Reasoning the Fast and Frugal Way: Models of Bounded Rationality

Gerd Gigerenzer and Daniel G. Goldstein
Max Planck Institute for Psychological Research and University of Chicago

Humans and animals make inferences about the world under limited time and knowledge. In contrast, many models of rational inference treat the mind as a Laplacean Demon, equipped with unlimited time, knowledge, and computational might. Following H. Simon's notion of satisficing, the authors have proposed a family of algorithms based on a simple psychological mechanism: one-reason decision making. These fast and frugal algorithms violate fundamental tenets of classical rationality: They neither look up nor integrate all information. By computer simulation, the authors held a competition between the satisficing "Take The Best" algorithm and various "rational" inference procedures (e.g., multiple regression). The Take The Best algorithm matched or outperformed all competitors in inferential speed and accuracy. This result is an existence proof that cognitive mechanisms capable of successful performance in the real world do not need to satisfy the classical norms of rational inference.

Organisms make inductive inferences. Darwin (1872/1965) observed that people use facial cues, such as eyes that waver and lids that hang low, to infer a person's guilt. Male toads, roaming through swamps at night, use the pitch of a rival's croak to infer its size when deciding whether to fight (Krebs & Davies, 1987). Stock brokers must make fast decisions about which of several stocks to trade or invest when only limited information is available. The list goes on. Inductive inferences are typically based on uncertain cues: The eyes can deceive, and so can a tiny toad with a deep croak in the darkness.

How does an organism make inferences about unknown aspects of the environment? There are three directions in which to look for an answer. From Pierre Laplace to George Boole to Jean Piaget, many scholars have defended the now classical view that the laws of human inference are the laws of probability and statistics (and to a lesser degree logic, which does not deal as easily with uncertainty). Indeed, the Enlightenment probabilists derived the laws of probability from what they believed to be the laws of human reasoning (Daston, 1988). Following this time-honored tradition, much contemporary research in psychology, behavioral ecology, and economics assumes standard

statistical tools to be the normative and descriptive models of inference and decision making. Multiple regression, for instance, is both the economist's universal tool (McCloskey, 1985) and a model of inductive inference in multiple-cue learning (Hammond, 1990) and clinical judgment (B. Brehmer, 1994); Bayes's theorem is a model of how animals infer the presence of predators or prey (Stephens & Krebs, 1986) as well as of human reasoning and memory (Anderson, 1990). This Enlightenment view that probability theory and human reasoning are two sides of the same coin crumbled in the early nineteenth century but has remained strong in psychology and economics.

In the past 25 years, this stronghold came under attack by proponents of the heuristics and biases program, who concluded that human inference is systematically biased and error prone, suggesting that the laws of inference are quick-and-dirty heuristics and not the laws of probability (Kahneman, Slovic, & Tversky, 1982). This second perspective appears diametrically opposed to the classical rationality of the Enlightenment, but this appearance is misleading. It has retained the normative kernel of the classical view. For example, a discrepancy between the dictates of classical rationality and actual reasoning is what defines a *reasoning error* in this program. Both views accept the laws of probability and statistics as normative, but they disagree about whether humans can stand up to these norms.

Many experiments have been conducted to test the validity of these two views, identifying a host of conditions under which the human mind appears more rational or irrational. But most of this work has dealt with simple situations, such as Bayesian inference with binary hypotheses, one single piece of binary data, and all the necessary information conveniently laid out for the participant (Gigerenzer & Hoffrage, 1995). In many real-world situations, however, there are multiple pieces of information, which are not independent, but redundant. Here, Bayes's theorem and other "rational" algorithms quickly become mathematically complex and computationally intractable, at least for ordinary human minds. These situations make neither of the two views look promising. If one would apply the classical view to such complex real-world environments, this

would suggest that the mind is a supercalculator like a Laplacean Demon (Wimsatt, 1976)—carrying around the collected works of Kolmogoroff, Fisher, or Neyman—and simply needs a memory jog, like the slave in Plato's *Meno*. On the other hand, the heuristics-and-biases view of human irrationality would lead us to believe that humans are hopelessly lost in the face of real-world complexity, given their supposed inability to reason according to the canon of classical rationality, even in simple laboratory experiments.

There is a third way to look at inference, focusing on the psychological and ecological rather than on logic and probability theory. This view questions classical rationality as a universal norm and thereby questions the very definition of "good" reasoning on which both the Enlightenment and the heuristics-and-biases views were built. Herbert Simon, possibly the best-known proponent of this third view, proposed looking for models of *bounded rationality* instead of classical rationality. Simon (1956, 1982) argued that information-processing systems typically need to *satisfice* rather than optimize. *Satisficing*, a blend of *sufficing* and *satisfying*, is a word of Scottish origin, which Simon uses to characterize algorithms that successfully deal with conditions of limited time, knowledge, or computational capacities. His concept of satisficing postulates, for instance, that an organism would choose the first object (a mate, perhaps) that satisfies its aspiration level—instead of the intractable sequence of taking the time to survey all possible alternatives, estimating probabilities and utilities for the possible outcomes associated with each alternative, calculating expected utilities, and choosing the alternative that scores highest.

Let us stress that Simon's notion of bounded rationality has two sides, one cognitive and one ecological. As early as in *Administrative Behavior* (1945), he emphasized the cognitive limitations of real minds as opposed to the omniscient Laplacean Demons of classical rationality. As early as in his *Psychological Review* article titled "Rational Choice and the Structure of the Environment" (1956), Simon emphasized that minds are adapted to real-world environments. The two go in tandem: "Human rational behavior is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor" (Simon, 1990, p. 7). For the most part, however, theories of human inference have focused exclusively on the cognitive side, equating the notion of bounded rationality with the statement that humans are limited information processors, period. In a Procrustean-bed fashion, *bounded rationality* became almost synonymous with *heuristics and biases*, thus paradoxically reassuring classical rationality as the normative standard for both biases and bounded rationality (for a discussion of this confusion see Lopes, 1992). Simon's insight that the minds of living systems should be understood relative to the environment in which they evolved, rather than to the tenets of classical rationality, has had little impact so far in research on human inference. Simple psychological algorithms that were observed in human inference, reasoning, or decision making were often discredited without a fair trial, because they looked so stupid by the norms of classical rationality. For instance, when Keeney and Raiffa (1993) discussed the lexicographic ordering procedure they had observed in practice—a procedure related to the class of satisficing algorithms we propose in this article—they concluded that this procedure "is naively simple" and "will rarely pass a test of

'reasonableness'" (p. 78). They did not report such a test. We shall.

Initially, the concept of bounded rationality was only vaguely defined, often as that which is not classical economics, and one could "fit a lot of things into it by foresight and hindsight," as Simon (1992, p. 18) himself put it. We wish to do more than oppose the Laplacean Demon view. We strive to come up with something positive that could replace this unrealistic view of mind. What are these simple, intelligent algorithms capable of making near-optimal inferences? How fast and how accurate are they? In this article, we propose a class of models that exhibit bounded rationality in both of Simon's senses. These satisficing algorithms operate with simple psychological principles that satisfy the constraints of limited time, knowledge, and computational might, rather than those of classical rationality. At the same time, they are designed to be fast and frugal without a significant loss of inferential accuracy, because the algorithms can exploit the structure of environments.

The article is organized as follows. We begin by describing the task the cognitive algorithms are designed to address, the basic algorithm itself, and the real-world environment on which the performance of the algorithm will be tested. Next, we report on a competition in which a satisficing algorithm competes with "rational" algorithms in making inferences about a real-world environment. The "rational" algorithms start with an advantage: They use more time, information, and computational might to make inferences. Finally, we study variants of the satisficing algorithm that make faster inferences and get by with even less knowledge.

The Task

We deal with inferential tasks in which a choice must be made between two alternatives on a quantitative dimension. Consider the following example:

Which city has a larger population? (a) Hamburg (b) Cologne.

Two-alternative-choice tasks occur in various contexts in which inferences need to be made with limited time and knowledge, such as in decision making and risk assessment during driving (e.g., exit the highway now or stay on); treatment-allocation decisions (e.g., who to treat first in the emergency room: the 80-year-old heart attack victim or the 16-year-old car accident victim); and financial decisions (e.g., whether to buy or sell in the trading pit). Inference concerning population demographics, such as city populations of the past, present, and future (e.g., Brown & Siegler, 1993), is of importance to people working in urban planning, industrial development, and marketing. Population demographics, which is better understood than, say, the stock market, will serve us later as a "drosophila" environment that allows us to analyze the behavior of satisficing algorithms.

We study two-alternative-choice tasks in situations where a person has to make an inference based solely on knowledge retrieved from memory. We refer to this as *inference from memory*, as opposed to *inference from givens*. Inference from memory involves search in declarative knowledge and has been investigated in studies of, inter alia, confidence in general knowledge (e.g., Juslin, 1994; Sniezek & Buckley, 1993); the

effect of repetition on belief (e.g., Hertwig, Gigerenzer, & Hoffrage, in press); hindsight bias (e.g., Fischhoff, 1977); quantitative estimates of area and population of nations (Brown & Siegler, 1993); and autobiographic memory of time (Huttenlocher, Hedges, & Prohaska, 1988). Studies of inference from givens, on the other hand, involve making inferences from information presented by an experimenter (e.g., Hammond, Hursch, & Todd, 1964). In the tradition of Ebbinghaus's nonsense syllables, attempts are often made here to prevent individual knowledge from impacting on the results by using problems about hypothetical referents instead of actual ones. For instance, in celebrated judgment and decision-making tasks, such as the "cab" problem and the "Linda" problem, all the relevant information is provided by the experimenter, and individual knowledge about cabs and hit-and-run accidents, or feminist bank tellers, is considered of no relevance (Gigerenzer & Murray, 1987). As a consequence, limited knowledge or individual differences in knowledge play a small role in inference from givens. In contrast, the satisficing algorithms proposed in this article perform inference from memory, they use limited knowledge as input, and as we will show, they can actually profit from a lack of knowledge.

Assume that a person does not know or cannot deduce the answer to the Hamburg–Cologne question but needs to make an inductive inference from related real-world knowledge. How is this inference derived? How can we predict choice (Hamburg or Cologne) from a person's state of knowledge?

