Psychonomic Bulletin & Review
2004, 11 (2), 343-352

Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models

MICHAEL D. LEE and TARRANT D. R. CUMMINS
University of Adelaide, Adelaide, South Australia, Australia

An evidence accumulation model of forced-choice decision making is proposed to unify the fast and frugal *take the best* (TTB) model and the alternative *rational* (RAT) model with which it is usually contrasted. The basic idea is to treat the TTB model as a sequential-sampling process that terminates as soon as any evidence in favor of a decision is found and the rational approach as a sequential-sampling process that terminates only when all available information has been assessed. The unified TTB and RAT models were tested in an experiment in which participants learned to make correct judgments for a set of real-world stimuli on the basis of feedback, and were then asked to make additional judgments without feedback for cases in which the TTB and the rational models made different predictions. The results show that, in both experiments, there was strong intraparticipant consistency in the use of either the TTB or the rational model but large interparticipant differences in which model was used. The unified model is shown to be able to capture the differences in decision making across participants in an interpretable way and is preferred by the minimum description length model selection criterion.

A simple but pervasive type of decision requires choosing which of two alternatives has the greater (or the lesser) value on some variable of interest. Examples of these forced-choice decisions range from the everyday (e.g., deciding whether a red or a green curry will taste better for lunch), to the moderately important (e.g., deciding whether Madrid or Rome will provide the more enjoyable holiday), to the very important (e.g., deciding whether cutting the red or the black wire is more likely to lead to the destruction of the world).

The Rational Approach

One approach to modeling the way people make these decisions, often referred to as the *rational approach*, assumes that all of the relevant available information is combined in some (near) optimal way. At lunchtime, this means that knowledge of the ingredients of the different curries, previous experiences with the two curry types, current sensory information about the available curries, and a range of other relevant information must be weighed and combined to give an overall preference. Simon (1976) described this approach as *substantively rational*, because its overarching goal is to maximize the utility of the decision made, regardless of the efficiency of the processes required to make the decision. This rational approach is often viewed as a normative theory of decision making and is central to the decision and utility theoretic frameworks widely used in the physical sciences and in the behavioral sciences, such as psychology and economics (see Doyle, 1999, for an overview).

A large number of well-known and successful models in cognitive psychology aim for substantive rationality. For example, the MINERVA 2 model of memory retrieval (Hintzman, 1984) uses the sum of every memory trace, weighted by the frequency of each trace, to remember or reconstruct stimuli. Similarly, the ALCOVE model of category learning (Kruschke, 1992) makes categorization decisions by potentially considering the weighted sum of evidence for each category alternative provided by every stimulus in a domain, and the same is true of the closely theoretically related context model (Medin & Schaffer, 1978) and generalized context model (Nosofsky, 1984). There are also various Bayesian cognitive models, including accounts of generalization (Myung & Shepard, 1996; Shepard, 1987) and concept learning (Tenenbaum, 1999; Tenenbaum & Griffiths, 2001), that integrate across prior-weighted probability densities to determine response probabilities and, so, strive for substantive rationality in a very direct way. Finally, there are substantively rational psychological models—most notably, Anderson's (1990, 1991, 1992) rational model—that introduce time and memory constraints into the criteria for decision making but continue to allow for the weighting and combination of all of the relevant available evidence to optimize decisions under these criteria.

This research was supported by Australian Research Council Grant DP0211406. We thank Helen Braithwaite, Jerry Busemeyer, Marcus Butavicius, Len Dalgleish, Gerd Gigerenzer, Daniel Navarro, Robert Nosofsky, Douglas Vickers, and Chris Woodruff for helpful comments and Jo Carbone for assistance with the experiments. Correspondence concerning this article should be addressed to M. D. Lee, Department of Psychology, University of Adelaide, Adelaide, SA 5005, Australia (e-mail: michael.lee@adelaide.edu.au).

The Fast-and-Frugal Approach

In developing their *fast-and-frugal* approach to modeling human decision making, Gigerenzer and Todd (1999; see also Gigerenzer & Goldstein, 1996; Todd & Gigerenzer, 2000) challenged the rational approach and emphasized what Simon (1976) describes as *procedural rationality.* They argued that because human decision-making processes evolved in competitive environments, they needed to be fast, and because they evolved in changeable environments, they needed to have the robustness that comes from simplicity. This emphasis on the role of the environment in shaping human decision processes follows more general ecological approaches to psychology (e.g., Brunswik, 1943; Simon, 1956, 1982) and suggests that understanding human decision making requires understanding not just mental processes, but also the external task environment and its interaction with mental processes.

As Gigerenzer and Todd (1999) have argued, the fact that environments are not arbitrary means that they can play a role in supporting (or confounding) human decision making. For example, in an environment in which one piece of information in a stimulus is highly predictive of the remaining pieces of information and the search for additional information is an effortful process, it is adaptive to consider only the first piece of information. Similarly, in an environment of diminishing returns, in which each successive piece of information provides less information than do previous pieces, it makes sense to base decisions on the first few pieces of information. Gigerenzer and Todd showed that many real-world stimulus domains have these sorts of information structures and developed a number of cognitive models—including the *take the best* (TTB) model of forced choice, the *QuickEst* model of value estimation, and the *categorization by elimination* model of categorization—that make inferences by assuming the presence of environmental regularities.

Unifying the TTB and the Rational Approaches

Gigerenzer and Todd (1999) have mounted an impressive theoretical case for the ecological rationality of their fast-and-frugal models, showing that they often match or outperform competing rational models for decision accuracy when tested in real-world domains and do so more quickly and with fewer cognitive resources. Empirical comparisons of rational and fast-and-frugal models (e.g., Bröder, 2000; Newell, Weston, & Shanks, 2003), however, have reached more equivocal conclusions. In these experiments, people make decisions for which the fast-and-frugal and the rational models give different predictions. The standard finding is that only subsets of people behave in a way consistent with using a noncompensatory fast-and-frugal decision-making process. From this finding, the theoretical conclusions drawn by Bröder focused on the need to understand how individual differences influence the selection of alternative decision-making strategies for different tasks.

