Psychological Review
2008, Vol. 115, No. 1, 199–213

0033-295X/08/$12.00 DOI: 10.1037/0033-295X.115.1.199

THEORETICAL NOTES

Psychological Plausibility of the Theory of Probabilistic Mental Models and the Fast and Frugal Heuristics

Michael R. Dougherty
University of Maryland, College Park

Ana M. Franco-Watkins
Auburn University

Rick Thomas
University of Oklahoma

The theory of probabilistic mental models (PMM; G. Gigerenzer, U. Hoffrage, & H. Kleinbölting, 1991) has had a major influence on the field of judgment and decision making, with the most recent important modifications to PMM theory being the identification of several fast and frugal heuristics (G. Gigerenzer & D. G. Goldstein, 1996). These heuristics were purported to provide psychologically plausible cognitive process models that describe a variety of judgment behavior. In this article, the authors evaluate the psychological plausibility of the assumptions upon which PMM were built and, consequently, the psychological plausibility of several of the fast and frugal heuristics. The authors argue that many of PMM theory's assumptions are questionable, given available data, and that fast and frugal heuristics are, in fact, psychologically implausible.

Keywords: judgment, decision making, fast and frugal, recognition heuristic, take the best

Research on heuristics and biases has dominated behavioral decision theory for the past several decades. This approach characterizes decision makers as having a limited set of general purpose cognitive heuristics that can be used to make quick and relatively effortless decisions (Tversky & Kahneman, 1973, 1974, 1982). Although this research led to many insights concerning the factors affecting judgment and decision making, it has been criticized for a number of reasons (Dougherty, Gettys, & Ogden, 1999; Gigerenzer, 1996). As a result, several researchers have argued that the focus should shift beyond describing heuristics and their associated biases to developing quantitatively specified cognitive process models of judgment and decision making (Dougherty et al., 1999; Gigerenzer, Hoffrage, & Kleinbölting, 1991).

One alternative approach that has received considerable attention was born out of the theory of probabilistic mental models (PMM; Gigerenzer et al., 1991; Gigerenzer & Goldstein, 1996). PMM provided a unified theory of judgment grounded in Brunswik's probabilistic functionalism and led to the development of a set of fast and frugal cognitive algorithms or heuristics (Gigerenzer & Goldstein, 1996). In contrast to the heuristics born out of the heuristics and biases program, which were assumed to be domain general, those proposed by Gigerenzer and Goldstein (1996) were assumed to be content and domain specific as well as ecologically grounded. The algorithms were labeled "fast and frugal" because they presumably required little time, knowledge, and computational ability. In addition to proposing that these algorithms are psychologically plausible, Gigerenzer and Goldstein (1996) illustrated that they performed as well as or better than more complex algorithms, such as multiple regression models.

Although the most recent instantiations of fast and frugal heuristics have abandoned their connection to PMM theory, the initial heuristics proposed by Gigerenzer and Goldstein (1996; Goldstein & Gigerenzer, 2002) were built on the original PMM theory, and many assumptions of PMM were carried over. Hence, the aim of the present article is to examine the assumptions of PMM and to assess the psychological plausibility of the fast and frugal heuristics they imply. Our criticisms are organized around four components of PMM and the fast and frugal heuristics proposed by Gigerenzer and Goldstein (1996). First, we argue that the automatic frequency-counter assumption adopted by Gigerenzer and Goldstein (1996) is not well supported by the existing literature. Second, the definition of cue validity, the foundation for the most widely studied heuristic (Take The Best or TTB), is fundamentally flawed. Third, given the problem with the formula for cue validity and the automatic frequency-counter assumption, validity-guided generation is

Michael R. Dougherty, Department of Psychology, University of Maryland, College Park; Ana Franco-Watkins, Department of Psychology, Auburn University; Rick Thomas, Department of Psychology, University of Oklahoma.

This material is based on work supported by National Science Foundation Grant SES-0134678 awarded to Michael R. Dougherty. We thank Thomas Wallsten, Amber Sprenger, David Huber, Eddy Davelaar, and Scott Gronlund for comments on an earlier draft of this article.

Correspondence concerning this article should be addressed to Michael R. Dougherty, Department of Psychology, University of Maryland, College Park, MD 20742. E-mail: mdougherty@psyc.umd.edu

highly unlikely. Instead, we argue that memory retrieval processes likely supersede cue validity as a general mechanism for cue generation. Fourth, the recognition principle, as implemented by Gigerenzer and Goldstein (1996; Goldstein & Gigerenzer, 2002) violates accepted principles of memory and also ignores essential assumptions of ecological theories (of which PMM theory is an example). We end by presenting the results of a set of simulations comparing the recognition heuristic with a familiarity model and a frequency-sampling model.

Description of PMM

PMM provide a framework for using probabilistic information from the natural environment to make judgments. They are invoked when local mental models (LMM) cannot effectively derive the solution for a specific judgment. In a two-alternative general knowledge task, LMM reflect the process of retrieving specific information directly from memory and/or obtaining it by the process of elementary logical operations to make the judgment.

In contrast to LMM, PMM require the interplay between the structure of the task and the structure of a person's environment and assume that inference about a criterion variable is based on cues that are probabilistically related to the criterion variable. Consider a two-alternative general knowledge question. First, a reference class (set of objects from a person's environment) pertaining to both items in the task is activated. Cue validities in the form of conditional probabilities are assumed to be based on the relative frequencies with which cues predict the outcome variable. The formula for cue validity is

$$v_i = p(t(a) > t(b) | c_i(a) = + \text{ } and \text{ } c_i(b) = -), \tag{1}$$

where the validity (v_i) of cue i (denoted c_i) on target variables a and b, $t(a)$ and $t(b)$, is given by the relative frequency with which $t(a) > t(b)$, given that a is positive on cue i and b is negative in reference class R. The terms $t(a)$ and $t(b)$ correspond to the value of the target variable for object a and b, respectively. For instance, if the task is to judge which city is more populous on the basis of whether the city has a soccer team, $t(a)$ and $t(b)$ correspond to the value of the population of City A and City B, conditioned on all pairs of cities in which one city in the pair has a soccer team ($c_i(a) = +$) and one city does not ($c_i(b) = -$). Therefore, for items a and b, the validity of cue c_i is given by the proportion of times that the cue would discriminate correctly between a and b. For example, if a cue has a validity of .90, then it will correctly discriminate between a and b on 90% of the trials in which that cue can be used.

The Fast and Frugal Heuristics

Recent developments of PMM theory have focused on identifying several fast and frugal heuristics. As Gigerenzer and Goldstein (1996) stated, these cognitive algorithms "are realizations of the framework for modeling inferences from memory, the theory of *probabilistic mental models* " (p. 652). They elaborated on three such algorithms: TTB, Take The Last (TTL), and minimalist. We begin by focusing on TTB because it has received the most attention in the literature and because it is the "basic algorithm of the PMM framework" (Gigerenzer & Goldstein, 1996, p. 653). Lesser space is devoted to evaluating the TTL and minimalist algorithms.

TTB consists of a five-step algorithm, as detailed in Figure 1 (adapted from Gigerenzer & Goldstein, 1996, Figure 2). Consider the decision where one must choose the more populous of two German cities. TTB proceeds first by implementing the recognition principle. If one of the two cities is recognized (cue value is +) and the other is not recognized (cue value is −), then choose the recognized city. If neither of the two cities is recognized, then choose randomly (guess). If both cities are recognized (both have + for the recognition cue), then proceed to Step 2: Search for cue validities and retrieve the cue values of the highest ranking cue from memory. Step 3 is the discrimination rule: The cue discriminates if one of the cities has a positive cue value and the other has a negative cue value or is unknown (cue value is '?'). Step 4 is the cue substitution principle. If the cue discriminates, then the search for cues stops and the decision process proceeds to Step 5. If the cue does not discriminate, then the decision maker is assumed to return to Step 2 and selects the next best cue. Step 5 is the maximizing rule for choice, which says that the decision maker should choose whichever city has the positive cue value. If none of the cues discriminate, then an alternative is chosen at random.