Theory

The cognitive algorithms we propose are realizations of a framework for modeling inferences from memory, the theory of *probabilistic mental models* (PMM theory; see Gigerenzer, 1993; Gigerenzer, Hoffrage, & Kleinbölting, 1991). The theory of probabilistic mental models assumes that inferences about unknown states of the world are based on probability cues (Brunswik, 1955). The theory relates three visions: (a) Inductive inference needs to be studied with respect to natural environments, as emphasized by Brunswik and Simon; (b) inductive inference is carried out by satisficing algorithms, as emphasized by Simon; and (c) inductive inferences are based on frequencies of events in a reference class, as proposed by Reichenbach and other frequentist statisticians. The theory of probabilistic mental models accounts for choice and confidence, but only choice is addressed in this article.

The major thrust of the theory is that it replaces the canon of classical rationality with simple, plausible psychological mechanisms of inference—mechanisms that a mind can actually carry out under limited time and knowledge and that could have possibly arisen through evolution. Most traditional models of inference, from linear multiple regression models to Bayesian models to neural networks, try to find some optimal integration of all information available: Every bit of information is taken into account, weighted, and combined in a computationally expensive way. The family of algorithms in PMM theory does not implement this classical ideal. Search in memory for relevant information is reduced to a minimum, and there is no integration (but rather a substitution) of pieces of information. These satisficing algorithms dispense with the fiction of the omniscient Laplacean Demon, who has all the time and knowledge

Figure 1. Illustration of bounded search through limited knowledge. Objects a, b, and c are recognized; object d is not. Cue values are positive (+) or negative (−); missing knowledge is shown by question marks. Cues are ordered according to their validities. To infer whether $a > b$, the Take The Best algorithm looks up only the cue values in the shaded space; to infer whether $b > c$, search is bounded to the dotted space. The other cue values are not looked up.

to search for all relevant information, to compute the weights and covariances, and then to integrate all this information into an inference.

Limited Knowledge

A PMM is an inductive device that uses limited knowledge to make fast inferences. Different from mental models of syllogisms and deductive inference (Johnson-Laird, 1983), which focus on the logical task of truth preservation and where knowledge is irrelevant (except for the meaning of connectives and other logical terms), PMMs perform intelligent guesses about unknown features of the world, based on uncertain indicators. To make an inference about which of two objects, a or b, has a higher value, knowledge about a reference class R is searched, with $a, b \in R$. In our example, knowledge about the reference class "cities in Germany" could be searched. The knowledge consists of probability cues C_i ($i = 1, \ldots, n$), and the cue values a_i and b_i of the objects for the ith cue. For instance, when making inferences about populations of German cities, the fact that a city has a professional soccer team in the major league (*Bundesliga*) may come to a person's mind as a potential cue. That is, when considering pairs of German cities, if one city has a soccer team in the major league and the other does not, then the city with the team is likely, but not certain, to have the larger population.

Limited knowledge means that the matrix of objects by cues has missing entries (i.e., objects, cues, or cue values may be unknown). Figure 1 models the limited knowledge of a person. She has heard of three German cities, a, b, and c, but not of d (represented by three positive and one negative recognition values). She knows some facts (cue values) about these cities with respect to five binary cues. For a binary cue, there are two cue values, positive (e.g., the city has a soccer team) or negative (it does not). *Positive* refers to a cue value that signals a higher value on the target variable (e.g., having a soccer team is correlated with high population). Unknown cue values are shown by a question mark. Because she has never heard of d, all cue values for object d are, by definition, unknown.

People rarely know all information on which an inference

could be based, that is, knowledge is limited. We model limited knowledge in two respects: A person can have (a) incomplete knowledge of the objects in the reference class (e.g., she recognizes only some of the cities), (b) limited knowledge of the cue values (facts about cities), or (c) both. For instance, a person who does not know all of the cities with soccer teams may know some cities with positive cue values (e.g., Munich and Hamburg certainly have teams), many with negative cue values (e.g., Heidelberg and Potsdam certainly do not have teams), and several cities for which cue values will not be known.

The Take The Best Algorithm

The first satisficing algorithm presented is called the *Take The Best* algorithm, because its policy is "take the best, ignore the rest." It is the basic algorithm in the PMM framework. Variants that work faster or with less knowledge are described later. We explain the steps of the Take The Best algorithm for binary cues (the algorithm can be easily generalized to many valued cues), using Figure 1 for illustration.

The Take The Best algorithm assumes a subjective rank order of cues according to their validities (as in Figure 1). We call the highest ranking cue (that discriminates between the two alternatives) the best cue. The algorithm is shown in the form of a flow diagram in Figure 2.

Step 1: Recognition Principle

The recognition principle is invoked when the mere recognition of an object is a predictor of the target variable (e.g., population). The recognition principle states the following: If only one of the two objects is recognized, then choose the recognized object. If neither of the two objects is recognized, then choose randomly between them. If both of the objects are recognized, then proceed to Step 2.

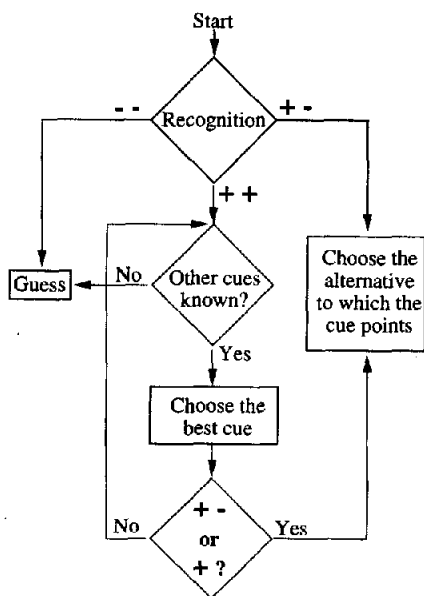*Example:* If a person in the knowledge state shown in Figure

Figure 2. Flow diagram of the Take The Best algorithm.

Object a

Figure 3. Discrimination rule. A cue discriminates between two alternatives if one has a positive cue value and the other does not. The four discriminating cases are shaded.

1 is asked to infer which of city *a* and city *d* has more inhabitants, the inference will be city *a*, because the person has never heard of city *d* before.

Step 2: Search for Cue Values

For the two objects, retrieve the cue values of the highest ranking cue from memory.

Step 3: Discrimination Rule

Decide whether the cue discriminates. The cue is said to discriminate between two objects if one has a positive cue value and the other does not. The four shaded knowledge states in Figure 3 are those in which a cue discriminates.

Step 4: Cue-Substitution Principle

If the cue discriminates, then stop searching for cue values. If the cue does not discriminate, go back to Step 2 and continue with the next cue until a cue that discriminates is found.

Step 5: Maximizing Rule for Choice

Choose the object with the positive cue value. If no cue discriminates, then choose randomly.

Examples: Suppose the task is judging which of city *a* or *b* is larger (Figure 1). Both cities are recognized (Step 1), and search for the best cue results with a positive and a negative cue value for Cue 1 (Step 2). The cue discriminates (Step 3), and search is terminated (Step 4). The person makes the inference that city *a* is larger (Step 5).

Suppose now the task is judging which of city *b* or *c* is larger. Both cities are recognized (Step 1), and search for the cue values cue results in negative cue value on object *b* for Cue 1, but the corresponding cue value for object *c* is unknown (Step 2). The cue does not discriminate (Step 3), so search is continued (Step 4). Search for the next cue results with positive and a negative cue values for Cue 2 (Step 2). This cue discriminates (Step 3), and search is terminated (Step 4). The person makes the inference that city *b* is larger (Step 5).

The features of this algorithm are (a) search extends through only a portion of the total knowledge in memory (as shown by the shaded and dotted parts of Figure 1) and is stopped imme-

diately when the first discriminating cue is found, (b) the algorithm does not attempt to integrate information but uses cue substitution instead, and (c) the total amount of information processed is contingent on each task (pair of objects) and varies in a predictable way among individuals with different knowledge. This fast and computationally simple algorithm is a model of bounded rationality rather than of classical rationality. There is a close parallel with Simon's concept of "satisficing": The Take The Best algorithm stops search after the first discriminating cue is found, just as Simon's satisficing algorithm stops search after the first option that meets an aspiration level.

The algorithm is hardly a standard statistical tool for inductive inference: It does not use all available information, it is noncompensatory and nonlinear, and variants of it can violate transitivity. Thus, it differs from standard linear tools for inference such as multiple regression, as well as from nonlinear neural networks that are compensatory in nature. The Take The Best algorithm is noncompensatory because only the best discriminating cue determines the inference or decision; no combination of other cue values can override this decision. In this way, the algorithm does not conform to the classical economic view of human behavior (e.g., Becker, 1976), where, under the assumption that all aspects can be reduced to one dimension (e.g., money), there exists always a trade-off between commodities or pieces of information. That is, the algorithm violates the Archimedian axiom, which implies that for any multidimensional object $a(a_1, a_2, \ldots, a_n)$ preferred to $b(b_1, b_2, \ldots, b_n)$, where a_1 dominates b_1, this preference can be reversed by taking multiples of any one or a combination of $b_2, b_3, \ldots, b_n$. As we discuss, variants of this algorithm also violate transitivity, one of the cornerstones of classical rationality (McClennen, 1990).

Empirical Evidence

Despite their flagrant violation of the traditional standards of rationality, the Take The Best algorithm and other models from the framework of PMM theory have been successful in integrating various striking phenomena in inference from memory and predicting novel phenomena, such as the confidence-frequency effect (Gigerenzer et al., 1991) and the less-is-more effect (Goldstein, 1994; Goldstein & Gigerenzer, 1996). The theory of probabilistic mental models seems to be the only existing process theory of the overconfidence bias that successfully predicts conditions under which overestimation occurs, disappears, and inverts to underestimation (Gigerenzer, 1993; Gigerenzer et al., 1991; Juslin, 1993, 1994; Juslin, Winman, & Persson, 1995; but see Griffin & Tversky, 1992). Similarly, the theory predicts when the hard–easy effect occurs, disappears, and inverts—predictions that have been experimentally confirmed by Hoffrage (1994) and by Juslin (1993). The Take The Best algorithm explains also why the popular confirmation-bias explanation of the overconfidence bias (Koriat, Lichtenstein, & Fischhoff, 1980) is not supported by experimental data (Gigerenzer et al., 1991, pp. 521–522).

