Journal of Experimental Psychology:
Learning, Memory, and Cognition
2007, Vol. 33, No. 1, 107–129

0278-7393/07/$12.00 DOI: 10.1037/0278-7393.33.1.107

A Response-Time Approach to Comparing Generalized Rational and Take-the-Best Models of Decision Making

F. Bryan Bergert and Robert M. Nosofsky
Indiana University Bloomington

The authors develop and test generalized versions of take-the-best (TTB) and rational (RAT) models of multiattribute paired-comparison inference. The generalized models make allowances for subjective attribute weighting, probabilistic orders of attribute inspection, and noisy decision making. A key new test involves a response-time (RT) approach. TTB predicts that RT is determined solely by the expected time required to locate the 1st discriminating attribute, whereas RAT predicts that RT is determined by the difference in summed evidence between the 2 alternatives. Critical test pairs are used that partially decouple these 2 factors. Under conditions in which ideal observer TTB and RAT strategies yield equivalent decisions, both the RT results and the estimated attribute weights suggest that the vast majority of subjects adopted the generalized TTB strategy. The RT approach is also validated in an experimental condition in which use of a RAT strategy is essentially forced upon subjects.

Keywords: decision making, response times, paired-comparison inference

Recent research in decision making has introduced heuristics that dramatically simplify processing while maintaining a high level of accuracy. Gigerenzer and Todd (1999) and Gigerenzer and Selten (2001) refer to these heuristics as "the adaptive toolbox." They theorize that much decision making results from an application of one or more of these simple heuristics, instead of from an application of the classically "rational" strategies that make exhaustive use of all available information. Such heuristics as *recognition* and *take-the-best* (TTB) are thought to be adaptations to natural information environments, and they constitute an alternative vision of rationality, called *ecological rationality* or *bounded rationality* (Gigerenzer & Goldstein, 1996; Gigerenzer & Selten, 2001; Goldstein & Gigerenzer, 2002; Simon, 1956).

A great deal of research has been devoted to comparing the predictions from the TTB model and classically rational models of decision making. In this article, however, we suggest that the compared models make quite strong assumptions. For example, within the framework of the compared models, it is often assumed that subjects attach certain ideal observer weights to the multiple attributes that compose the alternatives and also that the underlying choice mechanisms are fully deterministic in nature. The main purpose of the present research was to consider generalized versions of the models that relax these assumptions, leading to what we view as more psychologically plausible models. Furthermore, we introduce a new response-time method for distinguishing between the predictions from the generalized models.

The TTB Model

In making a decision about which of two alternatives is higher on some variable of interest, TTB considers the features of the alternatives in order of diagnosticity and makes a decision according to the first feature found that distinguishes between the alternatives. For example, suppose first that the presence of a professional basketball team provides a good indication of the size of a city, and then imagine that one city has a team and another does not. On the basis of this cue, in a pairwise comparison, TTB could make a decision immediately. If both cities have a team, or if both cities do not—that is, if they match on this cue—then TTB would consider the next cue in order of diagnosticity, and so on until a decision is made.

Researchers have argued that in numerous natural-world environments, TTB strikes an adaptive balance between the quality of an agent's final decision and the efficiency with which that decision is made (Gigerenzer & Todd, 1999; Simon, 1956, 1976). TTB maximizes speed by making its decision on the basis of a single discriminating cue, but, of course, the order of consideration of cues is important. TTB would make fast decisions if uninformative cues were examined first, but those decisions would be of poor quality. Therefore, to strike a balance between speed and decision quality, TTB inspects cues from the best predictor of the variable of interest to the worst.

The predictor value of a feature for a certain variable, also known as its *cue validity*, is defined as the proportion of distinctions a feature makes that are correct. For example, to determine the validity of having a professional basketball team for predicting city size, one would consider all pairs of cities and count the number of pairs in which one city has a team and the other does not. The cue validity would be the proportion of these pairs in

F. Bryan Bergert and Robert Nosofsky, Department of Psychology, Indiana University Bloomington.

This work was supported by National Institute of Mental Health Grant R01 MH48494. We thank Frances Fawcett of Cornell University for allowing us to use her drawings of beetles as stimuli in these experiments. We also thank Ben Newell for his helpful criticisms of an earlier version of this article and Peter Todd for numerous helpful discussions.

Correspondence concerning this article should be addressed to Robert M. Nosofsky, Department of Psychology, Indiana University Bloomington, Bloomington, IN 47405. E-mail: nosofsky@indiana.edu

which the city with a team is actually larger than the city without a team. In general, the validity of cue i (v_i) is defined as

$$v_i = \frac{correct\_distinctions_i}{total\_distinctions_i}. \quad (1)$$

One consequence of the definition given in Equation 1 is that a cue that makes 1 out of 1 distinctions correctly will have a cue validity identical to a cue that makes 100 out of 100 distinctions correctly. In fact, a cue that makes 1 out of 1 distinctions correctly will have a higher cue validity than one that makes 999 out of 1,000 distinctions correctly. The Bayesian modification (Lee, Chandrasena, & Navarro, 2002; Lee & Cummins, 2004) given in Equation 2 has been used to address this counterintuitive result:

$$v_i = \frac{1 + correct\_distinctions_i}{2 + total\_distinctions_i}. \quad (2)$$

Thus, cues that make a small number of distinctions have a low Bayesian cue validity. Following Lee and Cummins (2004), we adopt the Equation 2 definition of cue validity in the present article.

Once cue validities have been calculated, TTB begins by inspecting cues in order of validity, from most valid cue to least valid cue, until a feature is found for which the alternatives have different feature levels. TTB then uses this feature to make a decision.[1]

A Rational (RAT) Model

A weighted additive model (Payne & Bettman, 2001; Rieskamp & Otto, 2006)—which we will call *RAT*, after Lee and Cummins's (2004) "rational" model—is often chosen as the embodiment of rational decision making, for purposes of comparison with the heuristic approaches. RAT assigns to each feature a weight (w_i), calculates the summed total evidence in favor of each alternative, and chooses the alternative with more total evidence. The evidence in favor of each alternative is calculated only over those features that mismatch and therefore discriminate between the alternatives. Thus, the decision rule in RAT is to "choose A" if

$$\sum_{a \in FA} w_a > \sum_{b \in FB} w_b, \quad (3)$$

where FA and FB denote the sets of all discriminating features that favor alternatives A and B, respectively. Thus defined, RAT assumes that each feature makes an independent contribution to the evidence in favor of an alternative. Given this assumption, it turns out that the optimal value to use for w_i is the log odds of the cue validity, v_i, given in Equation 4:

$$w_i = \log\left(\frac{v_i}{1 - v_i}\right). \quad (4)$$

Lee and Cummins (2004) give a conceptual explanation of Equation 4, and Katsikopoulos and Martignon (in press) provide a proof.

Experiments on the Use of TTB and RAT

The existence and relative efficiency of the heuristics in the adaptive toolbox have naturally led to experiments to determine whether people actually use them in making decisions (for a review, see Payne & Bettman, 2001). A combination of process tracing and outcome approaches has been used (Rieskamp & Hoffrage, 1999; Rieskamp & Otto, 2006). In a typical process-tracing experiment, a person must actively uncover each cue to see its value, so the experimenter can observe the order, extent, and time course of information search (e.g., Broder, 2000; Johnson, Payne, Schkade, & Bettman, 1991; Newell, Rakow, Weston, & Shanks, 2004; Newell & Shanks, 2004; Rieskamp & Hoffrage, 1999; Rieskamp & Otto, 2006). In contrast, in an outcome-oriented experiment, certain decisions are predicted by certain models, so researchers can infer which strategy was used from the final decision made (Juslin, Jones, Olsson, & Winman, 2003; Lee & Cummins, 2004; Rieskamp & Hoffrage, 1999).

Numerous experiments have shown that people's choice of strategy is contingent upon the characteristics of the task and stimuli. If a decision task (Rieskamp & Otto, 2006) or a categorization task (Juslin et al., 2003) affords different strategies different levels of accuracy, people tend to adopt the most accurate strategy. Conversely, if strategies produce the same level of performance, people tend to adopt the simplest strategy (Juslin et al., 2003). However, increases in time pressure and information-acquisition costs, as well as presentation formats involving separate attribute listings as opposed to holistic images, lead to increased use of TTB-like strategies (Broder & Schiffer, 2003; Gigerenzer & Todd, 1999; Martignon & Hoffrage, 1999; Payne, Bettman, & Johnson, 1988, 1993).

The current research is most directly motivated by an experiment conducted by Lee and Cummins (2004), who used an outcome analysis to determine whether subjects would tend to adopt RAT or TTB in a domain for which either strategy would work equally well. The task for subjects in Lee and Cummins (2004) was to look at the molecular makeup of two gases, presented simultaneously and in their entirety, and to decide which one of them was more poisonous. To avoid biasing subjects toward the use of one strategy or the other, Lee and Cummins specifically chose stimuli so that, for every pair, the same choice was made by RAT and by TTB.[2] An illustration is provided in Figure 1A.

In the figure, we assume that the presence of a cue (represented by a value of 1) is diagnostic of poison and that cues are arranged from left to right in decreasing order of validity. RAT chooses the item with more total evidence, so it chooses the left item because most of the evidence is in that alternative's favor; TTB chooses the

[1] Martignon and Hoffrage (1999) have pointed out that this order of inspection of cues is not necessarily optimal but that to find the optimal ordering "there is no simpler procedure than testing all possible permutations of cues and comparing the performance of a lexicographic strategy [TTB] for each of these orderings" (p. 132). Inspecting cues in order of validity is a highly efficient, "approximately rational" compromise.

[2] By contrast, in previous related experiments, use of a TTB strategy would lead to markedly worse performance than use of a weighted additive strategy (e.g., Juslin et al., 2003). It is interesting that the stimulus structure that Lee and Cummins (2004) used was adapted from a real-world environment, previously considered by Czerlinski, Gigerenzer, and Goldstein (1999), that relates certain geographic features of the Galapagos Islands to a count of the number of species on each island. It is precisely for such real-world environments that Gigerenzer and colleagues have argued that fast and frugal heuristics, such as TTB, have been adapted.

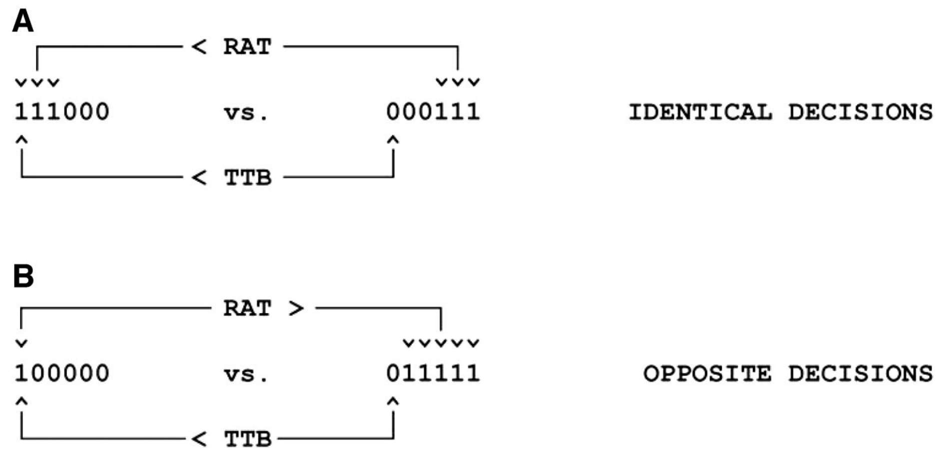

Figure 1. A: The rational model (RAT) and the take-the-best (TTB) heuristic yield identical decisions on training pairs. Cues are arranged from left to right in decreasing order of validity. RAT chooses the alternative with more total evidence; TTB chooses the same item because it has the most valid discriminating cue. B: In contrast, RAT and TTB yield opposite decisions on test pairs. RAT chooses the alternative to the right because it has more total evidence; TTB chooses the left item because it has the most valid discriminating cue.

same item because it bases its decision on the single most valid discriminating cue. Exposure to 119 different stimulus pairs with the property illustrated in Figure 1A constituted the training phase. The point of this training was to familiarize subjects with the decision task and to give subjects the opportunity to infer the validities of the various cues that made up the stimuli. After training, new pairs of stimuli were introduced for which RAT and TTB make opposite choices. As illustrated in Figure 1B, RAT chooses one item because it has a greater total amount of evidence, whereas TTB chooses the opposite item because it is favored by the single most valid cue.

Because the two strategies predict opposite choices on these stimulus pairs, one can use them to determine which strategy a subject adopted as a result of training. The results from these test pairs indicate that some subjects (35%) consistently made RAT decisions, others (13%) consistently made TTB decisions, and still others (52%) were inconsistent in their choice of strategy. In addition to analyzing subjects' performance on these test pairs, Lee and Cummins (2004) fit versions of the RAT and TTB models to subjects' choice data with a model fit criterion known as minimum description length (MDL; Rissanen, 1978). The average fits for RAT (MDL = 130.7) and TTB (MDL = 138.6) were similar, so neither model was conclusively favored.

Generalizing the TTB and RAT Models

As we describe in our General Discussion, Lee and Cummins's (2004) central motivation was to move toward the development of a unified approach to decision making that combines TTB and RAT within a common framework. To provide faithful tests, Lee and Cummins followed numerous other researchers by considering the TTB and RAT models in their strong forms. Our own view, however, is that the strong forms of TTB and RAT are based on assumptions that are psychologically implausible. A central purpose of the present work, therefore, is to consider generalizations of these models that relax these assumptions while preserving important component themes of the original models. By considering these generalized models, we investigated whether certain processing strategies embodied in the original models might indeed govern human performance, even if the strong special cases of the models are ruled out.

The first assumption that we find implausible is that all subjects learn and use the optimal feature weights. In RAT, it is assumed that the weight that is attached to each cue is the log odds of its objective cue validity, given by Equation 4. In TTB, the cues are examined in order of objective cue validity until a discriminating cue is found.

In our view, it is implausible that every subject learns and uses the objective cue validities from experience with the decision domain. We think it is more likely that subjects occasionally make mistakes in their assignment of weights to cues (e.g., see Newell & Shanks, 2003). Calculating cue validity for each dimension requires a great deal of memory capacity and experience with the domain (for similar arguments, see, e.g., Newell, 2005). For this reason, we consider natural generalizations of both RAT and TTB that relax the assumption that optimal feature weights are always learned and used. In these generalized models, the feature weights are simply free parameters; for each subject, the weights that yield the best fit to the subject's data are found, and the models' predictions depend on these weights. The original, strong versions of the models are special cases of these general models with weights set to cue validities or log odds of cue validities.