Although there is no inherent contradiction in arguing that different people make decisions in different ways under different conditions, it is a less than completely satisfying conclusion. The finding seems empty without some attempt to provide a unifying account of *why* people use different strategies, *when* they use different strategies, and *how* they manage to use different strategies with the same cognitive apparatus. Pinker (1997) put the concern more bluntly: "In psychology, invoking 'strategies' to explain funny data is the last refuge of the clueless" (p. 282).

The goal of this article is to develop and evaluate a unifying theoretical account of the TTB model and its rational alternative. The theoretical unification is achieved by viewing both as special cases of a sequential-sampling decision-making process. The basic idea is that the TTB model corresponds to the case in which the first piece of sampled evidence that favors one decision is sufficient, whereas the rational approach requires all of the available information to be sampled before a decision is made. By setting different threshold levels of evidence required for decision making, both the TTB and the rational models become special cases of a more general evidence accumulation account.

THE TWO DECISION MODELS

The Take the Best Model

In developing their TTB decision model, Gigerenzer and Todd (1999) followed a substantial body of other cognitive modeling (e.g., Medin & Schaffer, 1978; Tversky, 1977) by representing stimuli in terms of the presence or absence of a set of discrete features or properties that they called *cues*. This means that, for a stimulus domain with k cues, a Stimulus A is represented by a set of binary variables $a_1, \ldots, a_k$, where $a_i = 1$ if Stimulus A has the ith cue and $a_i = 0$ if it does not. The TTB decision model, like many described by Gigerenzer and Todd, is based on the validities of these cues. Validities estimate the rate at which a cue makes correct decisions, in those cases in which it distinguishes between the two alternatives.

Formally, this means that the validity, v_i of the ith cue is defined as

$$v_i \equiv p(A > B \mid a_i = 1, b_i = 0),$$

where the stimulus that has the ith cue has been denoted A and the stimulus that does not is B. Because either $A > B$ or $A < B$, the value $1 - v_i$ gives $p(A < B \mid a_i = 1, b_i = 0)$. It is also always possible to define cues so that every validity is greater than or equal to .5. If a cue ever has a validity less than .5, replacing it with its complement (i.e., defining stimuli in terms of the absence of the cue, rather than its presence) encodes exactly the same information and changes the validity to a value greater than .5.

Given a set of stimuli defined in terms of a set of cues and a known ranking of the stimuli in terms of the decision variable, it is possible to estimate cue validities by considering every possible pairing of stimuli. Gigerenzer and Todd (1999) adopted a frequentist approach to estimation, calculating cue validities as the proportion of correct inferences made across those stimulus pairs where a cue discriminates (i.e., one stimulus has the cue, and the other does not). This means that an estimate, $\hat{v}_i$, of cue validity is given by the proportion

$$\hat{v}_i = \frac{\text{number of correct decisions made by } i\text{th cue}}{\text{number of decisions made by } i\text{th cue}}.$$

Although this approach to estimating validities is often adequate, it does not take into account how often the cue discriminates when its validity is calculated. For example, a cue that gets its one and only decision right across all stimulus pairs (i.e., gets 1 correct out of 1) will have the same validity of 1 as a cue that makes 150 decisions and gets them all correct. It seems clear that the second of these cues, where a large data sample is available, is likely to be more valid than the first, but the frequentist approach is not sensitive to the difference.

To overcome this problem, Lee, Chandrasena, and Navarro (2002; see also Lee, Loughlin, & Lundberg, 2002) proposed a Bayesian approach to estimating cue validities. The basic idea is to establish a prior distribution for the validity of each cue and to modify these priors, using the evidence provided by a cue making correct or incorrect decisions. As a cue makes more correct decisions, higher values for its validity become more likely; as it makes more incorrect decisions, lower values for its validity become more likely. Bayes's theorem formally describes the way in which the prior beliefs are modified by data to give a probability distribution over the range [0, 1] of possible cue validities. Defining the estimated Bayesian validity of a cue as the mean of this distribution and assuming a uniform prior, a standard result from Bayesian statistics (e.g., Gelman, Carlin, Stern, & Rubin, 1995, p. 31) gives

$$\hat{v}_i = \frac{\text{number of correct decisions made by } i\text{th cue} + 1}{\text{number of decisions made by } i\text{th cue} + 2}.$$

As a cue makes more decisions, the frequentist and Bayesian validity estimates converge toward the same value. When the available data are limited, however, the Bayesian approach is sensitive to the sample size and provides a better measure. For example, the cue that made one correct decision has Bayesian validity $(1 + 1)/(1 + 2) \approx .67$, whereas the cue that made 150 correct decisions has Bayesian validity $(150 + 1)/(150 + 2) \approx .99$.

Once validities for each cue have been estimated, the TTB decision model is straightforward. For a given pair of stimuli, the cue with the highest validity is examined. If this cue discriminates between the stimuli (i.e., one stimulus has the cue, and the other does not), the favored stimulus is chosen, and no further cues are examined. If, however, the cue does not discriminate, the cue with the next highest validity is examined. This process continues until either a decision is made or all of the cues have been exhausted (because the stimuli have identical cue representations) and, so, a random guess must be made.

The Rational Model (RAT)

The motivation behind substantively rational decision models is to use all of the relevant available information. This can be done by evaluating the probability that Stimulus A is greater than Stimulus B, having considered whether or not the stimuli have each of the cues. If the probability is greater than .5, it is rational to choose Stimulus A; if the probability is less than .5, it is rational to choose Stimulus B; if the probability is exactly .5, it is rational to guess.