Unlike earlier accounts of these striking phenomena in confidence and choice, the algorithms in the PMM framework allow for predictions of choice based on each individual's knowledge. Goldstein and Gigerenzer (1996) showed that the recognition principle predicted individual participants' choices in about 90% to 100% of all cases, even when participants were taught information that suggested doing otherwise (negative cue values for the recognized objects). Among the evidence for the empirical validity of the Take-The-Best algorithm are the tests of a bold prediction, the less-is-more effect, which postulates conditions under which people with little knowledge make better inferences than those who know more. This surprising prediction has been experimentally confirmed. For instance, U.S. students make slightly more correct inferences about German city populations (about which they know little) than about U.S. cities, and vice versa for German students (Gigerenzer, 1993; Goldstein 1994; Goldstein & Gigerenzer, 1995; Hoffrage, 1994). The theory of probabilistic mental models has been applied to other situations in which inferences have to be made under limited time and knowledge, such as rumor-based stock market trading (DiFonzo, 1994). A general review of the theory and its evidence is presented in McClelland and Bolger (1994).

The reader familiar with the original algorithm presented in Gigerenzer et al. (1991) will have noticed that we simplified the discrimination rule.[1] In the present version, search is already terminated if one object has a positive cue value and the other does not, whereas in the earlier version, search was terminated only when one object had a positive value and the other a negative one (cf. Figure 3 in Gigerenzer et al. with Figure 3 in this article). This change follows empirical evidence that participants tend to use this faster, simpler discrimination rule (Hoffrage, 1994).

This article does not attempt to provide further empirical evidence. For the moment, we assume that the model is descriptively valid and investigate how accurate this satisficing algorithm is in drawing inferences about unknown aspects of a real-world environment. Can an algorithm based on simple psychological principles that violate the norms of classical rationality make a fair number of accurate inferences?

The Environment

We tested the performance of the Take The Best algorithm on how accurately it made inferences about a real-world environment. The environment was the set of all cities in Germany with more than 100,000 inhabitants (83 cities after German reunification), with population as the target variable. The model of the environment consisted of 9 binary ecological cues and the actual 9×83 cue values. The full model of the environment is shown in the Appendix.

Each cue has an associated validity, which is indicative of its predictive power. The *ecological validity* of a cue is the relative frequency with which the cue correctly predicts the target, defined with respect to the reference class (e.g., all German cities with more than 100,000 inhabitants). For instance, if one checks all pairs in which one city has a soccer team but the other city does not, one finds that in 87% of these cases, the city with the team also has the higher population. This value is the ecological validity of the soccer team cue. The validity v_i of the ith cue is

$$v_i = p[t(a) > t(b)|a_i \text{ is positive and } b_i \text{ is negative}],$$

---

[1] Also, we now use the term *discrimination rule* instead of *activation rule*.

Table 1
Cues, Ecological Validities, and Discrimination Rates

Cue	Ecological validity	Discrimination rate
National capital (Is the city the national capital?)	1.00	.02
Exposition site (Was the city once an exposition site?)	.91	.25
Soccer team (Does the city have a team in the major league?)	.87	.30
Intercity train (Is the city on the Intercity line?)	.78	.38
State capital (Is the city a state capital?)	.77	.30
License plate (Is the abbreviation only one letter long?)	.75	.34
University (Is the city home to a university?)	.71	.51
Industrial belt (Is the city in the industrial belt?)	.56	.30
East Germany (Was the city formerly in East Germany?)	.51	.27

where $t(a)$ and $t(b)$ are the values of objects a and b on the target variable t and p is a probability measured as a relative frequency in R.

The ecological validity of the nine cues ranged over the whole spectrum: from .51 (only slightly better than chance) to 1.0 (certainty), as shown in Table 1. A cue with a high ecological validity, however, is not often useful if its discrimination rate is small.

Table 1 shows also the *discrimination rates* for each cue. The discrimination rate of a cue is the relative frequency with which the cue discriminates between any two objects from the reference class. The discrimination rate is a function of the distribution of the cue values and the number N of objects in the reference class. Let the relative frequencies of the positive and negative cue values be x and y, respectively. Then the discrimination rate d_i of the ith cue is

$$d_i = \frac{2x_i y_i}{1 - \frac{1}{N}},$$

as an elementary calculation shows. Thus, if N is very large, the discrimination rate is approximately $2x_i y_i$.[2] The larger the ecological validity of a cue, the better the inference. The larger the discrimination rate, the more often a cue can be used to make an inference. In the present environment, ecological validities and discrimination rates are negatively correlated. The redundancy of cues in the environment, as measured by pairwise correlations between cues, ranges between $-.25$ and $.54$, with an average absolute value of .19.[3]

The Competition

The question of how well a satisficing algorithm performs in a real-world environment has rarely been posed in research on inductive inference. The present simulations seem to be the first to test how well simple satisficing algorithms do compared with standard integration algorithms, which require more knowl-

edge, time, and computational power. This question is important for Simon's postulated link between the cognitive and the ecological: If the simple psychological principles in satisficing algorithms are tuned to ecological structures, these algorithms should not fail outright. We propose a competition between various inferential algorithms. The contest will go to the algorithm that scores the highest proportion of correct inferences in the shortest time.

Simulating Limited Knowledge

We simulated people with varying degrees of knowledge about cities in Germany. Limited knowledge can take two forms. One is limited recognition of objects in the reference class. The other is limited knowledge about the cue values of recognized objects. To model limited recognition knowledge, we simulated people who recognized between 0 and 83 German cities. To model limited knowledge of cue values, we simulated 6 basic classes of people, who knew 0%, 10%, 20%, 50%, 75%, or 100% of the cue values associated with the objects they recognized. Combining the two sources of limited knowledge resulted in 6 × 84 types of people, each having different degrees and kinds of limited knowledge. Within each type of people, we created 500 simulated individuals, who differed randomly from one another in the particular objects and cue values they knew. All objects and cue values known were determined randomly within the appropriate constraints, that is, a certain number of objects known, a certain total percentage of cue values known, and the validity of the recognition principle (as explained in the following paragraph).

The simulation needed to be realistic in the sense that the simulated people could invoke the recognition principle. Therefore, the sets of cities the simulated people knew had to be carefully chosen so that the recognized cities were larger than the unrecognized ones a certain percentage of the time. We performed a survey to get an empirical estimate of the actual co-

---

[2] For instance, if $N = 2$ and one cue value is positive and the other negative ($x_i = y_i = .5$), $d_i = 1.0$. If N increases, with x_i and y_i held constant, then d_i decreases and converges to $2x_i y_i$.

[3] There are various other measures of redundancy besides pairwise correlation. The important point is that whatever measure of redundancy one uses, the resultant value does not have the same meaning for all algorithms. For instance, all that counts for the Take The Best algorithm is what proportion of correct inferences the second cue adds to the first in the cases where the first cue does not discriminate, how much the third cue adds to the first two in the cases where they do not discriminate, and so on. If a cue discriminates, search is terminated, and the degree of redundancy in the cues that were not included in the search is irrelevant. Integration algorithms, in contrast, integrate all information and, thus, always work with the total redundancy in the environment (or knowledge base). For instance, when deciding among objects a, b, c, and d in Figure 1, the cue values of Cues 3, 4, and 5 do not matter from the point of view of the Take The Best algorithm (because search is terminated before reaching Cue 3). However, the values of Cues 3, 4, and 5 affect the redundancy of the ecological system, from the point of view of all integration algorithms. The lesson is that the degree of redundancy in an environment depends on the kind of algorithm that operates on the environment. One needs to be cautious in interpreting measures of redundancy without reference to an algorithm.

variation between recognition of cities and city populations. Let us define the *validity* α of the recognition principle to be the probability, in a reference class, that one object has a greater value on the target variable than another, in the cases where the one object is recognized and the other is not:

$$\alpha = p[t(a) > t(b) | a_r \text{ is positive and } b_r \text{ is negative}],$$

where $t(a)$ and $t(b)$ are the values of objects a and b on the target variable t, a_r and b_r are the recognition values of a and b, and p is a probability measured as a relative frequency in R.

In a pilot study of 26 undergraduates at the University of Chicago, we found that the cities they recognized (within the 83 largest in Germany) were larger than the cities they did not recognize in about 80% of all possible comparisons. We incorporated this value into our simulations by choosing sets of cities (for each knowledge state, i.e., for each number of cities recognized) where the known cities were larger than the unknown cities in about 80% of all cases. Thus, the cities known by the simulated individuals had the same relationship between recognition and population as did those of the human individuals. Let us first look at the performance of the Take The Best algorithm.

Testing the Take The Best Algorithm

We tested how well individuals using the Take The Best algorithm did at answering real-world questions such as, Which city has more inhabitants: (a) Heidelberg or (b) Bonn? Each of the 500 simulated individuals in each of the 6 × 84 types was tested on the exhaustive set of 3,403 city pairs, resulting in a total of 500 × 6 × 84 × 3,403 tests, that is, about 858 million.

The curves in Figure 4 show the average proportion of correct inferences for each proportion of objects and cue values known. The x axis represents the number of cities recognized, and the y axis shows the proportion of correct inferences that the Take The Best algorithm drew. Each of the 6 × 84 points that make up the six curves is an average proportion of correct inferences taken from 500 simulated individuals, who each made 3,403 inferences.