The second assumption we find implausible is that subjects respond deterministically. The original RAT model makes a deterministic decision in favor of the alternative with more evidence. Similarly, in TTB, cues are always inspected in a deterministic, fixed order, and once a discriminating cue is found, the alternative with the correct discriminating cue value is chosen. Thus, there is no error theory associated with the models.

A simple generalization involves making allowance for probabilistic decision making. Our generalized versions of RAT and

TTB make responding probabilistic in several ways. First, we introduce a guessing parameter to all models under consideration, such that a subject is assumed to guess with probability g ($0 < g < 1$) on any trial and to use the model of interest (RAT or TTB) otherwise, with probability $1 - g$.

The guessing mechanism provides a rudimentary form of error theory,[3] but there is good reason to believe that it will be insufficient. Consider, for example, an observer who uses a RAT strategy. In one case, the summed evidence for alternative A may be dramatically greater than the summed evidence for alternative B, whereas in a second case the difference in evidence may be minuscule. Extending the deterministic decision rule of the RAT model with an all-or-none guessing parameter would still lead the model to predict identical choice probabilities in these two cases. Clearly, however, an observer is more likely to choose an alternative when the evidence that favors the alternative is dramatic than when it is minuscule.

Thus, in our generalization of RAT, we assume that the probability that alternative A is chosen from pair AB is given by

$$P(\text{“}A\text{”}|AB) = \frac{\left[\sum_{a \in FA} w_a\right]^\gamma}{\left[\sum_{a \in FA} w_a\right]^\gamma + \left[\sum_{b \in FB} w_b\right]^\gamma}, \quad (5)$$

where γ is a response-scaling parameter. First, note that this response rule allows the model to predict increases in choice probability as the degree of evidence in favor of alternative A increases. Second, note that the response rule incorporates the original, deterministic RAT model as a special case: If $\gamma = \infty$, the decision is deterministic in favor of the alternative with more evidence (compare with Equation 3). Another important special case arises when $\gamma = 1$. In this case, the decision is probabilistic and matches the theoretic response probability $P(\text{“A”}|\text{AB})$ to alternative A's proportion of the total evidence.

Likewise, the strong version of TTB considered by Lee and Cummins (2004) also assumed a form of strict deterministic behavior having to do with the order in which cues are inspected. Specifically, the original TTB model automatically inspected features in order of validity, from most valid to least valid cue. Our initial generalization, allowing for subjective feature weights, would allow for the cues to be inspected in an order other than that prescribed by cue validity. For example, if a subject assigned the highest weight to feature i, then the subject would inspect feature i first. However, even with free feature weights, the original assumption that features would always be inspected in some fixed order would remain.

We thought a reasonable further generalization of the TTB model would be to make allowances for a probabilistic order of inspection. Thus, for each step of the process, we assumed that the probability of inspecting a feature was simply proportional to its weight,

$$P(\text{inspect feature}_i) = \frac{w_i}{\sum w}. \quad (6)$$

Because features that have already been inspected are eliminated from further consideration, the total weight in the denominator is calculated only over features that have not yet been inspected.

The probability of a particular order being used is the product of the individual probabilities of feature inspections. For example, the probability of order (1, 2, 3, 4, 5, 6) would be given by

$$P(order = (1,2,3,4,5,6)) = \frac{w_1}{w_1 + \ldots + w_6} \times \frac{w_2}{w_2 + \ldots + w_6} \times \frac{w_3}{w_3 + \ldots + w_6} \times \frac{w_4}{w_4 + \ldots + w_6} \times \frac{w_5}{w_5 + w_6} \times \frac{w_6}{w_6}. \quad (7)$$

To find the overall probability of an A response for a given stimulus pair AB, one finds the probability of choosing an order that favors A, which is just the sum of the probabilities of all orders that favor A:

$$P(\text{“}A\text{”}|AB) = \sum_{order \in O_A} P(order), \quad (8)$$

where O_A is the set of orders of cue inspection that lead the TTB model to choose alternative A from pair AB.

This generalized, probabilistic choice of orders of cue inspection contains as a special case the original, deterministic inspection of cues in order of validity. We can visualize this relation by creating weights that are in the same rank order as cue validity and are as different in magnitude from each other as possible. As the weights get further apart in magnitude, the probability of inspecting the features in order of cue validity approaches 1, as is illustrated in Figure 2.

It is interesting that the straightforward generalizations of TTB described here—making feature weights free parameters and allowing for probabilistic orders of cue inspection—transform TTB into exactly Tversky's (1972) elimination by aspects (EBA) choice model. In EBA, a cue is chosen for inspection with probability equal to its proportion of the total weight (Equation 6). Any alternatives without this feature are eliminated from further consideration, and a second feature is chosen for inspection from among the remaining uninspected cues. Elimination continues until there is only one alternative left, or until all features have been inspected, whereupon a choice is made randomly.

A point of great importance is the relation between the generalized version of TTB, which we will call *gTTB*, and the generalized version of RAT, which we will call *gRAT*. Clearly, the underlying processes that are represented by the models are dramatically different. Remarkably, however, gTTB's prediction of the probability of choosing item A from among A and B (i.e., Equations 6–8) yields an expression that is formally identical to gRAT's prediction (Equation 5) when gRAT's response-scaling parameter γ is set equal to 1 (see Tversky, 1972, p. 287, Equation 6). We illustrate the reason for the formal identity in Appendix A for a simple case in which the two alternatives contain no matching features.

[3] As discussed in detail in the *Results* sections of our article, we use likelihood-based methods of evaluating the fits of the alternative models. If a subject makes a single response that is not in accord with the deterministic versions of TTB or RAT, then these likelihood-based methods would evaluate the models as being infinitely wrong. A rudimentary error theory is needed to avoid this difficulty.

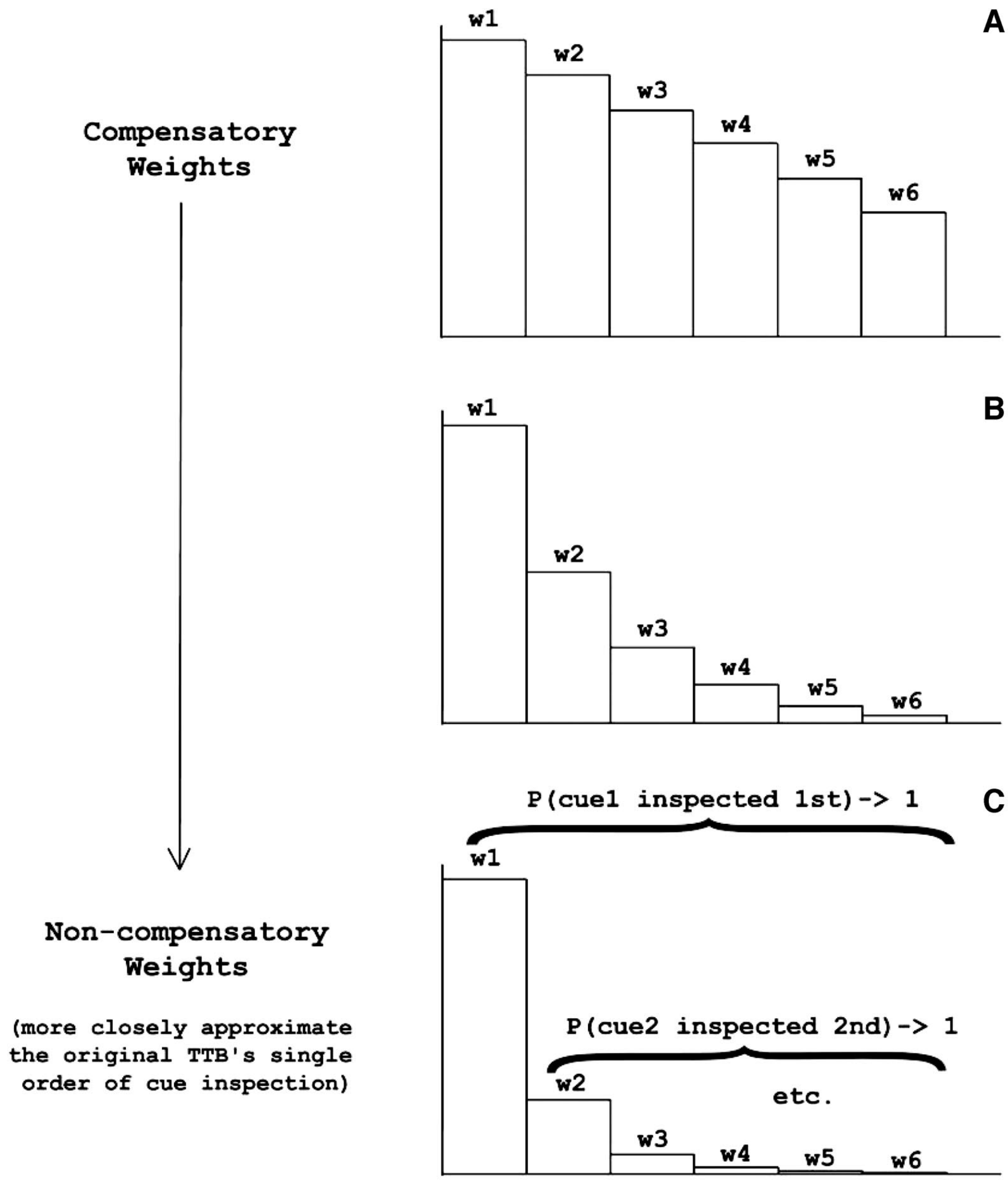

Figure 2. In the general version of take-the-best (TTB) with free weights and probabilistic choice of cue inspection order, the order of cue inspection can be made deterministic by making the weights sufficiently noncompensatory. As the weights become more and more dissimilar, the probability of choosing the highest one first increases to 1; if it fails to discriminate, the probability of choosing the second highest one next increases to 1; et cetera.

One important implication of this identity between the generalized models is that the test pairs used by Lee and Cummins (2004) to distinguish between the original RAT and TTB models do not, of course, distinguish between the general versions. Thus, for example, consider again the stimulus pair in Figure 1B, in which the difference between a TTB choice and a RAT choice is dramatic. The original strong version of TTB chooses the left item because it possesses the most valid cue, and the original strong version of RAT chooses the right item because it contains the greatest amount of summed evidence. Suppose, however, that a gTTB subject gave greatest weight to the second attribute instead of to the first. Then, such a subject would tend to choose the right item rather than the left. Thus, considering the predictions of only the strong versions of TTB and RAT could lead to potential misinterpretations of the subject's decision-making behavior.

Because the gTTB and gRAT ($\gamma = 1$) models are formally identical, we cannot tell them apart on the basis of fits to choice-probability data. Nevertheless, as we argue later in this article, certain types of parameter estimates derived from fits of the models are far more in the spirit of a TTB-like process than a

RAT-like process. The main purpose of Experiment 1 is to test the utility of the generalized models and to use the best-fitting parameter estimates to reach interpretations about subjects' behavior. Then, in Experiments 2 and 3 we move to the central goal of our research, which involves the introduction of a new response-time (RT) method for distinguishing between the predictions from the generalized models.

Experiment 1

The main initial goal of Experiment 1 is to address the question of whether the generalizations of RAT and TTB we have outlined offer an improvement over the original strong versions. We evaluate this question by conducting quantitative fit comparisons among the models, where the measure of fit penalizes the generalized models for their increase in the number of free parameters. Once we establish the usefulness of the generalizations, we then interpret the nature of subjects' decision strategies in terms of the best-fitting parameter estimates from the generalized models.

To test whether the generalizations of RAT and TTB were necessary, our first experiment replicated Lee and Cummins's (2004) experiment, with a few modifications. First, Lee and Cummins used the presence and absence of individual colored dots (representing "molecules") to instantiate their abstract stimulus structure. Thus, their physical stimuli were composed from a set of relatively homogeneous, present–absent features. In our experiments, we instead used pictures of insects that varied along six heterogeneous, binary-valued "substitutive" dimensions. (For binary-valued substitutive dimensions, one of two feature values is always present.) As discussed by Lee and Cummins, the use of present–absent features might bias subjects toward counting strategies involving the total number of features that are present on a stimulus. Thus, during the test phase, these researchers used stimulus pairs in which each member always had the same number of present features. By using substitutive dimensions, we avoided the need to impose this constraint because six heterogeneous features were always present on each stimulus. In addition, we were interested in testing the models in a domain involving integrated perceptual displays.

As in Lee and Cummins (2004), subjects were first trained on pairs for which both RAT and TTB made identical decisions. Then, to examine which strategy they had adopted during training, subjects were tested on new pairs for which the original strong versions of RAT and TTB made opposite decisions. A second difference between our experiment and the earlier study conducted by Lee and Cummins (2004) is that our transfer phase included a large number of test pair trials in order to obtain a data set suitable for quantitative fitting.

Method

Subjects. The subjects were 61 undergraduate students at Indiana University Bloomington who participated as part of a course requirement. Subjects were told at the beginning of the experiment that if they performed well on the test phase they would be paid a $3 bonus; those who achieved 80% correct or better were paid the bonus.

Stimuli. Table 1 contains the 16 abstract stimulus patterns used in the training phase and their corresponding poison levels. These abstract stimulus patterns were mapped onto pictures of poisonous

Table 1
Training Stimulus Patterns Used in Lee and Cummins (2004) and in Experiments 1 and 2

Stimulus number	Stimulus pattern	Decision variable (Poison)
1	0 0 0 1 0 0	16
2	0 1 0 0 1 0	18
3	0 0 1 0 0 1	21
4	0 0 0 1 1 0	25
5	0 0 0 0 1 0	31
6	1 0 0 0 1 1	40
7	0 0 1 1 1 1	44
8	1 1 0 1 0 0	51
9	1 1 1 0 0 1	62
10	1 1 0 0 1 0	70
11	1 1 0 1 1 1	97
12	1 1 1 1 0 0	104
13	1 1 1 1 1 1	280
14	1 1 1 1 0 1	285
15	1 1 1 0 1 0	347
16	1 1 1 1 1 0	444

insects, with six features corresponding to body parts: body, eyes, legs, antennae, fangs, and tail. The binary values of these features were different appearances of the same feature (e.g. long or short legs). The $2^6 = 64$ possible insects were created by separating the body, eyes, legs, antennae, fangs, and tail from two drawings of beetles and then recombining them in all possible combinations. Figure 3 provides examples of the stimuli used in the experiments. The mapping of the six physical features onto the six abstract features was randomized across subjects, so that, for example, Feature 1 corresponded to the eyes for one subject and the legs for another subject. Similarly, the level of each feature that was diagnostic of poison was randomized across subjects, so that long legs indicated poison for some subjects, whereas short legs indicated poison for other subjects.