Formally, the log-odds that Stimulus A is greater than Stimulus B, given their cue representations, is written as

$$L_{AB} = \ln \frac{p(A > B \mid a_1, \ldots, a_k, b_1, \ldots, b_k)}{p(A < B \mid a_1, \ldots, a_k, b_1, \ldots, b_k)}.$$

Using Bayes's theorem, this may be rewritten as

$$L_{AB} = \ln \frac{p(A > B)}{p(A < B)} + \ln \frac{p(a_1, \ldots, a_k, b_1, \ldots, b_k \mid A > B)}{p(a_1, \ldots, a_k, b_1, \ldots, b_k \mid A < B)}.$$

Since Stimulus A's being greater than Stimulus B is a priori equally as likely as Stimulus B being greater than Stimulus A [i.e., $p(A > B) = p(A < B) = .5$], the priors do not provide any information, which means the log-odds reduce to the evidence provided by the cues themselves, as follows:

$$L_{AB} = \ln \frac{p(a_1, \ldots, a_k, b_1, \ldots, b_k \mid A > B)}{p(a_1, \ldots, a_k, b_1, \ldots, b_k \mid A < B)}.$$

If the simplifying assumption is made that each of the cues provides an independent source of evidence in making a decision,[1] the required log-odds can be approximated by

$$L_{AB} = \sum_{i=1}^{k} \ln \frac{p(a_i, b_i \mid A > B)}{p(a_i, b_i \mid A < B)}.$$

This result has a straightforward interpretation. The log-odds are found by summing the evidence provided by each of the cues in favor of the alternative decisions. For cues that do not distinguish between the stimuli (i.e., $a_1 = b_i = 1$ or $a_i = b_i = 0$), this evidence will be zero. For those cues that do discriminate, it turns out that the evidence can be expressed in terms of cue validities. When the cue favors Stimulus A (i.e., $a_i = 1$ and $b_i = 0$), the log-odds for that cue are given by $\ln [v_i/(1 - v_i)]$. When the cue favors Stimulus B (i.e., $a_i = 0$ and $b_i = 1$), the log-odds are $\ln [(1 - v_i)/v_i]$, which is the same as $-\ln [v_i/(1 - v_i)]$.

Putting these results together allows the log-odds that Stimulus A is greater than Stimulus B to be written as

$$L_{AB} = \sum_{i \in \mathrm{FA}} \ln \frac{\hat{v}_i}{1 - \hat{v}_i} - \sum_{i \in \mathrm{FB}} \ln \frac{\hat{v}_i}{1 - \hat{v}_i}, \qquad (1)$$

where the first sum is across the cues favoring Stimulus A (the FA set), and the second sum is across the cues favoring Stimulus B (the FB set).

Having calculated these log-odds, the rational decision model is straightforward. If L_{AB} is positive, indicating that the cues provide more evidence for Stimulus *A*'s being greater than Stimulus *B*, this decision should be made. If L_{AB} is negative, Stimulus *B* should be chosen as being the greater. If L_{AB} is exactly zero, a random choice should be made.

A UNIFYING MODEL BASED ON EVIDENCE ACCUMULATION

Sequential-sampling processes have been extensively studied as models of human decision making (e.g., Busemeyer & Rapoport, 1988; Busemeyer & Townsend, 1993; Diederich, 1997; Laming, 1968; Link & Heath, 1975; Nosofsky & Palmeri, 1997; Ratcliff, 1978; Smith, 2000; Vickers, 1979; Wallsten & Barton, 1982), particularly in relation to elementary psychophysical tasks, such as judging which of two lines is longer. Although there are many variants, at their heart sequential-sampling models assume that stimuli are searched for information until sufficient evidence has been accrued to favor one decision or no more information is available.

Figure 1 shows a particular sequential-sampling process, known as a *random walk*, accruing information in making a comparison between two stimuli. Each of the cues is examined, from the highest validity to the lowest, and the evidence provided by that cue is used to update the state of the random walk in favor of choosing Stimulus *A* or Stimulus *B*. The evidence provided by a cue corresponds to its log-odds value, as defined in Equation 1. If Stimulus *A* has the cue and Stimulus *B* does not, the random walk moves toward choosing *A*. If Stimulus *B* has the cue and Stimulus *A* does not, the random walk moves toward choosing *B*. If both stimuli either have or do not have the cue, the state of the random walk is unchanged.

The important observation about Figure 1 is that the TTB and the RAT models correspond simply to different required levels of evidence being accrued before a decision is made. If a very small evidence threshold were set, the sequential-sampling process would choose Stimulus *A*, in agreement with the TTB choice. Alternatively, if a very large evidence threshold were set, the sequential-sampling process would eventually choose Stimulus *B* (because the final log-odds are in its favor), in agreement with the RAT model. In general, if a threshold is small enough that the first discriminating cue is guaranteed to have log-odds that exceed the threshold, sequential-sampling corresponds to the TTB decision model. If a threshold is large enough that it is guaranteed never to be reached, the final log-odds are used to make a forced decision, and sequential sampling corresponds to the RAT decision model.[2]

In this way, both the TTB and the RAT models can be considered as special cases of an evidence accrual model, corresponding to the use of low and high evidence thresholds, respectively. Thus, one unifying way of explaining experimental results in which people make decisions consistent with both fast-and-frugal and rational models, is that all of the people were using an evidence accrual decision process but that different people required different levels of evidence before making their decisions.