TTB makes four important assumptions: (a) Cue validities are prestored in the form of co-occurrence frequencies, (b) participants start by determining whether the choice can be made by recognition—if one alternative is recognized and the other is not, then choose the alternative that is recognized, (c) cues are ordered hierarchically from most to least predictive, and (d) cues are searched sequentially starting with the most predictive. In contrast to TTB, TTL assumes cues are generated sequentially from the most recently used to the least recently used. Thus, TTL does not require a hierarchical ordering of cue validities; it needs only a record of which cues discriminated in the past. Finally, the minimalist algorithm assumes cues are generated randomly until one is found that discriminates. Like TTL, minimalist does not require a hierarchy of cues; however, unlike TTL, it does not require a memory of past cues.

The Psychological Plausibility of PMM

As with any theory, the assumptions underlying PMM can be categorized as either primary or auxiliary. Primary assumptions are those that are foundational to the theory and central to its functioning. In contrast, auxiliary assumptions are those that are made out of convenience and are necessary for the implementation of a model but not foundational to the theory. In this section, we briefly discuss the primary assumptions of PMM and outline potential problems of these assumptions.

Automatic Frequency Counter

PMM and the fast and frugal algorithms assume that cue validities are based on a frequency-counter process, such as that proposed by Hasher and Zacks (1979). The basic idea was that environmental events or objects are registered in memory akin to an event counter such that whenever the same event is encountered, frequency information is incremented. One by-product of the frequency-counter process is that frequency encoding is as-

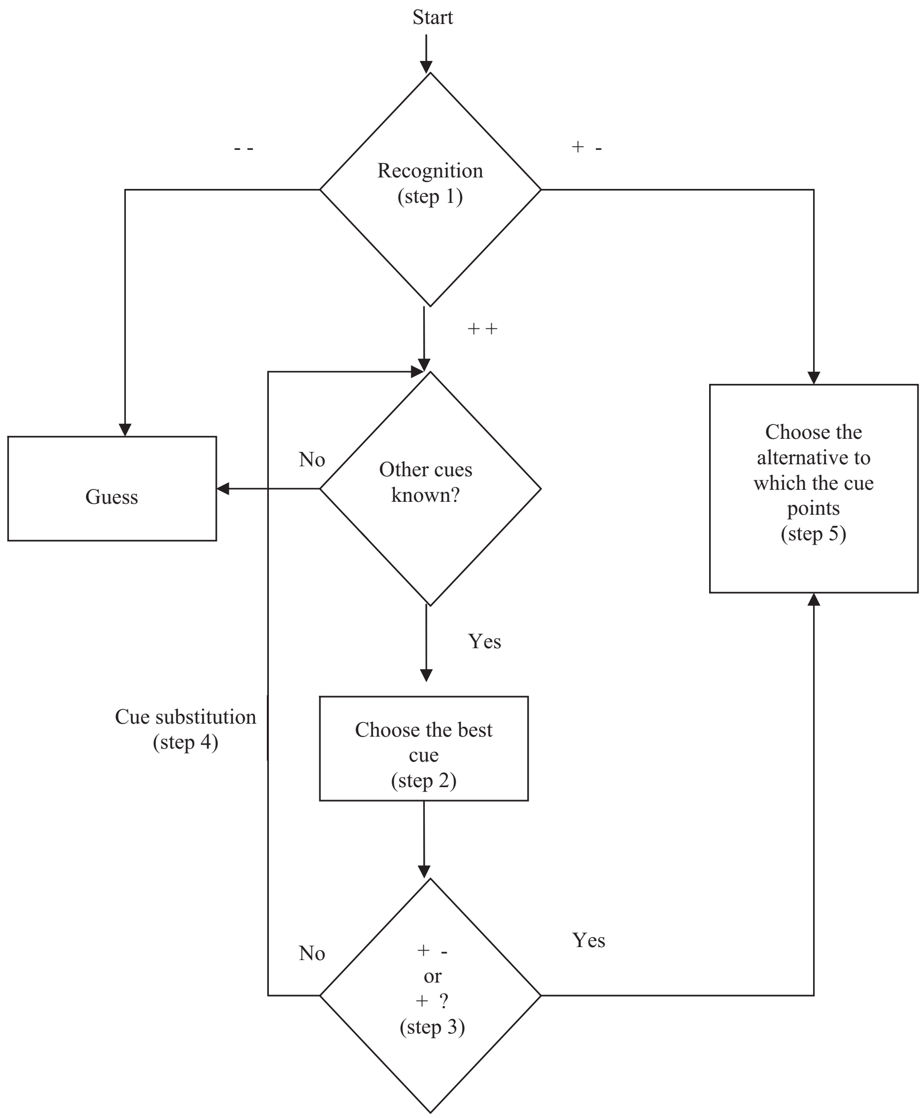

Figure 1. Flow diagram for the Take The Best heuristic. Adapted from "Reasoning the Fast and Frugal Way: Models of Bounded Reality," by G. Gigerenzer and D. G. Goldstein, 1996, *Psychological Review, 103*, p. 653. Copyright 1996 by the American Psychological Association.

sumed to be relatively accurate, regardless of the amount of attention devoted to the event.

How central is the frequency-counter mechanism to PMM theory? Gigerenzer et al. (1991, p. 510) seemed to be noncommittal with respect to the exact nature of the frequency mechanism, for they stated,

> whatever mechanism of frequency encoding, we use the following assumption for deriving our predictions: If subjects had repeated experience with a reference class, a target variable, and cues in their environment, we assume that cue validities correspond to the ecological validities.

This statement implies that the frequency-counter assumption is auxiliary. However, because the frequency-counter assumption is necessary (though not sufficient) for the functioning of some of the mechanisms proposed by PMM, such as the TTB algorithm (discussed below), it constitutes a primary assumption.

At issue is the distinction between online and retrospective models of frequency estimation. Online models assume that frequency information is updated dynamically over time. Accordingly, judgments of frequency merely require a read out of the prestored frequency. Hasher and Zacks's (1979) frequency-counter model is one instantiation of an online model. Retrospective models, such as Hintzman's (1988) Minerva 2 model and Brown's (1995) enumeration model, do not record frequency of occurrence per se. Rather, frequency is inferred from the representation of individual events stored in memory. Thus, frequency of occurrence is determined retrospectively by mapping a memory variable (e.g., familiarity) onto a response scale.

Why do PMM require an online model? The assumption that cues are accessed according to their validity necessitates that the cue validities are precomputed and stored in memory. If cue validities are not prestored, then the TTB algorithm cannot work because a hierarchy of cue validities would not exist. Given this as the case, the event-counter assumption is central to the functioning of the model, and data arguing against this assumption, which are quite widespread, pose serious problems for the theory (see Brown, 1995, 1997; Greene, 1986, 1988; Hanson & Hirst, 1988; Hintzman, 1988; Naveh-Benjamin & Jonides, 1986).[1]

In the absence of a frequency-counter process, how would one establish a cue hierarchy? Retrospective models require that cue validities are computed and ordered at the time of judgment: Participants would have to generate a list of cues, compute their validities, and order the list of cues by their validities before any of the cues could be tested. Such a process would be computationally intensive. Moreover, the assumption that cues are generated before assessing validity leads to circularity in the definition of TTB: Generation cannot be based on cue validity unless cue validity information is available prior to generation.

The event-counter mechanism is implausible. However, even if we accept the event-counter assumption, there are still problems with its implementation within PMM and TTB. One problem concerns the complexity of computing cue validities. Consider the classic city-size problem often used to demonstrate TTB: "Which city is larger? Bonn or Munich?" Gigerenzer and Goldstein (1996) demonstrated TTB within an ecology of 83 cities and nine cues. The computation of cue validity within this ecology would require 30,627 pairwise comparisons just to establish the cue validity hierarchy for predicting city size (see also Juslin & Persson, 2002). However, because any particular predictor cue can also serve as the criterion (e.g., one might want to use city size to predict which city is likely to have monuments), it would require $J(J-1)/2 \times K(K-1)/2$ pairwise comparisons, where J is the number of cues and criteria in the matrix and K is the number of objects (e.g., cities) in the reference class. For the city-size task with $J = 10$ (9 cues and 1 criterion) and $K = 83$ cities, this would entail 153,135 pairwise comparisons. In principle, there are no limits on the number of potential predictor and criterion variables (i.e., there are many more than nine cues that one might use to predict city size and many criterion variables of interest other than city size). Moreover, the ordering of the cues would be different depending on the criterion variable of interest, and they would need updates periodically in response to changes in the reference class, the addition of new cues to the ecology, and changes in one's ecological experience. These updates are necessary if TTB is to maintain its ecological adaptability.

