *ecological* and *validity* each have multiple meanings and interpretations in psychology, which make the term *ecological validity* prone to confusion. Nonetheless, to maintain consistency with a number of recent articles, we kept the term in this article, noting that ecological validity has been precisely defined in the context in which we used it (Gigerenzer & Goldstein, 1996; Gigerenzer, Hoffrage, & Kleinbölting, 1991). It is worth noting that all three information criteria that we considered are ecological, in that they all define properties of the ecology—that is, inherent properties of the cue-outcome relationship that are external to the decision maker.

The Relationship Between the Three Criteria

It can be shown that the Bayesian and correlational criteria for the cue-outcome relationship can be represented as functions of both the EV and DR of cues. Thus, the simple heuristics approach provides a scheme for decomposing ECB and Pearson's correlation into components that depend on the EV and DR of cues.

Working within this framework, Martignon and Hoffrage (1999, 2002) defined an additional measure of the cue-outcome relationship within a MCPL choice environment that they termed *success*. Success is the probability of a correct choice when only that cue is used, and in the context of two alternative forced-choice (2AFC) success (S), (S) is defined as follows:

$$S = EV * DR + (1 - DR) * 0.5, \quad (1a)$$

or equivalently, by rearrangement, as follows:

$$S = (EV - 0.5) * DR + 0.5. \quad (1b)$$

We have demonstrated in another article that ranking cues according to ECB is isomorphic to ranking them by success (Rakow, Hinest, Jackson, & Palmer, 2004). It should make intuitive sense that how much one expects to revise belief on examination of a cue depends on the probability that one will revise at all (dependent on DR) and the magnitude of revision given that one does revise one's belief (dependent on EV). Thus, ECB can be seen to represent the value of information as a mix of DR and EV.

Similarly, it can be shown that the correlation between the pattern of cue values and outcome can also be represented as a mix of DR and EV. To do so requires one to represent the pattern of cue values as a single variable with three (ordered) levels: $+1$ if the pair of cue values point to Object A, 0 if the cue does not discriminate, and -1 if the pair of cue values points to Object B. Then, the Pearson correlation coefficient for the relationship between cue pattern and the choice criterion (A is greater vs. B is greater) can be represented as follows:

$$\text{Pearson's } r = (2 * EV - 1) * \sqrt{DR}. \quad (2)$$

A demonstration of this result is given in the Appendix. Again, it should make some intuitive sense that this correlation coefficient should depend on both EV and DR. The EV reflects co-occurrences of discriminating cue patterns and choice criterion values, and the frequency of these co-occurrences influences the strength of correlation. The DR reflects the *variance* in cue patterns, which affects the strength of correlation (as correlation is a standardized measure of *covariance*).

Comparison of Equations 1b and 2 indicates that the ranking of cues according to success/ECB can be different to that defined by Pearson's r (if EV and DR vary across cues). Both sets of rankings reflect the DR and EV of cues, but, of the two sets of rankings, Pearson's r will be more influenced by variation in EV. (If cues are selected in descending order of Pearson's r , this order will be more similar to the descending order of EV than would be the case if cues are selected in descending order of success/ECB. For instance, in the first two experiments that we report, it is sufficient to switch the order of one pair of the four cues to rearrange the order of cues from descending order of Pearson's r to the descending order of EV. Effecting the equivalent rearrangement from success/ECB order into EV order requires two pairwise reversals.)

Researchers have previously found that success/ECB provides a better description of participants' cue search than EV or DR alone (Newell, Rakow, Weston, & Shanks, 2004). However, for the probabilistic environment used in those previous experiments, the rank order of cues by success/ECB was identical to that of Pearson's r , thus preventing Newell et al. (2004) from assessing the ability of the three criteria for describing behavior. Furthermore, Newell et al. (2004) did not manipulate any aspect of the task environment—cue costs were kept constant, and participants always operated without the imposition of any time pressure. In the experiments that follow, we created environments in which all three orders (prescribed by the accuracy, correlational, and Bayesian criteria) are different, and we investigated how well each of the three criteria describe search behavior under different conditions of time, information cost, and scope of information acquisition. By varying task characteristics, we created a varied "terrain" in which to test people's ability to make good use of the cue properties that they have learned. In the next section, we briefly review the task characteristics that we manipulated in the experiments.

Adapting to Changing Task Characteristics: Stopping Search and Deciding

The most commonly examined task characteristics in relation to information acquisition are information costs and time pressure. With respect to information costs, the standard summary is that participants do adapt appropriately to differences in cue costs but do so insufficiently. Relative to optimal purchasing, participants buy too much information when little is needed and underpurchase when higher levels are beneficial (Connolly & Serre, 1984; Rapoport & Wallsten, 1972). Much of this work was conducted under the Bayesian optional stopping paradigm (Edwards, 1965), in which a participant can purchase samples from a population. Conceptually, this is a little different from the multiple-cue paradigm that we consider—it corresponds to repeated sampling of a single cue rather than sampling multiple cues. We examined adaptation to cue cost in Experiment 1.

Not surprisingly, under time pressure people do search for less information prior to making a choice, though reductions in information acquisition can be less than might be anticipated as people can also adapt by speeding up their processing of information (Payne, Bettman, & Johnson, 1993). Payne et al. observed a tendency to use attribute-based strategies in preference to alternative-based strategies under time pressure. Such strategies allow one to eliminate options on the basis of partial information, obviating the need to fully evaluate options on the basis of all information. We examined adaptation to time constraints in Experiment 2.

The final task characteristic we considered, in Experiment 3, is changes in the scope of information search permitted. Specifically, we examined the effect on search patterns of changing the amount of information that can be acquired. This corresponds to fixing the budget for predecisional information (for a given decision) and, across decisions, considering the effects of different budget levels. In addition, we compared situations in which there was certainty about the number of pieces of information available to make a decision before a given trial commenced, with those on which there was uncertainty. Thus, we compared perfect knowledge of the limits on information acquisition with the imperfect knowledge

that prevails in many competitive markets in which the exact time at which a decision must be made is not predetermined. In such situations, one does not know in advance how much information can be acquired before being called on to make a decision (because one does not know how much time will be allowed for information acquisition).

How Might Participants Adapt?

In all three experiments, we expected adaptation to these changes in task characteristics primarily at the level of the amount of information acquired. If this is the only way in which participants adapt to changing task characteristics, then whichever of our three information criteria best predicts information acquisition should remain constant across experiments and across conditions. However, we have already set out how the three information criteria imply different levels of emphasis on accuracy and the amount of information required for decisions (because they mix EV and DR in differing proportions), and we will present rational analyses to show how switching which criterion is used to guide information search as task characteristics vary can be profitable to the decision maker. We are, therefore, curious to see whether, under different task conditions, the criterion that best predicts information search does indeed alter. How then should we measure information acquisition?

Inferring Importance From Observed Patterns of Search

Perhaps the most intuitive method for inferring search behavior from the data is to examine the extent to which participants look through cues in the strict deterministic orders predicted by the different conceptions of importance. However, using such deterministic orderings as a benchmark precludes the possibility that external factors may affect participants' search. For example, Albert et al. (1989) suggested that though important cues should be processed earlier than less important ones, in many settings this ordering may not be perfect as other factors such as reading habits, accessibility, salience, and availability could all play a role. Similarly, Tversky (1972) proposed a *probabilistic* search rule for the elimination-by-aspects model of choice (a model closely related to the lexicographic strategies we consider). In the 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 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).

In the following experiments, we used a computerized information board setup in which it is quite likely that external factors—such as ease of moving the mouse to particular buttons—will have an effect on the order of search. Consider a participant who decides to examine the two most ecologically valid cues first but, because of their position on the screen, always looks at the second most valid followed by the most valid. Such behavior would be inconsistent with a strict validity ordering, yet would seem to indicate that the participant was responding primarily to EV.

For these reasons, we examined the *frequency* with which particular cues were acquired. That is, the proportion of trials on which each cue was acquired aggregated across participants. We contend that an ordering based on aggregated frequency of acqui-

sition provides a valid indication of the extent to which particular cues are considered important by participants.²

Experiment 1: Varying the Cost of Information

Overview of Experiment 1

The task in Experiment 1 was to predict, using up to four binary cues, which of two objects would have the higher criterion value. The objects were fictional companies and the cues were "yes" or "no" answers to questions about aspects of the company's financial status (e.g., "Is it an established company?"). The aim for the participants was to make money by predicting which company would have the higher share price. The experiment was divided into three stages: training, practice, and test. In training, all cue values (i.e., "yes" or "no") were presented simultaneously without cost, but cue values had to be purchased in practice and test sessions. The cost of cue information was manipulated within subjects so that each participant experienced each of four information cost conditions. Each cost condition was run separately either in morning and afternoon sessions or on consecutive days. Our principal interest was to examine whether the relative cost of information would affect not only the amount of information purchased but also the frequency with which different cues were purchased.