When the proportion of cities recognized was zero, the proportion of correct inferences was at chance level (.5). When up to half of all cities were recognized, performance increased at all levels of knowledge about cue values. The maximum percentage of correct inferences was around 77%. The striking result was that this maximum was not achieved when individuals knew all cue values of all cities, but rather when they knew less. This result shows the ability of the algorithm to exploit limited knowledge, that is, to do best when not everything is known. Thus, the Take The Best algorithm produces the *less-is-more* effect. At any level of limited knowledge of cue values, learning more German cities will eventually cause a decrease in proportion correct. Take, for instance, the curve where 75% of the cue values were known and the point where the simulated participants recognized about 60 German cities. If these individuals learned about the remaining German cities, their proportion correct would decrease. The rationale behind the less-is-more effect is the recognition principle, and it can be understood best from the curve that reflects 0% of total cue values known. Here, all decisions are made on the basis of the recognition principle,
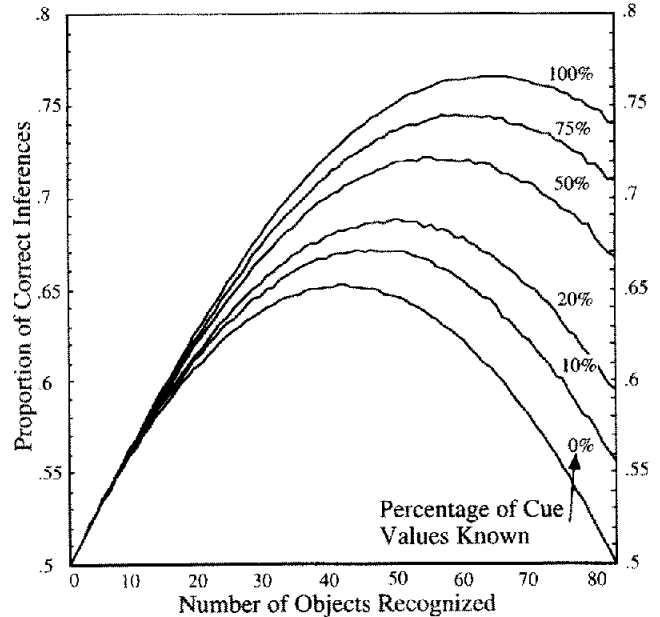

Figure 4. Correct inferences about the population of German cities (two-alternative-choice tasks) by the Take The Best algorithm. Inferences are based on actual information about the 83 largest cities and nine cues for population (see the Appendix). Limited knowledge of the simulated individuals is varied across two dimensions: (a) the number of cities recognized (x axis) and (b) the percentage of cue values known (the six curves).

or by guessing. On this curve, the recognition principle comes into play most when half of the cities are known, so it takes on an inverted-U shape. When half the cities are known, the recognition principle can be activated most often, that is, for roughly 50% of the questions. Because we set the recognition validity in advance, 80% of these inferences will be correct. In the remaining half of the questions, when recognition cannot be used (either both cities are recognized or both cities are unrecognized), then the organism is forced to guess and only 50% of the guesses will be correct. Using the 80% effective recognition validity half of the time and guessing the other half of the time, the organism scores 65% correct, which is the peak of the bottom curve. The mode of this curve moves to the right with increasing knowledge about cue values. Note that even when a person knows everything, all cue values of all cities, there are states of limited knowledge in which the person would make more accurate inferences. We are not going to discuss the conditions of this counterintuitive effect and the supporting experimental evidence here (see Goldstein & Gigerenzer, 1996). Our focus is on how much better integration algorithms can do in making inferences.

Integration Algorithms

We asked several colleagues in the fields of statistics and economics to devise decision algorithms that would do better than the Take The Best algorithm. The five integration algorithms we simulated and pitted against the Take The Best algorithm in a competition were among those suggested by our colleagues.

These competitors include "proper" and "improper" linear models (Dawes, 1979; Lovie & Lovie, 1986). These algorithms, in contrast to the Take The Best algorithm, embody two classical principles of rational inference: (a) complete search—they use all available information (cue values)—and (b) complete integration—they combine all these pieces of information into a single value. In short, we refer in this article to algorithms that satisfy these principles as "rational" (in quotation marks) algorithms.

Contestant 1: Tallying

Let us start with a simple integration algorithm: tallying of positive evidence (Goldstein, 1994). In this algorithm, the number of positive cue values for each object is tallied across all cues ($i = 1, \ldots, n$), and the object with the largest number of positive cue values is chosen. Integration algorithms are not based (at least explicitly) on the recognition principle. For this reason, and to make the integration algorithms as strong as possible, we allow all the integration algorithms to make use of recognition information (the positive and negative recognition values, see Figure 1). Integration algorithms treat recognition as a cue, like the nine ecological cues in Table 1. That is, in the competition, the number of cues (n) is thus equal to 10 (because recognition is included). The decision criterion for tallying is the following:

$$\text{If } \sum_{i=1}^{n} a_i > \sum_{i=1}^{n} b_i, \text{ then choose city } a.$$

$$\text{If } \sum_{i=1}^{n} a_i < \sum_{i=1}^{n} b_i, \text{ then choose city } b.$$

$$\text{If } \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i, \text{ then guess.}$$

The assignments of a_i and b_i are the following:

$$a_i, b_i = \begin{cases} 1 \text{ if the } i\text{th cue value is positive} \\ 0 \text{ if the } i\text{th cue value is negative} \\ 0 \text{ if the } i\text{th cue value is unknown.} \end{cases}$$

Let us compare cities a and b, from Figure 1. By tallying the positive cue values, a would score 2 points and b would score 3. Thus, tallying would choose b to be the larger, in opposition to the Take The Best algorithm, which would infer that a is larger. Variants of tallying, such as the frequency-of-good-features heuristic, have been discussed in the decision literature (Alba & Marmorstein, 1987; Payne, Bettman, & Johnson, 1993).

Contestant 2: Weighted Tallying

Tallying treats all cues alike, independent of cue validity. Weighted tallying of positive evidence is identical with tallying, except that it weights each cue according to its ecological validity, v_i. The ecological validities of the cues appear in Table 1. We set the validity of the recognition cue to .8, which is the empirical average determined by the pilot study. The decision rule is as follows:

$$\text{If } \sum_{i=1}^{n} a_i v_i > \sum_{i=1}^{n} b_i v_i, \text{ then choose city } a.$$

$$\text{If } \sum_{i=1}^{n} a_i v_i < \sum_{i=1}^{n} b_i v_i, \text{ then choose city } b.$$

$$\text{If } \sum_{i=1}^{n} a_i v_i = \sum_{i=1}^{n} b_i v_i, \text{ then guess.}$$

Note that weighted tallying needs more information than either tallying or the Take The Best algorithm, namely, quantitative information about ecological validities. In the simulation, we provided the real ecological validities to give this algorithm a good chance.

Calling again on the comparison of objects a and b from Figure 1, let us assume that the validities would be .8 for recognition and .9, .8, .7, .6, .51 for Cues 1 through 5. Weighted tallying would thus assign 1.7 points to a and 2.3 points to b. Thus, weighted tallying would also choose b to be the larger.

Both tallying algorithms treat negative information and missing information identically. That is, they consider only positive evidence. The following algorithms distinguish between negative and missing information and integrate both positive and negative information.

Contestant 3: Unit-Weight Linear Model

The unit-weight linear model is a special case of the equal-weight linear model (Huber, 1989) and has been advocated as a good approximation of weighted linear models (Dawes, 1979; Einhorn & Hogarth, 1975). The decision criterion for unit-weight integration is the same as for tallying, only the assignment of a_i and b_i differs:

$$a_i, b_i = \begin{cases} 1 \text{ if the } i\text{th cue value is positive} \\ -1 \text{ if the } i\text{th cue value is negative} \\ 0 \text{ if the } i\text{th cue value is unknown.} \end{cases}$$

Comparing objects a and b from Figure 1 would involve assigning 1.0 points to a and 1.0 points to b and, thus, choosing randomly. This simple linear model corresponds to Model 2 in Einhorn and Hogarth (1975, p. 177) with the weight parameter set equal to 1.

Contestant 4: Weighted Linear Model

This model is like the unit-weight linear model except that the values of a_i and b_i are multiplied by their respective ecological validities. The decision criterion is the same as with weighted tallying. The weighted linear model (or some variant of it) is often viewed as an optimal rule for preferential choice, under the idealization of independent dimensions or cues (e.g., Keeney & Raiffa, 1993; Payne et al., 1993). Comparing objects a and b from Figure 1 would involve assigning 1.0 points to a and 0.8 points to b and, thus, choosing a to be the larger.

Contestant 5: Multiple Regression

The weighted linear model reflects the different validities of the cues, but not the dependencies between cues. Multiple regression creates weights that reflect the covariances between

predictors or cues and is commonly seen as an "optimal" way to integrate various pieces of information into an estimate (e.g., Brunswik, 1955; Hammond, 1966). Neural networks using the delta rule determine their "optimal" weights by the same principles as multiple regression does (Stone, 1986). The delta rule carries out the equivalent of a multiple linear regression from the input patterns to the targets.

The weights for the multiple regression could simply be calculated from the full information about the nine ecological cues, as given in the Appendix. To make multiple regression an even stronger competitor, we also provided information about which cities the simulated individuals recognized. Thus, the multiple regression used nine ecological cues and the recognition cue to generate its weights. Because the weights for the recognition cue depend on which cities are recognized, we calculated 6 × 500 × 84 sets of weights: one for each simulated individual. Unlike any of the other algorithms, regression had access to the actual city populations (even for those cities not recognized by the hypothetical person) in the calculation of the weights.[4] During the quiz, each simulated person used the set of weights provided to it by multiple regression to estimate the populations of the cities in the comparison.

There was a missing-values problem in computing these 6 × 84 × 500 sets of regression coefficients, because most simulated individuals did not know certain cue values, for instance, the cue values of the cities they did not recognize. We strengthened the performance of multiple regression by substituting unknown cue values with the average of the cue values the person knew for the given cue.[5] This was done both in creating the weights and in using these weights to estimate populations. Unlike traditional procedures where weights are estimated from one half of the data, and inferences based on these weights are made for the other half, the regression algorithm had access to all the information in the Appendix (except, of course, the unknown cue values)—more information than was given to any of the competitors. In the competition, multiple regression and, to a lesser degree, the weighted linear model approximate the ideal of the Laplacean Demon.

Results

Speed

The Take The Best algorithm is designed to enable quick decision making. Compared with the integration algorithms, how much faster does it draw inferences, measured by the amount of information searched in memory? For instance, in Figure 1, the Take The Best algorithm would look up four cue values (including the recognition cue values) to infer that a is larger than b. None of the integration algorithms use limited search; thus, they always look up all cue values.

Figure 5 shows the amount of cue values retrieved from memory by the Take The Best algorithm for various levels of limited knowledge. The Take The Best algorithm reduces search in memory considerably. Depending on the knowledge state, this algorithm needed to search for between 2 (the number of recognition values) and 20 (the maximum possible cue values: Each city has nine cue values and one recognition value). For instance, when a person recognized half of the cities and knew 50% of their cue values, then, on average, only about

4 cue values (that is, one fifth of all possible) are searched for. The average across all simulated participants was 5.9, which was less than a third of all available cue values.