Adding to the plausibility of this account is that in many real-world environments, cue validities show sharply diminishing returns (Martignon & Hoffrage, 1999) and, so, only small changes in required evidence will be sufficient to switch between TTB and RAT decision making. For example, the sequential-sampling de-

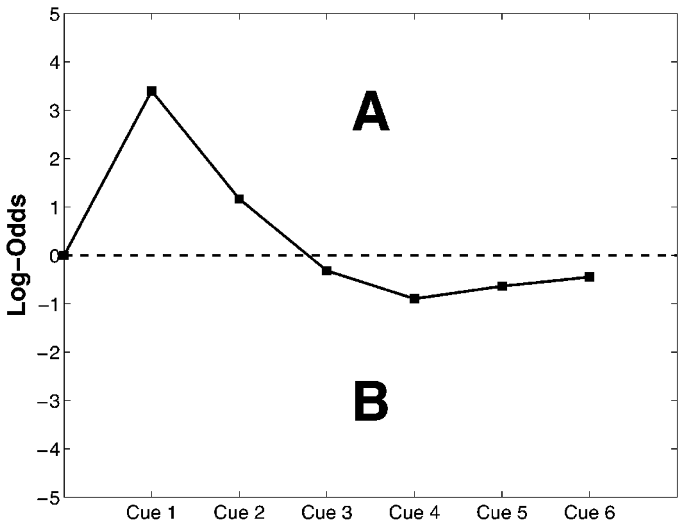

Figure 1. The unified model, using a random walk evidence accumulation decision process. Successive values are shown for the log-odds as each cue is examined from highest validity to lowest.

cision in Figure 1 will be consistent with the TTB model for any evidence threshold less than about 3.5 and will be consistent with the RAT model for any larger value.

EXPERIMENT

The methodology used to evaluate the unified model takes the form of a two-phase decision-making experiment. In the first phase, called the *training* phase, participants learn to make correct decisions, using feedback. During this phase, the participants are never required to make decisions that discriminate between the TTB and the RAT models. In this way, the participants are able to learn about the validities of the various stimulus cues, without the feedback encouraging them to adopt either decision strategy. In the second phase, called the *test* phase, the participants are asked to make a number of decisions for new combinations of stimuli drawn from the same domain but do not receive feedback. These test phase comparisons are designed so that the TTB and the RAT models now do make different predictions about which decision will be made.

Method

Participants. Forty adult participants completed the experiment. There were 16 males and 24 females, with a mean age of 28 years.

Stimulus domain. One potential criticism of previous empirical comparisons (e.g., Bröder, 2000; Newell et al., 2003) is that the stimulus domains used were artificially constructed, despite the emphasis placed by Gigerenzer and Todd (1999) on the role of the structured environment in allowing fast-and-frugal heuristics to operate effectively. For this reason, we used a stimulus domain previously considered in fast-and-frugal modeling (Czerlinski, Gigerenzer, & Goldstein, 1999), based on data originally reported in Johnson and Raven (1973). These data relate to the Galapagos Islands, forming cues by quantizing a number of continuous measures, such as land area, elevation, and so on. The decision variable is a count of the number of species on each island.

Several of the stimuli in this real-world domain had exactly the same cue representation, which would have made a forced-choice decision impossible when they were presented together. To avoid this difficulty, in each of these cases, the stimulus that had the greatest decision variable value was retained, and all of the others were removed. A number of cues were also replaced with their complement to ensure that every cue validity was greater than .5. The final stimulus domain resulting from these modifications is shown in Table 1.

Table 1
The Stimulus Domain, Detailing the Assignment of the Six Cues to Each of the 16 Stimuli (With the Decision Variable Values for Each Stimulus and the Bayesian Validities of the Cues)

Stimulus Number	Cue 1 (.97)	Cue 2 (.90)	Cue 3 (.82)	Cue 4 (.64)	Cue 5 (.56)	Cue 6 (.55)	Decision Variable
1				×			16
2		×			×		18
3			×			×	21
4				×	×		25
5					×		31
6	×				×	×	40
7			×	×	×	×	44
8	×	×		×			51
9	×	×	×			×	62
10	×	×			×		70
11	×	×		×	×	×	97
12	×	×	×	×			104
13	×	×	×	×	×	×	280
14	×	×	×	×		×	285
15	×	×	×		×		347
16	×	×	×	×	×		444

Stimulus presentation. In both the training and the test phases, it is necessary to present the stimuli in a way that conveys information about which cues they have but is not subject to individual differences arising from varying prior exposure. This means that the stimuli must have a surface presentation that is artificial, while maintaining the underlying cue structure that mimics the real environment. This was achieved by depicting each stimulus as a gas in a canister, where the cues corresponded to gas molecules with different colors, and describing the decision variable as the *poisonousness* of the gas. The colors for the cues were selected to encourage interpretation on a nominal level of data scaling, so that no natural ordering would be implied. In addition, the gas displays were animated, so that the molecules moved randomly within the confines of their canister. The animation prevents configural effects from emerging from the superposition of different cues in the same locations within a static display and means that there are no fixed spatial relationships among the cues. For this reason, the gas display is a useful alternative to artificial stimuli, such as faces, trains, robots, and other static displays of featural stimulus representations that have previously been used (e.g., Vandierendonck & Rosseel, 2000). The animated gases present graphically all of the cue information in the underlying representation but do not convey any other information as a by-product of the presentation.

Training phase procedure. If all of the stimulus pairs in Table 1 are considered, the cue validities take values that lead to the TTB and the RAT models making a different decision for one of the pairs. This is the pairing of Stimulus 2 and Stimulus 7, where the TTB model chooses Stimulus 2 because it has the high-validity second cue that Stimulus 7 does not have, whereas the RAT model chooses Stimulus 7, because the log-odds are 0.02 in its direction. To avoid the learning phase favoring either decision model, this stimulus pair was removed from the training set, and the validities were recalculated in relation to the remaining 119 pairs. It is these validities that are shown in Table 1, and their use leads to the TTB and the RAT models making the same decisions for all of the stimulus pairs presented during training.

The fact that the two models make the same decisions does not, of course, mean that these decisions are necessarily correct. The correctness of a decision, as provided in feedback to the participants, is determined by which stimulus has the larger poison value. As it turns out, the decision models are correct for 102 out of 119 stimulus pairs, which corresponds to a proportion correct of .86.