These computational complexities aside, a much more fundamental problem lies in the definition of cue validity itself. Namely, the definition requires complementary knowledge of events that are present and events that are absent from the environment.

The Definition of Cue Validity and Reliance on Missing Information

Cue validities are assumed to be based on co-occurrences derived from a set of pairwise comparisons between objects within the reference class. At this point, it is necessary to highlight three assumptions of PMM. First, as an ecological theory, PMM theory assumes that people learn probabilistic relationships in the environment through ecological sampling, that is, randomly sampling from the ecology. The basic idea is that individuals learn statistical relationships in the environment merely by interacting in that environment. More experience is assumed to lead to cognitive adjustment, such that cognitive representations more accurately reflect the statistical structure of the environment. These are foundational assumptions of ecological psychology, which we view as relatively uncontroversial.

Second, the idea of ecological sampling and the computation of cue validity necessitates that one identify a well-defined reference class. Although this assumption may well be instantiated at the individual subject level, for theory-testing purposes, one needs to define a priori the reference class over which cue validities are computed so that predictions can be made and proper experimental tests carried out.

A third assumption, which builds on the first two, relates directly to the definition of cue validity. Equation 1 can be restated as

$$p(t(a) > t(b) | c_i(a) = + \textit{ and } c_i(b) = -)$$
$$= \frac{F(t(a) > t(b) | c_i(a) = + \cap c_i(b) = -)}{F(t(a) > t(b) | c_i(a) = + \cap c_i(b) = -) + F(t(a) \leq t(b) | c_i(a) = + \cap c_i(b) = -)}, \quad (2)$$

where F corresponds to the count of the conditional event. Equation 2 is a straightforward conditional probability that can be computed if one has knowledge of both the positive and the negative cue values, as could be obtained by referencing an almanac.

Figure 2 presents a graphical depiction of Equation 2 but expressed in terms of state tables. The tables consist of frequency counts derived by making all $K(K-1)/2$ pairwise comparisons of a set of objects (e.g., cities) within a reference class. These pairwise comparisons can be divided into two sets: those for which $t(a) > t(b)$ (the left box) and those for which $t(a) \leq t(b)$ (the right box). Within each set, we can further compute the frequency for which (cue $i = +$ for object a) $\cap$ (cue $i = +$ for object b), (cue $i = -$ for object a) $\cap$ (cue $i = +$ for object b), and so forth. The relevant cells for computing cue validity are given by Cells B and B′: $v_i = p(t(a) > t(b) | c_i(a) = + \textit{ and } c_i(b) = -) = \mathrm{B}/(\mathrm{B} + \mathrm{B}')$. In words, Cell B corresponds to all pairs of cities for which the larger city has a monument and the smaller city does not, and Cell B′ corresponds to all pairs of cities for which the smaller of the two has a monument and the larger one does not.

Implementation of Equation 2 within the context of a cognitive model requires several key assumptions. First, the formula is assumed to operate on the decision maker's memory representation as constructed through ecological sampling, not information gleaned from an almanac. Thus, any implementation of the formula needs to take into account the psychological and sampling constraints of the memory system and how these cues might be represented mentally. Second, as illustrated in Figure 2, cue va-

[1] It is important to note also that the application in PMM of the frequency-counter model operates on co-occurrences, not single events, where the co-occurrences involve missing information. We return to this point in our discussion of the definition of cue validity.

$t(a) > t(b)$

Value of cue i on object a	Value of Cue i on object b: +	Value of Cue i on object b: -
+	A	B
-	C	D

$t(a) \leq t(b)$

Value of cue i on object a	Value of Cue i on object b: +	Value of Cue i on object b: -
+	A′	B′
-	C′	D′

$$v_i = p(t(a) > t(b) \mid c_i(a) = + \text{ and } c_i(b) = -) = B/(B+B')$$

Figure 2. State tables illustrating the input required for computing cue validity. Cells in the table represent the co-occurrence of the cue's presence (denoted as +) and absence (denoted as −) in the environment. For example, Cells A and A′ correspond to co-occurrences of cue *i* that is present for both object *a* and object *b*. Cells B, B′, C, and C′ correspond to co-occurrences where a cue is present for one object but absent for the other object. Cells D and D′ correspond to the co-occurrence of cues that were absent for both objects *a* and *b* in the pairwise comparisons. The equation illustrates the relevant cells for computing cue validity.

lidities are derived from the co-occurrence of objects with and without the cue (e.g., pairs of cities, where one city has a monument and the other does not). Finally, because cue values with minus signs correspond to the absence of the cue with an object in the environment, the definition assumes that people can register the co-occurrence of missing information. For example, if a city does not have any monuments, then for the purposes of computing cue validity, one would need to register the co-occurrence of the city and that absence. The assumption that we maintain a record of missing information that can be fed into a validity calculation is tantamount to assuming that we encode information in memory that is absent from the environment. As an example, you probably did not encode that the word "and" was not used in the prior sentence, nor were the words "xylophone," "pepper," "you," or millions of other words or nonwords. Certainly it is possible to build a memory representation that can register the presence of events (and even the co-occurrence of events that are present) in the environment, but we see no way that a memory representation can register the absence of information, let alone infer a co-occurrence matrix based on the presence and absence of information. However, registration in memory of cues absent from the environment is as necessary as the registration of cues present, and both present and absent cues are afforded the same information value. As a litmus test, consider building an artificial intelligent agent that could learn cue validities through unsupervised learning by interacting in its environment. How would this agent perceive and register in memory the absence of information in its environment?

Not only does the registration of nonoccurrences fail the test of logic, but considerable evidence within both the probability learning (Hearst, 1991) and the memory (Brewer & Treyens, 1981) literatures also indicates that people generally are poor at recognizing information absence. The inability to encode the co-occurrence of two events, where one event is present and the other is absent, is the basis of the feature positive effect—the empirical finding that people and other animals generally are poor at detecting the absence of information (Lea, 1974). Indeed, this result is highly robust, even when such absences are predictive of an outcome variable, and has been argued to be fundamental to behaviors ranging from discrimination learning to concept learning to decision making (for an excellent review, see Newman, Wolff, & Hearst, 1980).[2] Although people can learn correlations between events when information is explicitly presented to them, work on the feature positive effect indicates that participants fail to learn in probability learning tasks when it is the absences (as opposed to the information that is present) that are predictive of the criterion variable.

Generation of Cues

The three algorithms outlined by Gigerenzer and Goldstein (1996), TTB, TTL, and minimalist, all assume that cues are generated from memory. The first of these, TTB, assumes that the process involves sampling from a hierarchy of cue validities starting with the cue that has the highest validity. As argued above, we view the notion of a prestored hierarchy as implausible.

Accepting, for the moment, the possibility that one has access to missing information and can compute cue validity, how might TTB be implemented in the absence of a prestored hierarchy? One possibility that does not necessitate a cue hierarchy but still requires that cue validity be computed according to Equation 1 involves (a) generating a set of candidate cues from long-term memory, (b) computing the validities of each cue in the candidate set, and (c) selecting the cues from most to least valid within the candidate set. However, this set of processes is computationally intensive and, by necessity, must be governed by domain-general memory retrieval variables: The candidate set would still have to be driven by retrieval variables.

Several issues arise within the context of cue generation and its relationship to memory retrieval processes. When does one decide to terminate the cue generation process? Is the process influenced by primacy, recency, and/or other memorial or motivational factors (cf. Dougherty & Harbison, in press)? Gigerenzer et al. (1991) addressed the former problem by stating that cue generation can be terminated after only one cue is generated that adequately discriminates between task items. Certainly it seems plausible that one would terminate cue generation after generating a discriminating cue, but the task goal, in particular with TTB, is to choose the cue with the highest validity because it has the highest likelihood of leading to a correct judgment. However, if only one cue is generated, then how does one know that it has the highest cue validity? Again, such a position requires that the decision maker maintain a hierarchy of cue validities that can be generated from best to worst.