Accuracy

Given that it searches only for a limited amount of information, how accurate is the Take The Best algorithm, compared with the integration algorithms? We ran the competition for all states of limited knowledge shown in Figure 4. We first report the results of the competition in the case where each algorithm achieved its best performance: When 100% of the cue values were known. Figure 6 shows the results of the simulations, carried out in the same way as those in Figure 4.

To our surprise, the Take The Best algorithm drew as many correct inferences as any of the other algorithms, and more than some. The curves for Take The Best, multiple regression, weighted tallying, and tallying are so similar that there are only slight differences among them. Weighted tallying performed about as well as tallying, and the unit-weight linear model performed about as well as the weighted linear model—demonstrating that the previous finding that weights may be chosen in a fairly arbitrary manner, as long as they have the correct sign (Dawes, 1979), is generalizable to tallying. The two integration algorithms that make use of both positive and negative information, unit-weight and weighted linear models, made considerably fewer correct inferences. By looking at the lower-left and upper-right corners of Figure 6, one can see that all competitors do equally well with a complete lack of knowledge or with complete knowledge. They differ when knowledge is limited. Note that some algorithms can make more correct inferences when they do not have complete knowledge: a demonstration of the less-is-more effect mentioned earlier.

What was the result of the competition across all levels of limited knowledge? Table 2 shows the result for each level of limited knowledge of cue values, averaged across all levels of recognition knowledge. (Table 2 reports also the performance of two variants of the Take The Best algorithm, which we discuss later: the Minimalist and the Take The Last algorithm.) The values in the 100% column of Table 2 are the values in Figure 6 averaged across all levels of recognition. The Take The Best algorithm made as many correct inferences as one of the competitors (weighted tallying) and more than the others. Because it was also the fastest, we judged the competition goes to the Take The Best algorithm as the highest performing, overall.

To our knowledge, this is the first time that it has been demonstrated that a satisficing algorithm, that is, the Take The Best algorithm, can draw as many correct inferences about a real-

---

[4] We cannot claim that these integration algorithms are the best ones, nor can we know a priori which small variations will succeed in our bumpy real-world environment. An example: During the proof stage of this article we learned that regressing on the ranks of the cities does slightly better than regressing on the city populations. The key issue is what are the structures of environments in which particular algorithms and variants thrive.

[5] If no single cue value was known for a given cue, the missing values were substituted by .5. This value was chosen because it is the midpoint of 0 and 1, which are the values used to stand for negative and positive cue values, respectively.
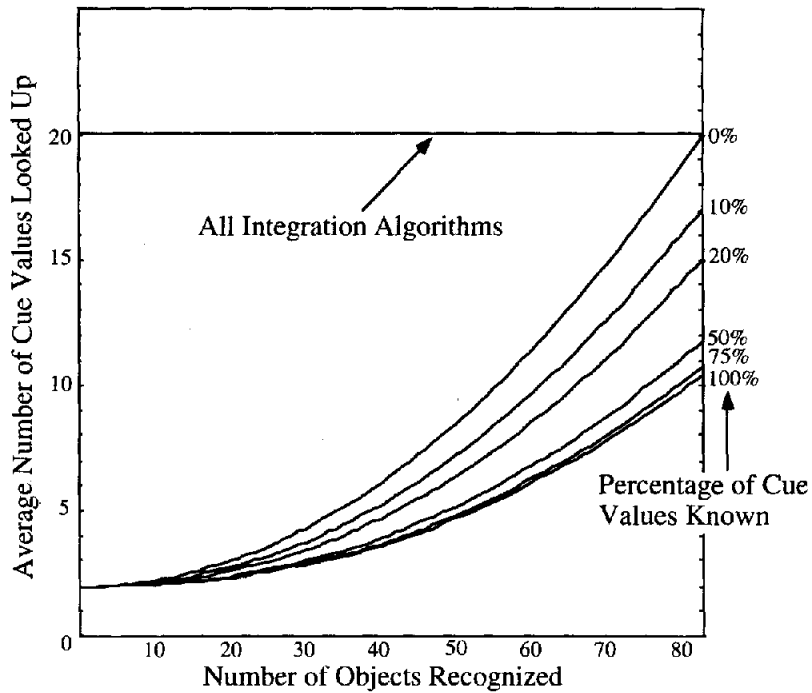
Figure 5. Amount of cue values looked up by the Take The Best algorithm and by the competing integration algorithms (see text), depending on the number of objects known (0–83) and the percentage of cue values known.
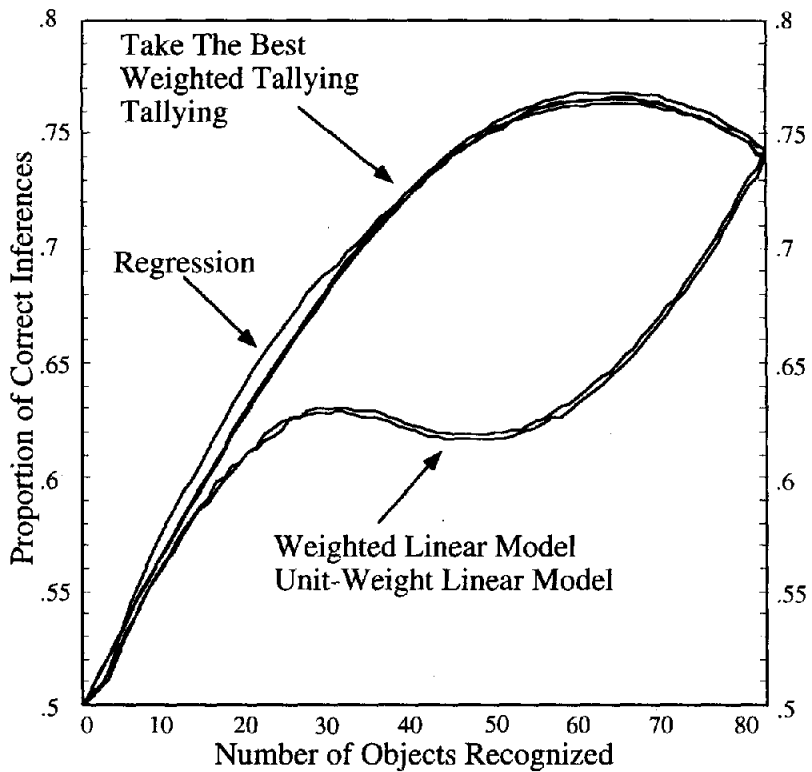
Figure 6. Results of the competition. The curve for the Take The Best algorithm is identical with the 100% curve in Figure 4. The results for proportion correct have been smoothed by a running median smoother, to lessen visual noise between the lines.

Table 2
Results of the Competition: Average Proportion of Correct Inferences

Algorithm	Percentage of cue values known					
	10	20	50	75	100	Average
Take The Best	.621	.635	.663	.678	.691	.658
Weighted tallying	.621	.635	.663	.679	.693	.658
Regression	.625	.635	.657	.674	.694	.657
Tallying	.620	.633	.659	.676	.691	.656
Weighted linear model	.623	.627	.623	.619	.625	.623
Unit-weight linear model	.621	.622	.621	.620	.622	.621
Minimalist	.619	.631	.650	.661	.674	.647
Take The Last	.619	.630	.646	.658	.675	.645

Note. Values are rounded; averages are computed from the unrounded values. Bottom two algorithms are variants of the Take The Best algorithm.

world environment as integration algorithms, across all states of limited knowledge. The dictates of classical rationality would have led one to expect the integration algorithms to do substantially better than the satisficing algorithm.

Two results of the simulation can be derived analytically. First and most obvious is that if knowledge about objects is zero, then all algorithms perform at a chance level. Second, and less obvious, is that if all objects and cue values are known, then tallying produces as many correct inferences as the unit-weight linear model. This is because, under complete knowledge, the score under the tallying algorithm is an increasing linear function of the score arrived at in the unit-weight linear model.[6] The equivalence between tallying and unit-weight linear models under complete knowledge is an important result. It is known that unit-weight linear models can sometimes perform about as well as proper linear models (i.e., models with weights that are chosen in an optimal way, such as in multiple regression; see Dawes, 1979). The equivalence implies that under complete knowledge, merely counting pieces of positive evidence can work as well as proper linear models. This result clarifies one condition under which searching only for positive evidence, a strategy that has sometimes been labeled *confirmation bias* or *positive test strategy,* can be a reasonable and efficient inferential strategy (Klayman & Ha, 1987; Tweney & Walker, 1990).

Why do the unit-weight and weighted linear models perform markedly worse under limited knowledge of objects? The reason is the simple and bold recognition principle. Algorithms that do not exploit the recognition principle in environments where recognition is strongly correlated with the target variable pay the price of a considerable number of wrong inferences. The unit-weight and weighted linear models use recognition information and integrate it with all other information but do not follow the recognition principle, that is, they sometimes choose unrecognized cities over recognized ones. Why is this? In the environment, there are more negative cue values than positive ones (see the Appendix), and most cities have more negative cue values than positive ones. From this it follows that when a recognized object is compared with an unrecognized object, the (weighted) sum of cue values of the recognized object will often be smaller than that of the unrecognized object (which is −1 for

the unit-weight model and −.8 for the weighted linear model). Here the unit-weight and weighted linear models often make the inference that the unrecognized object is the larger one, due to the overwhelming negative evidence for the recognized object. Such inferences contradict the recognition principle. Tallying algorithms, in contrast, have the recognition principle built in implicitly. Because tallying algorithms ignore negative information, the tally for an unrecognized object is always 0 and, thus, is always smaller than the tally for a recognized object, which is at least 1 (for tallying, or .8 for weighted tallying, due to the positive value on the recognition cue). Thus, tallying algorithms always arrive at the inference that a recognized object is larger than an unrecognized one.

Note that this explanation of the different performances puts the full weight in a psychological principle (the recognition principle) explicit in the Take The Best algorithm, as opposed to the statistical issue of how to find optimal weights in a linear function. To test this explanation, we reran the simulations for the unit-weight and weighted linear models under the same conditions but replacing the recognition cue with the recognition principle. The simulation showed that the recognition principle accounts for all the difference.

Can Satisficing Algorithms Get by With Even Less Time and Knowledge?