Key Questions

Consistent with prior investigation of information purchase, we anticipated attendant changes in the number of cues bought prior to choice as cue cost increases. However, our discussion of the different criteria that can be used to order the search through cues demonstrates that the cue-outcome relationship can be decomposed into components that depend on the DR and EV of cues. We can therefore examine whether changes in the patterns of cue purchase will reflect these dimensions and mediate changes in the amount of information bought. For instance, when information becomes more expensive, will participants simply truncate their search, or will they search for cues with higher DRs (those more likely to discriminate between options) at the expense of cues with higher EVs (those more likely to indicate the "correct" option)? To illustrate the rationality of these adaptations, we present a rational analysis of the probabilistic environment used in Experiment 1.

Rational Analysis of the Environment for Experiment 1

Table 2 summarizes the cue-outcome relationships according to the measures that could define cue importance. (See the Method section for details of the design that generated these relationships and for a description of the "share prediction" cover story given to participants.) We considered five classes of search strategy, which previous work had suggested would yield differing levels of accuracy and use differing amounts of information: (a) A (compensatory) *weighted additive rule* (WADD) derived using multiple

² Despite the problems associated with examining the order in which cues are acquired, we did use the proportion of trials in which each cue was the first cue acquired as an auxiliary indicator of the importance given to each cue.

Table 2
Experimental Environment in Experiment 1

Measure	Cue A	Cue B	Cue C	Cue D	Defines the search order	Strategy searching in this order
DR	.219	.375	.500	.469	C > D > B > A	
EV	.825	.606	.671	.768	A > D > C > B	Accuracy
Success/ECB	.571	.540	.586	.626	D > C > A > B	Bayesian
Pearson's <i>r</i>	.304	.130	.242	.367	D > A > C > B	Correlational

Note. DR = discrimination rate; EV = ecological validity; ECB = expected change in belief.

linear regression—this strategy always searches through all cues. The next three classes of strategy all use one-reason stopping and decision rules (“search until the one cue discriminates, and choose the object that the cue points to”) but search through cues according to three different hierarchies of importance (defined by the three criteria for the cue-outcome relationship that we have discussed). (b) The *accuracy strategy* searches cues in descending order of EV and is, therefore, the TTB strategy described by Gigerenzer and Goldstein (1996). We consider both the “full” strategy, which searches through as many cues as are available before guessing (if no cues discriminate), as well as the “truncated” strategies that stop and guess after a maximum of one, two, or three cues if none have discriminated by that point (accuracy-T1, accuracy-T2, and accuracy-T3, respectively). (c) The *Bayesian strategy* searches through cues in descending order of success/ECB, and again we consider the entire family of full and truncated strategies (Bayesian-full, Bayesian-T3, Bayesian-T2, and Bayesian-T1). (d) The *correlational strategy* searches cues in descending order of Pearson's *r*, and, similarly, we consider the entire family of correlational strategies. (e) *Random choice* (RC) searches for no information and chooses one of the two objects randomly (and therefore expects to make the correct choice on 50% of occasions). This provides a baseline against which other classes of strategy can be assessed. Readers should note that each of Strategies (2) through (4) are “one-reason decision strategies,” which stop searching through cues as soon as a discriminating cue has been found. It is an empirical question whether people do adopt such a stopping rule (which we investigate in our first

experiment), though previous research indicates that they often, if not always, do so, when information has a relatively high cost (Bröder, 2000; Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003). Applying the one-reason stopping rule in this rational analysis serves to provide a like-for-like comparison of the three criteria for the value of information that we are examining.

Expected payoffs for each strategy were computed, setting the reward for a correct choice at 100 points and allowing the cost (in points) of examining each cue to vary. Figure 1 shows that the strategy yielding the greatest expected payoff varies as the cue costs increase—reflecting the trade-off between the accuracy achieved and the amount of information used by these strategies. Each intercept on the vertical axis shows the expected accuracy (percentage of correct choices) of that strategy. The gradient of each line reflects the amount of information searched for, with more frugal strategies having more shallow gradients indicating that payoffs for these strategies are less sensitive to changes in the cost of information. For a given cost (corresponding to a point on the horizontal axis), the best strategy from among those considered can be found by observing which line is uppermost. For clarity, only those strategies that yielded the highest expected payoff over some range of costs (however short) are shown in Figure 1 (i.e., dominated strategies are not plotted). Naturally, when information is free, the most accurate strategy (WADD) provides the highest expected payoff. However, even with minimal search costs, the more frugal (and only slightly less accurate) accuracy-full strategy returns a higher expected payoff. When search costs are moderate, the correlational and then Bayesian strategies yield the greatest

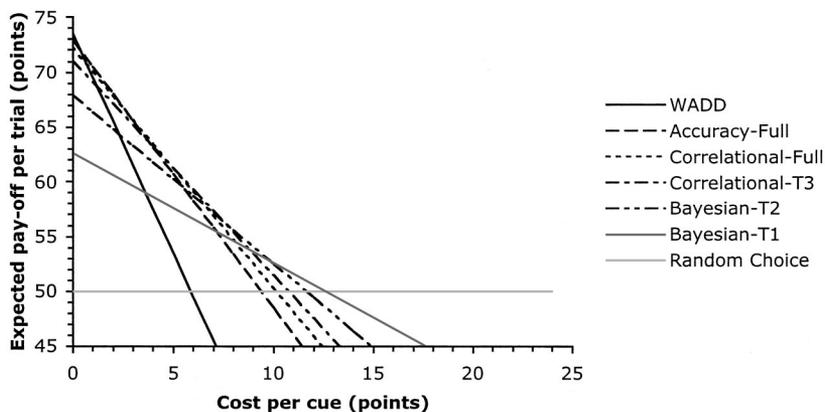


Figure 1. Expected payoff per trial for different strategies (100 points for a correct choice). The uppermost line corresponding to any point along the horizontal axis indicates the best strategy for a given cue cost. WADD = weighted additive rule.

rewards, though the maximum number of cues one is prepared to search should decrease with increasing costs. When search costs are high, no cues are sufficiently informative to improve on the expected payoff from random choice. Thus, as information costs increase, two types of adaptation to search are seen to be of benefit. First, changing the order in which cues are searched—switching from orders more akin to EV to orders that are more influenced by the DR of cues (i.e., switching progressively from the accuracy to the correlational and then to the Bayesian search orders). Second, reducing the maximum number of cues that one is prepared to search through prior to making a choice.

Taking data from previous similar experiments into account (Newell et al., 2004), this rational analysis was used to guide the cost of information that was set in four different cost conditions (3, 10, 17, and 24 points per cue). In these earlier experiments, participants differed in the amount of information they purchased, and, on average, tended to overpurchase information relative to what would be expected given purely financial considerations (Newell et al., 2004). For instance, with cue costs set at a level equivalent to 17 points, just over half of the participants examined more than two cues per trial (on average). Consequently, it was felt necessary to set costs up to 24 points per cue if we were to examine search behavior in situations in which participants opt to restrict their search significantly.

There is no reason to preclude the validity of nonfinancial motivations in experiments such as this (e.g., some intrinsic reward for accuracy) and that such motivations might legitimately vary between individuals. Therefore, we cannot specify that the strategy our analysis indicates is best for a given cost condition is necessarily the one that should be adopted by all participants. However, over a range of cost conditions (such as those set), we might expect to see adaptations in search behavior consistent with some, or all, of the sequence of strategy changes evident in the rational analysis above. Thus, with a more qualitative interpretation of our rational analysis, we would anticipate that as information costs increase there would be a tendency to switch or adapt strategies. Switching from compensatory decision making to strategies that are prepared to guess even if additional information could help improve accuracy. In addition, a qualitative interpretation of our analysis would anticipate an EV-based search rule (accuracy strategy) with lower information costs, but a Pearson's r (correlational) or ECB-/success-based (Bayesian) search rule with higher costs. Thus, our rational analysis shows that adaptations in both the stopping rule and the search rule can be used to enhance the pay-off from choices.