During the training phase, the participants were presented with six contiguous blocks of trials. The first of these blocks contained 19 stimulus comparisons, whereas the remaining five contained 20 comparisons. On each trial, the two gases were presented in their canisters adjacent to one another on the screen, and the mouse pointer was initially placed halfway between them. The participants chose the gas they believed to be the most poisonous by moving the pointer into the appropriate canister. Once a decision was made, feedback was provided by highlighting the canister of the most poisonous gas in red. The participants could view this feedback for as long as they wanted and pressed the mouse button to commence the next trial. The 119 trials were presented in a random order for each participant, and the placement of each gas in the left or the right canister was also randomized on every trial.

Test phase procedure. It is important during the test phase that both stimuli have the same number of cues, so that the participants cannot use this information as the basis for making their decisions. Although a cue corresponding to the number of gas molecules has a validity of only .88 and, so, is less informative than Cue 1 and Cue 2 from Table 1, it is easy to detect differences in the number of molecules perceptually, and so it could form part of the basis for decision making if equal numbers were not used during the test phase.

Given the cue validities in Table 1, there are only five diagnostic stimulus comparisons in which each stimulus has the same number of cues. These comparisons are {Cue 1, Cue 6} versus {Cue 2, Cue 3}, {Cue 1, Cue 5} versus {Cue 2, Cue 3}, {Cue 1, Cue 5, Cue 6} versus {Cue 2, Cue 3, Cue 4}, {Cue 1, Cue 4, Cue 5} versus {Cue 2, Cue 3, Cue 4}, and {Cue 1, Cue 4, Cue 5, Cue 6} versus {Cue 2, Cue 3, Cue 4, Cue 5}. In each comparison, the TTB model favors the first stimulus, because it has a high-validity cue that the second does not, whereas the rational model favors the second stimulus.

In the test phase, the participants moved the mouse pointer in the same way to make a decision as in the training phase. Once a decision had been made, feedback was not provided, but the participants were instead required to express their confidence on a 5-point scale, where 1 represented *random guess* and 5 represented *completely certain*. The time taken for the participants to make a decision (but not to rate their confidence) was also collected in this phase. The five comparisons were presented in random order for each participant, and the placement of each gas in the left or the right canister was again randomized.

Results

Training phase learning. Figure 2 shows the proportion of correct decisions for each of the six training blocks, averaged across all the participants, together with one standard error in each direction. The monotonic increase in accuracy, together with the small error bars, indicates that most, if not all, of the participants learned effectively from the feedback. Performance was well above the chance level of .5 after the first block of only 19 trials, suggesting that the validities of at least some of the cues were not difficult to learn. The asymptotic level of performance evident in Blocks 5 and 6 is consistent with the theoretical limit of .86 imposed by the TTB and the RAT decision models, which is shown by the broken line.

Test phase decisions. Table 2 shows the raw decision and confidence data from the test phase, giving the model consistent with each participant's decision on each comparison and the confidence of the participants in each of their decisions. It can be seen that, across all the participants, each of the comparisons produced decisions that were consistent with both decision models. The number of decisions consistent with the TTB model out of 40 was 11, 18, 14, 17, and 12 for the five comparisons. On these grounds, it seems that each comparison provided roughly equal ambiguous evidence about whether people used the TTB or the RAT decision model. There also was not much difference between the confidence the participants had in their decision for the five comparisons, with mean values of 2.93, 2.88, 2.95, 3.15, and 3.23.

A more useful analysis of the data considers decisions within participants, rather than across participants. Table 2 shows in bold those participants whose decisions for each comparison were all consistent with either the TTB or the RAT model. Nineteen participants, which is almost half of the sample, met this criterion, with 5 following TTB predictions and 14 following RAT predictions. What is remarkable about the number of consistent participants is that it was very unlikely to have happened by chance. If the participants were making random choices and, so, favoring each model with a probability of .5 independently on each comparison, only 6.25% should show consistency. This would correspond to only 2 or 3 of the 40 in the sample being consistent. The prob-

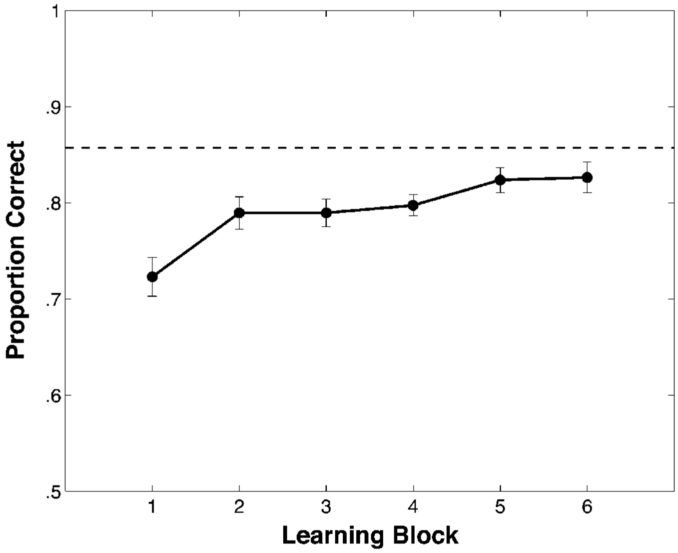

Figure 2. The mean percentage correct across all the participants over the six training blocks. One standard error is shown about the mean. The maximum possible percentage correct using the take-the-best and the rational decision models is shown by the broken line.