As in Lee and Cummins (2004), all possible pairs of stimuli were shown except pair (2, 7), which was the only pair for which RAT and TTB made opposite choices. As described earlier, the point of the training phase was to train subjects on the stimulus domain without biasing them toward one model or the other by having one model make more accurate decisions. Therefore, pair (2, 7) was eliminated so that both models would make identical choices for every training pair. Table 2 shows the calculation of Bayesian cue validity for each cue (from Equation 2), along with the optimal feature weight (from Equation 4). (The calculations are slightly different for Lee & Cummins [2004] and the present experiments because we used twice as many training trials—see the present *Procedure* section.)

Table 3 contains the five pairs used in the test phase to distinguish RAT from TTB choices, by virtue of the fact that the original strong versions of RAT and TTB make opposite decisions for these pairs.

Procedure. The experimental design was a training phase followed by a testing phase, with each phase consisting of a number of decision trials. During each trial of the training phase, a subject was shown a pair of insects, one to the left and one to the right, on a computer screen and was asked to decide which of the two was more poisonous. A subject gave his or her response by pressing the

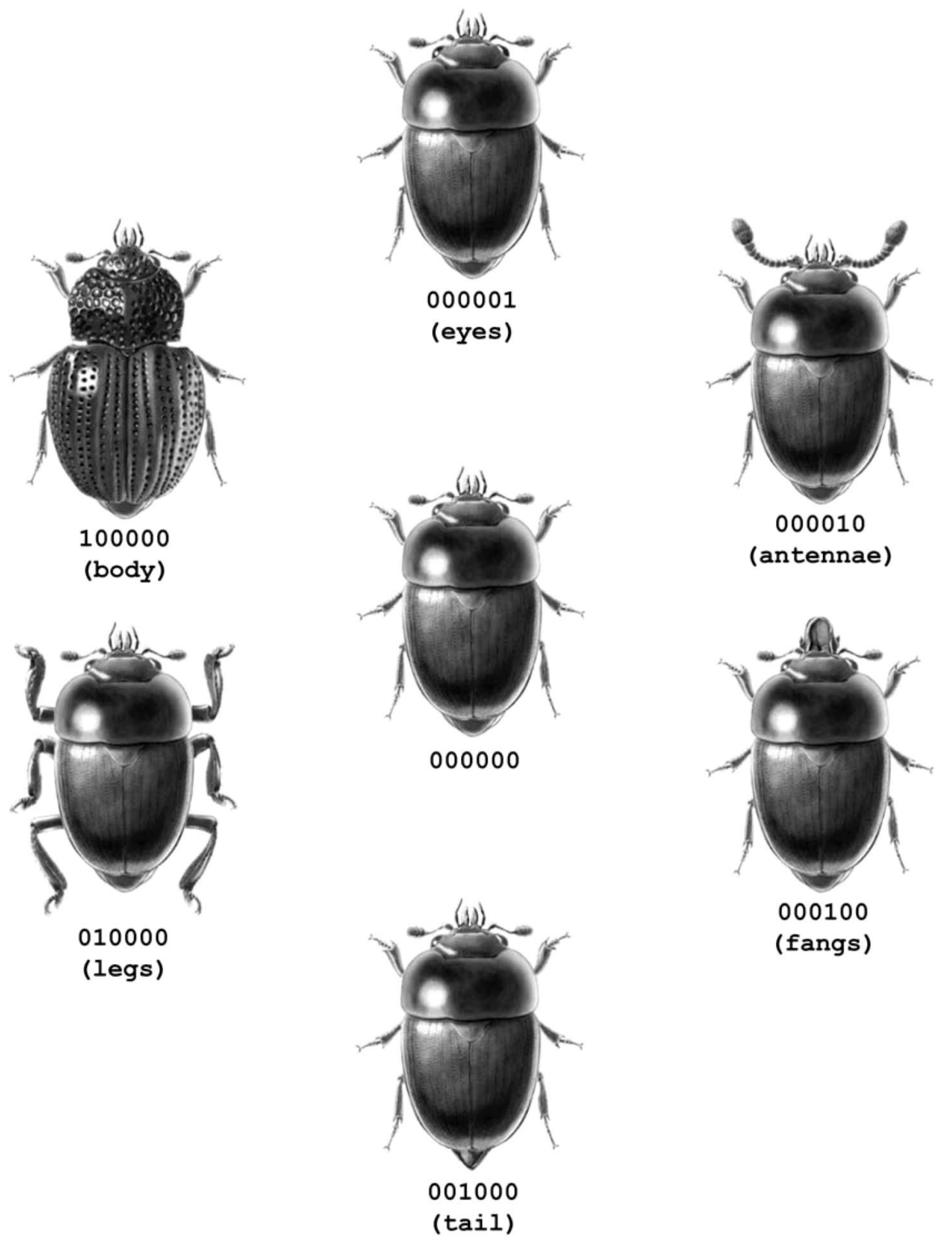

Figure 3. Examples of stimuli used in Experiments 1–3. Note that the correspondence between the underlying stimulus pattern (e.g., 001000) and a particular appearance was completely randomized for each subject. To see the six features, notice that each stimulus in the outer ring differs from the central stimulus by one feature. (Eye color = blue or green on color monitor.)

F key (labeled *Left*) or the *J* key (labeled *Right*). After responding, the subject was given feedback about which response was correct in the form of a red rectangular border appearing around the more poisonous insect. The word *Correct* or *Incorrect* also appeared below the pair of insects during this feedback. The subject was allowed to study the feedback and the pair of insects for as long as he or she wished. To move on to the next trial, the subject pressed the space bar, making each trial self-paced.

Training was designed to familiarize subjects with the cue validities and with the decision task. Our training phase differed from Lee and Cummins's (2004) design in that stimulus pairs were shown twice each, instead of once, for a total of 238 training trials. The increased training was intended to improve the chances that observers would learn the correct cue-validity ordering of the attribute values.

The test phase included presentations of each of the original training pairs and each of the critical transfer pairs (see Table 3) from Lee and Cummins (2004). The test phase also differed from the training phase in that no feedback was provided; instead, following Lee and Cummins (2004), subjects were asked to rate

Table 2
Calculation of Cue Validities (v) and Optimal Feature Weights (w) for Lee and Cummins (2004) and Experiments 1 and 2

Feature_i	Dcorrect_i	Dtotal_i	v_i	$w_i = \ln(v_i/[1 - v_i])$
Lee and Cummins (2004)				
1	59	60	60/62 = .968	3.40
2	55	59	56/61 = .918	2.42
3	52	63	53/65 = .815	1.49
4	40	62	41/64 = .641	0.578
5	34	60	35/62 = .565	0.260
6	34	62	35/64 = .547	0.188
Experiments 1 and 2				
1	118	120	119/122 = .975	3.68
2	110	118	111/120 = .925	2.51
3	104	126	105/128 = .820	1.52
4	80	124	81/126 = .643	0.588
5	68	120	69/122 = .566	0.264
6	68	124	69/126 = .548	0.191

Note. Dcorrect_i reports how many correct discriminations Feature i makes out of all 119 (Lee & Cummins, 2004) or 238 (Experiments 1 and 2) training trials; Dtotal_i reports the total number of discriminations, correct and incorrect; and v_i is the Bayesian validity, calculated by using Equation 2. The optimal weights, which are the log of the odds of the Bayesian validities, are shown in the last column.

how confident they were about every decision. As was the case in training, subjects initiated the next trial by pressing the space bar. We showed the five critical transfer pairs eight times each, instead of once each as in Lee and Cummins. The main reason we increased the number of test pairs is because we wished to create a rich data set suitable for quantitative model fitting at the individual subject level. For the same reason, we showed the 119 training pairs once each during testing (whereas Lee & Cummins, 2004, did not show the training pairs). We also showed the excluded pair (2, 7) once during testing, for a total of 160 test trials. Lee and Cummins's test phase consisted of only the five test pairs, shown once each.

Results

We start by describing the general pattern of results in the data. Table 4 shows the individual subjects' choices on the critical test pairs designed to distinguish strong RAT from strong TTB behavior. Each data column shows how many times the TTB item was chosen for the test pairs, each of which was shown eight times. Most subjects were extremely consistent in their patterns of choices, both within each test pair and across test pairs.

Among both consistent and inconsistent subjects, TTB choices dominate. This result contrasts with Lee and Cummins's (2004) finding of more RAT choices than TTB choices. Although not the focus of the present research, we should note several possible reasons for this contrasting result. First, one possibility is that Lee and Cummins's use of stimuli composed of a set of homogeneous present–absent features may have promoted more use of "summing" strategies than is the case when heterogeneous, substitutive dimensions are involved. A second possibility is that, because Lee and Cummins's design used fewer training trials, perhaps fewer subjects in their study learned that the first attribute had the greatest cue validity. As explained previously in our article, even if subjects had adopted a TTB-like strategy, they might still make a RAT choice if they failed to inspect the most diagnostic dimension first. Still another possibility involves the differing nature of the transfer tasks used in the two studies. Whereas Lee and Cummins's design used only a single presentation of each of the five critical test pairs, in our design there were 160 test pair presentations. Thus, in Lee and Cummins's design, subjects may have been more inclined to consider more of the evidence in making their decisions. A TTB procedure would provide a reasonable shortcut in our much longer and more demanding transfer task. We consider this possibility in greater depth later in our article.

Next, we address the main initial question of Experiment 1: Were our generalizations of RAT and TTB necessary? That is, do gRAT and gTTB provide better explanations of subjects' choice probabilities than do RAT and TTB? The original versions of RAT and TTB are parameter-free and predict completely deterministic responding. Recall that to provide a rudimentary form of error theory, we allow each model a guessing parameter g $(0 < g < 1)$. Thus, for example, the probability that TTB chooses "A" given pair AB is given by $P(\text{"A"}|\text{AB}) = g/2 + (1 - g)\text{TTB}_\text{A}$, where $\text{TTB}_\text{A} = 1$ if the original version of TTB chooses A, and $\text{TTB}_\text{A} = 0$ otherwise.

For purposes of comparability, we add the guessing parameter g to the generalized versions of the models as well. In addition, gRAT and gTTB have free parameters for feature weights and, in the case of gRAT, for γ, which determines how deterministic the response rule is. Because only the relative magnitudes of the weights are relevant, the weights can be normalized to sum to 1 without changing the models' predictions. Thus, the full version of gRAT has seven free parameters (guessing, five free weights, and γ), and gTTB has six (guessing and weights).

To evaluate the utility of the generalized models, we used the Bayesian information criterion (BIC; Schwarz, 1978) as a measure of fit. As described in Appendix B, the BIC includes a term that penalizes a model for its number of free parameters. The BIC gets smaller as the probability of observing the data given the model increases; however, the BIC gets larger as the number of free parameters increases. The model that yields a smaller BIC is considered to provide a more likely account of the data. We used a computer-search algorithm (Hooke & Jeeves, 1961) to locate the

Table 3
Transfer Stimulus Pairs Used by Lee and Cummins (2004)

Transfer stimulus pair	TTB item	RAT item
1	1 0 0 0 0 1	0 1 1 0 0 0
2	1 0 0 0 1 0	0 1 1 0 0 0
3	1 0 0 0 1 1	0 1 1 1 0 0
4	1 0 0 1 1 0	0 1 1 1 0 0
5	1 0 0 1 1 1	0 1 1 1 1 0

Note. The classic rational (RAT) and take-the-best (TTB) models make opposite predictions for these pairs. In each pair, TTB chooses the left item because it has the most valid cue, and RAT chooses the right item because it has more total evidence (assuming the use of optimal feature weights).

Table 4
Response Patterns for Consistent Take-the-Best (TTB) Subjects, Inconsistent Subjects, and Consistent Classic Rational (RAT) Model Subjects From Experiment 1

Subject number	Test pair 1	2	3	4	5
TTB subjects					
1	8	8	8	8	8
3	8	8	8	8	8
6	8	8	8	8	8
13	8	8	8	8	8
14	8	8	8	8	8
17	8	8	8	8	8
25	8	8	8	8	8
28	8	8	8	8	8
31	8	8	8	8	8
33	8	8	8	8	8
34	8	8	8	8	8
38	8	8	8	8	8
39	8	8	8	8	8
40	8	8	8	8	8
41	8	8	8	8	8
52	8	8	8	8	8
54	8	8	8	8	8
58	8	8	8	8	8
59	8	8	8	8	8
60	8	8	8	8	8
Inconsistent subjects					
51	8	8	8	8	7
18	8	8	8	7	8
44	8	8	8	7	7
23	8	8	8	6	8
10	8	8	7	8	8
37	8	8	7	7	8
21	8	7	8	8	8
26	8	7	8	8	8
32	8	7	8	8	8
43	8	7	8	8	8
30	8	7	7	7	7
24	7	8	8	8	6
19	7	8	8	7	7
55	7	8	8	7	6
7	7	8	7	7	8
61	7	8	7	7	5
22	7	7	8	8	8
9	7	6	5	7	6
35	5	5	5	7	3
27	3	3	5	1	3
57	3	3	3	3	2
45	2	7	4	7	7
15	2	5	4	4	0
56	2	3	0	0	0
50	2	2	6	5	4
2	2	1	0	0	0
49	1	4	5	1	2
42	1	2	3	0	0
47	1	2	0	1	1
12	1	1	2	1	6
46	1	0	0	1	1
53	0	6	2	6	0
48	0	2	5	6	1
5	0	0	1	2	1
RAT subjects					
4	0	0	0	0	0
8	0	0	0	0	0
11	0	0	0	0	0
16	0	0	0	0	0
20	0	0	0	0	0
29	0	0	0	0	0
36	0	0	0	0	0

Note. Each data column indicates the number of trials, out of eight, on which the subject chose the TTB item from test pair.

values of the free parameters that minimized the BIC for each individual subject.

Table 5 shows the fits of the strong and general versions of the models, averaged across subjects. Rows 1 and 2 of Table 5 show that RAT and TTB achieve average BIC fits of 181 and 140, respectively. (Not surprisingly, TTB better fit the subjects who made TTB choices on the critical test pairs, whereas RAT better fit the subjects who made RAT choices.) More important, Rows 3 and 4 show that gRAT (with $\gamma = 1$) and gTTB, which are equivalent in terms of choice predictions, achieve average fits of 111. Thus

Table 5
Average Bayesian Information Criterion (BIC) Fits of the Strong and General Versions of RAT and TTB in Experiment 1, Along With the Assumptions Made in Each Model

Model	Guess	Weights	Gamma	Inspection order	Parameters	BIC
RAT	Free	Optimal	∞		1	181
TTB	Free	Optimal		Fixed	1	140
gRAT	Free	Free	1		6	111
gTTB	Free	Free		Probabilistic	6	111
gRAT	Free	Free	Free		7	113

Note. RAT = classic rational model; TTB = take-the-best model; gRAT = general version RAT; gTTB = general version TTB.

the generalizations led to a marked improvement in fit. Row 5 shows that allowing γ to vary freely in the gRAT model yielded a slightly worse average BIC of 113. Thus, in general, the extra free parameter, γ, did not yield a sufficient improvement in fit to offset its increase in model complexity.