The Take The Best algorithm produced a surprisingly high proportion of correct inferences, compared with more computationally expensive integration algorithms. Making correct inferences despite limited knowledge is an important adaptive feature of an algorithm, but being right is not the only thing that counts. In many situations, time is limited, and acting fast can be as important as being correct. For instance, if you are driving on an unfamiliar highway and you have to decide in an instant what to do when the road forks, your problem is not necessarily making the best choice, but simply making a quick choice. Pressure to be quick is also characteristic for certain types of verbal interactions, such as press conferences, in which a fast answer indicates competence, or commercial interactions, such as having telephone service installed, where the customer has to decide in a few minutes which of a dozen calling features to purchase. These situations entail the dual constraints of limited knowledge and limited time. The Take The Best algorithm is already faster than any of the integration algorithms, because it performs only a limited search and does not need to compute weighted sums of cue values. Can it be made even faster? It can, if search is guided by the recency of cues in memory rather than by cue validity.

The Take The Last Algorithm

The Take The Last algorithm first tries the cue that discriminated the last time. If this cue does not discriminate, the algo-

---

[6] The proof for this is as follows. The tallying score t for a given object is the number n^+ of positive cue values, as defined above. The score u for the unit weight linear model is $n^+ - n^-$, where n^- is the number of negative cue values. Under complete knowledge, $n = n^+ + n^-$, where n is the number of cues. Thus, $t = n^+$, and $u = n^+ - n^-$. Because $n^- = n - n^+$, by substitution into the formula for u, we find that $u = n^+ - (n - n^+) = 2t - n$.

rithm then tries the cue that discriminated the time before last, and so on. The algorithm differs from the Take The Best algorithm in Step 2, which is now reformulated as Step 2':

Step 2': Search for the Cue Values of the Most Recent Cue

For the two objects, retrieve the cue values of the cue used most recently. If it is the first judgment and there is no discrimination record available, retrieve the cue values of a randomly chosen cue.

Thus, in Step 4, the algorithm goes back to Step 2'. Variants of this search principle have been studied as the "Einstellung effect" in the water jar experiments (Luchins & Luchins, 1994), where the solution strategy of the most recently solved problem is tried first on the subsequent problem. This effect has also been noted in physicians' generation of diagnoses for clinical cases (Weber, Böckenholt, Hilton, & Wallace, 1993).

This algorithm does not need a rank order of cues according to their validities; all that needs to be known is the direction in which a cue points. Knowledge about the rank order of cue validities is replaced by a memory of which cues were last used. Note that such a record can be built up independently of any knowledge about the structure of an environment and neither needs, nor uses, any feedback about whether inferences are right or wrong.

The Minimalist Algorithm

Can reasonably accurate inferences be achieved with even less knowledge? What we call the *Minimalist* algorithm needs neither information about the rank ordering of cue validities nor the discrimination history of the cues. In its ignorance, the algorithm picks cues in a random order. The algorithm differs from the Take The Best algorithm in Step 2, which is now reformulated as Step 2":

Step 2": Random Search

For the two objects, retrieve the cue values of a randomly chosen cue.

The Minimalist algorithm does not necessarily speed up search, but it tries to get by with even less knowledge than any other algorithm.

Results

Speed

How fast are the fast algorithms? The simulations showed that for each of the two variant algorithms, the relationship between amount of knowledge and the number of cue values looked up had the same form as for the Take The Best algorithm (Figure 5). That is, unlike the integration algorithms, the curves are concave and the number of cues searched for is maximal when knowledge of cue values is lowest. The average number of cue values looked up was lowest for the Take The Last algorithm (5.29) followed by the Minimalist algorithm (5.64) and the Take The Best algorithm (5.91). As knowledge becomes more and more limited (on both dimensions: recognition and cue values known), the difference in speed becomes

smaller and smaller. The reason why the Minimalist algorithm looks up fewer cue values than the Take The Best algorithm is that cue validities and cue discrimination rates are negatively correlated (Table 1); therefore, randomly chosen cues tend to have larger discrimination rates than cues chosen by cue validity.

Accuracy

What is the price to be paid for speeding up search or reducing the knowledge of cue orderings and discrimination histories to nothing? We tested the performance of the two algorithms on the same environment as all other algorithms. Figure 7 shows the proportion of correct inferences that the Minimalist algorithm achieved. For comparison, the performance of the Take The Best algorithm with 100% of cue values known is indicated by a dotted line. Note that the Minimalist algorithm performed surprisingly well. The maximum difference appeared when knowledge was complete and all cities were recognized. In these circumstances, the Minimalist algorithm did about 4 percentage points worse than the Take The Best algorithm. On average, the proportion of correct inferences was only 1.1 percentage points less than the best algorithms in the competition (Table 2).

The performance of the Take The Last algorithm is similar to Figure 7, and the average number of correct inferences is shown in Table 2. The Take The Last algorithm was faster but scored slightly less than the Minimalist algorithm. The Take The Last algorithm has an interesting ability, which fooled us in an earlier series of tests, where we used a systematic (as opposed to a random) method for presenting the test pairs, starting with the largest city and pairing it with all others, and so on. An integration algorithm such as multiple regression cannot "find out" that it is being tested in this systematic way, and its inferences are accordingly independent of the sequence of presentation. However, the Take The Last algorithm found out and won this first round of the competition, outperforming the other competitors by some 10 percentage points. How did it exploit systematic testing? Recall that it tries, first, the cue that discriminated the last time. If this cue does not discriminate, it proceeds with the cue that discriminated the time before, and so on. In doing so, when testing is systematic in the way described, it tends to find, for each city that is being paired with all smaller ones, the group of cues for which the larger city has a positive value. Trying these cues first increases the chances of finding a discriminating cue that points in the right direction (toward the larger city). We learned our lesson and reran the whole competition with randomly ordered of pairs of cities.

Discussion

The competition showed a surprising result: The Take The Best algorithm drew as many correct inferences about unknown features of a real-world environment as any of the integration algorithms, and more than some of them. Two further simplifications of the algorithm—the Take The Last algorithm (replacing knowledge about the rank orders of cue validities by a memory of the discrimination history of cues) and the Minimalist algorithm (dispensing with both) showed a compara-
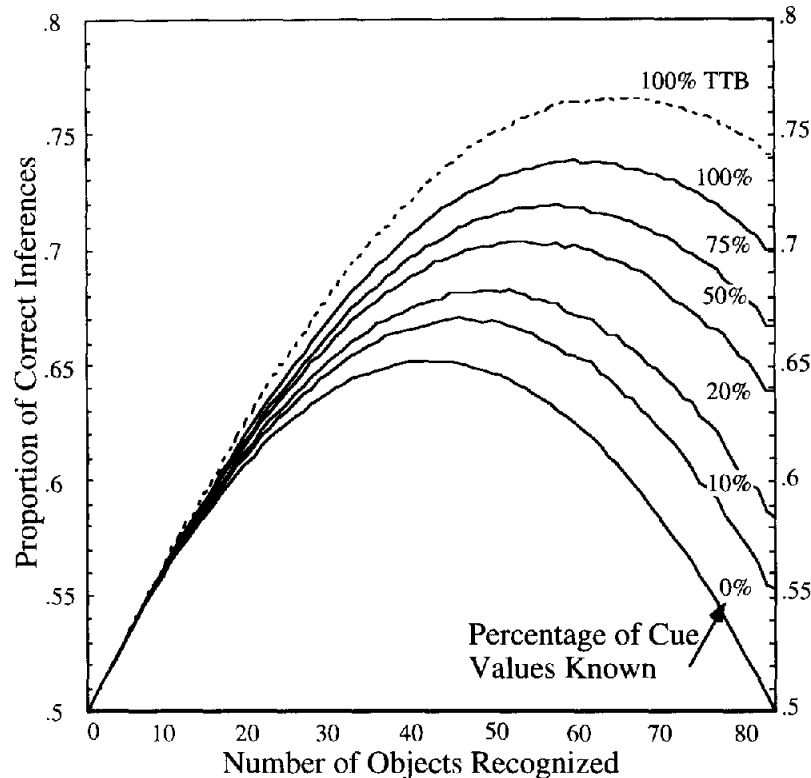
Figure 7. Performance of the Minimalist algorithm. For comparison, the performance of the Take The Best algorithm (TTB) is shown as a dotted line, for the case in which 100% of cue values are known.

tively small loss in correct inferences, and only when knowledge about cue values was high.

To the best of our knowledge, this is the first inference competition between satisficing and "rational" algorithms in a real-world environment. The result is of importance for encouraging research that focuses on the power of simple psychological mechanisms, that is, on the design and testing of satisficing algorithms. The result is also of importance as an existence proof that cognitive algorithms capable of successful performance in a real-world environment do not need to satisfy the classical norms of rational inference. The classical norms may be sufficient but are not necessary for good inference in real environments.

Cognitive Algorithms That Satisfice

In this section, we discuss the fundamental psychological mechanism postulated by the PMM family of algorithms: one-reason decision making. We discuss how this mechanism exploits the structure of environments in making fast inferences that differ from those arising from standard models of rational reasoning.

One-Reason Decision Making

What we call *one-reason decision making* is a specific form of satisficing. The inference, or decision, is based on a single, good reason. There is no compensation between cues. One-reason decision making is probably the most challenging feature of the

PMM family of algorithms. As we mentioned before, it is a design feature of an algorithm that is not present in those models that depict human inference as an optimal integration of all information available (implying that all information has been looked up in the first place), including linear multiple regression and nonlinear neural networks. One-reason decision making means that each choice is based exclusively on one reason (i.e., cue), but this reason may be different from decision to decision. This allows for highly context-sensitive modeling of choice. One-reason decision making is not compensatory. Compensation is, after all, the cornerstone of classical rationality, assuming that all commodities can be compared and everything has its price. Compensation assumes commensurability. However, human minds do not trade everything, some things are supposed to be without a price (Elster, 1979). For instance, if a person must choose between two actions that might help him or her get out of deep financial trouble, and one involves killing someone, then no amount of money or other benefits might compensate for the prospect of bloody hands. He or she takes the action that does not involve killing a person, whatever other differences exist between the two options. More generally, hierarchies of ethical and moral values are often noncompensatory: True friendship, military honors, and doctorates are supposed to be without a price.