Method

Participants

Twelve members of the University College London community took part in the experiment. Six were male and six were female. They had a mean age of 21.9 years (range = 18–33 years, $SD = 4.3$). Each participant completed four separate testing sessions.

Stimuli, Design, and Procedure

A share profitability prediction task similar to the one used in a number of previous studies was used (Bröder, 2000; Newell et al., 2004; Newell & Shanks, 2003; Newell et al., 2003; Rieskamp & Hoffrage, 1999). Parti-

cipants 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 ("Was the share trend positive over the last few months?" "Does the company have financial reserves?" "Does the company invest in new projects?" "Is it an established company?"). Each cue had an associated validity, discrimination rate, success rate, and Pearson's r .

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 one eighth, two eighths, three eighths, or one half 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. Twelve blocks were created in total, and these were rotated such that each block was used an equal number of times in the training, practice, and test phases of the experiment (see below).

Table 2 displays the validities, discrimination rates, success rates, and Pearson's r 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. By counterbalancing the position of cues, we remove any systematic effect of the location in which cues appear on the screen. By counterbalancing cues across the four labels, any systematic effect of prior beliefs about the supposed differential value of cues (that participants might infer from the labels) is also eliminated.

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's 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.

Each testing session was divided into three stages: a training block, a practice block, and a test block. In the training block, consisting of 64 trials, cue information was supplied automatically, simultaneously, and at no cost to participants. On each trial, participants were motivated to learn about the cues by winning 20 points for every correct prediction. Participants were told that every 500 points was equal to U.K. £1.00 (approximately U.S. \$1.65; € 1.43) and that their points would be converted into money at the end of the fourth and final testing session. Following training, participants were given a practice block in which cue information had to be bought. 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. Participants bought information by clicking on a screen button that uncovered the values of each piece of information for the two shares (e.g., a *Yes* for Share A and a *No* for Share B in answer to the question "Is it an established company?").

The cost of each piece of information varied across testing sessions (see below), but the payoff for a correct prediction was held constant at 100 points. Participants were told that any points spent or won during the practice block would not affect their final earnings. They were encouraged to use the 64 trials of the practice block to work out how to earn the most money in the test block that followed practice. The information cost and payoff structure of the 64 trials in the test block were the same as those in the practice block, but points won were converted into final earnings. The trials in each block were presented in an order randomly selected for each participant.

Information cost conditions. Our principal interest was to examine whether the relative cost of information would affect not only the amount of information purchased but also the frequency with which different cues were purchased. To this end, we examined conditions in which the cost of information was 3, 10, 17, or 24 points per cue (see the rational analysis for

Table 3

Experiment 1: Mean (SD) Proportion of Trials on Which Each Cue Is Purchased and z for Specified Trends in Cue Proportions Purchased

Cost condition	Proportions purchased					Order examined by Page Trend Test				
	Cue A	Cue B	Cue C	Cue D	All cues	Search frequency	DR	EV	Success/ECB	Pearson r
3 points	.65 (.43)	.32 (.29)	.56 (.29)	.83 (.20)	.59 (.18)	D > A > C > B	0.50	2.50**	2.70**	3.00**
10 points	.49 (.40)	.18 (.19)	.47 (.27)	.83 (.28)	.49 (.13)	D > A > C > B	0.85	2.45**	3.15***	3.30***
17 points	.52 (.45)	.09 (.12)	.42 (.34)	.61 (.34)	.41 (.13)	D > A > C > B	0.45	2.95**	2.75**	3.10***
24 points	.46 (.48)	.11 (.26)	.33 (.37)	.51 (.47)	.35 (.11)	D > A > C > B	0.50	1.70*	1.90*	2.00*
Combined	.53 (.40)	.18 (.17)	.44 (.24)	.69 (.25)		D > A > C > B	0.90	2.30*	3.10***	3.20***

Note. DR = discrimination rate; EV = ecological validity; ECB = expected change in belief.

* $p < .05$. ** $p < .01$. *** $p < .001$.

an explanation of why these particular values were chosen). In a single testing session, the cost of information was held constant (i.e., if information cost 3 points in the practice block, it cost 3 points in the test block). The order in which participants were run in the four cost conditions was counterbalanced. Testing sessions were run either on 4 separate days or over 2 days with a session in the morning and a session in the afternoon.

Procedure. 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, and on a handout placed on the table next to the computer. Participants were told that they had to complete four testing sessions (over the course of 2 or 4 days) and that the cost of information would be different in the different sessions. Participants were also 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. Outcome feedback was provided after each prediction during all stages of the experiment (training, practice, and test). (A figure depicting the screen layout and a more detailed description of the procedure can be found in Newell et al., 2004). At the end of the fourth testing session, participants were thanked, debriefed, and paid their total earnings.

Results and Discussion

Purchase Proportions

The frequency with which each cue was purchased during the test blocks was recorded for each participant in each cost condition (summarized in Table 3). The left side of Table 3 indicates that the rank order of purchase proportions fell in the order prescribed by Pearson's r in all cost conditions. A two-way (Cost condition \times Cue) repeated measures analysis of variance (ANOVA) was performed with the purchase proportion as the dependent variable. A significant main effect of cost, $F(3, 33) = 8.08, p < .001$, and a significant linear trend, $F(1, 11) = 13.88, p = .003$, indicated that participants did acquire less information when it was more expensive.³

Tukey's honestly significant difference (HSD) post hoc tests indicated that purchase levels in the 3-point condition were significantly higher than those in the 24-point ($p < .001$) and 17-point ($p < .01$) conditions, and that more information was purchased in the 10-point condition than in the 24-point condition ($p < .05$). Thus, purchase levels in adjacent cost conditions did not differ significantly from each other, whereas all pairs of nonadjacent cost conditions did differ in the expected direction. There was a significant main effect of cue, $F(3, 33) = 6.15, p = .002$.

Tukey's HSD post hoc tests indicated that Cue D (highest success/ECB and Pearson's r) was bought significantly more often than Cue B (lowest EV, success/ECB, and Pearson's r ; $p < .01$), and that Cue A (greatest EV) was also bought significantly more often than Cue B ($p < .05$). There was no significant two-way interaction, $F(9, 99) = 0.83, p = .595$. Patterns of cue purchase did not differ significantly by cost condition, and changes in purchase frequency across different cost conditions did not differ significantly from one cue to another. Thus, there is clear evidence that people search for less information when costs are higher, but no evidence that changes in cost prompt them to change the relative frequency with which they purchase particular cues.

A simple tally of which cue was examined most frequently by each of the participants showed that Cue A (highest EV) and Cue D (highest success/ECB and highest Pearson's r) were the most favored cue for 5 participants each. Cues B and C were examined with greatest frequency by only 1 participant each.

Purchase proportions were further examined using Page's non-parametric trend test to determine which of the four specified cue hierarchies (EV, DR, success/ECB, or Pearson's r) described participants' search patterns best. Purchase proportions were positively related to each of the four trend orders examined (i.e., $z > 0$ in each case) and always significantly related to the orders predicted by EV, success/ECB, and Pearson's r (see the right side of Table 3). Purchase proportions were always best predicted by the order determined by Pearson's r . In three of the four cost conditions, purchase proportions were better predicted by success/ECB order than by EV order, with the reverse being true in the 17-point cost condition.

First Cue Purchased

Which cue was acquired first was examined for each trial and gave a similar, though not identical, ranking of cues to the frequency of purchase data. Cue A (highest EV) was acquired first on 41% of trials, and Cue D (highest success/ECB and highest Pearson's r) was acquired first on 38% of trials. Cue C was acquired first on 16% of trials and Cue B on 5% of trials. Thus, on this measure, the discrepancy between most favored and least favored cues is even more distinct than for the frequency of purchase data.

³ There were no significant effects of cost condition order or session order on the proportion of information purchased.

This ranking for which cue was acquired first was consistent across the 3-, 17-, and 24-point conditions. In contrast, in the 10-point condition, Cue D was acquired first on 60% of trials, and Cue A was acquired first on 26% of trials.