A comparison of fits at the individual subject level further supports the conclusion that the generalized models provided an improvement. The leftmost column of Table 6 shows the fit of the equivalent gRAT and gTTB, the middle column shows the fit of RAT, and the rightmost column shows the fit of TTB. The generalized model fit better than RAT for 85% of the subjects, better than TTB for 80% of the subjects, and simultaneously better than both RAT and TTB for 70% of the subjects.

Next, we turn from the overall fits of the models to an analysis of the best-fitting feature weights from gRAT/gTTB, which are shown in Table 7 for each subject. One way to describe the pattern of results is to note that 57 (93%) of the 61 subjects gave over half the total weight to some single feature. This pattern of weighting is noncompensatory in the sense that, if the single feature discriminated between the alternatives, then it would dictate the direction of choice, regardless of the values on the remaining five features. In other words, if the single feature pointed to alternative A, and the remaining five features pointed to alternative B, then the subject would still tend to choose A with probability greater than one half.

Furthermore, 50 (82%) of the 61 subjects gave noncompensatory weights to both of their two most highly weighted cues; for these subjects, the highest weight was greater than all other weights combined, and the second highest weight was greater than all remaining weights combined. This pattern implies that if a pair of alternatives mismatched on either the most highly weighted feature or the second most highly weighted feature, the direction of choice was determined by the value of a single cue.

Perhaps more indicative of the noncompensatory nature of most subjects' responding was that 41 (67%) of the 61 subjects gave over 99% of their weight to some single feature. In this case, not only was the direction of responding dictated by the single feature, it was essentially completely determined by that single feature.

Table 8 lists the proportions of subjects weighting each feature most highly, with the features listed in decreasing order of objective validity. There were 38 subjects (62%) who assigned the highest weight to the most valid cue, leaving 23 subjects (38%) who made a mistake in ordering the most diagnostic cue's validity. Inspection of Table 7 reveals that the vast majority of subjects ordered the remaining cues suboptimally as well.

Discussion

In summary, the generalized versions of RAT and TTB provided a substantially better fit to subjects' choice data than did the original strong versions of these models, even when the generalized models were penalized for their extra parameters by BIC. The improvements in fits are due, at least in part, to the generalized models' allowance of subjective feature weights. The strong version of RAT assumes that subjects use a compensatory set of decision weights (see Table 2); however, 93% of subjects used their most highly weighted feature in a noncompensatory manner (see Table 7). Furthermore, virtually all subjects assigned weights to the dimensions in a manner that departed from their rank order of objective cue validity. This result poses a challenge to the strong version of TTB.

Because gRAT and gTTB are formally identical in their predictions of choice probabilities, we cannot tell them apart on the basis of goodness-of-fit to subjects' choice data. However, in our view, the pattern of estimated feature weightings is far more in the spirit of the TTB process than the RAT process. To elaborate, because the feature weightings tended to be noncompensatory, the value of the single most highly weighted feature dictated the direction of choice between alternatives, regardless of the values on the remaining features. Indeed, for two-thirds of subjects, the single most highly weighted feature received over 99% of the total weight, meaning that their choices were essentially determined by that feature value alone, at least for trials on which that feature discriminated between the alternatives. This type of process is as envisioned by the TTB model and can be viewed as only a rather degenerate case of RAT. In this degenerate case, the observer is presumed to evaluate and integrate all cues, but the choice is determined by the value of just a single cue. Martignon (2001) echoes this view: "A linear model with a non-compensatory set of weights ends up making exactly the same inferences as TTB" (p. 156).

Recall also that, in terms of the gTTB model, the finding of highly noncompensatory weights implies that most subjects followed an essentially fixed order in inspecting the cues that composed the stimulus pairs (see Figure 2C). Thus, although the particular order that most individual subjects used did not conform precisely to the cue-validity ordering, it nevertheless tended to be fixed.

For a much smaller subset of subjects, the feature weights showed a compensatory pattern. Although this result may point to subjects who adopted a RAT-like strategy, it is also consistent with another possibility. In particular, such subjects might have inspected the cues in a probabilistic order but still might have responded on the basis of the first single cue that discriminated between the alternatives, as envisioned in the gTTB model. This possibility can be examined with the RT method that we introduce in Experiment 2.

Although noncompensatory weights provided the best fits to the vast majority of subjects' data, it is possible that comparable fits could be achieved with compensatory weights. A computer-based parameter-search algorithm follows any improvement in fit through the parameter space, however small. It is possible that less extreme weights could yield similar fits and that these less extreme weights would be compensatory. Therefore, though the finding of noncompensatory weights seems to point toward a process in the spirit of the gTTB model, it is important to seek converging evidence for this finding. In Experiments 2 and 3, we use an RT approach to seeking such evidence.

Experiment 2

Although gTTB and gRAT yield formally identical predictions of choice probabilities, they embody dramatically different decision-making processes. One approach to distinguishing between the models would be to use the process-tracing techniques that have proved valuable in past work, in which subjects uncover cues one by one to inspect their values. As acknowledged by other investigators, however, the decision-making behavior of subjects

Table 6
Comparison of Individual Subjects' Model Fits for Experiment 1

Subject	gRAT/gTTB + Guess	RAT + Guess	TTB + Guess
1	65.2	201.5	95.2
2	189.2	188.2	216.0
3	82.4	189.7	62.6
4	155.3	165.7	224.4
5	191.0	185.5	211.8
6	135.9	210.8	121.7
7	77.3	180.1	87.4
8	145.8	143.9	219.6
9	133.7	164.2	117.3
10	94.6	200.7	109.2
11	106.4	126.0	216.7
12	173.3	160.1	189.9
13	77.6	197.8	85.2
14	88.1	197.8	85.2
15	125.3	157.9	177.9
16	112.8	136.8	217.7
17	42.9	193.9	74.2
18	82.0	197.7	100.4
19	107.2	189.7	109.2
20	80.4	77.7	200.4
21	78.1	197.7	100.4
22	110.4	202.9	128.5
23	102.4	186.2	83.6
24	97.8	198.9	129.5
25	60.3	213.4	129.5
26	67.0	196.0	95.8
27	199.6	182.9	189.3
28	67.9	191.8	68.6
29	66.9	98.8	206.8
30	135.0	193.5	130.4
31	52.0	193.9	74.3
32	112.6	197.5	91.1
33	70.6	203.2	100.0
34	102.7	204.8	104.6
35	160.8	177.5	164.2
36	149.2	150.6	221.2
37	95.3	204.1	132.0
38	72.9	213.4	129.5
39	97.0	222.9	162.4
40	74.1	195.9	79.8
41	57.7	201.5	95.2
42	148.3	147.9	196.8
43	151.1	220.5	171.5
44	92.0	195.4	109.3
45	141.2	175.6	155.1
46	136.9	160.2	210.1
47	117.9	120.8	189.1
48	152.2	183.9	197.1
49	169.8	180.0	192.8
50	212.1	190.3	190.3
51	75.1	200.8	109.2
52	77.9	213.4	129.5
53	133.6	142.5	177.0
54	86.6	201.5	95.2
55	99.5	177.6	85.0
56	97.7	101.2	184.0
57	189.1	167.9	182.3
58	113.1	212.1	125.7
59	60.3	197.8	85.1
60	115.7	217.8	136.8
61	104.0	170.4	101.1
Average	111.0	181.5	140.4

Note. Column 2 contains the fit for the general version of the rational model (gRAT) and the general version of the take-the-best model (gTTB), which are formally identical in predicting choice probabilities and therefore achieve the same fit. Column 3 contains each subject's fit of the original RAT with optimal weights and deterministic responding. Column 4 contains the fit of the original TTB model with inspection of cues in order of cue validity.

under such overt monitoring conditions may not conform to their strategies under more natural, covert conditions. The central theme of the present research was to pursue an alternative, complementary avenue to contrasting the models that relies on their RT predictions.

Consider stimulus pairs such as those shown in Figure 4, in which the two stimuli to be compared differ on every feature. All such pairs yield an equally efficient decision process for gTTB. An observer might inspect the cues in any order, depending on the feature weights, but regardless of which order is used the first cue examined would discriminate between the alternatives and therefore allow gTTB to make a decision. Pairs like those shown in Figure 4 guarantee gTTB an identical, maximally fast decision after only one inspection, regardless of the order of cue inspection.

In contrast, gRAT will have a much easier time with pair AB than with pair CD. The gRAT decision is based solely on the total evidence in favor of each alternative. The top pair in Figure 4, AB, is a "RAT-easy" pair, in that the difference in evidence between the alternatives is large: All of the evidence points to A being more poisonous. In contrast, the bottom pair, CD, is a "RAT-hard" pair, in that the difference in evidence between the alternatives is small: Half of the features favor C, whereas the other half favor D. Thus, assuming that responding is faster for easy decisions than for hard decisions, gRAT would predict fast RTs for RAT-easy pairs and slow RTs for RAT-hard pairs.

This definition of RAT-hard and RAT-easy pairs assumes that gRAT is using compensatory weights. With extremely noncompensatory weights, a single feature would dominate the contribution to each item's evidence, so the same evidence difference would exist for every pair of stimuli with completely mismatching features. Thus, the goal here is to distinguish gTTB from compensatory versions of gRAT.[4]

In addition to the types of RAT-easy/RAT-hard pairs illustrated in Figure 4, we also tested RAT-easy/RAT-hard pairs of the form shown in Figure 5. In this case, the two alternatives match on the first cue but mismatch on all other cues. Once again, gTTB predicts no difference in RTs for RAT-easy versus RAT-hard pairs. With some probability, the observer might first check Cue 1 and find that it fails to discriminate between the two alternatives. Thus, the decision will be made upon checking the next cue. However, the probability of first checking Cue 1 is identical for RAT-easy and RAT-hard pairs, so gTTB predicts identical RTs for such pairs. By contrast, gRAT (with compensatory weights) again predicts faster RTs for RAT-easy pairs than for RAT-hard pairs.

Finally, although gTTB predicts no difference in RTs between RAT-easy and RAT-hard pairs, note that it does predict a difference between the six mismatch pairs (see Figure 4) and the five

[4] Note that even if the weights are compensatory (in the sense described earlier in this article), gTTB still predicts no difference in RT between RAT-easy and RAT-hard pairs. If the weights are compensatory, then the order in which gTTB inspects the cues will be probabilistic and will vary across trials. Regardless of the order in which the cues are inspected, however, the model predicts identical RTs because the first cue that is inspected will lead to a decision. Furthermore, even if the time to make a decision varies as a function of which particular cue is inspected (e.g., due to encoding-time differences), the predicted RTs are still identical because the probability of inspecting any given cue is identical for RAT-easy and RAT-hard pairs.

Table 7
Individual Subjects' Best-Fitting Weights for the Equivalent gRAT and gTTB Models, Fit to Experiment 1's Test Phase Choice Probabilities

Subject	w_1	w_2	w_3	w_4	w_5	w_6
1	1.00E+00	3.43E−13	4.91E−07	1.04E−10	1.57E−07	1.00E−15
2	1.35E−13	7.20E−01	2.80E−01	1.35E−13	1.87E−09	3.07E−08
3	1.00E+00	6.44E−09	2.08E−08	1.86E−09	1.00E−12	1.00E−15
4	1.00E−15	1.00E+00	4.69E−07	7.56E−07	1.00E−15	3.46E−07
5	1.00E−15	1.00E+00	5.75E−08	1.00E−15	3.07E−08	1.00E−15
6	1.00E+00	1.00E−15	1.76E−08	1.00E−15	1.00E−15	1.00E−15
7	8.66E−01	5.12E−02	4.25E−02	1.50E−02	2.51E−02	1.76E−10
8	7.23E−05	1.00E+00	4.29E−08	4.29E−08	1.00E−15	4.91E−12
9	9.17E−01	1.26E−05	8.31E−02	2.05E−09	7.59E−11	7.59E−11
10	1.00E+00	1.00E−15	7.52E−08	2.75E−08	2.39E−08	2.74E−12
11	2.50E−11	7.75E−01	4.76E−02	1.41E−01	3.67E−02	2.50E−11
12	9.66E−02	1.87E−01	3.98E−01	2.23E−01	6.01E−12	9.57E−02
13	1.00E+00	1.49E−05	1.49E−05	1.00E−15	2.00E−09	7.29E−13
14	1.00E+00	3.31E−09	9.42E−05	6.75E−09	8.36E−09	2.10E−09
15	9.72E−02	4.55E−11	5.43E−01	3.25E−02	3.28E−01	7.11E−07
16	1.00E−15	1.00E+00	1.37E−06	6.40E−14	1.33E−12	1.25E−13
17	1.00E+00	1.00E−12	1.18E−07	1.69E−09	9.66E−09	1.00E−15
18	1.00E+00	7.29E−13	1.41E−07	1.00E−09	6.64E−09	1.00E−15
19	1.00E+00	3.66E−08	5.15E−08	6.94E−08	2.41E−08	1.00E−15
20	1.13E−07	1.00E+00	2.16E−13	8.00E−15	8.00E−15	1.00E−15
21	1.00E+00	6.86E−09	1.00E−15	1.00E−15	1.00E−15	9.41E−10
22	1.00E+00	1.00E−15	4.43E−05	1.76E−10	9.41E−10	4.74E−09
23	1.00E+00	6.99E−08	1.41E−07	2.75E−08	3.65E−09	1.00E−15
24	9.85E−01	1.48E−02	6.95E−10	1.54E−09	7.52E−06	9.85E−16
25	1.00E+00	1.25E−10	4.44E−08	1.00E−15	1.82E−04	3.43E−13
26	9.99E−01	1.29E−15	8.35E−04	1.29E−15	1.03E−14	3.49E−14
27	1.09E−07	2.28E−04	1.00E+00	1.00E−15	5.06E−11	5.06E−11
28	1.00E+00	2.20E−12	3.32E−05	8.74E−09	1.73E−12	1.00E−15
29	1.23E−09	1.00E+00	9.97E−07	1.00E−12	4.89E−07	1.00E−15
30	1.00E+00	8.62E−07	4.97E−04	1.00E−15	1.00E−09	1.00E−09
31	1.00E+00	9.73E−11	8.10E−05	3.43E−13	2.62E−08	1.00E−15
32	1.00E+00	1.03E−09	1.03E−09	1.00E−15	1.00E−15	1.00E−15
33	1.00E+00	1.00E−15	2.16E−07	3.28E−11	3.90E−08	2.16E−13
34	1.00E+00	6.59E−10	2.07E−05	1.23E−09	7.05E−10	5.51E−10
35	3.29E−01	1.60E−01	1.03E−01	7.15E−11	4.07E−01	2.34E−06
36	2.32E−11	9.96E−01	3.32E−03	2.21E−04	2.32E−11	2.32E−11
37	9.61E−01	6.56E−07	4.91E−03	5.40E−03	2.86E−02	1.60E−10
38	1.00E+00	2.00E−09	1.30E−09	2.05E−09	5.84E−08	1.00E−15
39	5.16E−01	4.58E−11	4.58E−11	7.20E−02	2.20E−01	1.91E−01
40	1.00E+00	2.98E−11	4.57E−06	9.26E−12	7.41E−11	1.00E−15
41	1.00E+00	1.00E−15	2.31E−07	3.43E−13	1.15E−07	7.41E−11
42	1.88E−03	9.98E−01	5.71E−05	8.07E−08	6.06E−05	6.06E−11
43	1.00E+00	4.74E−09	2.20E−12	7.79E−08	1.00E−15	3.34E−08
44	1.00E+00	1.00E−15	3.39E−07	3.43E−13	9.80E−08	6.89E−11
45	2.20E−01	3.74E−06	1.28E−01	3.92E−01	2.60E−01	5.84E−11
46	9.23E−08	1.00E+00	1.00E−15	2.70E−08	3.59E−07	8.46E−08
47	6.42E−02	7.80E−01	2.65E−02	9.73E−02	3.24E−02	4.10E−11
48	4.97E−08	1.86E−01	1.80E−01	6.81E−11	6.34E−01	1.20E−06
49	3.80E−11	5.93E−05	5.97E−01	3.80E−11	4.03E−01	2.25E−06
50	9.96E−16	3.68E−03	1.94E−07	1.38E−11	9.96E−01	9.96E−16
51	1.00E+00	1.00E−15	8.62E−09	2.08E−08	1.19E−08	1.00E−12
52	9.75E−01	3.48E−11	3.22E−07	1.63E−02	3.48E−11	8.99E−03
53	1.32E−01	8.23E−02	4.50E−01	7.48E−02	2.61E−01	6.51E−11
54	1.00E+00	2.74E−12	5.67E−05	1.91E−09	1.19E−08	1.00E−15
55	9.59E−01	8.48E−03	3.08E−02	2.87E−04	1.14E−03	4.34E−04
56	9.25E−02	8.09E−01	6.70E−02	6.27E−03	2.57E−02	6.74E−11
57	1.01E−01	5.77E−01	9.90E−02	2.77E−06	4.27E−07	2.23E−01
58	1.00E+00	6.86E−12	4.47E−05	6.86E−12	1.00E−15	1.60E−08
59	1.00E+00	2.16E−13	1.06E−04	3.93E−11	4.51E−08	1.00E−15
60	1.00E+00	1.26E−04	1.00E−15	2.18E−07	3.43E−13	2.27E−10
61	9.26E−01	4.59E−04	7.29E−02	1.31E−04	7.12E−04	4.19E−05
Average	6.27E−01	2.19E−01	6.82E−02	1.76E−02	6.00E−02	8.52E−03