Table 2
Raw Data From the Test Phase, Detailing the Decision and Confidence for Each Participant on Each Comparison

Participant	Q1	Q2	Q3	Q4	Q5
1	R3	T4	R3	R4	R4
2	R4	T3	R3	T5	R3
3	**R4**	**R3**	**R4**	**R4**	**R3**
4	R1	R3	T1	R1	R1
5	R2	T3	T2	T3	R3
6	**R3**	**R3**	**R4**	**R3**	**R3**
7	T3	T2	R2	T3	T3
8	**T3**	**T4**	**T3**	**T4**	**T3**
9	**R5**	**R2**	**R1**	**R4**	**R4**
10	**R3**	**R3**	**R3**	**R3**	**R4**
11	**R5**	**R5**	**R4**	**R5**	**R4**
12	**R3**	**R3**	**R3**	**R2**	**R3**
13	R2	T1	T2	T4	T2
14	R4	T4	R2	T2	R4
15	**T3**	**T3**	**T3**	**T3**	**T3**
16	T5	T4	R3	R4	R5
17	R2	T2	T3	T4	R4
18	**T3**	**T5**	**T5**	**T5**	**T4**
19	R3	T2	R4	R2	R4
20	T1	R2	R1	T2	T1
21	**R2**	**R2**	**R2**	**R3**	**R1**
22	R1	R1	T3	T3	R3
23	R4	R3	R3	R3	T3
24	R3	T3	R4	R4	T3
25	T4	T5	T4	T5	R5
26	R1	R1	R1	R1	T1
27	**T3**	**T5**	**T4**	**T4**	**T4**
28	**R3**	**R2**	**R2**	**R2**	**R4**
29	**R2**	**R3**	**R5**	**R4**	**R5**
30	T3	R4	T5	T5	R4
31	**R4**	**R3**	**R3**	**R3**	**R4**
32	R4	T1	R4	R3	R3
33	**R3**	**R3**	**R4**	**R3**	**R4**
34	**R2**	**R2**	**R4**	**R3**	**R4**
35	**T2**	**T4**	**T2**	**T2**	**T2**
36	**R3**	**R4**	**R4**	**R4**	**R3**
37	R1	R1	R1	T1	R1
38	**R4**	**R4**	**R2**	**R3**	**R3**
39	R4	R1	T3	T2	R3
40	T2	T2	T2	R1	T4

Note—Decisions are shown as favoring the TTB (T) or the RAT (R) model, followed by the confidence value (1–5). Participants who were consistent with one of the models for all five comparisons are shown in bold type.

ability of 19 participants being consistent by chance is $\binom{40}{19}(1/16)^{19}(15/16)^{21}$, which corresponds to odds of about 2,231,908,215,164 to 1 against. This strongly suggests that many of the participants were using a consistent decision strategy during the test phase but that different participants were using different strategies. It also makes it clear that the participants did not use only the number of molecules as a basis for making decisions during training.

Confidence and response times. Statistical analysis of the confidence ratings and response times further supports this conclusion. The Bayesian information criterion (Schwarz, 1978) approximation to Bayes factors (Kass & Raftery, 1995) was used to compare the confidence ratings of consistent and inconsistent participants. Under the assumption that confidence follows a Gaussian distribution, the most likely account is that consistent and inconsistent confidence ratings have different means and variances. This account, which corresponds to the assumption the confidence ratings are *different* between the groups, was found to be 60.6 times more likely than the *same* account, which uses the same mean and variance for both groups. This can be interpreted as strong evidence that the participants who made consistent decisions were more confident. Use of the same analysis to compare the confidence distributions of TTB and RAT decisions, however, showed that the *same* account was 67.0 times more likely than the *different* account. This can be interpreted as providing strong evidence that the participants who made consistent decisions were equally confident, regardless of whether they made decisions consistent with the TTB or the RAT model.

The response times, which suggested that the TTB participants were faster than the RAT participants, were also examined for statistical differences. We made the commonly used (e.g., Cousineau & Larochelle, 1997; Ida, 1980; Logan, 1992; Van Zandt, 2000), if not universally accepted, assumption that response times follow a two-parameter Weibull distribution for the practical purpose of estimating Bayes factors. Under this approach, the *different* account, in which both parameters were allowed to vary, turned out to be 1,970 times more likely than the *same* account. This can be interpreted as providing overwhelming evidence that the response times for TTB decisions are shorter than those for RAT decisions. This makes sense, since the TTB decision model requires only that a cue be found that discriminates between stimuli, whereas the RAT decision requires all cues to be considered.

Unified model evaluation. Given that many participants made decisions consistent with both the TTB and the RAT models, the ability of the unified model to encompass both approaches should enable it to account for many of the data. Formally, this can be achieved by allowing the unified model to assume different evidence threshold parameterizations for different participants or for different groups of participants and examining the ability of these *families* of models to account for all the decisions made.

To fit the model in this way, the 40 participants were divided into two groups, on the basis of whether they made more TTB or more RAT decisions. That is, the participants who made three, four, or five TTB decisions were placed in one group, and the remaining participants were placed in the other group. Best-fitting evidence thresholds, in terms of maximizing the number of decisions correctly accounted for, were then found for both groups. Because of the pattern of cue validities used in the experiment, there is a range of parameterizations that are indistinguishable in their predictions, and so any threshold less than the evidence of the first cue (i.e., less than about 3.5) leads to TTB decisions, whereas larger thresholds lead to RAT decisions. With these parameterizations for the two participant groups, the unified model accounts for 84.5% of the decisions made by all the participants. This compares with an accuracy of 64%, under the assumption that all the participants used the RAT decision model, and of 36% for the TTB model.

Of course, the unified model, through its use of two parameters, is more complicated than both the TTB and the RAT decision models, which are parameter free. This raises the issue of whether the extra complexity is warranted by the improved accuracy of the unified model (e.g., Roberts & Pashler, 2000). This concern can be addressed substantially by using recent developments in psychological model selection theory (e.g., Myung & Pitt, 1997; Pitt, Myung, & Zhang, 2002), which provides Bayesian and information theoretic criteria for choosing between models in ways that consider both accuracy and complexity.