Stopping Search

We examined two types of departure from the one-reason stopping rule (“stop search at the first cue that discriminates”): the purchase of unnecessary information (obtaining further cues after the first cue that discriminates), and unforced guessing (choosing a share when no cue has discriminated and not all cues have been examined). Figure 2 displays these two types of departure. Taking both types together, there are departures from the one-reason stopping rule on half of the test trials in the highest cost condition and about one third of the test trials in the other three conditions. There is a significant effect of cost condition on the purchase of unnecessary information, $F(3, 33) = 3.87, p = .031$, reflecting the lower levels of purchase in higher cost conditions, though the linear trend did not reach significance, $F(1, 11) = 4.05, p = .069$. As can be seen in Figure 2, unnecessary information is rarely purchased in the two highest cost conditions. The proportion of unforced guessing increases steadily with information cost, with a significant linear trend, $F(1, 11) = 15.67, p = .002$.

Summary

Taken together, the frequency and stopping rule analyses indicate that people adapt (appropriately) to higher information costs by truncating their sequential search through cues rather than by changing the frequency of purchase of particular cues to a strategy that is more likely to yield discriminating information. We consider the reasons for this behavior in the General Discussion section.

Experiment 2: Varying Time Constraints

Financial costs are not the only information that people have to contend with. Arguably, in everyday situations it is more common for people to experience information costs in terms of the time taken or effort expended in searching for information. We investigate the effect of time constraints by examining a fixed number of choices made within an overall time limit for a set of choices.

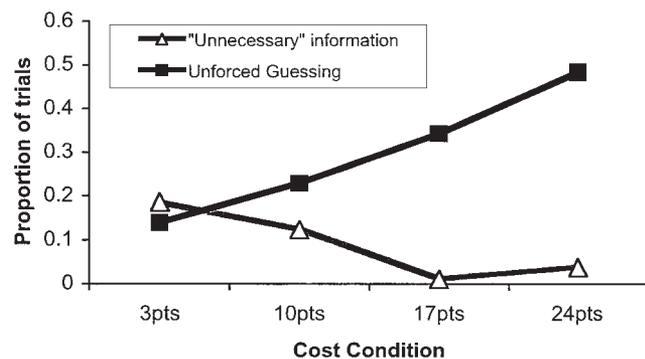


Figure 2. Departures from a one-reason stopping rule with varying cue cost. pts = points.

We manipulate time constraints by altering the length of time participants must wait after acquiring a cue before they may acquire the next cue. Cues are acquired without financial cost, but correct decisions are rewarded as before, and choices not made within the overall time limit incur a financial penalty. Thus, this experiment examines choice when time is of the essence—those difficult situations in which additional information makes the right choice more likely, but the time taken to acquire information risks costly delays in reaching a decision. This is the dilemma faced by the emergency medicine doctor or the financial trader. In examining a patient for additional signs and symptoms, the doctor risks delaying necessary treatment (for this and for other patients). By researching the wisdom of an investment, the trader risks missing the opportunity to invest at the price that was initially quoted or may be indisposed when other investment opportunities arise.

Time and effort in acquisition can genuinely be considered to be search costs (Fried & Peterson, 1969; Payne, 1982); however, this time-limited paradigm differs in one important respect from the financial cost paradigm used in Experiment 1 and previously (e.g., Newell et al., 2004). In Experiment 1, the cue-cost function was linear—each additional cue acquired incurred the same incremental cost (within a given cost condition). When working within a time limit, the cue-cost function is a positively increasing one. If the first few cues acquired can be examined sufficiently quickly that there is little danger of failing to keep to the time limit, these can be considered to have relatively little cost. If acquiring further cues jeopardizes one’s ability to make a sufficient number of choices within the time limit, these are high cost cues.

Key Questions

In addition to anticipating the truncation of information search under high time constraints (i.e., when less information can be collected within the allotted time), we can again examine whether people adapt their patterns of cue search with respect to the mix of DR and EV and determine which cues are examined with relatively greater frequency. Specifically, under more severe time constraints, will cues of high DR (those more likely to discriminate between options) be examined relatively more frequently than under less stringent constraints? If so, we would expect a shift from search strategies in keeping with the accuracy criterion under low time constraints, but following the prescriptions of the Bayesian and correlational criteria under high time constraints.

Method

Participants

Thirty-two members of the University of Essex community took part in the experiment. Ten were male and 22 were female. They had a mean age of 25.5 years (range = 17–40 years, $SD = 5.3$).

Stimuli, Design, and Procedure

The share prediction task as described in Experiment 1 was used. Cue EVs (and, consequently, success/ECB and Pearson’s r) were altered slightly from Experiment 1 such that only on those trials with no discriminating cues was the probability of both shares being the better share equal (both with probability of 0.5). This ensured that as long as there were cues that discriminated, there was always one share that had a strictly higher probability of being the better share (probability >0.5). The relative

Table 4
Experimental Environment in Experiment 2

Measure	Cue A	Cue B	Cue C	Cue D	Defines the search order	Strategy searching in this order
DR	.469	.375	.500	.219	C > A > B > D	
EV	.776	.608	.677	.840	D > A > C > B	Accuracy
Success/ECB	.629	.541	.589	.574	A > C > D > B	Bayesian
Pearson's <i>r</i>	.377	.134	.252	.316	A > D > C > B	Correlational

Note. DR = discrimination rate; EV = ecological validity; ECB = expected change in belief.

ranking of cues with respect to the parameters remained as it had been in Experiment 1 (e.g., the cue with highest EV had the second highest Pearson's *r* and the third highest success/ECB in either experiment). Cues were placed (randomly) into a new cyclical sequence (see Table 4).

Participants again earned points on each trial according to the payment function $100(p - 0.5)$, where p is the probability of a correct choice. Thus, a participant performing at chance would earn no points. Points were incremented to a running total after each trial.⁴

Time constraint was manipulated within participants with each participant encountering three blocks of 64 trials: a training block, a low time constraint block, and a high time constraint block. In the training block, all cue information was provided simultaneously, without cost, and without any time limit. In the instructions given at the start of the experiment, participants were informed that they would later have to make decisions under time pressure and may not have time to examine all the available information—they were advised to use the training trials to learn which pieces of information were more informative. Trials in each time constraint block were divided randomly into two fixed phases of 32 trials: a practice phase followed by a test phase. For each phase, there was a time limit of 320 s, in which the 32 trials were to be completed. In other words, time constraints were applied at the level of the session, not at the level of the individual trial (participants had an average of 10 s—320 s/32 trials—to complete each trial). At the end of every trial, information on screen was updated to show how many trials and how much time remained. On-screen messages also indicated when one half and when three quarters of the time limit had elapsed. Each trial not completed within the limit incurred a penalty of 25 points. Points earned in the two test phases were converted into money at a rate of 400 points to U.K. £1.00 and were added to the participants "turn-up" fee of U.K. £2.00.

In the low time constraint condition, following the acquisition of a cue, the program did not permit participants to acquire another cue for 2 s. In the high time constraint condition, this delay was programmed at 3 s. Thus, time constraint was manipulated (within subjects) by changing the length of time it took to acquire information, not the length of time in which choices had to be made.

Pilot work had indicated that these lengths in delay would result in a difference in the average amount of information that could be acquired in excess of half a cue per trial. A rational analysis of the environment had indicated that a difference of this magnitude was sufficient to alter which search strategy performs best. Obtaining the cues required for the accuracy–full strategy (an average of 2.4 cues per trial) was estimated to be relatively easy in the low time constraint condition but relatively hard under high time constraints. In contrast, the Bayesian–full strategy requiring an average search string of only 2.0 cues per trial could be used effectively even under high time constraints.

The screen position of cues was counterbalanced across four configurations as per Experiment 1, and this was fully crossed with the counterbalancing of the order of time constraint conditions. Consequently, 4 participants were assigned to each of the eight possible combinations of information position and condition order, eliminating any systematic effects of screen location (of cues), prior beliefs, and task learning.

Results and Discussion

Acquisition Proportions

The left side of Table 5 summarizes the proportion of information examined and indicates that the rank order of acquisition proportions fell in the order prescribed by Pearson's *r* in both time constraint conditions. A two-way (Time Pressure × Cue) repeated measures ANOVA with proportion examined as the dependent variable revealed significant main effects of time constraint, $F(1, 31) = 8.60$, $p = .006$, and of cue, $F(3, 93) = 13.12$, $p < .001$, but no significant two-way interaction, $F(3, 93) = 0.43$, $p = .731$. Tukey's HSD post hoc tests indicated that Cue A (highest success/ECB and greatest Pearson's *r*) was examined more frequently than each of the other cues ($p < .001$ for Cues B and C, $p < .01$ for Cue D). No other pairs of cues were significantly different. Thus, participants were sensitive to the time constraint manipulation and did not examine cues with equal frequency.⁵

Tallies of each participant's most favored cue found that 17.5 participants (55%) chose Cue A (highest success/ECB and greatest Pearson's *r*) most frequently (.5 denotes two cues acquired with equal frequency). Cue D (highest EV and second highest Pearson's *r*) was the most favored cue for 8.5 participants, with Cues C and B being the most frequently chosen cues for only 3.5 and 2.5 participants, respectively.