Note. In comparison, optimal normalized weights (w) are .4204, .2870, .1735, .0671, .0301, and .0218. Parameter values are represented in scientific notation, such that 1.57E−07 represents 1.57×10^{-7} or .000000157. This level of precision is required to determine the extent to which noncompensatory weighting strategies were used by subjects.

Table 8
Proportion of Subjects Weighting Each Feature Most Highly in Experiment 1

Feature	Proportion of subjects
1	.62
2	.23
3	.08
4	.02
5	.05
6	.00

Note. Features are listed in decreasing order of cue validity.

mismatch pairs (see Figure 5), particularly for subjects who give high weight to Cue 1. Such subjects will tend to inspect Cue 1 first, thereby making an immediate decision for the six mismatch pairs. However, because Cue 1 fails to discriminate between the alternatives for the five mismatch pairs, gTTB predicts that decision making will be delayed for those pairs.

Method

Subjects. The subjects were 114 undergraduates at Indiana University Bloomington who participated as part of a course requirement. Subjects were told at the beginning of the experiment that if they performed well on the test phase of the experiment they would be paid a $3 bonus; those who achieved 80% correct or better were paid the bonus.

Stimuli. The stimuli used in the training phase were the same 119 training stimulus pairs (each presented twice) that were used in Experiment 1. Testing introduced two new types of stimulus pairs: some with all six features mismatching and some with five out of six features mismatching. In the latter case, the most diagnostic feature, Cue 1, always matched. The six-mismatch pairs all have the property, described above, of identical gTTB-predicted RTs but different gRAT-predicted RTs, assuming that gRAT uses compensatory weights. The same is true of the five-mismatch pairs. Stimulus pairs of these types were ranked in order of difference in total evidence, assuming optimal weights. From the six-mismatch pairs, the 10 pairs with the largest difference in total evidence (RAT-easy), and the 10 pairs with the smallest difference in total evidence (RAT-hard), were included in the experiment. We used the same procedure to choose the five-mismatch pairs. This procedure yielded 20 six-mismatch pairs and 20 five-mismatch pairs with the most diagnostic feature matching. The complete list of these critical RT test pairs is presented in Table 9. The test phase also included the various pairs used in Experiment 1.

Procedure. The training phase was identical to the one described in Experiment 1.

The test phase included one presentation of each of the 119 training pairs, plus one trial of the untrained pair (2, 7), plus eight presentations of each of the five diagnostic test pairs used by Lee and Cummins (2004), plus three presentations of each of the 40 RAT-easy and RAT-hard RT pairs, for a total of 280 test trials. The order of presentation of pairs was randomized for each subject.

The procedure for the test phase was identical to the one used in Experiment 1, except that subjects were instructed that they should try to answer as quickly as possible without making mistakes. Furthermore, we imposed a 15-s time limit per response. If a subject took longer than 15 s to respond, a message would appear on the screen reminding the subject of the time limit. We felt that this time limit was sufficiently long to allow subjects to inspect all cues, yet it also established the RT context of the experiment. Trials in which subjects exceeded the 15-s deadline (less than 1% of the trials) were excluded from the analyses. RT was measured to the nearest millisecond from the time that the test pair first appeared on the screen to the time that the subject completed his or her button press.

Results

Using the methods described previously, we fitted gRAT/gTTB to the choice-probability data of each individual subject. The best-fitting feature weights are reported in Table 10.

The general pattern of estimated weights was the same as observed in Experiment 1. Of the 114 subjects, 103 (90%) gave over half their weight to a single feature. In addition, of these 103 subjects, 92 (89%) gave over half their remaining weight to the next most highly weighted feature. Thus, once again, an analysis of the best-fitting parameters points to a highly noncompensatory

	w							
	3.68	2.51	1.52	0.59	0.26	0.19	Evidence	
A	1	1	1	1	1	1	8.75	RAT-EASY
B	0	0	0	0	0	0	0.00	
C	1	0	0	1	0	1	4.46	RAT-HARD
D	0	1	1	0	1	0	4.29	

Figure 4. Two pairs of alternatives (AB and CD) that differ (mismatch) on every feature, along with the optimal feature weights and the total objective evidence for each alternative. Pair AB has a large difference in total evidence (8.75 – 0.00 = 8.75), whereas pair CD has a small difference (4.46 – 4.29 = 0.17). The general version of take-the-best predicts identical response times (RTs) for pairs AB and CD because the expected number of cue inspections is the same for both pairs. In contrast, the general version of the rational (RAT) model predicts a faster RT for pair AB because of its larger difference in evidence.

	w							
	3.68	2.51	1.52	0.59	0.26	0.19	Evidence	
A	1	1	1	1	1	1	5.07	RAT-EASY
B	1	0	0	0	0	0	0.00	
C	1	1	0	0	0	0	2.51	RAT-HARD
D	1	0	1	1	1	1	2.56	

Figure 5. Two pairs of alternatives (AB and CD) that differ (mismatch) on every feature except the first, along with the optimal feature weights and the total objective evidence for each alternative. Pair AB has a large difference in total evidence (5.07 − 0.00 = 5.07), whereas pair CD has a small difference (2.56 − 2.51 = 0.05). The general version of take-the-best predicts identical response times (RTs) for pairs AB and CD because the expected number of cue inspections is the same for both pairs. In contrast, the general version of the rational model (RAT) predicts a faster RT for pair AB because of its larger difference in evidence.

pattern of weighting. Nevertheless, only 65 (57%) of the subjects assigned the highest weight to the most valid cue, leaving 43% of the subjects who made a mistake in ordering the cues' validities. Once again, the pattern of feature weighting challenges the strong versions of both RAT and TTB.

We turn now to the central question of interest, namely the RT contrast for the RAT-easy and RAT-hard pairs. First, we constructed a histogram of subjects' proportions correct on the test phase in order to discard subjects who performed at a very low level due to lack of motivation or inability to learn the task. The histogram showed what appeared to be a boundary at 65% correct between a distribution of high-performing subjects and a distribution of low-performing subjects. We therefore excluded from analysis subjects with less than 65% correct on trained pairs during the test phase. Out of 114 total subjects, this left 98 (86%) good performers.

In one approach to analyzing the RT data, we conducted independent samples t tests for each individual subject comparing log RTs for all RAT-easy and RAT-hard trials. For the five-mismatch pairs, only 16 (16%) of the 98 high-performing subjects showed a significant difference ($p < .05$) in the expected direction (RAT-easy trials faster than RAT-hard trials); for the six-mismatch pairs, only 21 (21%) showed a significant difference; and for the five- and six-mismatch pairs taken together, only 21 (21%) showed a significant difference. Thus, less than one-quarter of subjects showed a pattern of RTs indicative of gRAT with compensatory weights. Instead, the vast majority of subjects had an RT signature that points toward gTTB behavior.

In a second method, we used a Bayesian modeling approach to analyze the RT results. In this Bayesian hypothesis test, we treated the null hypothesis of "no difference" between RAT-easy and RAT-hard trials as a model to be fit to a subject's data and the alternative hypothesis of "difference" as another model. The no-difference model assumes that the RTs for both easy and hard trials come from a single distribution. In contrast, the difference model assumes that easy and hard trials' RTs come from two separate distributions. The Bayesian approach then determines which hypothesis is more likely. The details of the analysis are reported in Appendix C. We conducted the analysis separately for the five-mismatch pairs, the six-mismatch pairs, and the five- and six-mismatch pairs taken together. The two-distribution model was favored for the five-mismatch pairs by only 8% of the subjects, for the six-mismatch pairs by only 9% of the subjects, and for the five- and six-mismatch pairs taken together by only 10% of the subjects. Thus, the results of the Bayesian analysis point even more strongly to the gTTB interpretation of the data.

Another RT test of the models concerns the five- versus six-mismatch pairs. The five-mismatch pairs contain one matching feature (Feature 1, the most diagnostic feature), so gTTB predicts a total of two feature inspections whenever Feature 1 is inspected first. By contrast, according to gTTB, only one feature will ever be inspected for the six-mismatch pairs, so the six-mismatch pairs should tend to be faster than the five-mismatch pairs. Furthermore, subjects who attach a large weight to Feature 1 will inspect Feature 1 first very often. Therefore, we would expect these subjects to almost always make two feature inspections when confronted with five-mismatch pairs and therefore to show a pronounced difference in average RT between five- and six-mismatch pairs. In contrast, subjects who attach a small weight to Feature 1 should rarely inspect Feature 1 first and therefore show little or no difference between five- and six-mismatch pairs.

Note that an RT difference between the five- and six-mismatch pairs is also predicted by the gRAT model because of the greater average evidence difference for the six-mismatch pairs compared with the five-mismatch pairs (see Table 9). However, although the finding of such a difference would not distinguish between gTTB and gRAT, it would still be important for two reasons. First, it would provide clear evidence of the ability of the RT measure to reveal such predicted differences. Thus, it would suggest that the null results involving the earlier RAT-easy versus RAT-hard comparisons were not due to a noisy performance measure. Second, if the RT difference held primarily for subjects giving high weight to Feature 1, it would provide evidence of the "psychological reality" of the estimated feature weights. That is, the weights estimated by fitting the models to the choice-probability data would be making correct independent predictions of the pattern of RTs.

In accordance with these predictions, among the 43 subjects with over .99 of the total weight assigned to Feature 1, 41 (95%) showed a significant difference in RT between the five- and six-mismatch pairs. By contrast, among the 35 subjects with less than .50 of the total weight assigned to Feature 1, only 3 (9%) showed a significant difference between the five- and six-mismatch pairs.

Table 9
RAT-Easy and RAT-Hard Test Pairs Used in Experiment 2

A1	A2	A3	A4	A5	A6	EvA	B1	B2	B3	B4	B5	B6	EvB	EvDIFF
Six-mismatch pairs: RAT-Easy														
0	0	0	0	0	0	0.00	1	1	1	1	1	1	8.75	8.75
0	0	0	0	0	1	0.19	1	1	1	1	1	0	8.56	8.37
0	0	0	0	1	0	0.26	1	1	1	1	0	1	8.49	8.23
0	0	0	0	1	1	0.45	1	1	1	1	0	0	8.30	7.84
0	0	0	1	0	0	0.59	1	1	1	0	1	1	8.17	7.58
0	0	0	1	0	1	0.78	1	1	1	0	1	0	7.98	7.20
0	0	0	1	1	0	0.85	1	1	1	0	0	1	7.90	7.05
0	0	0	1	1	1	1.04	1	1	1	0	0	0	7.71	6.67
0	0	1	0	0	0	1.52	1	1	0	1	1	1	7.24	5.72
0	0	1	0	0	1	1.71	1	1	0	1	1	0	7.04	5.33
Six-mismatch pairs: RAT-Hard														
0	1	0	1	1	0	3.36	1	0	1	0	0	1	5.39	2.03
0	1	1	1	1	1	5.07	1	0	0	0	0	0	3.68	1.39
0	1	0	1	1	1	3.55	1	0	1	0	0	0	5.20	1.64
0	1	1	1	1	0	4.88	1	0	0	0	0	1	3.87	1.01
0	1	1	1	0	1	4.81	1	0	0	0	1	0	3.94	0.87
0	1	1	1	0	0	4.62	1	0	0	0	1	1	4.14	0.48
0	1	1	0	0	0	4.03	1	0	0	1	1	1	4.72	0.69
0	1	1	0	1	1	4.49	1	0	0	1	0	0	4.27	0.22
0	1	1	0	0	1	4.22	1	0	0	1	1	0	4.53	0.31
0	1	1	0	1	0	4.29	1	0	0	1	0	1	4.46	0.16
Five-mismatch pairs: RAT-Easy														
1	0	0	0	0	0	0.00	1	1	1	1	1	1	5.07	5.07
0	0	0	0	0	0	0.00	0	1	1	1	1	1	5.07	5.07
0	0	0	0	0	1	0.19	0	1	1	1	1	0	4.88	4.69
1	0	0	0	0	1	0.19	1	1	1	1	1	0	4.88	4.69
1	0	0	0	1	0	0.26	1	1	1	1	0	1	4.81	4.55
0	0	0	0	1	0	0.26	0	1	1	1	0	1	4.81	4.55
0	0	0	0	1	1	0.45	0	1	1	1	0	0	4.62	4.16
1	0	0	0	1	1	0.45	1	1	1	1	0	0	4.62	4.16
1	0	0	1	0	0	0.59	1	1	1	0	1	1	4.49	3.90
0	0	0	1	0	0	0.59	0	1	1	0	1	1	4.49	3.90
Five-mismatch pairs: RAT-Hard														
1	0	1	0	1	1	1.97	1	1	0	1	0	0	3.10	1.13
0	0	1	0	1	1	1.97	0	1	0	1	0	0	3.10	1.13
0	0	1	1	0	0	2.11	0	1	0	0	1	1	2.97	0.86
1	0	1	1	0	0	2.11	1	1	0	0	1	1	2.97	0.86
1	0	1	1	0	1	2.30	1	1	0	0	1	0	2.78	0.48
0	0	1	1	0	1	2.30	0	1	0	0	1	0	2.78	0.48
0	0	1	1	1	0	2.37	0	1	0	0	0	1	2.70	0.33
1	0	1	1	1	0	2.37	1	1	0	0	0	1	2.70	0.33
0	0	1	1	1	1	2.56	0	1	0	0	0	0	2.51	0.05
1	0	1	1	1	1	2.56	1	1	0	0	0	0	2.51	0.05

Note. RAT = classic rational model; EvA = total evidence for A; EvB = total evidence for B; EvDIFF = |EvA − EvB|. The average difference in total evidence is 7.27 for six-mismatch, RAT-Easy pairs; 0.88 for six-mismatch, RAT-hard pairs; 4.47 for five-mismatch, RAT-easy pairs; and 0.57 for five-mismatch, RAT-hard pairs.