One interesting challenge in doing this is that the unified, TTB, and RAT models are deterministic and do not specify an error theory. This means that various probabilistic model selection criteria, such as Bayes factors, minimum description length (MDL; e.g., Grünwald, 2000) or normalized maximum likelihood (Rissanen, 2001), are not immediately applicable. Recently, however, Grünwald (1999; see also Myung, Pitt, & Kim, in press) has developed a model selection methodology that overcomes these difficulties. He has provided a principled technique for associating deterministic models with probability distributions, through a process called *entropification*, that allows MDL criteria for competing models to be calculated. Applying this method to the unified, TTB, and RAT models for the decision data resulted in MDL values of 87.6, 138.6, and 130.7, respectively. The much smaller MDL value for the unified model indicates that it provides a better account of the data, even allowing for its additional complexity. This finding is not surprising, since only the unified model can capture the basic regularity in the human data and produce decisions consistent with both the TTB and the RAT models.

Discussion

One potential criticism of the experiment is that it does not make the visual search for information an effortful process. A central tenet of the fast-and-frugal approach is that the need to search an environment for information places strong constraints on cognitive processes. It is possible that the nature of the stimulus presentation is responsible for the relatively greater number of RAT than of TTB decision makers. Because of the ready availability of all the gas molecules, there would seem to be little disincentive for people not to gather complete cue information before making a decision and, so, behave in a way compatible with the RAT model.[3]

The unified model explicitly incorporates a search strategy and, accordingly, makes predictions about what decisions will be made under task conditions that encourage different levels of search. In particular, if the information needed to make a decision is difficult to find, the unified model predicts the use of fast-and-frugal strategies, achieved by reducing the required evidence to make a decision. In a follow-up experiment, reported only briefly here, we asked participants to make judgments about the poisonousness of the gases under conditions in which only part of the canister was visible, so not all of the molecules were visible at once. Seven of 20 participants consistently made TTB decisions, but none consistently followed the RAT model. Once again, the odds of the TTB consistency's arising by change are negligible (8,130 to 1 against), and an analysis of confidence rating showed that the consistent participants were significantly more confident. The unified, TTB, and RAT models accounted for 78%, 64%, and 36% of the decisions, respectively, and had MDL values of 54.1, 65.3, and 69.3. The unified model, therefore, also provided the best account of human performance in this second experiment, in which more effortful information search was required.

More generally, it is useful to interpret the unified model as a natural extension of the fast-and-frugal approach. Effectively, what the evidence accrual model does is extend the TTB account of *one-reason* decision making to *two-reason*, *three-reason*, or *many-reason* decision making. In these extensions of the TTB model, it may be necessary for a stimulus to have two or more high-validity cues that its alternative does not, rather than just relying on the first discriminating cue. This seems reasonable: Whereas choosing a red curry may demand only one good reason, many people would prefer to buy a plane ticket to Madrid only after they had established several advantages over Rome (Allen, 2000). By assuming that people must reach some threshold of evidence before making a decision, it is possible to capture the different importance of different decisions in a natural way that begins at the fast-and-frugal and ends at the rational. Furthermore, the evidence accrual formulation gives a sensible account of why decisions take longer in some environments than in others. In environments in which relatively little strong evidence is available to support any decision, more information will need to be found before a decision is made. In environments in which there is readily available information that provides strong evidence for a particular decision, response times may still be short even when the decision is a very important one.

More comprehensive data are needed to test the adequacy of the simple random walk sequential-sampling process assumed in the version of the unified model presented here. It may be that variants, such as multiple accumulator sampling processes (e.g., Lee & Corlett, 2003; Vickers, 1979), are needed to capture human performance. Similarly, whether or not it is useful to incorporate asymmetric or dynamic evidence thresholds (e.g., Ashby, 1983; Busemeyer & Townsend, 1993; Vickers, 1979) or memory models to describe the retention of information (e.g., Pietsch & Vickers, 1997; Smith, 2000; Usher & McClelland, 2001) or an interval of uncertainty in evidence accumulation (e.g., Juslin & Olsson, 1997;

Vickers, 2001; Vickers & Pietsch, 2001), or to employ any of a range of other theoretical mechanisms from the sequential-sampling process literature requires more detailed experimentation.

The obvious way to generate the necessary additional empirical constraints is to examine how the two factors described above—the setting of internal cognitive evidence thresholds based on the utility and consequences of a decision and the supply of information available through searching an external environment—interact in human decision making. There are a variety of experimental methods from the decision-making literature by which to manipulate the evidence thresholds used by people in making decisions, including manipulating the costs and benefits of making correct and incorrect decisions and demanding different target levels of confidence through speed–accuracy tradeoffs (e.g., Vickers, 1979). There has also been previous work in which the decisions made in environments with different significance structures have been studied (e.g., Bullock & Todd, 1999). Putting these experimental possibilities together and attempting to model how quickly, confidently, and accurately people make decisions in differently structured environments under different task demands is a priority for future research.

REFERENCES

Allen, C. (2000). The evolution of rational demons [Abstract]. *Behavioral & Brain Sciences*, **23**, 742.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, **98**, 409-429.

Anderson, J. R. (1992). Is human cognition adaptive? *Behavioral & Brain Sciences*, **14**, 471-517.

Ashby, F. G. (1983). A biased random walk model of two choice reaction times. *Journal of Mathematical Psychology*, **27**, 277-297.

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, & Cognition*, **26**, 1332-1346.

Brunswik, E. (1943). Organismic achievement and environmental probabilities. *Psychological Review*, **50**, 255-272.

Bullock, S., & Todd, P. M. (1999). Made to measure: Ecological rationality in structured environments. *Minds & Machines*, **9**, 497-541.

Busemeyer, J. R., & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology*, **32**, 91-134.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic cognition approach to decision making. *Psychological Review*, **100**, 432-459.