As shown in the right side of Table 5, Page trend tests indicated that the frequencies with which each cue was examined are best predicted by the orders of importance prescribed by Pearson's *r* (correlational criterion) and success/ECB (Bayesian criterion), with EV (accuracy criterion) predicting search less effectively (though still significantly). Pairwise differences in the strength of trend for cue orders specified by the three criteria were not statistically significant. However, the difference in trend strength between Pearson's *r* and EV did approach significance (two-tailed p for the difference of .070 and .139 under low and high time

⁴ As payment in this experiment was determined by the probability of outcomes and not the outcome itself (which is subject to random variation), no feedback was given on the outcome of each trial. Feedback of the probability that the chosen share was the better share was retained.

⁵ Though statistically significant, the mean difference between conditions in the proportion of the available information examined (equivalent to 0.27 cues per trial) was smaller than we had anticipated. Notably, the mean (*SD*) time taken per trial was 7.61 s (1.62) under low time constraints and 9.05 s (1.21) under high time constraints. Given that the time limit given to participants allowed them to take an average of 10 s per trial, there was a clear tendency for participants to collect less information than the time permitted them to in the low time constraints condition.

Table 5
 Experiment 2: Mean (SD) Proportion of Trials on Which Each Cue Is Examined and *z* for Specified Trends in Cue Proportions Examined

Time constraint condition	Proportions examined					Order examined by Page trend test				
	Cue A	Cue B	Cue C	Cue D	All cues	Search frequency	DR	EV	Success/ECB	Pearson <i>r</i>
Low	.87 (.22)	.44 (.25)	.57 (.27)	.61 (.37)	.62 (.13)	A > D > C > B	1.13	2.97**	4.50***	4.78***
High	.81 (.26)	.36 (.29)	.54 (.25)	.56 (.34)	.55 (.08)	A > D > C > B	1.29	2.84**	4.23***	4.32***
Combined	.84 (.22)	.40 (.24)	.56 (.25)	.59 (.35)		A > D > C > B	1.10	2.89**	4.29***	4.53***

Note. *z* values differing by more than 1.96 are significantly different ($p < .05$) from each other, hence purchase proportions are significantly more closely related to success/expected change in belief (ECB) or Pearson’s *r* than to discrimination rate (DR). EV = ecological validity.
 ** $p < .01$. *** $p < .001$.

constraints, respectively, and .101 when both conditions are combined). The difference in trend strength between success/ECB and EV was a little smaller (two-tailed *p* for the difference of .126, .165, and .165 under low, high, and combined time constraints, respectively). Thus, there is good evidence that the frequency with which a cue is examined is related to its importance as defined by correlational and Bayesian criteria. There is also clear evidence that people examine less information under higher time constraints. However, there is no clear evidence favoring adaptive changes in the mix of DR and EV that people respond to as a function of time constraints.

First Cue Acquired

Analysis of which cue was acquired first on each trial followed the same pattern of results. Cue A (highest success/ECB and greatest Pearson’s *r*) was acquired first on 50% of trials, Cue D on 28% of trials, Cue C on 15% of trials, and Cue B on 7% of trials. These percentages barely differed between the two time constraint conditions—the largest discrepancy being for Cue C, which was acquired first on 13.4% of trials under low time constraints and on 17.7% of trials under high time constraints.

Summary

Consistent with Experiment 1, adaptation to costs takes the form of acquiring less information, but this is not achieved by changing which cues are acquired relatively more, or less, frequently. Taking frequency data and first-cue-acquired data together points to the truncation of information acquisition as the main influence of time pressure, as, although the number of cues acquired did change, the first-cue-acquired data were barely affected

Experiment 3: Varying the Number of Cues Available

Experiment 3 examines one final constraint on information acquisition: fixed resources for each individual decision. This was achieved by placing a limit on the number of cues that could be acquired before a choice was made. Thus, the constraint on information search applies to each individual trial, rather than being averaged across a block of trials (as was the case in Experiments 1 and 2). This models decisions that have a fixed time or fixed budget for information acquisition. Any choice with a strict deadline for selecting your option has this characteristic with respect to predecisional information acquisition.

On different trials in this experiment, participants could acquire one, two, or three cues. In the absence of financial costs for cues or time limitations, we anticipated that participants would acquire the amount of information that they were allowed to. However, which cues are acquired remains of interest.⁶

In addition to limiting the amount of information participants could acquire, we manipulated whether or not participants knew in advance how many cues they could acquire before making a decision. This models decisions in which, after initiating the search for predecisional information, the deadline for selecting your option may be expedited (ceasing information acquisition) or delayed (allowing further search through cues). In one block of trials, participants knew at the start of the trial exactly how many cues they could acquire before making a decision (*known limit condition*). In another block, participants discovered the limitation to their cue acquisition only at the point when a decision was required (*unknown limit condition*).

Key Questions

If participants are sensitive to the different informational properties of the cues and adapt accordingly to these properties, then one might observe a greater frequency of acquisition of higher EV, low DR cues when participants know in advance that they can obtain two or three cues. This is because when there is an opportunity to examine other cues, participants can take the risk of acquiring a cue that has a lower probability of being useful for the decision (i.e., low DR). Such a risk is less attractive in the case of an unknown limit for acquisition because one might be left with only a nondiscriminating cue when the choice must be made. Alternatively, a participant may simply follow the same order of cue acquisition, irrespective of whether the constraints on their search are known in advance. These competing hypothesis are examined in this third experiment.

⁶ With respect to the manipulation of cue cost, this manipulation is a little different from those used in the first two experiments. In Experiments 1 and 2, cue costs were manipulated (financially or via time pressure) and we observed the effect on cue acquisition and found that levels of cue acquisition varied considerably between individuals. One way to think of this limitation to one, two, or three cues is by specifying that it is as if cue costs have been manipulated such that participants opt to select one, two, or three cues on a given trial. (The costs required to achieve this would inevitably vary from one participant to another.)

Method

Participants

Thirty-two members of the University of Essex community took part in the experiment. Ten were male and 22 were female, and they had a mean age of 22.8 years (range = 16–46 years, $SD = 6.6$).

Stimuli, Design, and Procedure

The task and the measures obtained were the same as Experiment 2. A new environment was created for this experiment, with cues placed randomly into the cyclical order shown in Table 6. The DRs were slightly more widely and more evenly distributed—this was achieved by setting the proportion of positive cue values to one ninth, two ninths, three ninths, and four ninths for the four cues. This configuration of cue values resulted in experimental blocks of 81 trials. (With cue patterns derived by dichotomizing cues with nine, rather than eight, levels per cue, SPSS orthoplan requires a larger block of cases to ensure cues are uncorrelated—81 here compared with 64 previously.) In this new probabilistic environment, the cue with the highest EV had the lowest success (Cue B) giving a clearer distinction between the orders prescribed by success/ECB and EV. Cue parameters (shown in Table 6) were such that for one or two cues examined, a Bayesian strategy was more accurate than an accuracy strategy (with absolute differences of 6% and 2%, respectively, calculated from rational analysis). In contrast, if three cues are examined, the advantage for the Bayesian strategy disappears, and the accuracy strategy outperforms Bayesian strategy by a small margin (1%). This indicates that there is a small advantage to EV-based search when you know you can search for more information, with search directed by success/ECB being clearly advantageous when cue search is highly limited.

Each participant encountered three blocks of 81 trials: a training block and the two test blocks that composed the known limit and unknown limit conditions. In the training block, all cue information was provided simultaneously. Participants were told that they would later have to make choices without having all the information available and that they should use the training trials to learn which information would be most useful for this. Trials in each test block were divided into three levels of *amount of information*; participants could acquire one, two, or three cues before making their choice. Equal numbers (27) of each trial were experienced, and the order of different types was assigned randomly but fixed across participants. In the known limit condition, participants were told how many pieces of information they could acquire via a message shown on the screen from the start of each trial. In the unknown limit condition, when the maximum number of cues permitted on that trial had been acquired, the message “Choose now” was shown on the screen. The programming of the experiment did not permit participants to acquire more than the permitted number of cues. Points earned in the two test blocks were converted into money at a rate of 1,000 points to U.K. £1.00, and were added to the participants’ turn-up fee of U.K. £1.00.