Thus, the pattern of results comparing the five- and six-mismatch pairs is as predicted by gTTB.

One concern that may be raised is that, because of the extended number of trials in the test phase, subjects who initially adopted a RAT strategy may have eventually shifted to a TTB strategy to ease their processing burden. To address this concern, we also conducted all of the aforementioned analyses on just the first 70 trials of the test phase (i.e., the first quarter of test phase data). The results were essentially identical to those that we have already reported for the complete set of 280 test trials. For example, among the full set of 98 learners, only 9 (9%) showed a significant difference in RT between the RAT-easy and the RAT-hard pairs. By contrast, among the 43 learners who gave highly noncompensatory weight to Cue 1, 41 (95%) showed significantly faster RTs to the six-mismatch pairs than to the five-mismatch pairs. Thus, the main pattern of

Table 10
Individual Subjects' Best-Fitting Weights for the Equivalent gRAT and gTTB, Fit to Choice Probabilities in Experiment 2

Subject	BIC	w_1	w_2	w_3	w_4	w_5	w_6
1	215	4.94E−01	2.15E−01	9.18E−03	7.12E−03	2.74E−01	5.88E−05
2	176	9.86E−06	9.86E−01	3.90E−03	3.18E−03	4.04E−03	2.80E−03
3	412	9.36E−01	5.96E−02	9.36E−06	9.26E−04	8.42E−05	3.72E−03
4	228	9.90E−01	2.03E−03	3.37E−03	1.35E−03	1.35E−03	1.81E−03
5	113	7.92E−01	1.07E−03	1.21E−02	9.50E−05	1.95E−01	7.92E−06
6	284	9.93E−01	1.38E−03	1.11E−03	1.61E−03	1.62E−03	1.36E−03
7	252	2.23E−02	4.70E−01	3.18E−01	2.23E−03	1.87E−01	2.23E−04
8	267	1.70E−04	2.34E−01	5.09E−01	1.70E−04	2.56E−01	1.70E−04
9	145	9.55E−01	3.04E−03	4.16E−02	5.73E−05	5.16E−04	9.55E−06
10	333	1.46E−01	4.92E−03	5.42E−01	3.02E−01	5.08E−03	1.69E−04
11	312	2.55E−03	9.94E−06	9.94E−01	1.22E−03	1.12E−03	1.10E−03
12	146	9.93E−06	9.93E−01	2.82E−03	1.76E−03	2.20E−03	9.93E−06
13	267	5.57E−01	1.14E−01	8.60E−02	1.62E−01	8.05E−02	5.51E−04
14	113	9.91E−01	1.29E−04	4.45E−03	2.14E−03	2.01E−03	9.91E−06
15	242	9.52E−02	1.55E−04	3.16E−01	2.63E−01	3.22E−01	3.56E−03
16	212	4.55E−01	1.04E−01	2.55E−01	6.95E−02	5.62E−02	6.00E−02
17	196	5.78E−01	4.07E−03	3.02E−01	6.49E−02	5.08E−02	2.72E−04
18	159	2.25E−02	5.16E−01	1.85E−01	2.44E−03	2.74E−01	2.44E−04
19	233	1.78E−01	2.66E−01	2.74E−01	1.51E−01	1.31E−01	1.47E−04
20	118	6.64E−01	1.13E−01	2.22E−01	2.71E−04	5.42E−04	2.71E−04
21	248	3.06E−01	1.35E−01	1.82E−01	1.57E−01	1.58E−01	6.20E−02
22	410	9.89E−01	9.89E−06	9.89E−06	1.68E−04	5.69E−03	5.19E−03
23	200	2.36E−01	3.48E−01	3.92E−01	1.66E−03	2.26E−02	2.07E−04
24	108	9.97E−01	7.98E−05	1.14E−03	8.57E−04	9.97E−04	9.97E−06
25	208	2.02E−01	2.48E−01	1.89E−01	2.05E−01	1.55E−01	1.20E−04
26	140	9.88E−01	9.88E−06	3.84E−03	2.24E−03	3.66E−03	2.04E−03
27	202	9.93E−01	9.93E−06	1.89E−03	9.93E−05	3.07E−03	2.14E−03
28	291	1.80E−04	2.63E−01	1.82E−01	1.55E−01	2.18E−01	1.83E−01
29	140	9.94E−01	5.96E−05	3.84E−03	2.98E−04	2.29E−03	9.94E−06
30	111	3.01E−02	6.59E−01	2.99E−01	3.04E−04	4.56E−03	6.39E−03
31	415	1.65E−04	8.43E−03	3.49E−01	3.59E−01	2.83E−01	1.65E−04
32	174	5.96E−01	1.81E−01	1.14E−01	4.08E−03	1.04E−01	2.92E−04
33	203	2.72E−02	3.95E−01	5.48E−01	2.75E−02	2.36E−03	2.62E−04
34	259	6.03E−01	2.64E−04	2.64E−04	3.96E−01	2.64E−04	2.64E−04
35	109	9.97E−06	9.97E−01	7.98E−05	9.97E−06	1.84E−03	1.13E−03
36	137	9.93E−01	9.93E−06	1.43E−03	9.93E−06	5.76E−03	9.93E−06
37	268	9.87E−01	4.05E−03	2.45E−03	2.64E−03	9.87E−06	4.10E−03
38	420	9.65E−06	1.85E−03	1.35E−04	9.65E−06	9.65E−01	3.30E−02
39	417	6.40E−06	2.56E−04	8.19E−03	6.40E−01	6.40E−06	3.52E−01
40	297	3.11E−01	2.36E−01	8.25E−02	2.19E−01	1.19E−01	3.32E−02
41	265	3.86E−01	1.99E−01	6.35E−03	4.05E−01	3.98E−03	8.47E−05
42	144	9.54E−01	9.54E−06	3.34E−02	7.35E−04	1.18E−02	8.59E−05
43	286	4.62E−01	2.26E−04	2.44E−01	9.00E−02	2.00E−01	3.84E−03
44	211	1.90E−04	5.08E−01	3.01E−01	8.13E−02	5.75E−02	5.20E−02
45	251	2.33E−01	5.00E−03	3.03E−01	1.61E−04	2.33E−01	2.26E−01
46	166	2.07E−01	7.91E−01	9.36E−04	7.20E−05	1.14E−03	8.00E−06
47	218	5.78E−01	2.62E−04	2.03E−01	3.14E−03	3.40E−03	2.13E−01
48	223	9.91E−01	9.91E−06	1.98E−03	1.96E−03	3.61E−03	1.31E−03
49	156	9.41E−01	9.41E−06	1.43E−02	9.41E−05	4.39E−02	9.59E−04
50	273	9.90E−01	9.90E−06	2.32E−03	2.97E−03	2.99E−03	1.95E−03
51	215	4.11E−01	2.29E−04	5.08E−01	3.14E−02	2.93E−02	2.04E−02
52	82	2.72E−04	5.87E−01	1.85E−01	8.52E−02	1.43E−01	2.72E−04
53	239	9.93E−01	2.15E−03	9.93E−06	1.97E−03	1.88E−03	1.25E−03
54	412	1.74E−04	2.93E−01	4.36E−01	9.23E−03	1.74E−04	2.61E−01
55	108	6.27E−01	2.33E−04	1.39E−01	1.63E−03	2.20E−01	1.14E−02
56	96	9.85E−01	9.85E−06	9.81E−03	1.76E−03	3.28E−03	1.18E−04
57	162	8.38E−01	8.05E−03	1.53E−01	8.30E−04	8.38E−06	8.38E−05
58	177	9.89E−01	9.89E−06	3.93E−03	2.76E−03	2.33E−03	2.11E−03
59	162	9.97E−01	1.79E−04	1.90E−03	3.99E−05	1.01E−03	9.97E−06
60	419	2.96E−01	6.74E−06	1.99E−03	2.84E−02	6.74E−01	1.01E−04
61	121	9.58E−01	9.58E−06	3.73E−02	2.16E−03	2.16E−03	1.15E−04
62	238	9.83E−01	3.60E−03	3.34E−03	3.06E−03	3.18E−03	3.91E−03
63	202	9.88E−01	9.88E−06	4.17E−03	3.22E−03	2.70E−03	1.85E−03
64	289	2.09E−01	1.91E−01	2.01E−01	2.24E−01	1.48E−04	1.75E−01
65	262	5.37E−01	1.84E−04	3.11E−01	1.01E−02	1.40E−01	1.29E−03
66	248	4.92E−01	3.87E−01	3.28E−02	1.87E−02	4.48E−02	2.49E−02

Table 10 (*continued*)

Subject	BIC	w_1	w_2	w_3	w_4	w_5	w_6
67	376	1.77E−04	3.07E−01	2.41E−01	1.97E−01	2.54E−01	1.77E−04
68	390	9.84E−01	2.36E−04	9.84E−06	7.85E−03	7.55E−03	2.36E−04
69	358	2.56E−01	2.38E−01	2.94E−01	5.66E−03	8.40E−03	1.98E−01
70	214	1.84E−04	3.43E−01	2.94E−01	1.41E−01	2.21E−01	1.84E−04
71	341	4.43E−01	5.01E−03	1.16E−01	2.30E−01	2.06E−01	2.09E−04
72	406	3.13E−01	4.29E−01	1.57E−04	1.32E−02	2.43E−01	1.89E−03
73	150	9.89E−01	1.38E−04	5.05E−03	1.96E−03	4.14E−03	9.89E−06
74	146	1.00E+00	1.00E−05	4.00E−05	1.00E−05	1.00E−05	1.00E−05
75	121	2.26E−02	9.75E−01	8.97E−04	1.14E−03	9.75E−06	8.78E−05
76	367	3.15E−01	1.45E−02	3.45E−01	1.52E−03	3.23E−01	1.90E−04
77	266	9.94E−01	1.22E−03	1.85E−03	1.27E−03	1.16E−03	9.94E−06
78	419	5.99E−06	2.40E−04	5.99E−01	3.91E−01	5.99E−06	1.02E−02
79	416	9.43E−06	1.60E−04	5.48E−02	9.43E−01	1.98E−03	9.43E−06
80	109	3.71E−02	9.56E−01	2.89E−03	9.56E−06	3.45E−03	1.43E−04
81	261	4.90E−01	2.69E−03	4.80E−01	1.32E−02	1.32E−02	2.24E−04
82	208	3.35E−01	4.23E−01	1.22E−01	2.33E−04	1.16E−01	3.72E−03
83	185	9.92E−01	9.92E−06	4.73E−03	9.92E−06	3.09E−03	9.92E−06
84	103	9.96E−01	9.96E−06	1.69E−03	4.68E−04	1.17E−03	6.37E−04
85	222	9.88E−01	1.88E−04	2.16E−03	9.88E−06	4.09E−03	5.82E−03
86	130	9.96E−01	1.28E−03	3.12E−03	1.99E−05	1.99E−05	9.96E−06
87	419	5.60E−06	4.28E−01	5.60E−01	5.60E−06	2.74E−04	1.20E−02
88	145	9.94E−01	8.95E−05	3.38E−03	1.23E−03	9.15E−04	9.94E−06
89	214	9.91E−01	1.59E−04	4.70E−03	9.91E−06	2.09E−03	1.98E−03
90	198	1.74E−03	9.92E−01	1.91E−03	1.10E−03	2.90E−03	9.92E−06
91	185	9.88E−06	9.88E−01	5.12E−03	2.06E−03	2.75E−03	2.24E−03
92	417	2.14E−04	2.11E−01	2.52E−01	2.14E−04	2.59E−01	2.77E−01
93	219	5.62E−01	1.24E−02	3.94E−01	1.19E−02	1.36E−02	6.89E−03
94	184	3.29E−03	9.90E−01	1.57E−03	2.00E−03	1.98E−03	1.41E−03
95	168	2.94E−03	5.10E−01	3.54E−01	2.10E−04	1.28E−01	4.20E−03
96	229	9.89E−01	2.52E−03	2.77E−03	3.30E−03	2.20E−03	9.89E−06
97	176	9.93E−01	1.83E−03	3.07E−03	9.93E−05	2.02E−03	9.93E−06
98	84	9.53E−01	5.72E−05	4.36E−02	2.77E−04	2.58E−03	9.53E−06
99	158	6.83E−01	4.54E−03	1.68E−01	4.58E−02	9.80E−02	2.06E−04
100	175	9.89E−01	5.94E−05	5.81E−03	3.36E−04	4.48E−03	9.89E−06
101	245	6.43E−06	4.82E−01	5.15E−01	1.18E−03	1.31E−03	6.56E−04
102	318	5.80E−01	2.74E−04	2.10E−01	4.11E−03	1.39E−01	6.66E−02
103	114	3.48E−04	9.93E−01	3.50E−03	5.96E−05	2.73E−03	9.93E−06
104	144	6.68E−01	1.10E−01	6.40E−02	4.72E−02	6.83E−02	4.29E−02
105	186	9.85E−01	3.34E−03	5.11E−03	3.67E−03	9.85E−06	2.85E−03
106	324	9.82E−01	9.82E−06	5.37E−03	4.33E−03	4.04E−03	3.79E−03
107	416	3.50E−05	4.17E−01	2.50E−01	3.50E−03	3.26E−01	3.50E−03
108	235	5.41E−01	1.76E−04	1.12E−01	7.09E−02	1.92E−01	8.43E−02
109	260	2.89E−05	9.63E−01	1.88E−03	9.63E−06	3.52E−02	1.35E−04
110	119	9.89E−01	2.48E−03	6.76E−03	1.09E−04	2.02E−03	9.89E−06
111	191	1.38E−02	8.39E−01	1.44E−01	1.17E−03	1.04E−03	7.64E−04
112	143	2.16E−04	5.61E−01	4.17E−01	1.73E−03	2.03E−02	2.16E−04
113	134	9.93E−01	9.93E−06	1.24E−03	3.83E−03	2.01E−03	8.94E−05
114	166	9.91E−01	9.91E−06	4.71E−03	4.33E−03	9.91E−06	9.91E−06
Average	228	6.35E−01	1.99E−01	7.29E−02	4.23E−02	4.06E−02	1.10E−02

Note. Optimal normalized weights (w) are .4204, .2870, .1735, .0671, .0301, and .0218. Parameter values are represented in scientific notation, such that 1.57E−07 represents 1.57×10^{-7}, or .000000157. BIC = Bayesian information criterion.

results appears to have arisen fairly early during the testing sequence.