There are many alternative mechanisms that might guide (or influence) cue generation. For example, drawing on memory theory, one might hypothesize that the order in which cues are generated is related to how frequently they occur in one's environment (see work on the word-frequency effect; Gregg, 1976),

[2] Note that our criticism of the inability to use missing information is specific to the computation of cue validity and not to whether people use binary cues as the basis of judgment. People may well use the presence or absence of cues as the basis of judgment, but we argue that using validity as the basis of cue choice is implausible because of its reliance on missing information.

how easily they can be brought to mind, or other memory retrieval variables (Koriat, 1993).[3]

How does one reconcile data illustrating that participants generate cues from best to worst if, as we argue, the hierarchy cannot be constructed? In some empirical studies cited as evidence for TTB, participants were explicitly told the cue validities or provided with hints regarding which cues were best (e.g., Bröder, 2000, 2002; Bröder & Schiffer, 2003; B. R. Newell, Rakow, Weston, & Shanks, 2004; Rieskamp & Hoffrage, 1999; Rieskamp & Otto, 2006). Such tasks do not provide evidence that participants can learn cue validities through experience or that they construct the cue hierarchy. Rather, they reflect one's ability to remember which cues the experimenter told or hinted to him or her were good cues. Although some studies have required participants to learn cue validities through experience, no attempt is made to control for the differential retrievability of the cues. It is possible that cue validity in these studies is correlated with cue accessibility or retrievability. Although this possibility offers an intriguing connection between cue generation processes and models of memory, it suggests that valid tests of TTB require that one explicitly control for memory retrieval variables. Moreover, it suggests that more emphasis needs to be placed on understanding the memory-theoretic alternatives to validity-guided cue generation.

TTL and minimalist fare better than TTB in that neither requires the cue hierarchy. However, problems persist, as both still ignore basic retrieval processes. For example, TTL assumes a sequential search of cues based on past use: The rule is to choose the cue that discriminated at Time $t - 1$, and if that does not discriminate, then choose the cue that discriminated at Time $t - 2$, and so forth. Although TTL capitalizes on recency effects, it ignores the fact that the retrieval of cues will become more difficult (and less successful) the longer one has to search for a discriminating cue. Moreover, implementation of TTL is conditioned on a prior cue selection algorithm, such as TTB, minimalist, or even a prior implementation of TTL. Minimalist is the simplest of the three algorithms proposed by Gigerenzer and Goldstein (1996), as it proposes that cue generation is random. However, as with TTB and TTL, minimalist ignores memory retrieval variables. Minimalist is particularly problematic, in the sense that randomness cannot be empirically validated.

Thus far, we have argued three main points: (a) The frequency-counter mechanism is implausible, (b) the formula for cue validity cannot be implemented cognitively both because it is computationally intensive and because it requires access to missing information, and (c) cue generation is likely driven by memory retrieval processes, not cue validity. We now turn our attention to the recognition heuristic and Gigerenzer and Goldstein's (1996; Goldstein & Gigerenzer, 2002) implementation of TTB.

Recognition Principle

A major component of TTB is the recognition principle, which apparently accounts for a fair number of judgments within TTB (Goldstein & Gigerenzer, 1999; Goldstein & Gigerenzer, 2002). We question the validity of the recognition heuristic, as implemented in Goldstein and Gigerenzer (2002), on three grounds.[4]

1. The recognition heuristic treats recognition as an all-or-none process, which is counter to the literature on recognition memory.

2. Their implementation of the recognition heuristic model failed to incorporate the assumption of ecological sampling—a foundational assumption of ecological theories and one that is explicitly embodied in PMM.

3. They claimed that the so-called "less-is-more" effect arises from an increase in knowledge. However, as we demonstrate below, there is at least one alternative interpretation of their experimental data that does not imply that too much knowledge is bad.

All-or-none recognition. Goldstein and Gigerenzer (2002; Gigerenzer & Goldstein, 1996) assumed that recognition is all or none. Specifically, they stated, "Thus, with the term *recognition,* we divide the world into the novel and the previously experienced" (Goldstein & Gigerenzer, 2002, p 77). Moreover, they explicitly separate their use of the term "recognition" from the concepts of familiarity and availability:

> Unlike availability, the recognition heuristic does not address comparisons between items in memory, but rather the difference between items in and out of memory (Goldstein, 1997). The term *familiarity* is typically used in the literature to denote the degree of knowledge (or amount of experience) a person has of a task or object. The recognition heuristic, in contrast, treats recognition as a binary, all-or-none distinction; further knowledge is irrelevant. (Goldstein & Gigerenzer, 2002, p. 77)

Inspection of the simulations implemented by Gigerenzer and Goldstein (1996) and Goldstein and Gigerenzer (2002) are consistent with the interpretation that recognition is all or none. This view contrasts with most current theories, which assume recognition is based on a continuous underlying memory variable, often referred to as "memory strength" or familiarity. In most memory experiments, dichotomizing the memory variable works well because participants typically are presented with single-item presentation (i.e., they are not making a choice of which of two items is more familiar but rather whether a single item exceeds the threshold). However, for any single recognition memory judgment involving nonnovel stimuli, even items not studied on a study trial will elicit some level of familiarity (hence, false alarms often are observed). The task for the participant in a memory experiment therefore becomes one of deciding whether the feeling of familiarity can be attributed to the recent exposure of the stimulus on the study list or to having been exposed to the item prior to the

[3] There is ample empirical evidence across a variety of tasks that the cues people use in forming judgments are partly influenced by memory retrieval variables (Adelman, Bresnick, Christian, Gualtieri, & Minionis, 1997; Anderson & Norman, 1964; Chapman, Bergus, & Elstein, 1996; Hoch, 1984; Hogarth & Einhorn, 1992).

[4] B. R. Newell and Fernandez (2006; see also B. R. Newell & Shanks, 2004) showed that the recognition heuristic is sometimes used in conjunction with other cues in a compensatory fashion. Thus, the assumption that the recognition heuristic is independent of cue use in the TTB heuristic may be unwarranted. In addition, we argue that one's knowledge of environmental cues likely is not independent of the number of objects one recognizes. That is, participants who happen to know a lot of cities would also presumably know a lot about those cities (i.e., they would be able to sample cues with higher validities).

experiment: Is the familiarity high enough to decide that it is "old"?

In contrast, when presented with an *n*-alternative choice task, the task becomes one of deciding which alternative has the highest familiarity, not whether a single item's familiarity exceeds a threshold (Nosofsky, 1992). In decision tasks such as those used by Gigerenzer and colleagues (e.g., city size; see, e.g., Gigerenzer & Goldstein, 1996; Goldstein & Gigerenzer, 2002), there exists no explicit learning within the experiment. Therefore, any recognition decision is based on whatever information was learned naturalistically prior to the experiment: The task does not require the decision maker to make an attribution of source. This is problematic for the recognition principle because it means that any pair of alternatives in a two-alternative (or greater) forced-choice task in which both alternatives elicit a feeling of familiarity will not be solvable by the recognition principle: By definition, the recognition principle can be triggered only when one item is recognized and the other is not.

We suspect that recognition memory processes can be used in a much wider number of circumstances, even when participants recognize both alternatives. In this case, choice can be based on whichever alternative is most familiar (a continuous variable) rather than the all-or-none process proposed by Gigerenzer and Goldstein (1996).[5] Characterizing the recognition principle in terms of the underlying continuous variable and using memory models has a number of advantages. First, it allows us to describe the underlying memory processes and to derive novel predictions regarding decision-making behavior (see Dougherty, 2001; Dougherty et al., 1999; Juslin & Persson, 2002). Second, it allows us to test the boundary conditions of seemingly nonintuitive predictions, such as the less-is-more effect (e.g., Pleskac, 2007), and to elucidate alternative accounts of these effects. In our case, we show that the data supporting the seemingly nonintuitive prediction that less knowledge leads to better inference can be accounted for by a simpler alternative account.

Where is the ecology? PMM and the fast and frugal heuristics embody elements of ecological psychology and, as such, assume that the internal cognitive representation is based on sampling information from one's environment. Presumably, the more experience one has within a particular ecology, the better the internal representation will reflect the ecological structure (Juslin, Olsson, & Bjorkman, 1997). As stated previously, we agree with this assumption but wonder why the authors did not incorporate it into their simulations of the recognition heuristic.