As previously, the screen position of cues was counterbalanced across four configurations and was fully crossed with the counterbalancing of the order of known–unknown limit conditions.

Results and Discussion

Proportions of Information Examined: The Effects of Cue and Certainty

To determine whether patterns of cue examination varied between known and unknown conditions we examined the effect of cue (four cues, A–D) and certainty (known and unknown limits) for each level of amount of information (separate analyses for one, two, or three cues). All main effects of certainty were nonsignificant ($p > .27$), as were all two-way interactions ($p > .15$). Hence, in later analyses data were collapsed across both certainty conditions (and, for brevity, Table 7 summarizes the mean acquisition proportions accordingly).

Trials with one cue. The main effect of cue was significant, $F(3, 93) = 6.94, p < .001$, indicating that cues differed in how frequently they were examined. Tukey’s HSD post hoc tests revealed that Cue C (greatest success/ECB and Pearson’s r) was examined significantly more frequently than each of the other three cues (all $p < .01$).

Trials with two cues. The main effect of cue was significant, $F(3, 93) = 3.63, p = .016$. Tukey post hoc tests revealed that Cue C was bought significantly more frequently than Cue A (lowest Pearson’s r and EV; $p < .05$).

Trials with three cues. The main effect of cue was significant, $F(3, 93) = 3.50, p = .019$. Tukey post hoc tests revealed that Cue C was bought significantly more frequently than Cue A ($p < .01$).

Proportions of Information Examined: The Effect of Cue and Amount of Information

Appropriate adaptation to the probabilistic structure of the environment would predict that participants would alter which cues they examine most when they are forced to examine fewer (or allowed to examine more) cues. Under such circumstances, the amount of information would interact with cue, such that cues with high EV were more likely to be chosen (in comparison with other cues) when more cues could be examined. Two-way ANOVA was used to examine the pattern of cue examination as a function of the amount of information that participants could examine. The proportion of information examined was the dependent variable, and cue (four levels, A–D) and amount of information (three levels, one to three cues) were the independent variables. Amount of information is uninteresting in its own right (as proportions must increase if more cues are examined); however, including this factor permits examining the two-way interaction to determine whether patterns of cue examination change meaningfully as more information is examined.

Table 6
Experimental Environment for Experiment 3

Measure	Cue A	Cue B	Cue C	Cue D	Defines the search order	Strategy searching in this order
DR	.494	.198	.444	.346	A > C > D > B	
EV	.697	.829	.785	.747	B > C > D > A	Accuracy
Success/ECB	.597	.565	.627	.585	C > A > D > B	Bayesian
Pearson’s r	.276	.292	.381	.289	C > B > D > A	Correlational

Note. DR = discrimination rate; EV = ecological validity; ECB = expected change in belief.

Table 7
 Experiment 3: Mean (SD) Proportion of Trials on Which Each Cue Is Examined and z for Specified Trends in Cue Proportions Examined

Amount of information	Proportions purchased				Order examined by Page trend test				
	Cue A	Cue B	Cue C	Cue D	Search frequency	DR	EV	Success/ECB	Pearson r
One cue	.16 (.23)	.19 (.29)	.48 (.38)	.17 (.24)	C > B > D > A	0.74	0.92	2.51**	2.63**
Two cues	.38 (.35)	.43 (.34)	.68 (.33)	.47 (.37)	C > D > B > A	0.12	1.23	1.90**	2.63**
Three cues	.59 (.34)	.70 (.29)	.86 (.22)	.70 (.30)	C > B > D > A	0.37	1.38	2.42**	3.09**
All trials	.38 (.27)	.44 (.26)	.68 (.28)	.45 (.25)	C > D > B > A	0.34	1.53	2.60**	3.40***

Note. DR = discrimination rate; EV = ecological validity; ECB = expected change in belief.
 ** $p < .01$. *** $p < .001$.

Consistent with prior analyses, the main effect of cue was significant, $F(3, 93) = 5.89$, $p = .001$. Tukey's HSD post hoc tests revealed that Cue C (highest Pearson's r and success/ECB) was examined more frequently than each of the other three cues ($p < .05$ for Cues B and D; $p < .01$ for Cue A). The two-way interaction was not significant, $F(6, 186) = .972$, $p = .446$, thus, patterns of cue examination were relatively constant no matter how many cues could be examined.

Tallies of participants' most favored cue indicated that Cue C (highest Pearson's r and success/ECB) was the most frequently examined cue for 19.5 participants (61%). Cues A, B, and D were examined most frequently by 3.5, 5, and 4 participants, respectively.

Patterns of cue examination were further examined using Page's nonparametric trend test (see Table 7). For any number of cues acquired, cue examination is best described by Pearson's r , then success/ECB, then EV, and then DR. Acquisition proportions are significantly related to the orders prescribed by the correlational and Bayesian criteria, but not to other orders. The difference in trend strength between Pearson's r and EV approached significance—two-tailed p for the difference in trend strength being .087, .162, .087, and .061 for one, two, and three cues, and all trials combined, respectively. Thus, the evidence is good that the correlational criterion predicts the pattern of cue acquisition more effectively than the accuracy criterion.

Summary

Consistent with Experiments 1 and 2, there is clear evidence that the frequency with which a cue is examined is related to its importance as defined by correlational and Bayesian criteria. However, there is no evidence favoring adaptation to the mix of DR and EV as a function of the number of cues available, or whether or not it is known in advance how many cues can be acquired.

General Discussion

When making decisions we are often faced with the need to trade-off the financial, cognitive, or opportunity costs of acquiring information against the perceived improvements in accuracy from obtaining extra information. The results of three experiments suggest that people are sensitive to the (sometimes competing) requirements to acquire information to enhance the accuracy of their

decisions while limiting the amount of information that they obtain.

Using Cue-Outcome Relationships to Order the Search Through Cues: Which Criterion Is Best?

We took a novel and integrative approach by examining three separate criteria for assessing the value of information carried by a cue: the Bayesian, correlational, and accuracy criteria. Results indicated that each one has some explanatory power in accounting for the frequency with which cues are searched. Specifically, if cues are ordered from highest to lowest by the frequency with which they are examined, this ranking is related to the descending order of informativeness specified by each of these criteria. All three criteria are of some value, but which is best? A reliable pattern was seen across the three experiments. The correlational criterion was best at describing patterns of search, followed by the Bayesian criterion, and then the accuracy criterion (EV-directed search).

The different rank orders specified by the three criteria were nonindependent. This was an inevitable feature of our design given that Pearson's r and success/ECB are monotonic increasing functions of EV and DR. Nonetheless, the ranking of explanatory power for these criteria (correlational > Bayesian > accuracy) was apparent in three experiments using different cue-outcome environments and different task characteristics, suggesting that the results indicate a real and robust effect. (Under the null hypothesis that all three criteria are equally good, the probability of observing an identical ranking in three experiments is 1 in 36)

The implication of the success of the correlational and Bayesian criteria in predicting the search through cues implies that both EV and DR are important features for determining how people search for information in probabilistic environments. The advantage of the correlational criterion over the Bayesian criterion implies that the variation in EV across cues seems to carry greater impact than the variation in DR. (Pearson's r is more closely related to EV than is success/ECB. Success/ECB increases with EV and DR, whereas Pearson's r increases with EV and the square root of DR [see Equations 1b and 2]. Hence, variation in DR has less impact on Pearson's r than on success/ECB, leaving the ranking of cues by Pearson's r more closely allied to that for EV.) EV can be thought of as the accuracy component of the information carried by the cue, whereas DR can be considered as the speed component (if

cues with high DRs are examined, informed decisions can be made more quickly because such cues are more likely to discriminate between options). Thus, the search through cues is sensitive to cue properties that enhance accuracy and reduce the amount of information that is considered. Our rational analyses show that if there is a cost to acquiring cues, both considerations are entirely appropriate.