Discussion

As in Experiment 1, the best-fitting weight patterns indicate that about 80% of subjects adopted a noncompensatory strategy. Again, the decision making of a vast majority of subjects seems to have been based on single discriminating features, a finding which seems to rule out the classic conception of a rational strategy and points instead toward the use of gTTB.

Turning to the RT data, we see as well that the results follow the predictions of gTTB. That is, the model predicted no difference in RT on RAT-hard versus RAT-easy pairs that differed on every feature or that matched on one particular feature and mismatched on all others. Compensatory versions of gRAT, on the other hand, predicted differences in RT on the basis of differences in total evidence, a result not supported by our data. Moreover, we corroborated gTTB's prediction that, for those subjects giving noncompensatory weight to Feature 1, decision making would be faster for the six-mismatch pairs than for the five-mismatch pairs.

Nevertheless, one might still argue that the lack of a difference in RT between RAT-easy and RAT-hard trials could be attributed to a lack of statistical power or to a noisy experiment. To address this issue, we conducted a new experiment in which subjects were "forced" to use a form of compensatory RAT to perform well on the task. If subjects who performed well showed a significant difference in RT between RAT-easy and RAT-hard trials, it would suggest that our RT measure is capable of detecting RAT-compensatory behavior when it exists.

Experiment 3

The goal of Experiment 3 was to "force" subjects to use a form of compensatory RAT and then compare the RAT-easy/RAT-hard RT results to the previous results from Experiment 2. To promote the use of compensatory RAT, we defined the "poisonousness" of an insect during training to be simply the number of poisonous features it had. Furthermore, in our design, a subject could perform the task perfectly (100% correct) by counting the number of poisonous features each insect had and then choosing the one with the larger number. This strategy is a special case of RAT with equal weights, which Gigerenzer and Todd (1999) call "Dawes's rule" (Dawes, 1979; Dawes & Corrigan, 1974) and Payne and Bettman (2001) call the equal weight strategy. In contrast, TTB's choices would be correct only when it happened to inspect one of the cues in the more poisonous insect first, which would happen only 72% of the time during training (see the present *Method* section). Thus, to achieve a high level of performance, it would be necessary to use gRAT with compensatory weights. Additionally, for reasons explained in the *Method* section, subjects were also provided with explicit instructions to use the RAT strategy.

Once subjects had been trained on this stimulus domain with equal cue validities, an RT transfer task measured performance on RAT-easy and RAT-hard pairs. As in Experiment 2, the use of gTTB would produce no difference in RT between RAT-easy and RAT-hard pairs, whereas use of gRAT with compensatory weights would produce a difference.

Method

Subjects. The subjects were 119 undergraduate students at Indiana University Bloomington who participated as part of a course requirement. Subjects were told at the beginning of the experiment that if they performed well on the test phase of the experiment they would be paid $3; those who achieved 80% correct or better were paid the bonus.

Stimuli. The stimuli were the same poisonous insects used in Experiments 1 and 2, but the underlying stimulus structure differed. During training, all pairs had the property that one feature matched and the other five features mismatched. The feature that matched was constant across all training pairs; for example, 1 subject might see the body matching on every training trial, but across trials the body would take on both of its two possible values.

Although all training pairs had five mismatching features, the distribution of poisonous features varied between pairs. Within the total set of five-mismatch pairs, there were three types: those in which one alternative had five poisonous features and the other had zero (5/0), those in which one alternative had four poisonous features and the other had one (4/1), and those which one alternative had three poisonous features and the other had two (3/2). There were 2 possible 5/0 pairs, 10 possible 4/1 pairs, and 20 possible 3/2 pairs. (Recall that the sixth dimension always matched for these pairs but could take on either of the two possible values.)

Notice that gRAT predicts different RTs for these three types of stimulus pairs (see Table 11) on the basis of their differences in evidence: gRAT predicts 5/0 pairs to be fastest, as a result of their large difference in evidence, and 3/2 pairs to be slowest because of their small difference in evidence. However, because all three types of pairs have one particular matching feature and five mismatching features, gTTB predicts the same average number of cue inspections for all of them, and therefore the same RT, regardless of the distribution of poisonous features. Therefore, gTTB predicts that 5/0, 4/1, and 3/2 pairs should all have the same RT.

Instructions. The experiment reported here was the fourth in a series of experiments that attempted to force subjects to adopt RAT as their decision strategy. The three earlier attempts failed, insofar as most subjects performed poorly on the task, which would be solved by the use of gRAT with approximately equal weights. The critical feature of the present version that allowed for successful performance, we think, was the inclusion of instructions explaining the RAT strategy and explicitly telling subjects to use it. The inclusion of these instructions seemed reasonable because the point of Experiment 3 was to collect RT data from subjects actually using RAT, not to observe what strategy subjects would adopt by default.

Subjects were provided with the same RT instructions in Experiment 3 as those given in Experiment 2. As was the case in Experiment 2, subjects exceeded the 15-s time limit on less than 1% of the trials.

Procedure. Because 5/0 pairs were considered very easy, training consisted of two blocks of 4/1 and 3/2 pairs only. During the first block, each of the ten 4/1 pairs was shown twelve times, for a total of 120 trials. During the second block, each of the ten 4/1 pairs was shown two times, and each of the twenty 3/2 pairs was shown five times, for a total of $20 + 100 = 120$ trials. The order of trials in each block was randomized. The initial training block was used to train subjects on what feature values tended to indicate poison, and the second block was used to motivate subjects to attend to all five diagnostic features.

Testing consisted of 60 RAT-easy trials (thirty 5/0 and thirty 4/1), 60 RAT-hard trials (3/2), and 120 trials with randomly constructed pairs. Note that use of a deterministic RAT strategy with equal feature weights predicts 100% correct on 5/0, 4/1, and 3/2 items during testing. In contrast, TTB predicts 100% correct on the thirty 5/0 trials (because an observer would choose a feature possessed by the more poisonous insect five out of five times), 80% correct on the thirty 4/1 trials, and 60% correct on the sixty 3/2 trials, for 75% correct overall on these RT pairs during testing. Thus, because the point of this experiment was to look at RT differences for subjects actually using a form of compensatory RAT, we considered data only from subjects performing at or above 80% correct on the critical pairs during the test phase. Among the 119 subjects, 76 (64%) met this criterion. In all other respects, the procedure was the same as used in Experiment 2.

Table 11
Stimulus Pairs Used in Experiment 3

A1	A2	A3	A4	A5	A6	EvA	B1	B2	B3	B4	B5	B6	EvB	EvDIFF	Type
Easy pairs															
1	1	1	1	1	0	4.60	0	0	0	0	0	0	0.00	4.60	5/0
1	1	1	1	1	1	4.60	0	0	0	0	0	1	0.00	4.60	5/0
1	1	1	1	0	0	3.68	0	0	0	0	1	0	0.92	2.76	4/1
1	1	1	1	0	1	3.68	0	0	0	0	1	1	0.92	2.76	4/1
1	1	1	0	1	0	3.68	0	0	0	1	0	0	0.92	2.76	4/1
1	1	1	0	1	1	3.68	0	0	0	1	0	1	0.92	2.76	4/1
1	1	0	1	1	0	3.68	0	0	1	0	0	0	0.92	2.76	4/1
1	1	0	1	1	1	3.68	0	0	1	0	0	1	0.92	2.76	4/1
1	0	1	1	1	0	3.68	0	1	0	0	0	0	0.92	2.76	4/1
1	0	1	1	1	1	3.68	0	1	0	0	0	1	0.92	2.76	4/1
0	1	1	1	1	0	3.68	1	0	0	0	0	0	0.92	2.76	4/1
0	1	1	1	1	1	3.68	1	0	0	0	0	1	0.92	2.76	4/1
Hard pairs															
1	1	1	0	0	0	2.76	0	0	0	1	1	0	1.84	0.919	3/2
1	1	1	0	0	1	2.76	0	0	0	1	1	1	1.84	0.919	3/2
1	1	0	1	0	0	2.76	0	0	1	0	1	0	1.84	0.919	3/2
1	1	0	1	0	1	2.76	0	0	1	0	1	1	1.84	0.919	3/2
1	0	1	1	0	0	2.76	0	1	0	0	1	0	1.84	0.919	3/2
1	0	1	1	0	1	2.76	0	1	0	0	1	1	1.84	0.919	3/2
0	1	1	1	0	0	2.76	1	0	0	0	1	0	1.84	0.919	3/2
0	1	1	1	0	1	2.76	1	0	0	0	1	1	1.84	0.919	3/2
1	1	0	0	1	0	2.76	0	0	1	1	0	0	1.84	0.919	3/2
1	1	0	0	1	1	2.76	0	0	1	1	0	1	1.84	0.919	3/2
1	0	1	0	1	0	2.76	0	1	0	1	0	0	1.84	0.919	3/2
1	0	1	0	1	1	2.76	0	1	0	1	0	1	1.84	0.919	3/2
0	1	1	0	1	0	2.76	1	0	0	1	0	0	1.84	0.919	3/2
0	1	1	0	1	1	2.76	1	0	0	1	0	1	1.84	0.919	3/2
1	0	0	1	1	0	2.76	0	1	1	0	0	0	1.84	0.919	3/2
1	0	0	1	1	1	2.76	0	1	1	0	0	1	1.84	0.919	3/2
0	1	0	1	1	0	2.76	1	0	1	0	0	0	1.84	0.919	3/2
0	1	0	1	1	1	2.76	1	0	1	0	0	1	1.84	0.919	3/2
0	0	1	1	1	0	2.76	1	1	0	0	0	0	1.84	0.919	3/2
0	0	1	1	1	1	2.76	1	1	0	0	0	1	1.84	0.919	3/2

Note. EvA = total evidence for A; EvB = total evidence for B; EvDIFF = $|\text{EvA} - \text{EvB}|$. All pairs matched on Feature 6 and mismatched on all other features. The distribution of poisonous features (represented by 1s) differed across pairs, with three types: those in which one insect had five poisonous features and the other had zero (5/0 pairs), those in which one insect had four poisonous features and the other had one (4/1 pairs), and those in which one insect had three poisonous features and the other had two (3/2 pairs). The 5/0 pairs and the 4/1 pairs were considered easy, and the 3/2 pairs were considered hard. The average difference in total evidence is 3.06 for easy pairs (5/0 and 4/1) and 0.919 for hard pairs (3/2).

Results

For each individual subject, we conducted an independent samples *t* test comparing log RT for the 60 RAT-easy (5/0 and 4/1 pairs) and 60 RAT-hard (3/2 pairs) trials. Significant differences ($p < .05$) were observed for 69 (91%) of the 76 high-performing subjects.

We also used the Bayesian modeling approach to analyze these data (see Appendix C). BIC model fits favored the two-distribution model in 61 (80%) of the 76 subjects meeting our performance criterion. Thus the *t* tests and Bayesian analysis both indicate that the vast majority of subjects who used gRAT to perform well showed a significant RAT-easy/RAT-hard difference in RT.

A potential confound is that RAT-easy pairs received more training than did RAT-hard pairs. Specifically, the ten 4/1 RAT-easy pairs received 14 training trials each, for a total of 140 trials, whereas the twenty 3/2 RAT-hard pairs received only 5 training trials each, for a total of 100 trials. This extra training for RAT-easy pairs was included to ease subjects into the training phase because previous attempts to run the experiment had encountered great difficulty in getting subjects to learn the task.

To address this potential confound, we conducted a separate analysis comparing RAT-easy RTs with RAT-hard RTs, excluding 4/1 pairs. Therefore, this analysis included 15 trials for each of the two 5/0 pairs, which were not trained at all, for a total of 30 RAT-easy test trials. It included the three test trials for each of the twenty 3/2 pairs, the same as the original test, for a total of 60 RAT-hard test trials. An independent samples *t* test for each subject compared the distributions of log(RT) for the 30 RAT-easy (5/0 only) and 60 RAT-hard (3/2) trials. Significant differences were found for 71 (93%) of the 76 subjects meeting our performance criterion. The Bayesian analysis with 4/1 pairs excluded favored the two-distribution model in 68 (89%) of the subjects.

Discussion

In Experiment 2, RAT-easy/RAT-hard RT differences were significant for only 21% of subjects, a pattern of null effects that we interpreted as providing support for the use of gTTB by the vast majority of subjects. In Experiment 3, in which high performance required the use of gRAT, the vast majority of high-performing subjects did show a significant RT difference between RAT-easy and RAT-hard trials. This result suggests that the design used in Experiment 2 was capable of producing a significant effect if subjects were actually using gRAT. Thus, the null results of the hard–easy manipulation of Experiment 2 do not seem to reflect a lack of experimental or statistical power.

General Discussion

Summary and Conclusions

One of the main issues in the field of multiattribute probabilistic inference has involved a contrast between models derived from the "fast and frugal heuristics" framework and models that embody classical rationality. A major representative of the former is TTB, and a major representative of the latter is RAT. Numerous experiments have been conducted to determine the conditions in which each processing strategy tends to be used. However, in most cases, quite strong versions of each model have been compared and contrasted with one another. For example, it is usually assumed that subjects use attribute weights based on the objective cue validities of the attribute values. In addition, deterministic versions of the models are generally applied, with no allowance for any noise in the decision process.