Noncompensatory inference algorithms—such as lexicographic, conjunctive, and disjunctive rules—have been discussed in the literature, and some empirical evidence has been reported (e.g., Einhorn, 1970; Fishburn, 1988). The closest rel-

ative to the PMM family of satisficing algorithms is the lexicographic rule. The largest evidence for lexicographic processes seems to come from studies on decision under risk (for a recent summary, see Lopes, 1995). However, despite empirical evidence, noncompensatory lexicographic algorithms have often been dismissed at face value because they violate the tenets of classical rationality (Keeney & Raiffa, 1993; Lovie & Lovie, 1986). The PMM family is both more general and more specific than the lexicographic rule. It is more general because only the Take The Best algorithm uses a lexicographic procedure in which cues are ordered according to their validity, whereas the variant algorithms do not. It is more specific, because several other psychological principles are integrated with the lexicographic rule in the Take The Best algorithm, such as the recognition principle and the rules for confidence judgment (which are not dealt with in this article; see Gigerenzer et al., 1991).

Serious models that comprise noncompensatory inferences are hard to find. One of the few examples is in Breiman, Friedman, Olshen, and Stone (1993), who reported a simple, noncompensatory algorithm with only 3 binary, ordered cues, which classified heart attack patients into high- and low-risk groups and was more accurate than standard statistical classification methods that used up to 19 variables. The practical relevance of this noncompensatory classification algorithm is obvious: In the emergency room, the physician can quickly obtain the measures on one, two, or three variables and does not need to perform any computations because there is no integration. This group of statisticians constructed satisficing algorithms that approach the task of classification (and estimation) much like the Take The Best algorithm handles two-alternative choice. Relevance theory (Sperber, Cara, & Girotto, 1995) postulates that people generate consequences from rules according to accessibility and stop this process when expectations of relevance are met. Although relevance theory has not been as formalized, we see its stopping rule as parallel to that of the Take The Best algorithm. Finally, optimality theory (Legendre, Raymond, & Smolensky, 1993; Prince & Smolensky, 1991) proposes that hierarchical noncompensation explains how the grammar of a language determines which structural description of an input best satisfies well-formedness constraints. Optimality theory (which is actually a satisficing theory) applies the same inferential principles as PMM theory to phonology and morphology.

Recognition Principle

The recognition principle is a version of one-reason decision making that exploits a lack of knowledge. The very fact that one does not know is used to make accurate inferences. The recognition principle is an intuitively plausible principle that seems not to have been used until now in models of bounded rationality. However, it has long been used to good advantage by humans and other animals. For instance, advertisement techniques as recently used by Benetton put all effort into making sure that every customer recognizes the brand name, with no effort made to inform about the product itself. The idea behind this is that recognition is a strong force in customers' choices. One of our dear (and well-read) colleagues, after seeing a draft of this article, explained to us how he makes inferences about which books are worth acquiring. If he finds a book about a great topic but does not recognize the name of the author, he makes the inference that it is probably not worth buying. If, after an inspection of the references, he does not recognize most of the names, he concludes the book is not even worth reading. The recognition principle is also known as one of the rules that guide food preferences in animals. For instance, rats choose the food that they recognize having eaten before (or having smelled on the breath of fellow rats) and avoid novel foods (Gallistel, Brown, Carey, Gelman, & Keil, 1991).

The empirical validity of the recognition principle for inferences about unknown city populations, as used in the present simulations, can be directly tested in several ways. First, participants are presented pairs of cities, among them critical pairs in which one city is recognized and the other unrecognized, and their task is to infer which one has more inhabitants. The recognition principle predicts the recognized city. In our empirical tests, participants followed the recognition principle in roughly 90% to 100% of all cases (Goldstein, 1994; Goldstein & Gigerenzer, 1996). Second, participants are taught a cue, its ecological validity, and the cue values for some of the objects (such as whether a city has a soccer team or not). Subsequently, they are tested on critical pairs of cities, one recognized and one unrecognized, where the recognized city has a negative cue value (which indicates lower population). The second test is a harder test for the recognition principle than the first one and can be made even harder by using more cues with negative cue values for the recognized object, and by other means. Tests of the second kind have been performed, and participants still followed the recognition principle more than 90% of the time, providing evidence for its empirical validity (Goldstein, 1994; Goldstein & Gigerenzer, 1996).

The recognition principle is a useful heuristic in domains where recognition is a predictor of a target variable, such as whether a food contains a toxic substance. In cases where recognition does not predict the target, the PMM algorithms can still perform the inference, but without the recognition principle (i.e., Step 1 is canceled).

Limited Search

Both one-reason decision making and the recognition principle realize limited search by defining stopping points. Integration algorithms, in contrast, do not provide any model of stopping points and implicitly assume exhaustive search (although they may provide rules for tossing out some of the variables in a lengthy regression equation). Stopping rules are crucial for modeling inference under limited time, as in Simon's examples of satisficing, where search among alternatives terminates when a certain aspiration level is met.

Nonlinearity

Linearity is a mathematically convenient tool that has dominated the theory of rational choice since its inception in the mid-seventeenth century (Gigerenzer et al., 1989). The assumption is that the various components of an alternative add up independently to its overall estimate or utility. In contrast, nonlinear inference does not operate by computing linear sums of (weighted) cue values. Nonlinear inference has many varieties, including simple principles such as in the conjunctive and

disjunctive algorithms (Einhorn, 1970) and highly complex ones such as in nonlinear multiple regression and neural networks. The Take The Best algorithm and its variants belong to the family of simple nonlinear models. One advantage of simple nonlinear models is transparency; every step in the PMM algorithms can be followed through, unlike fully connected neural networks with numerous hidden units and other free parameters.

Our competition revealed that the unit-weight and weighted versions of the linear models lead to about equal performance, consistent with the finding that the choice of weights, provided the sign is correct, does often not matter much (Dawes, 1979). In real-world domains, such as in the prediction of sudden infant death from a linear combination of eight variables (Carpenter, Gardner, McWeeny & Emery, 1977), the weights can be varied across a broad range without decreasing predictive accuracy: a phenomenon known as the "flat maximum effect" (Lovie & Lovie, 1986; von Winterfeldt & Edwards, 1982). The competition in addition, showed that the flat maximum effect extends to tallying, with unit-weight and weighted tallying performing about equally well. The performance of the Take The Best algorithm showed that the flat maximum can extend beyond linear models: Inferences based solely on the best cue can be as accurate as any weighted or unit-weight linear combination of all cues.

Most research in psychology and economics has preferred linear models for description, prediction, and prescription (Edwards, 1954, 1962; Lopes, 1994; von Winterfeldt & Edwards, 1982). Historically, linear models such as analysis of variance and multiple regression originated as tools for data analysis in psychological laboratories and were subsequently projected by means of the "tools-to-theories heuristic" into theories of mind (Gigerenzer, 1991). The sufficiently good fit of linear models in many judgment studies has been interpreted that humans in fact might combine cues in a linear fashion. However, whether this can be taken to mean that humans actually use linear models is controversial (Hammond & Summers, 1965; Hammond & Wascoe, 1980). For instance, within a certain range, data generated from the (nonlinear) law of falling bodies can be fitted well by a linear regression. For the data in the Appendix, a multiple linear regression resulted in $R^2 = .87$, which means that a linear combination of the cues can predict the target variable quite well. But the simpler, nonlinear, Take The Best algorithm could match this performance. Thus, good fit of a linear model does not rule out simpler models of inference.

Shepard (1967) reviewed the empirical evidence for the claim that humans integrate information by linear models. He distinguished between the perceptual transformation of raw sensory inputs into conceptual objects and properties and the subsequent inference based on conceptual knowledge. He concluded that the perceptual analysis integrates the responses of the vast number of receptive elements into concepts and properties by complex nonlinear rules but once this is done, "there is little evidence that they can in turn be juggled and recombined with anything like this facility" (Shepard, 1967, p. 263). Although our minds can take account of a host of different factors, and although we can remember and report doing so, "it is seldom more than one or two that we consider at any one time" (Shepard, 1967, p. 267). In Shepard's view, there is little evi-

Figure 8. Limited knowledge and a stricter discrimination rule can produce intransitive inferences.

dence for integration, linear or otherwise, in what we term *inferences from memory*—even without constraints of limited time and knowledge. A further kind of evidence does not support linear integration as a model of memory-based inference. People often have great difficulties in handling correlations between cues (e.g., Armelius & Armelius, 1974), whereas integration models such as multiple regression need to handle intercorrelations. To summarize, for memory-based inference, there seems to be little empirical evidence for the view of the mind as a Laplacean Demon equipped with the computational powers to perform multiple regressions. But this need not be taken as bad news. The beauty of the nonlinear satisficing algorithms is that they can match the Demon's performance with less searching, less knowledge, and less computational might.

Intransitivity

Transitivity is a cornerstone of classical rationality. It is one of the few tenets that the Anglo-American school of Ramsey and Savage shares with the competing Franco-European school of Allais (Fishburn, 1991). If we prefer a to b and b to c, then we should also prefer a to c. The linear algorithms in our competition always produce transitive inferences (except for ties, where the algorithm randomly guessed), and city populations are, in fact, transitive. The PMM family of algorithms includes algorithms that do not violate transitivity (such as the Take The Best algorithm), and others that do (e.g., the Minimalist algorithm). The Minimalist algorithm randomly selects a cue on which to base the inference, therefore intransitivities can result. Table 2 shows that in spite of these intransitivities, overall performance of the algorithm is only about 1 percentage point lower than that of the best transitive algorithms and a few percentage points better than some transitive algorithms.

An organism that used the Take The Best algorithm with a stricter discrimination rule (actually, the original version found in Gigerenzer et al., 1991) could also be forced into making intransitive inferences. The stricter discrimination rule is that search is only terminated when one positive and one negative cue value (but not one positive and one unknown cue value) are encountered. Figure 8 illustrates a state of knowledge in which this stricter discrimination rule gives the result that a dominates b, b dominates c, and c dominates a.[7]

---

[7] Note that missing knowledge is necessary for intransitivities to occur. If all cue values are known, no intransitive inferences can possibly result. The algorithm with the stricter discrimination rule allows precise predictions about the occurrence of intransitivities over the course of knowledge acquisition. For instance, imagine a person whose knowledge is described by Figure 8, except that she does not know the value of Cue 2 for object c. This person would make no intransitive judgments

Biological systems, for instance, can exhibit systematic intransitivities based on incommensurability between two systems on one dimension (Gilpin, 1975; Lewontin, 1968). Imagine three species: a, b, and c. Species a inhabits both water and land; species b inhabits both water and air. Therefore, the two only compete in water, where species a defeats species b. Species c inhabits land and air, so it only competes with b in the air, where it is defeated by b. Finally, when a and c meet, it is only on land, and here, c is in its element and defeats a. A linear model that estimates some value for the combative strength of each species independently of the species with which it is competing would fail to capture this nontransitive cycle.