The rational analyses for all experiments also indicated that a switch from an accuracy to a correlational or Bayesian search order was advantageous when (a) the cost of information increased (Experiment 1), (b) the time available for decisions decreased (Experiment 2), and (c) the availability of information decreased (Experiment 3). Despite these prescribed changes in the optimal strategy we did not see concomitant shifts in the acquisition frequency measure. A plausible explanation for this resistance to shift can be framed in terms of setting an aspiration level for earnings. The manipulation of task characteristics (cue cost, time, uncertainty) was within subjects in the three experiments, all experiments involved numerous repeated choices, and participants' earnings were present on-screen throughout experimental sessions. The combination of these factors may have led participants to set a personal aspiration level and then to "lock-in" to a particular pattern of acquisition that enabled them to maintain such a level. Indeed, one of the earliest choice heuristics proposed in the decision-making literature—*satisficing* (Simon, 1955)—incorporates an aspiration level as a key component.

We examined this possibility and found some support for it within the data. For Experiment 1, we examined cue purchase for the two lower cost conditions (3- and 10-point conditions combined) as a function of whether participants did one of these two conditions first or not. Those that did so bought more cues in these lower cost conditions (61% vs. 47% of the information available to them) than those who did not. We also examined cue purchase for the two higher cost conditions (17- and 24-point conditions combined) as a function of whether participants did one of these two conditions first or not. Those that did so bought a slightly higher proportion of the cues available to them (39% vs. 36%). Thus, those that began the experiment with a condition in which information was expensive, only modestly increased their purchase of information when it became cheaper (from 39% to 47%), whereas participants who started with cheap information were far more responsive to an increase in information cost (dropping from 61% to 36%). In Experiment 2, an equivalent pattern was observed. Those that did the high time constraints condition first examined 54% of the cues in that condition and only increased their acquisition to 58% under low time constraints. In contrast, those that experienced conditions in the reverse order examined 67% of the information under low time constraints and 57% under high time constraints. Thus, those that began with high time constraints acquired less information overall and were less sensitive to the time constraint manipulation. (The equivalent analysis of the amount of information acquired is, of course, precluded in Experiment 3 – as the number of cues that could be examined was fixed on any given trial.) Thus, in Experiments 1 and 2, participants who first experienced costly or difficult-to-acquire information were content to acquire only a little more information when information became cheaper or easier to acquire. In contrast, when these cheaper or easier conditions were the first to be experienced, participants devoted more of their resources (time or money) to acquiring predecisional information. These findings are consistent with the notion that the initial

conditions that participants experience play an important role in guiding where they set their aspiration level for the amount of information that they need to acquire. Crucially, the setting of this level impacts on the accuracy and the payoff that they can expect to obtain. These order effects seem to be confined to the gross amount of information that participants acquire—no dependable changes in the relative popularity or importance of cues as a function of condition order were apparent.

Because suboptimal search orders only had a relatively small impact on overall accuracy, and hence amount earned (e.g., see Figure 1), it is likely that a given aspiration level would have been unaffected by the switches between search orders. For example, switching from the accuracy–full to the correlational–full search strategy in the most expensive cue condition of Experiment 1 would have increased earnings by only 5.3 points per trial—a difference in magnitude of outcome that may only affect a few participants.

Much more likely to impact on a given aspiration level would be failure to adapt at the gross level of the raw amount of information acquired. For example, buying all the cues instead of only the cue of highest EV on each trial (i.e., not even the optimal one in such circumstances) in the 24-point cue cost condition of Experiment 1 would lead to a decrease in earnings of 55.6 points per trial. Only the most unambitious participant could afford to ignore a difference of this magnitude. Consistent with this explanation, we did indeed find adaptation in the amount of information acquired under different cost conditions. We discuss below the theoretical implications of finding changes in the amount of information acquired, but first, we consider the possibility that participants did not switch search strategies simply because they did not learn enough about individual cue parameters.

It is true that the differences between cues on our parameters of interest were often quite small. Search patterns were not identical across participants, which would most naturally be interpreted as evidence of considerable noise in cue learning resulting from the difficulty of distinguishing between cues. Nonetheless, participants' patterns of search indicated that they were able to distinguish between cues. Group level tests of significance showed that participants distinguished between cues that differed by no more than 0.2 in their correlation with the criterion. In Experiments 2 and 3, cues that differed by only 0.1 in their correlation with the criterion were differentiated. We find this ability to differentiate between cues quite impressive, and consider it unsurprising that there should be some noise in learning which cues are more useful than others.

It remains to be seen whether any of the variation in search orders between different participants represents stable individual differences (as opposed to random noise in learning). It is not inconceivable that some individuals will pay greater attention to the accuracy characteristics of cues (i.e., be more influenced by variation in EV), whereas others will be more concerned to enhance the speed of decisions (i.e., be more influenced by variation in DR). Consistent with such a view, in Experiments 1 and 2, the cue with highest EV but lowest DR showed the greatest variation in frequency of acquisition among the cues. This would be anticipated if participants have qualitatively different reactions to EV and DR. Such individual differences are plausible, as cognitive dispositions have been shown to explain significant variance in the performance of a range of thinking and reasoning tasks (Stanovich & West, 1998). Specifically, individual

differences in impulsivity and willingness to seek disconfirming evidence described by Stanovich and West (1998) might account for some variation in the number of cues that participants examine. Such an effect could be mediated by a propensity to search for cues with higher, or lower, DRs.

Our rational analyses indicated that there was a clear (though small) financial cost in failing to alter cue search patterns as task conditions change. Do such failures reflect a general limitation of human behavior, and, if so, what implications might they have for some of the real-world judgment tasks that we have alluded to throughout the article? Consider again the example of assessing the insurance risk of two clients (depicted in Table 1). An insurance company with unlimited time and resources would be advised to search first for information about a client's previous convictions, because this cue is likely to have high ecological validity and thus afford an accurate claims prediction. Contrast this with a company that is under pressure to process a high number of clients each day—where should they begin their search? Clearly the sex cue would be a better place to start. Although it may not be as accurate a predictor, it is very likely to discriminate between clients and thus provide a basis for making the assessment.

Our results suggest that, in both circumstances, there is a tendency to maintain the same search pattern, beginning by examining a cue with a good mix of EV and DR—perhaps the age cue in this example. With such an applied example, it is obviously difficult to calculate the impact of failing to adjust search on the overall accuracy of claim predictions; nevertheless, we can speculate on the effect of locking in to a particular pattern. Presumably for insurance companies there will always be a trade-off between the number of clients processed and the accuracy of the predicted claims. A company that exhaustively researched every client could not sustain itself. It would therefore be important to know as much as possible about the cue-outcome relations in a given environment. This would allow rational analyses to be conducted which would facilitate the development of predecisional search strategies that would balance accuracy, workload, and profit for the company. Our findings can be seen as a first step toward understanding how such search strategies might be characterized and systematically investigated. An intriguing next step would be to examine the extent to which the search patterns we have observed in our experimental tasks generalize to real-world decisions like the one illustrated in Table 1.

Switches Between Strategies or Adjustments of a Threshold?

Although it was not our principal focus, we were interested in the effect of our various task conditions on the gross amount of information acquired. Here, we briefly consider two competing interpretations of the effects we observed. In Experiments 1 and 2, we found evidence that the increase in the cost (monetary and time) of information led to a decrease in the average amount of information acquired. Despite the considerable amount of consistent evidence demonstrating changes in the amount of information acquired when task conditions alter (e.g., Bröder, 2000; Connolly & Serre, 1984; Newell & Shanks, 2003; Payne et al., 1993), there is tension between two distinct approaches to explaining these changes. On the one hand, changes in behavior are described in terms of qualitative changes in the choice strategy or heuristic that

is used (Beach & Mitchell, 1978; Gigerenzer et al., 1999; Payne et al., 1993; Rieskamp & Hoffrage, 1999). On the other, changes in behavior can be viewed as adaptations or adjustments in the application of a general mechanism (Newell, 2005).

Under the heuristic account, there is thought to be a cost–benefit trade-off involved in selection between strategies that are more labor intensive (e.g., WADD) and those that are more frugal but only sometimes less accurate (e.g., a one-reason choice strategy; Gigerenzer & Goldstein, 1996). Under the single-mechanism account, the assumption is that rather than taking a predetermined quantity of information (e.g., one discriminating cue in the case of TTB), sequential sampling of each option occurs until evidence sufficient to exceed a given threshold has been accumulated (e.g., Busemeyer & Townsend, 1993; Lee & Cummins, 2004; Nosofsky & Palmeri, 1997; Ratcliff, 1978).