Although this research strategy allows one to develop elegant qualitative contrasts between the predictions of competing models, our view is that such strong models are not psychologically plausible. Furthermore, a class of models may be correct in spirit, even if the results of certain qualitative tests falsify a strong special case. For example, observers may indeed follow a TTB-like process in making their decisions, even if they make errors in precisely estimating the cue validities of the attributes in the experienced environment.

For this reason, an initial purpose of the present research was to consider generalized versions of the TTB and RAT models with greater psychological plausibility and to study performance from the perspective of these generalized models. The generalizations that we considered made allowances for the attribute weights to be free parameters and for probabilistic mechanisms to enter into the decision process. Although this research strategy has the obvious disadvantage of requiring quantitative parameter estimation, our view is that, at least in the present situation, some reasonably clear-cut insights were achieved.

First, despite penalizing the generalized models for their increase in the number of free parameters, the quantitative fit indices provided clear evidence that the generalizations were needed. Furthermore, the fits to the choice probability data yielded parameter estimates of the attribute weights that departed markedly from those prescribed for an ideal observer. Subjective feature weights differed from the ideal-observer weights not only in their magnitudes, but also in their ordering. Thus, the results challenge the very strong assumptions made by TTB and RAT. In these respects, there is a considerable difference between the generalized models and the original formulations of both the TTB and RAT models of decision making.

Nevertheless, a second clear-cut result was that, for the vast majority of subjects, the estimated weight parameters followed a noncompensatory pattern (at least for the two most highly weighted dimensions). We argue that such a pattern is far more in the spirit of a TTB-like decision process than a RAT-like process. It implies that the most highly weighted discriminating attribute decides the direction of choice, regardless of the values of the remaining attributes. Indeed, for the majority of subjects, the magnitudes of the weight estimates implied an essentially fixed order of inspection of the cues and a deterministic choice based on the value of a single attribute.

Consider a subject who compared alternatives on an attribute by attribute basis, stopped when a single discriminating cue was located, and then chose the alternative with the positive cue value but that had attribute weight values that did not conform precisely to the cue-validity ordering. Whereas previous research approaches, such as the one adopted by Lee and Cummins (2004), would likely have classified such an observer as a non-TTB decision maker, the present strategy would indicate that there are strong elements of TTB-like decision making in operation. Thus, consideration of the generalized versions of these models can provide insights that would be missed by consideration of the strong versions alone.

A second major contribution of this work was the introduction of an RT method for discriminating between the predictions from the generalized models. In our view, this RT method provides an important form of evidence that complements alternative process-tracing techniques. In the process-tracing techniques, observers are generally required to uncover attribute information one attribute at a time until a decision is made, and the subjects' overt attribute inspections are monitored. Although such methods provide important insights about behavior, investigators have acknowledged potential limitations of the methods as well. In particular, the process-tracing techniques provide an example in which a measurement method can possibly influence the very behavior it is intended to measure. For example, on the one hand, requiring subjects to inspect alternatives one attribute at a time might promote strategies such as TTB. On the other hand, in these situations, subjects have knowledge that their attribute-inspecting behavior is being monitored. Perhaps such knowledge gives rise to demand characteristics, leading subjects to inspect attributes that they might not inspect otherwise.

The RT method therefore provides a potentially useful complement to the process-tracing techniques because it provides a window into the underlying cognitive processes without altering the nature of the testing situation. In the present case, the method provided strong converging evidence that the vast majority of subjects in our experiments adopted a TTB-like decision-making process. Observers' RTs were well predicted by the expected number of inspections required to find a single discriminating cue. Furthermore, the RTs were not influenced by whether a decision was "hard" or "easy" from a classic rationality perspective. On the other hand, in an environment in which the summing strategy of classic rationality was essentially forced upon subjects, the RT predictions from the gRAT model were finally observed. Thus, the method appears to provide an important source of converging evidence bearing on the nature of people's decision-making strategies.

The experimental conditions tested in this work provide only a small sample from a much larger body of research in the published literature. Various aspects of our experimental conditions may have promoted TTB-like strategies, whereas others might be viewed as favoring a RAT-like summing of evidence. The training environment from Experiments 1 and 2 was designed such that the strong TTB and RAT strategies lead to identical choices. Furthermore, the requirement that observers learn the diagnosticity of the cues through extensive pairwise training demands a great deal of cognitive effort, as does completing the extensive number of test trials in our design. Because TTB is presumably more cognitively efficient than is RAT, such factors might reasonably be expected to lead to the adoption of TTB-like strategies. On the other hand, all cues were simultaneously present in integrated, holistic perceptual displays, a condition that other researchers have found to promote RAT-like strategies (Broder & Schiffer, 2003). Careful and systematic research is needed to understand the interplay among such factors in determining performance. Such future inquiries may be fostered by consideration of the generalized models and the RT methods introduced here.

Categorization and Decision Making

It is of interest to consider the relation between the present decision-making tasks and tasks of categorization. In both cases, observers make judgments of a criterion variable based on probabilistic, multiattribute information. However, in categorization, observers typically learn to classify objects into a few nominally labeled groups or equivalence classes. By contrast, in the present tasks, observers needed to learn an ordering of each of the individual stimuli along a continuous-valued criterion. Note that the gTTB model considered in the present work is basically an ordered, single-dimension rule model. The observer is presumed to inspect each dimension, one at a time, and to make a decision as soon as a discriminating cue along any single dimension is found. It is interesting that reliance on single-dimension sources of information is also often observed in categorization. For example, there is a good deal of evidence that in free-sorting tasks, observers often divide objects into classes on the basis of values along a single dimension (e.g., Ahn & Medin, 1992; Medin, Wattenmaker, & Hampson, 1987). Likewise, in tasks of classification learning, there is evidence for an initial reliance on single-dimension rules (e.g., Nosofsky, Palmeri, & McKinley, 1994; Nosofsky & Zaki, 2002). In many categorization situations, however, sole reliance on a single-dimension rule is insufficient to yield adequate classification performance, so at later stages observers learn to combine multiattribute information into integrated units, such as exceptions to the rule, exemplars, or prototypes. Analogously, for the multiattribute decision-making tasks tested in Experiments 1 and 2, observers could rely on an ordered sequence of single-dimension rules, and the evidence suggests that the vast majority did so. By contrast, for environments in which such strategies yield inadequate performance, such as Experiment 3, observers learn to adopt more compensatory, information-integration strategies (e.g., Juslin et al., 2003). Thus, similar principles may operate across the categorization and decision-making domains.

Future Research Directions

Finally, an exciting new direction of research has been aimed at unifying the alternative TTB and RAT strategies within a common framework. The evidence accumulation via the random walk model of Lee and Cummins (2004; Newell, 2005) searches for information cue by cue, as in TTB, with the evidence provided by each cue (the cue's weight) serving as the walk's step size. The difference between TTB and RAT from this perspective is simply a difference in the evidence thresholds individuals require for making a decision: TTB has a threshold lower than the evidence provided by the least weighted cue, and RAT has a threshold higher than the evidence provided by all cues combined. The TTB threshold leads to a decision once a single mismatching cue is found, whereas the RAT threshold leads to a decision only after all cues are inspected. Through the model selection criterion of MDL, Lee and Cummins (2004) found that this unified model fit the pattern of subjects' decisions better (MDL = 87.6) than either TTB (MDL = 138.6) or RAT (MDL = 130.7).

Although Lee and Cummins's (2004) model provides a simple and elegant description of strategy differences across conditions, as well as individual differences within conditions, it remains to be seen whether subjects do conform to its specification of cuewise, as opposed to alternativewise, search (e.g., Juslin & Persson, 2002). In addition, as currently specified, Lee and Cummins's (2004) unified model makes the same strong assumptions that we address in this article, namely that subjects use optimal feature weights and behave deterministically. Thus, an important direction for future research will be to generalize this model in the same ways that we generalize TTB and RAT in this article, by allowing for probabilistic choice of cue inspection orders based on individual differences in cue weights. Such a model has promise for providing a rigorous joint account of choice probabilities and the time course of processing in tasks of multiattribute inference.

References

Ahn, W. K., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science, 16,* 81–121.

Broder, 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.

Broder, A., & Schiffer, S. (2003). Take-the-best versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General, 132,* 277–293.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571–582.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81,* 95–106.

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

Gigerenzer, G., & Selten, R. (2001). *Bounded rationality: The adaptive toolbox.* Cambridge, MA: MIT Press.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 3–34). New York: Oxford University Press.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109,* 75–90.

Hooke, R., & Jeeves, T. A. (1961). "Direct search" solution of numerical and statistical problems. *Journal of the Association for Computing Machinery, 8,* 212–229.

Johnson, E. J., Payne, J. W., Schkade, D. A., & Bettman, J. R. (1991). *Monitoring information processing and decisions: The mouselab system.* Unpublished manuscript, Center for Decision Studies, Fuqua School of Business, Duke University.

Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 924–941.

Juslin, P., & Persson, M. (2002). PROBabilities from EXemplars (PROBEX): A "lazy" algorithm for probabilistic inference from generic knowledge. *Cognitive Science, 26,* 563–607.

Katsikopoulos, K., & Martignon, L. (in press). Naive heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology.*

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., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the "take the best" and "rational" models. *Psychonomic Bulletin and Review, 11,* 343–352.

Martignon, L. (2001). Comparing fast and frugal heuristics and optimal models. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 147–171). Cambridge, MA: MIT Press.

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.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19,* 242–279.

Newell, B. R. (2005). Re-visions of rationality? *Trends in Cognitive Sciences, 9,* 11–15.

Newell, B. R., Rakow, T., Weston, N., & 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., & Shanks, D. R. (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 923–935.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 924–940.

Payne, J. W., & Bettman, J. R. (2001). Preferential choice and adaptive strategy use. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 123–145). Cambridge, MA: MIT Press.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* New York: Cambridge University Press.

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.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General, 135,* 207–236.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica, 14,* 465–471.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464.

Simon, H. A. (1956). Rational choice and the structure of the environment. *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). New York: Cambridge University Press.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review, 79,* 281–299.

Appendix A

Proof of Formal Equivalence of gRAT ($\gamma = 1$) and gTTB's Predictions of Response Probabilities

Because the pair of stimuli contains no matching features, the first cue that is inspected will lead gTTB to make a decision. The probability that gTTB chooses alternative A is therefore found by computing the probability that the first cue inspected favors alternative A—that is, that the first cue inspected is one in which alternative A has a positive cue value. Letting cue_1 denote the first cue inspected, FA the set of cues favoring alternative A, FA_i the ith cue favoring alternative A, and N_{FA} the number of cues favoring alternative A, then

$$
\begin{aligned}
P(\text{“A”}|AB) &= P(cue_1\ favors\ A) \\
&= P(cue_1 \in FA) \\
&= P[(cue_1 = FA_1) \vee (cue_1 = FA_2) \\
&\qquad \vee \ldots \vee (cue_1 = FA_{N_{FA}})] \\
&= P(cue_1 = FA_1) + P(cue_1 = FA_2) \\
&\qquad + \ldots + P(cue_1 = FA_{N_{FA}}) \\
&= \frac{w_{FA_1}}{\sum w} + \frac{w_{FA_2}}{\sum w} + \ldots + \frac{w_{FA_{N_{FA}}}}{\sum w} \\
&= \frac{w_{FA_1} + w_{FA_2} + \ldots + w_{FA_{N_{FA}}}}{\sum w} \\
&= \frac{\sum_{i \in FA} w_i}{\sum_{i \in FA} w_i + \sum_{i \in FB} w_i},
\end{aligned}
$$

which is the predicted probability of $P(\text{A}|\text{AB})$ from gRAT when $\gamma = 1$ (compare with Equation 5).

Appendix B

Method of Model Evaluation

The BIC fit for a model is given by

$$\mathrm{BIC} = -2\ln(\mathrm{L}) + \mathrm{N_{par}}\ln(\mathrm{N_{obs}}), \quad \text{(B1)}$$

where ln(L) is the log-likelihood of the data given the model, $\mathrm{N_{par}}$ is the number of free parameters in the model, and $\mathrm{N_{obs}}$ is the number of data observations on which the fit is based.

In the present case, the likelihood of the data set is given by

$$\mathrm{L} = \prod_{m=1}^{M} \binom{N_m}{f_m} p_m^{f_m}(1 - p_m)^{(N_m - f_m)}, \quad \text{(B2)}$$

where M = total number of stimulus pairs being modeled, N_m = number of observations for pair m, f_m = frequency of observing an "A" response for pair m, and p_m = predicted probability of an "A" response for pair m. This likelihood function assumes that the response choices are binomially distributed and that the observations are independent. The best-fitting parameters for a model are the ones that maximize the Equation B2 likelihood function and that thereby minimize the BIC expressed in Equation B1.

Appendix C

Bayesian Hypothesis Testing

In the Bayesian t test, we treat the null hypothesis of no difference between RAT-easy and RAT-hard trials as a model to be fit to a subject's data and the alternative hypothesis of difference as another model. The no-difference model assumes that the RTs for both easy and hard trials' RTs come from a single distribution. In contrast, the difference model assumes that easy and hard trials' RTs come from two separate distributions.

As in the classical t tests we conducted, we make the standard assumption that log(RT) follows a normal distribution. Fitting the single-distribution model involves the estimation of two free parameters, the mean (μ) and the standard deviation (σ) of the normal distribution. To evaluate its fit, one calculates the probability density of observing a subject's entire data set of easy and hard trial RTs from a normal distribution with the given μ and σ. The probability density of a single observed $\ln(\mathrm{RT}_i)$ with value x is therefore given by

$$P(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad \text{(C1)}$$

The probability density of observing a subject's entire data set $\vec{x}$ of RTs is then simply the product of the individual probability densities:

$$P(\vec{x}|\mu,\sigma) = \prod_i P(x_i|\mu,\sigma). \quad \text{(C2)}$$

For the two-distribution model, the procedure is very similar, except that the probability density of observing an individual RT depends on whether that RT was for an easy or a hard pair. If it was for an easy pair, its probability is calculated by using Equation C1, but with $\mu = \mu_{\mathrm{easy}}$ and $\sigma = \sigma_{\mathrm{easy}}$; if it was for a hard pair, its probability is calculated by using $\mu = \mu_{\mathrm{hard}}$ and $\sigma = \sigma_{\mathrm{hard}}$. Once the individual RTs' probability densities have been calculated, the probability density of the entire data set of all easy and hard trials is calculated as the product of the individual trials' probability densities. Finally, one uses a computer search method to find the values of the free parameters that maximize the likelihood for each model, and one compares the fits by using the BIC statistic described in Appendix B.

Received January 17, 2006
Revision received April 10, 2006
Accepted May 7, 2006 ■