Cousineau, D., & Larochelle, S. (1997). PASTIS: A program for curve and distribution analyses. *Behavior Research Methods, Instruments, & Computers*, **29**, 542-548.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. C. (1999). How good are simple heuristics? In G. Gigerenzer & P. M. Todd (Eds.), *Simple heuristics that make us smart* (pp. 97-117). New York: Oxford University Press.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, **41**, 260-274.

Doyle, J. (1999). Rational decision making. In R. A. Wilson & F. C. Keil (Eds.), *MIT encyclopedia of the cognitive sciences* (pp. 701-703). Cambridge, MA: MIT Press.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* London: Chapman & Hall.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, **103**, 650-669.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Grünwald, P. (1999). Viewing all models as "probabilistic." In *Proceedings of the Twelfth Annual Conference on Computational Learning Theory (COLT '99)* (pp.171-182). New York: ACM Press.

Grünwald, P. (2000). Model selection based on minimum description length. *Journal of Mathematical Psychology*, **44**, 133-152.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, **16**, 96-101.

Ida, M. (1980). The application of the Weibull distribution to the analysis of the reaction time data. *Japanese Psychological Research*, **22**, 207-212.

Johnson, M. P., & Raven, P. H. (1973). Species number and endemism: The Galapagos archipelago revisited. *Science*, **179**, 893-895.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review*, **104**, 344-366.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**, 773-795.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.

Laming, D. R. J. (1968). *Information theory and choice-reaction time.* London: Academic Press.

Lee, M. D., Chandrasena, L. H., & Navarro, D. J. (2002). Using cognitive decision models to prioritize e-mails. In W. G. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 478-483). Mahwah, NJ: Erlbaum.

Lee, M. D., & Corlett, E. Y. (2003). Sequential sampling models of human text classification. *Cognitive Science*, **27**, 159-193.

Lee, M. D., Loughlin, N., & Lundberg, I. B. (2002). Applying one reason decision making: The prioritization of literature searches. *Australian Journal of Psychology*, **54**, 137-143.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, **40**, 77-105.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 883-914.

Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? In G. Gigerenzer & P. Todd (Eds.), *Simple heuristics that make us smart* (pp. 119-140). New York: Oxford University Press.

Martignon, L., & Laskey, K. B. (1999). Bayesian benchmarks for fast and frugal heuristics. In G. Gigerenzer & P. M. Todd (Eds.), *Simple heuristics that make us smart* (pp. 169-188). New York: Oxford University Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification. *Psychological Review*, **85**, 207-238.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, **4**, 79-95.

Myung, I. J., Pitt, M. A., & Kim, W. J. (in press). Model evaluation, testing and selection. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition.* Thousand Oaks, CA: Sage.

Myung, I. J., & Shepard, R. N. (1996). Maximum entropy inference and stimulus generalization. *Journal of Mathematical Psychology*, **40**, 342-347.

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 & Human Decision Processes*, **91**, 82-96.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 104-114.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, **104**, 266-300.

Pietsch, A., & Vickers, D. (1997). Memory capacity and intelligence: Novel techniques for evaluating rival models of a fundamental infor-

mation processing mechanism. *Journal of General Psychology*, **124**, 229-339.
Pinker, S. (1997). *How the mind works*. New York: Norton.
Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, **109**, 472-491.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.
Rissanen, J. (2001). Strong optimality of the normalized ML models as universal codes and information in data. *IEEE Transactions on Information Theory*, **47**, 1712-1717.
Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, **107**, 358-367.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.
Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.
Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review*, **63**, 129-138.
Simon, H. A. (1976). From substantive to procedural rationality. In S. J. Latsis (Ed.), *Method and appraisal in economics* (pp. 129-148). London: Cambridge University Press.
Simon, H. A. (1982). *Models of bounded rationality.* Cambridge, MA: MIT Press.
Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, **44**, 408-463.
Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems* (Vol. 11, pp. 59-65). Cambridge, MA: MIT Press.
Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral & Brain Sciences*, **24**, 629-640.
Todd, P. M., & Gigerenzer, G. (2000). Précis of *Simple heuristics that make us smart*. *Behavioral & Brain Sciences*, **23**, 727-780.
Tversky, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.
Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, **108**, 550-592.
Vandierendonck, A., & Rosseel, Y. (2000). Interaction of knowledge-driven and data-driven processing in category learning. *European Journal of Cognitive Psychology*, **72**, 37-63.
Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, **7**, 424-465.
Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.
Vickers, D. (2001, September). *The elusive "interval of uncertainty": Evidence for a dynamic normalizing of sensory magnitudes.* Paper presented at the 32nd Meeting of the European Mathematical Psychology Group, Lisbon.
Vickers, D., & Pietsch, A. (2001). Decision-making and memory: A critique of Juslin and Olsson's (1997) sampling model of sensory discrimination. *Psychological Review*, **108**, 789-804.
Wallsten, T. S., & Barton, C. (1982). Processing probabilistic multidimensional information for decisions. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 361-384.

NOTES

1. There is a sense in which this independence assumption, by avoiding the full Bayesian treatment, is consistent with the fast-and-frugal idea of "taking a bet" on the structure of the environment (see Martignon & Laskey, 1999). It is reasonable to argue, nevertheless, that this model is "rational" in relation to the TTB model, because it explicitly integrates the information available in all stimulus cues.

2. Advocates of fast-and-frugal heuristics could argue that this unification relies on a generous interpretation of the rational model, in the sense that it really does not make any processing assumptions but simply requires all information to be available. This is true, but it remains the case that the complete validity-ordered sampling of cues under the sequential-sampling process will produce exactly the same decisions as the rational model.

3. It is also possible to argue, however, that using training examples in which the decision strategies always agree biases people toward using the simpler TTB strategy.

(Manuscript received December 5, 2002;
revision accepted for publication March 10, 2003.)