Although with our data we are not able to discriminate conclusively between these alternative accounts, we note that an evidence-accumulation model recently proposed by Lee and Cummins (2004) goes some way to formally defining the advantages of single models over “toolboxes” of heuristics. A common rebuttal to the single-mechanism approach is that the use of parameterized thresholds increases their complexity in comparison with simple-heuristic models, and it is this complexity which leads to improvements in accuracy (e.g., Roberts & Pashler, 2000). The Lee and Cummins model uses a Bayesian approach for estimating cue importance, combined with a random walk sequential sampling mechanism for accumulating evidence. Although the model has more parameters than a one-reason strategy (e.g., TTB), it is more successful in capturing interparticipant individual differences as well as being preferred over TTB by a standard measure of model selection (minimum description length; see Pitt, Myung, & Zhang, 2002). We believe that more work of this nature is necessary to determine the empirical validity of these two contrasting approaches to understanding and explaining human inferential judgment (see Newell, 2005).

Conclusion

In this article, we have provided the means to integrate previous research on predecisional information search. We have shown that of the methods used previously for assessing the value of information, cue-outcome correlation best predicts which cues are acquired most frequently, though, perhaps inevitably, learning seems subject to noise. Rational analysis indicates that sensitivity to this characteristic is beneficial when information has some cost, as both the speed and accuracy with which decisions can be made is reflected by cue-outcome correlations. We observed that even when changes in task characteristics prompted participants to alter the number of cues they acquired, the cue-outcome correlation always provided the best criterion for predicting the search through cues.

References

- Albert, D., Aschenbrenner, K. M., & Schmalhofer, F. (1989). Cognitive choice processes and the attitude-behavior relation. In A. Upmeyer (Ed.), *Attitudes and behavioral decisions* (pp. 61–99). New York: Springer.
- Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the

- selection of decision strategies. *Academy of Management Review*, 3, 439–449.
- 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.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62, 193–217.
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic–cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100, 432–459.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Connolly, T., & Serre, P. (1984). Information search in judgment tasks: The effects of unequal cue validity and cost. *Organizational Behavior and Human Performance*, 34, 387–401.
- Edwards, W. (1965). Optimal strategies for seeking information. *Journal of Mathematical Psychology*, 2, 312–329.
- Fried, L. S., & Peterson, C. R. (1969). Information seeking: Optional versus fixed stopping. *Journal of Experimental Psychology*, 80, 525–529.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.
- 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., & ABC Research Group (Eds.). (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Hammond, K. R., & Stewart, T. R. (Eds.). (2001). *The essential Brunswik: Beginnings, explications, applications*. New York: Oxford University Press.
- Holzworth, J. R. (2001). Multiple cue probability learning. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications*. (pp. 348–350). New York: Oxford University Press.
- Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury Press.
- Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211–228.
- Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the ‘take-the-best’ and the ‘rational’ models. *Psychonomic Bulletin & Review*, 11, 343–352.
- Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, 27, 986–1005.
- 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.
- Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, 52, 29–71.
- Newell, B. R. (2005). Re-visions of rationality? *Trends in Cognitive Sciences*, 9, 11–15.
- Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies in decision making: The success of success. *Journal of Behavioral Decision Making*, 17, 117–137.
- 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. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104, 266–300.
- Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world: Essays on the cognitive science of human reasoning*. Hove, United Kingdom: 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*. New York: Cambridge University Press.
- Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, 109, 472–491.
- Rakow, T., Hinvest, N., Jackson, E., & Palmer, M. (2004). Simple heuristics from the adaptive toolbox: Can we perform the requisite learning? *Thinking & Reasoning*, 10, 1–29.
- Rapoport, A., & Wallsten, T. S. (1972). Individual decision behavior. *Annual Review of Psychology*, 23, 131–179.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- 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). New York: Oxford University Press.
- Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107, 358–367.
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99–118.
- Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 6, 649–744.
- Smedslund, J. (1955). *Multiple-probability learning*. Oslo, Norway: Akademisk Forlag.
- Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, 127, 161–188.
- Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral & Brain Sciences*, 23, 727–780.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281–299.

Appendix

Expressing Pearson's r as a Function of Ecological Validity (EV) and Discrimination Rate (DR)

Definitions

For a cue with discrimination rate (d) and ecological validity (v), the success (s) can be given by

$$s = (v - 0.5)d + 0.5.$$

Purpose

Below we demonstrate that Pearson's r and success define similar hierarchies of cues, as it can be shown that

$$r = \frac{2s - 1}{\sqrt{d}},$$

or equivalently that

$$r = (2v - 1)(\sqrt{d}).$$

To demonstrate this, we first derive an intermediate result (Result A below).

Demonstration of Result A

Consider P paired choices between Objects A and B, where correct choices (R) and incorrect choices (W) are made. If Objects A and B are defined arbitrarily, we expect both correct and incorrect choices to be divided equally between the two objects. Define X as the direction the cue points to, which takes values $+1$ (cue points to Object A), 0 (cue points to neither object), or -1 (cue points to Object B). Define Y as the outcome of the target event, which takes values $+1$ (Object A has the higher criterion value) or -1 (Object B has the higher criterion value).

The general expression for Pearson's r is (Siegel & Castellan, 1988)

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} \tag{1}$$

where $x = X - \bar{X}$, and $y = \bar{Y} - Y$.

By symmetry both \bar{X} & \bar{Y} are 0, giving the following cross-tabulated frequencies for x and y .

Values of x	Values of y		Total
	+1	-1	
+1	$R/2$	$W/2$	$(R + W)/2$
0	$(P - R - W)/2$	$(P - R - W)/2$	$(P - R - W)$
-1	$W/2$	$R/2$	$(W + R)/2$
Total	$P/2$	$P/2$	P

Using Equation 1:

$$r = \frac{(R/2)(+1) + ((P - R - W)/2)(0) + (W/2)(-1) + (W/2)(-1) + ((P - R - W)/2)(0) + (R/2)(+1)}{\sqrt{\{((R + W)/2)(+1)^2 + (P - R - W)(0)^2 + ((W + R)/2)(-1)^2\}\{(P/2)(+1)^2 + (P/2)(-1)^2\}}}$$

$$r = \frac{(R/2) - (W/2) - (W/2) + (R/2)}{\sqrt{\{((R + W)/2) + ((W + R)/2)\}\{(P/2) + (P/2)\}}}$$

$$r = \frac{R - W}{\sqrt{(P(R + W))}} \tag{Result A}$$

Result A also defines Spearman's rank correlation coefficient (ρ) for the environments that we consider, as ρ can be calculated by applying the formula for Pearson's r to the ranks of the scores (Howell, 2002; Siegel & Castellan, 1988). (r and ρ would only differ were we to score the direction that the cue points to in some asymmetrical fashion, presumably reflecting that pointing to Object A was somehow more, or less, significant than pointing to Object B. We can think of no instance where this would seem reasonable.)

(Appendix continues)

Demonstration of Main Result

Starting with Martignon and Hoffrage's (1999, 2002) definition of success in terms of the number of correct (R) and incorrect (W) choices from a set of P pairs:

$$\begin{aligned} s &= \frac{R + 0.5(P - R - W)}{P}, \\ \Rightarrow 2s &= \frac{2R + P - R - W}{P}, \\ \Rightarrow 2s &= \frac{P + R - W}{P}, \\ \Rightarrow 2s &= 1 + \frac{R - W}{P}, \\ \Rightarrow P(2s - 1) &= R - W. \end{aligned} \tag{2}$$

Substituting Equation 2 into Result A:

$$\begin{aligned} r &= \frac{P(2s - 1)}{\sqrt{(P(R + W))}}, \\ \Rightarrow r &= \frac{(2s - 1)\sqrt{P}}{\sqrt{(R + W)}}. \end{aligned} \tag{3}$$

By definition, the discrimination rate is

$$\begin{aligned} d &= \frac{R + W}{P}, \\ \Rightarrow \frac{1}{\sqrt{d}} &= \frac{\sqrt{P}}{\sqrt{(R + W)}}. \end{aligned} \tag{4}$$

Substituting Equation 4 into 3 above:

$$r = \frac{2s - 1}{\sqrt{d}}. \tag{5a}$$

Using the definition of success above, $s = (v - 0.5)d + 0.5$, and substituting this into Equation 5a:

$$\begin{aligned} r &= \frac{2\{(v - 0.5)d + 0.5\} - 1}{\sqrt{d}}, \\ r &= \frac{(2v - 1)d + 1 - 1}{\sqrt{d}}, \\ r &= (2v - 1)(\sqrt{d}). \end{aligned} \tag{5b; Equation 2 in the Introduction}$$

Received August 5, 2004
Revision received February 7, 2005
Accepted March 8, 2005 ■