Gigerenzer and Goldstein (1996; Goldstein & Gigerenzer, 2002) illustrated that University of Chicago students' recognition of German cities was correlated with number of newspaper citations in the *Chicago Tribune* ($r = .79$) and that the number of newspaper citations was correlated with city population ($r = .70$). Goldstein and Gigerenzer (2002) used these correlation statistics to support the assumption that environmental mediators can be used as the basis of inference. We agree that mediators can be used as the basis of inference but argue that the mediator, rather than the recognition data, should be the basis of modeling.

Goldstein and Gigerenzer (2002) implemented their model by relying on the recognition data rather than on newspaper citations and, in so doing, sidestepped an important assumption of ecological theories. In their simulation, Goldstein and Gigerenzer used a Guttman scale to rank order German cities from most recognized to least recognized and entered cities into the simulation according to this rank ordering.[6] After each city was entered, the model computed the number of correct inferences that would be made if relying on recognition. If neither city was recognized, then the model chose randomly between the cities; if one city was recognized but the other one was not, then the model chose the recognized city as the largest; and if both cities were recognized, then the model implemented the TTB strategy with knowledge validities manipulated from .5 to .9. Using these simulations, Goldstein and Gigerenzer demonstrated what they refer to as the less-is-more effect, where more knowledge purportedly led to poorer performance. This result is illustrated as an inverted U-shaped function between number of cities recognized and the probability of choosing the city with the highest population. Peak accuracy is achieved when about 50% of the cities are recognized, so long as recognition validity is greater than the knowledge validity. Performance begins to decline as the percentage of cities recognized increases from about 50% to 100%. Thus, according to Goldstein and Gigerenzer, the more cities one knows, the less applicable the recognition heuristic becomes and (if the knowledge validity is less than the validity of recognition heuristic) accuracy declines—a less-is-more prediction. As empirical support for the less-is-more prediction, they demonstrated that American students are more accurate at judging the size of German cities ($Mdn = 73\%$ correct) than the size of American cities ($Mdn = 71\%$ correct), and German students are more accurate at judging American cities than German cities.

We take issue with Goldstein and Gigerenzer's (2002) implementation of the recognition heuristic. First, they treated participants from different ecologies as points on the same curve, as exemplified by their use of the three-sister analogy and its comparison to American students' estimates of American and German cities. The analogy is set up by assuming that three sisters differ on the amount of knowledge they have about various cities: The youngest has little knowledge, the middle sister has moderate knowledge, and the oldest sister has a high amount of knowledge. They then argued that the sister with a lot of knowledge would perform less well than the sister with moderate knowledge because she cannot implement the recognition heuristic. Although Goldstein and Gigerenzer (2002) pointed out that German and American students belong to different ecologies and that German and American cities are distinct reference classes (see Goldstein & Gigerenzer, 2002, p. 83), their simulations treat American and

[5] One could suppose that the recognition heuristic operates on a continuous underlying memory variable, where the participant makes two single-item recognition decisions to determine whether the activation of either or both alternatives exceeds a threshold criterion. However, this type of mechanism necessitates that the recognition heuristic be conceptualized within the context of a threshold model, such as signal-detection theory, where the contributions of d', criterion setting, and experience can be explicitly modeled (see Pleskac, 2007).

[6] Of importance, Nosofsky (1992) showed that one cannot infer *n*-alternative choice data from single-item recognition data. That is, there is empirical evidence that attentional processes and the decision rule differ between *n*-alternative and single-item recognition tasks. Thus, it seems unjustifiable for Goldstein and Gigerenzer (2002) to use the single-item recognition test as a measure of the contribution of recognition in the *n*-alternative task.

Table 1
Ecologies and Reference Classes Operating in the City-Size Task From Goldstein and Gigerenzer (2002)

Ecology	Reference class: German cities	Reference class: American cities
German students	German students' inferences about German cities	German students' inferences about American cities
American students	American students' inferences about German cities	American students' inferences about American cities

German students as if they are analogous to the middle and oldest sisters. Because American students have more knowledge of American cities than of German cities, their judgments of American cities are analogous to those of the oldest sister whereas their judgments of German cities are analogous to those of the middle sister. By consequence, American students are expected to make better inferences for German cities than for American cities. This same logic is applied in reasoning that German students should predict American cities more accurately than should American students and vice versa. Indeed, Gigerenzer and Goldstein (1996) reported such results. When conceptualized within the context of the three-sister analogy, these findings appear to support the less-is-more prediction.

However, there are several problems with interpreting these data as evidence of a less-is-more effect. First, the recognition heuristic predicts a nonmonotonic relationship between number of cities recognized and proportion correct. Thus, a valid test of the model requires a minimum of three data points.

The second problem is that the three-sister analogy implies a single reference class within a single ecology. However, the comparison of German and American students requires that we model two ecologies: one corresponding to German newspapers and one corresponding to American newspapers (presumably, Germans read German newspapers and Americans read American newspapers). In addition, there are two reference classes within each ecology: newspaper citations for German cities and newspaper citations for American cities. Thus, to compare predictions for Americans versus Germans judging American cities, we need distinct accuracy curves constructed from German and American newspaper citations about American cities. In a similar vein, comparing American students' inferences of American and German cities implies one ecology but two reference classes: The ecology is the American newspaper database, and the two reference classes are citation rates of German and American cities, respectively.

Inherent in our criticism of the recognition principle is the difficulty of identifying the reference class over which cues are computed. The data cited by Goldstein and Gigerenzer (2002) in support of the less-is-more effect actually imply four separate accuracy curves corresponding to the crossing of ecologies (American newspapers and German newspapers) with reference classes (German cities and American cities). This is illustrated in Table 1. Clearly, German and American students must be modeled by different ecologies (rows in Table 1), and, arguably, there needs to be separate reference classes for American and German cities (columns in Table 1). Model predictions need to be derived by comparing different combinations of ecologies and reference classes. The three-sister analogy pertains to how much experience one has with objects within a single reference class in a single ecology and cannot be extended to comparing judgments in different ecologies and/or reference classes. In essence, the proper test of the recognition model involves three sisters with different levels of knowledge all raised in the same ecology making inferences about objects within the same reference class. To our knowledge, no such data has been collected to test this prediction.

A third criticism is that the simulation methodology used by Gigerenzer and Goldstein (1996; Goldstein & Gigerenzer, 2002) ignores ecological sampling. If we take seriously the assumption of ecological sampling, then simulations of the recognition heuristic should be based on the mediating variable of newspaper citations, not the recognition performance of participants. Given this as the case, if we ask American students to judge American cities and German cities, then the simulation model needs to be run twice—once using the citations of the American cities and once using the citations of German cities, where citation frequencies are tabulated from an American database. Of importance, the ecological sampling assumption entails sampling processes. Thus, in modeling the recognition heuristic, knowledge of city names should be assumed to arise from a sampling process, where city names are sampled from the mediator (the distribution of newspaper citations) and where the sample size can be varied from 1 to N.

Finally, to infer a less-is-more effect within this paradigm, the ecological correlation between newspaper citations and German city size needs to be identical to the ecological correlation between newspaper citations and American city size. Note that we found this not to be the case for the newspaper citation rates of German and American cities. We evaluated the ecological correlations between city population and citation frequency of the 83 largest German cities and the 83 largest American cities in the *Chicago Tribune* from 1985 to 1997 (the database and time period used by Goldstein & Gigerenzer, 2002). The Kendal's τ correlation was .50 for the German cities and .39 for the American cities.[7] If the correlations between the environmental mediators and the criterion are not equal, any differences in accuracy could be the result of ecological structure and not experience or increases in knowledge of cities. In fact, as we show through simulations, one can account for the finding that Americans make slightly less accurate inferences regarding the population of American cities than German cities without assuming differences in experience, number of recognized cities, or differences in cue knowledge.

Three questions naturally suggest themselves from the above discussion. First, do the main predictions of the recognition heuristic hold when one allows recognition knowledge to be determined by a sampling process? Second, does the recognition heuristic discriminate between different ecologies (i.e., different distributions of citation frequencies)? Third, can one develop an

[7] The Pearson correlation between the citation frequencies with actual population was .85 for the German cities and .38 for the American cities.

alternative model that can predict the finding that Americans are more accurate judging German cities compared with American cities? We propose an alternative model below that does not require that one accept the idea that too much knowledge is bad.

Contrasting the Recognition Heuristic With a Familiarity Model

To address these three questions, we implemented three models: the original recognition heuristic model, a familiarity-based model, and a frequency model. Our simulations compare predictions in the city-size task for American students judging American and German city size (bottom row in Table 1).

Our implementation of the recognition heuristic differed somewhat from that of Goldstein and Gigerenzer (2002) but is consistent with the underlying assumptions of PMM. Rather than feed the city names into the model according to the rank order in which they were recognized in the single-item recognition task, we determined which cities were recognized by randomly sampling (with replacement) from the distribution of newspaper citations (cf. Schooler & Hertwig, 2005). Because Berlin has the highest citation rate, it has the highest probability of being sampled. We modeled two sampling distributions on the basis of newspaper citations in the *Chicago Tribune*, one representing the citations for the 83 largest German cities and the other representing the citations for the 83 largest American cities. We sampled cities according to the actual environmental ratios as represented in the *Chicago Tribune*, where the ratio of American to German city citations was 226:1; for every one German city mentioned in the newspaper database, 226 American cities were mentioned. We varied the sample size across 18 levels from $N = 1$ to infinity. It is important to note that in this model, the total sample size consists of 226 times more American cities than German cities (e.g., for $N = 1{,}135{,}000$, there are 5,000 German city names and 1,130,000 American city names in the sample). Thus, with $N = 1$, the model would have recognition knowledge of only one city. With $N = 10$, the model would sample 10 city names with replacement and therefore could have knowledge of anywhere from 1 (if 10 instances of one city were sampled) to 10 (if 10 different cities were sampled) cities. The probability that a particular city is sampled is determined by the distribution of newspaper citations.

So that we could capture the all-or-none nature of the recognition heuristic, the model did not distinguish between cities that were sampled once compared with more than once. Thus, if the sample included 100 instances of Berlin and 1 instance of Essen, both were modeled as being recognized. It is important to note that this process ignores the metric properties of the sampling distribution. As with Goldstein and Gigerenzer's (2002) implementation, we allowed the recognition heuristic to give way to TTB when both cities were recognized. The knowledge validity (denoted as B in Figure 3) was varied across two levels (B = 0.6 and B = 0.85). B corresponds to the probability of choosing the correct alternative given that the model recognizes both alternatives. Note that the probability of recognizing both cities increases as sample size increases. Thus, the recognition heuristic would become less applicable as experience within the environment increases.

The choice rule for deciding which of two cities is most populous was as follows: (a) If no cities are recognized, then choose randomly; (b) if one city is recognized, then choose the city that is recognized; and (c) if both cities are recognized, then implement TTB and choose whichever city the cue points to.

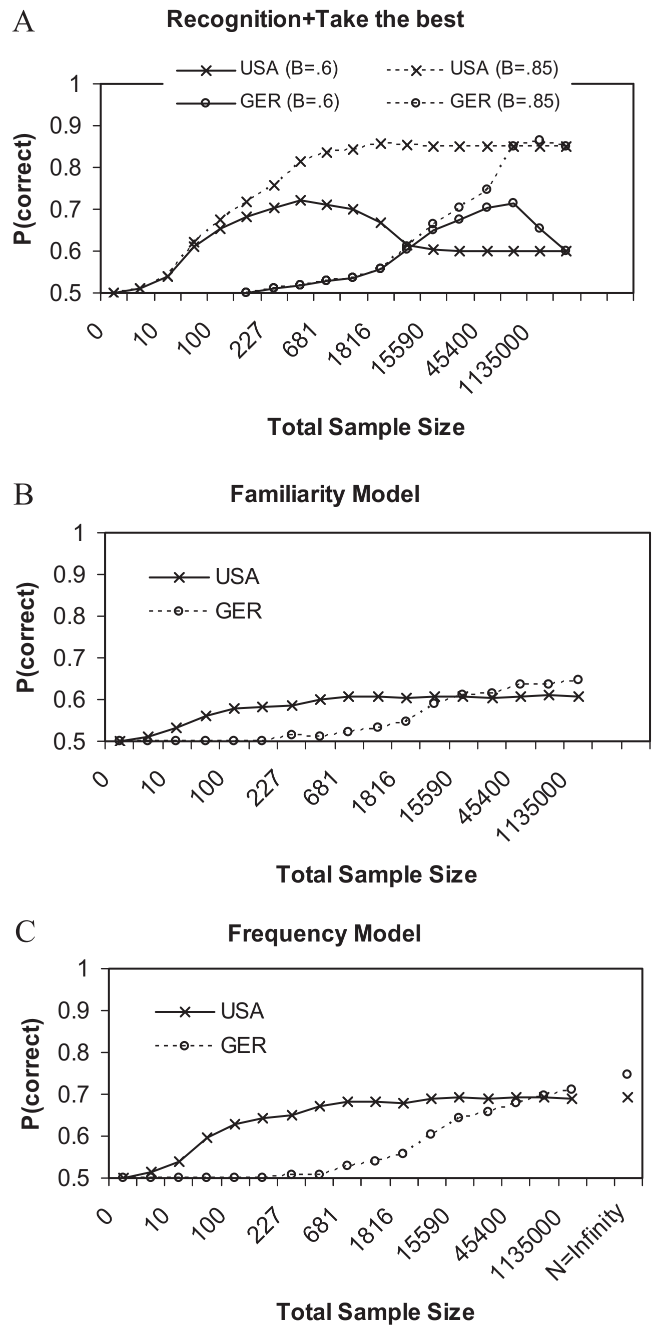

Figure 3. Predicted proportion correct when the reference class is defined on the basis of sample size (83 largest cities) for the three models: (a) the recognition + Take The Best model, (b) the familiarity model, and (c) the frequency model. Dashed lines in Panel A correspond to simulation results when knowledge validity B = .85; solid lines correspond to simulations when the knowledge validity B = .60. USA = United States of America (Americans); GER = Germany (Germans); P = proportion.

Our implementation of the familiarity model involved the application of the Minerva 2 memory model (Hintzman, 1988). We adopted the same sampling assumptions as were adopted for the recognition heuristic. However, rather than representing sampled items as a binary variable, the model encoded as many traces of the city as were in the sample (each trace was represented as a 15 element vector). Thus, if 100 instances of Berlin and 1 instance of Essen were included in the sample, we assumed that 100 traces of Berlin and only 1 trace of Essen were stored in memory. Encoding was varied across four levels of L (L = .25, .50, .75, and 1.0), where L corresponds to the proportion of features in the environmental event that are retained in the corresponding memory trace. As all four values of L produced the same ordinal pattern, we report only the simulation results for $L = 1.0$ (the only difference is how quickly the familiarity model reaches asymptotic performance). At test, the model compared the echo intensity (as determined by Minerva 2's global matching process; see Hintzman, 1988) of all pairs of cities. The rule used for deciding which of two cities was larger was as follows: Choose the city with the largest echo intensity. If two cities had the same echo intensities, then the model chose at random between the two cities (ties occur relatively infrequently because echo intensity is a continuous variable).[8]

Finally, we also implemented a frequency model that embodied all the properties of the familiarity model but operated on the raw frequencies. The choice rule in this case was to choose the alternative with the higher frequency in memory. As an aside, this model most closely resembles an automatic frequency-counter process, where the frequencies were learned through ecological sampling.[9] All other details of the simulations were identical to those for the recognition heuristic. In fact, the models were implemented in tandem so that the computations involved for the familiarity and frequency models were based on exactly the same sample of traces as the recognition heuristic for each simulated participant. For all three models, we ran 50 simulated participants for each sample size, with each simulated participant making $K(K - 1)/2$ pairwise comparisons, where K is the number of cities in the reference class.

Three points are important at this juncture. First, only the familiarity and frequency models preserve the metric properties of the ecology (i.e., the sample probability distribution approaches the population probability distribution as N approaches infinity). The only sensitivity to the ecology afforded by the recognition heuristic comes about through the sampling process.[10] However, it is important to note that the recognition heuristic's sensitivity to the ecology disappears with increases in sample size: As N increases, the probability that all cities are included in the sample approaches 1.0. Second, the familiarity model is not limited to cases in which only one city is recognized: It can operate even when both cities are recognized by exploiting the differences in echo intensity. Third, the familiarity model does not require explicit knowledge of cues or cue validities and, in fact, can operate even in the absence of such knowledge by exploiting the relative frequencies inherent in the mediator variable.

How do the models compare in the city-size task? Which models predict that Americans will be better at estimating German than American city size? Does that empirical finding require that we accept the suggestion that too much knowledge is bad?

Figure 3 plots the simulation results. Several things are noteworthy. First, the curves representing the German and American city ecologies diverge considerably: All three models predict that American students are better at estimating American than German relative city size when relying on small to moderate samples. Second, all three models show a crossover as sample size increases such that at some point Americans are predicted to be more accurate at estimating relative sizes of German than of American cities. Both the familiarity and frequency models show a single crossover toward the right-hand side of the graph. These are asymptotic predictions and invariant to increases in sample size. In fact, the asymptotic predictions for the frequency model actually correspond to the ecological choice probabilities: The expected percentage correct if using the newspaper frequencies themselves (without a sampling process) would be 74.6% for German cities and 69.2% for the American cities.

The plots for the recognition heuristic model are a bit more complicated. It is true that the recognition heuristic predicts more accurate relative judgments of German than of American cities given moderate sample sizes, which is consistent with Gigerenzer and Goldstein's (1996) less-is-more effect. However, there are two caveats to consider. First, this difference goes away when sample size is large and when the knowledge validity is relatively high (e.g., the accuracy for the American cities at B = .85 is almost always greater than the accuracy for the German cities at B = .85). Moreover, if we allow B to vary (e.g., to model the reasonable assumption that people with knowledge of more cities also have knowledge of cues with higher validity), then the model predicts the American students (as modeled with B = .85) will always be more accurate than German students (as modeled with B = .60). This is a particularly important caveat because it indicates that the recognition heuristic + TTB model can account for both a monotonic increase or a nonmonotonic (inverted U-shaped) pattern of results within each ecology. Second, the recognition + TTB model is a two-process model, compared with the one-process familiarity model. Thus, the familiarity model can account for the finding that American students are more accurate at predicting German city size than American city size, can do so within the context of a single-process model (as opposed to two), and does not require one to accept the counterintuitive assumption that too much knowledge can lead to poor judgment. Ironically, the familiarity and frequency models suggest that the so-called less-is-more effect coincides with two more-is-more curves with different slopes and asymptotes: Increased experience leads to increased accuracy for both American and German city-size predictions, with predictions of German city size having a higher asymptote because of the ecological correlations.

---

[8] Details of the Minerva 2 model are available in Hintzman's (1988) article. The codes used for our simulations are available at http://www.bsos.umd.edu/psyc/dougherty/ or by contacting either Rick Thomas or Michael R. Dougherty.

[9] Note that this model most closely resembles the assumption that events are automatically encoded in memory—an assumption explicitly made by Gigerenzer et al. (1991). We include the frequency model here because it serves as an ecological benchmark against which to compare the familiarity and recognition + TTB models.

[10] Note that Goldstein and Gigerenzer's (2002) implementation of the recognition heuristic did not include a sampling process. Their implementation was based on recognition performance of a sample of participants and therefore captures the structure of the ecology (as represented by the citation frequencies) only to the extent that recognition follows the probability distribution of the citation frequencies.

At a more general level, the simulations presented in Figure 3 demonstrate that the recognition + TTB model is relatively flexible, in that it can predict both a monotonic increase in judgment accuracy with sample size as well as a nonmonotonic relationship between judgment and sample size. In contrast, both the familiarity and frequency models necessarily predict monotonic increases with sample size, with asymptotic performance determined by the ecological correlation (these are given as the points to the far right, with N = infinity, in the bottom panel of Figure 3). It is important to note that regardless of whether one endorses the Minerva 2 framework, there exist alternatives to the recognition + TTB model that (a) can account for the main empirical findings comparing American's predictions of German and American city sizes and (b) do not require the added assumption that too much knowledge can lead to poor inference.[11]

Summary of Criticisms

We argued that an event-counter mechanism is needed for the TTB algorithm if one assumes that cue validity is the basis of cue generation (principally because the validities need to be prestored). We argued further that such an assumption (as well as others) is untenable and psychologically implausible. Specifically, data on frequency encoding does not support the automatic event-counter mechanism. Rather, the available data suggest that a retrospective implementation of frequency estimation would make PMM more psychologically plausible. However, we argued that an event-counter mechanism is needed if one assumes that cue validity is the basis of cue generation (principally because the validities need to be prestored).

Next, we argued that the definition of cue validity is fundamentally flawed. It requires that one record the absence of information (or nonoccurrences) in the environment using an automatic and unintentional encoding process. This assumption is logically unjustifiable, especially in light of the considerable data indicating that it is psychologically untenable.

Third, we argued that memory retrieval variables are likely to supersede cue validity as a mechanism for cue generation. The assumption that cues are generated according to their cue validity requires that people compute cue validity on missing information, that the validities are precomputed and stored in a hierarchy, and that the basis of generation is cue validity. In the absence of having a prestored hierarchy, validity would have to be computed online, which would entail an initial retrieval process for generating the cues before validity could be computed.

Finally, we argued that the implementation of the recognition principle in Goldstein and Gigerenzer (2002) is flawed. Moreover, we presented two alternative accounts of the empirical finding that Americans are less accurate at judging American cities than at judging German cities, neither of which requires one to accept that more knowledge leads to less accurate inference.

Conclusions

Two messages can be gleaned from our review. On the one hand, our analysis suggests a note of caution for consumers of fast and frugal heuristics, a note echoed in several recent articles addressing the psychological plausibility of the recognition and TTB (Bergert & Nosofsky, 2007; Chater, Oaksford, & Nakisa, 2003; Juslin & Persson, 2002; Lee & Cummings, 2004; B. R. Newell, 2005) heuristics: Although heuristic mechanisms like TTB and TTL have intuitive appeal, application of these mechanisms as models of human judgment can be misleading if we do not simultaneously scrutinize whether the assumptions underlying the models are met. Our analysis suggests that several of the primary assumptions of PMM are flawed, leading us to question the psychological plausibility of PMM and the heuristics proposed by Gigerenzer and Goldstein (1996). In addition, our analysis suggests that empirical tests on the cue selection process need to consider how the cue search processes might be constrained (or even dictated) by the underlying memory processes. It is our intuition that corroborating evidence for TTB will be hard to come by once experiments are designed that explicitly control for memory retrieval variables. Indeed, initial investigations of TTB in well-designed experiments have shown that it is far from a universal heuristic (Bröder, 2000; B. R. Newell et al., 2004; B. R. Newell, Weston, & Shanks, 2003). Whether there will be clear-cut evidence in favor of TTB once memory retrieval variables are controlled is an empirical question that needs to be addressed.

On the other hand, we are quick to recognize that any approach worth criticizing is a worthwhile approach. There is much to applaud about the fast and frugal program of research. In many ways, it has moved us closer to A. Newell's (1973) call for complete (and integrative) theories of cognition. In Newell's view, a complete model of cognition requires that we model the psychological control processes as well as the memories and primitives on which the control processes operate. To this we add that one also needs a model of the ecology. Work on the fast and frugal heuristics has called attention to two of these three: the control processes involved with cue search and the importance of the ecology in shaping the underlying memory representation. Although there have been attempts to integrate ecological models with memory models (see Dougherty, 2001; Juslin & Persson, 2002; Thomas, Dougherty, Sprenger, & Harbison, 2008), little work has examined the control processes that guide how people generate information or cues from memory and how they decide to terminate memory search (Dougherty & Harbison, in press). Although we view TTB, TTL, and minimalist as psychologically implausible, we think research on control processes is overdue.

Of interest, most of our criticisms stem from the link between the ecology and the control processes, a linkage that, in our view, depends on the memory system and its primitives. Indeed, it is this link that is not well instantiated within the fast and frugal framework. In the absence of specifying the memory system and prim-

[11] The above simulations used sample size as the basis of defining the reference class: the 83 largest German and American cities. An equally appropriate way to define the reference class is on the basis of the criterion variable: all cities with populations above 100,000. There are 83 German and 195 American cities over 100,000. Not only do simulations based on this definition of the reference class yield qualitatively different predictions, but also the ecological correlation between city size and citation rates changes: $\tau = .49$ (Pearson $r = .39$) when computed over the 195 American largest cities, but $\tau = .39$ (Pearson $r = .38$) when computed over the 83 German largest cities. Given this result, it is clear that how one defines the reference class affects the predictions of the model. In some ways, the definition of the reference class is a free parameter. The results of these simulations are available at http://www.bsos.umd.edu/psyc/dougherty/

itives, all-or-none recognition seems like a trivial simplifying assumption, the definition of cue validity appears justifiable, and the generation of cues according to validity makes sense. Our criticisms regarding the definition of cue validity, the use of TTB as a search rule, and the assumption of all-or-none recognition become obvious only when one attempts to instantiate PMM and TTB in the context of a memory model capable of unsupervised learning. Thus, in many ways, our analysis should be taken as a paradigm example of A. Newell's (1973) argument that tight theorizing requires that models specify the underlying memory representation and primitives as well as the control processes. In the absence of specifying these underlying processes, there are few constraints on the number of potential fast and frugal heuristics that can be developed and few constraints on their specification. As a result, one risks a proliferation of heuristic mechanisms that are either implausible or incompatible with the functioning of the underlying processes on which they are assumed to operate.

As a final note, we recognize that work on the fast and frugal heuristics has gone well beyond the original heuristics proposed by Gigerenzer and Goldstein (1996; see Gigerenzer, Todd, & The ABC Research Group, 1999). Although the bulk of our specific comments do not apply directly to the vast majority of these new heuristics (e.g., the priority heuristic, QuickEst, and the fluency heuristic, among others), the general criticism that tight theorizing requires specification of control processes, as well as the memory system and primitives, does apply.[12] As the number of heuristics in the fast and frugal tool kit grows, it becomes increasingly necessary to integrate them within a common process model to ensure that they are compatible with the underlying cognitive processes, the ecological constraints, and one another. Such an approach is likely to have an enormous payoff in terms of both enabling predictions of when a particular heuristic will be used and of moving the approach toward an integrative model of judgment and decision making.

[12] To our knowledge, the only model within the fast and frugal tool kit to specify the underlying memory processes and primitives is the fluency heuristic, which is specified within the ACT-R framework (Schooler & Hertwig, 2005).

References

Adelman, L., Bresnick, T. A., Christian, M., Gualtieri, J., & Minionis, D. (1997). Demonstrating the effect of context on order effects for an Army air defense task using the Patriot simulator. *Journal of Behavioral Decision Making, 10,* 327–342.

Anderson, N. H., & Norman, A. (1964). Order effects in impression formation in four classes of stimuli. *Journal of Abnormal and Social Psychology, 69,* 467–471.

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take the best models of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 107–129.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13,* 207–230.

Bröder, A. (2000). Assessing the empirical validity of the "take the best" heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1332–1346.

Bröder, A. (2002). Take the best, Dawes' rule, and compensatory decision strategies: A regression based classification method. *Quality and Quantity, 36,* 219–238.

Bröder, 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.

Brown, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1539–1553.

Brown, N. R. (1997). Context memory and the selection of frequency estimation strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 898–914.

Chapman, G. B., Bergus, G. R., & Elstein, A. S. (1996). Order of information affects clinical judgment. *Journal of Behavioral Decision Making, 9,* 201–211.

Chater, N., Oaksford, M., & Nakisa, R. (2003). Fast, frugal, and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes, 90,* 63–86.

Dougherty, M. R. P. (2001). Integration of the ecological and error models of overconfidence using a multiple-trace memory model. *Journal of Experimental Psychology: General, 130,* 579–599.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. (1999). Minerva-DM: A memory process model for judgments of likelihood. *Psychological Review, 106,* 180–209.

Dougherty, M. R. P., & Harbison, J. I. (in press). Motivated to retrieve: How often are you willing to go back to the well when the well is dry? *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review, 103,* 592–596.

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

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528.

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

Goldstein, D. G., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 37–58). New York: Oxford University Press.

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

Greene, R. L. (1986). Effects of intentionality and strategy on memory for frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 489–495.

Greene, R. L. (1988). Generation effects in frequency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 298–304.

Gregg, V. H. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183–216). Chichester, England: Wiley.

Hanson, C., & Hirst, W. (1988). Frequency encoding of token and type information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 289–297.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–388.

Hearst, E. (1991). Psychology and nothing. *American Scientist, 79,* 432–443.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple trace model. *Psychological Review, 95,* 528–551.

Hoch, S. J. (1984). Availability and interference in predictive judgment.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 649–662.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24,* 1–55.

Juslin, P., Olsson, H., & Bjorkman, M. (1997). Brunswikian- and Thurstonian-origins of bias in probability assessment: On the interpretation of stochastic components of judgment. *Journal of Behavioral Decision Making, 10,* 189–209.

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

Koriat, A. (1993). How do we know that we know? The accessibility model of feeling of knowing. *Psychological Review, 100,* 609–639.

Lea, S. E. (1974). The non-occurrence of a stimulus as a signal. *Quarterly Journal of Experimental Psychology, 26,* 616–621.

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

Naveh-Benjamin, M., & Jonides, J. (1986). On the automaticity of frequency coding: Effects of competing task load, encoding strategy, and intention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 378–386.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). New York: Academic Press.

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

Newell, B. R., & Fernandez, D. (2006). On the binary quality of recognition and the inconsequentiality of further knowledge: Two critical tests of the recognition heuristic. *Journal of Behavioral Decision Making, 19,* 333–346.

Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies in decision making: The success of "success." *Journal of Behavioral Decision Making, 17,* 117–137.

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

Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone "takes-the-best." *Organizational Behavior and Human Decision Processes, 91,* 82–96.

Newman, J. P., Wolff, W. T., & Hearst, E. (1980). The feature-positive effect in adult human subjects. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 630–650.

Nosofsky, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363–393). Hillsdale, NJ: Erlbaum.

Pleskac, T. (2007). A signal detection analysis of the recognition heuristic. *Psychonomic Bulletin & Review, 14,* 379–391.

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. 141–168). 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.

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review, 112,* 610–628.

Thomas, R., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review, 115,* 155–185.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probably. *Cognitive Psychology, 5,* 207–232.

Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1130.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgments under certainty: Heuristics and biases* (pp. 84–98). New York: Cambridge University Press.

Received November 8, 2004
Revision received March 23, 2007
Accepted March 26, 2007 ■

Postscript: Vague Heuristics Revisited

Michael R. Dougherty
University of Maryland, College Park

Rick Thomas
University of Oklahoma

Ana M. Franco-Watkins
Auburn University

Although we continue to disagree with Gigerenzer, Hoffrage, and Goldstein (2008) on a number of points, including the interpretation of data said to be consistent with the use of Take The Best (TTB), their reply has led to considerable clarity regarding the underlying assumptions of probabilistic mental models (PMM), TTB, and the recognition heuristic. Ironically, this new found clarity has led to more questions about the functioning of their heuristics rather than fewer and has exposed PMM, TTB, and the recognition heuristic to the much more damning criticism that they are nothing more than vaguely specified process models or heuristics. Moreover, Gigerenzer et al.'s (2008) reply expounds some of the problems with the fast and frugal approach, namely that there are no constraints on how large the fast and frugal toolbox can grow. In the subsequent sections, we detail the components of PMM, TTB, and the recognition heuristic that require specification and how this lack of specification leads us to conclude that they reduce to vague heuristics of the sort Gigerenzer (1996) argued against in his critique of Kahneman and Tversky's (1996) work.

TTB Does Not Assume an Ordering Based on Ecological Validity

Gigerenzer et al. (2008) portray our belief that cue validity is the same as ecological validity as a misconception (see their Table 1). To be sure, our apparent misconception is not without basis, as there seems to be a general misuse of the term TTB in the literature. For example, Dieckmann and Todd (2004, p. 310) explicitly stated that TTB requires knowledge of ecological cue validities:

> Although TTB is a very simple heuristic to apply, the set-up of its search rule requires knowledge of the ecological validities of cues. This knowledge is probably not usually available in an explicit precomputed form in the environment, and so must be computed from stored or ongoing experience.