Inferences Without Estimation

Einhorn and Hogarth (1975) noted that in the unit-weight model "there is essentially no estimation involved in its use" (p. 177), except for the sign of the unit weight. A similar result holds for the algorithms reported here. The Take The Best algorithm does not need to estimate regression weights, it only needs to estimate a rank ordering of ecological validities. The Take The Last and the Minimalist algorithms involve essentially no estimation (except for the sign of the cues). The fact that there is no estimation problem has an important consequence: An organism can use as many cues as it has experienced, without being concerned about whether the size of the sample experienced is sufficiently large to generate reliable estimates of weights.

Cue Redundancy and Performance

Einhorn and Hogarth (1975) suggested that unit-weight models can be expected to perform approximately as well as proper linear models when (a) R^2 from the regression model is in the moderate or low range (around .5 or smaller) and (b) predictors (cues) are correlated. Are these two criteria necessary, sufficient, or both to explain the performance of the Take The Best algorithm? The Take The Best algorithm and its variants certainly can exploit cue redundancy: If cues are highly correlated, one cue can do the job.

We have already seen that in the present environment, $R^2 = .87$, which is in the high rather than the moderate or low range. As mentioned earlier, the pairwise correlations between the nine ecological cues ranged between $-.25$ and $.54$, with an absolute average value of .19. Thus, despite a high R^2 and only moderate-to-small correlation between cues, the satisficing algorithms performed quite successfully. Their excellent performance in the competition can be explained only partially by cue redundancy, because the cues were only moderately correlated. High cue redundancy, thus, does seem sufficient but is not nec-

essary for the successful performance of the satisficing algorithms.

A New Perspective on the Lens Model

Ecological theorists such as Brunswik (1955) emphasized that the cognitive system is designed to find many pathways to the world, substituting missing cues by whatever cues happen to be available. Brunswik labeled this ability *vicarious functioning*, in which he saw the most fundamental principle of a science of perception and cognition. His proposal to model this adaptive process by linear multiple regression has inspired a long tradition of neo-Brunswikian research (B. Brehmer, 1994; Hammond, 1990), although the empirical evidence for mental multiple regression is still controversial (e.g., A. Brehmer & B. Brehmer, 1988). However, vicarious functioning need not be equated with linear regression. The PMM family of algorithms provides an alternative, nonadditive model of vicarious functioning, in which cue substitution operates without integration. This gives a new perspective of Brunswik's lens model. In a one-reason decision making lens, the first discriminating cue that passes through inhibits any other rays passing through and determines judgment. Noncompensatory vicarious functioning is consistent with some of Brunswik's original examples, such as the substitution of behaviors in Hull's habit–family hierarchy, and the alternative manifestation of symptoms according to the psychoanalytic writings of Frenkel-Brunswik (see Gigerenzer & Murray, 1987, chap. 3).

It has been reported sometimes that teachers, physicians, and other professionals claim that they use seven or so criteria to make judgments (e.g., when grading papers or making a differential diagnosis) but that experimental tests showed that they in fact often used only one criterion (Shepard, 1967). At first glance, this seems to indicate that those professionals make outrageous claims. But it need not be. If experts' vicarious functioning works according to the PMM algorithms, then they are correct in saying that they use many predictors, but the decision is made by only one at any time.

What Counts as Good Reasoning?

Much of the research on reasoning in the last decades has assumed that sound reasoning can be reduced to principles of internal consistency, such as additivity of probabilities, conformity to truth-table logic, and transitivity. For instance, research on the Wason selection task, the "Linda" problem, and the "cab" problem has evaluated reasoning almost exclusively by some measure of internal consistency (Gigerenzer, 1995, 1996a). Cognitive algorithms, however, need to meet more important constraints than internal consistency: (a) They need to be psychologically plausible, (b) they need to be fast, and (c) they need to make accurate inferences in real-world environments. In real time and real environments, the possibility that an algorithm (e.g., the Minimalist algorithm) can make intransitive inferences does not mean that it will make them all the time or that this feature of the algorithm will significantly hurt its accuracy. What we have not addressed in this article are constraints on human reasoning that emerge from the fact that *Homo sapiens* is a social animal (Gigerenzer, 1996b). For instance, some choices (e.g., who to treat first in an emergency

---

comparing objects a, b, and c. If she were to learn that object c had a negative cue value for Cue 2, she would produce an intransitive judgment. If she learned one piece more, namely, the value of Cue 1 for object c, then she would no longer produce an intransitive judgment. The prediction is that transitive judgments should turn into intransitive ones and back, during learning. Thus, intransitivities do not simply depend on the amount of limited knowledge but also on what knowledge is missing.

room) need to be justified (Tetlock, 1992). Going with the single best reason, the strategy of the Take The Best algorithm, has an immediate appeal for justification and can be more convincing and certainly easier to communicate than some complicated weighting of cues.

Further Research

Among the questions that need to be addressed in future research are the following. First, how can we generalize the present satisficing algorithm from two-alternative-choice tasks to other inferential tasks, such as classification and estimation? The reported success of the classification and regression tree models (Breiman et al., 1993), which use a form of one-reason decision making, is an encouraging sign that what we have shown here for two-alternative-choice tasks might be generalizable. Second, what is the structure of real-world environments that allows simple algorithms to perform so well? We need to develop a conceptual language that can capture important aspects of the structure of environments that simple cognitive algorithms can exploit. The traditional proposal for understanding the structure of environments in terms of ecological validities defined as linear correlations (Brunswik, 1955) may not be adequate, as the power of the nonlinear satisficing algorithms suggests.

Can Reasoning Be Rational and Psychological?

At the beginning of this article, we pointed out the common opposition between the rational and the psychological, which emerged in the nineteenth century after the breakdown of the classical interpretation of probability (Gigerenzer et al., 1989). Since then, rational inference is commonly reduced to logic and probability theory, and psychological explanations are called on when things go wrong. This division of labor is, in a nutshell, the basis on which much of the current research on judgment under uncertainty is built. As one economist from the Massachusetts Institute of Technology put it, "either reasoning is rational or it's psychological" (Gigerenzer, 1994). Can not reasoning be both rational and psychological?

We believe that after 40 years of toying with the notion of bounded rationality, it is time to overcome the opposition between the rational and the psychological and to reunite the two. The PMM family of cognitive algorithms provides precise models that attempt to do so. They differ from the Enlightenment's unified view of the rational and psychological, in that they focus on simple psychological mechanisms that operate under constraints of limited time and knowledge and are supported by empirical evidence. The single most important result in this article is that simple psychological mechanisms can yield about as many (or more) correct inferences in less time than standard statistical linear models that embody classical properties of rational inference. The demonstration that a fast and frugal satisficing algorithm won the competition defeats the widespread view that only "rational" algorithms can be accurate. Models of inference do not have to forsake accuracy for simplicity. The mind can have it both ways.

References

Alba, J. W., & Marmorstein, H. (1987). The effects of frequency knowledge on consumer decision making. *Journal of Consumer Research*, 14, 14–26.

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

Armelius, B., & Armelius, K. (1974). The use of redundancy in multiple-cue judgments: Data from a suppressor-variable task. *American Journal of Psychology*, 87, 385–392.

Becker, G. (1976). *The economic approach to human behavior.* Chicago: University of Chicago Press.

Brehmer, A., & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing? In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 75–114). Amsterdam: North-Holland.

Brehmer, B. (1994). The psychology of linear judgment models. *Acta Psychologica*, 87, 137–154.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1993). *Classification and regression trees.* New York: Chapman & Hall.

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review*, 100, 511–534.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62, 193–217.

Carpenter, R. G., Gardner, A., McWeeny, P. M., & Emery, J. L. (1977). Multistage scoring system for identifying infants at risk of unexpected death. *Archives of Disease in Childhood*, 53, 606–612.

Darwin, C. (1965). *The expressions of the emotions in man and animal.* Chicago: University of Chicago Press. (Original work published 1872)

Daston, L. (1988). *Classical probability in the Enlightenment.* Princeton, NJ: Princeton University Press.

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

DiFonzo, N. (1994). *Piggybacked syllogisms for investor behavior: Probabilistic mental modeling in rumor-based stock market trading.* Unpublished doctoral dissertation, Temple University, Philadelphia.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, 51, 380–417.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human Factors*, 4, 59–73.

Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision-making. *Psychological Bulletin*, 73, 221–230.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, 13, 171–192.

Elster, J. (1979). *Ulysses and the sirens: Studies in rationality and irrationality.* Cambridge, England: Cambridge University Press.

Fischer Welt Almanach [Fischer World Almanac]. (1993). Frankfurt, Germany: Fischer.

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 349–358.

Fishburn, P. C. (1988). *Nonlinear preference and utility theory.* Baltimore: Johns Hopkins University Press.

Fishburn, P. C. (1991). Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty*, 4, 113–134.

Gallistel, C. R., Brown, A. L., Carey, S., Gelman, R., & Keil, F. C. (1991). Lessons from animal learning for the study of cognitive development. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 3–36). Hillsdale, NJ: Erlbaum.

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98, 254–267.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. I. Manktelow & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 284–313). London: Routledge.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is relevant for psychology (and vice versa).

In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129–161). New York: Wiley.

Gigerenzer, G. (1995). The taming of content; Some thoughts about domains and modules. *Thinking and Reasoning, 1*, 324–333.

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

Gigerenzer, G. (1996b). Rationality: Why social context matters. In P. Baltes & U. M. Staudinger (Eds.), *Interactive minds: Life-span perspectives on the social foundation of cognition* (pp. 319–346). Cambridge, England: Cambridge University Press.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684–704.

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

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics.* Hillsdale, NJ: Erlbaum.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life.* Cambridge, England: Cambridge University Press.

Gilpin, M. E. (1975). Limit cycles in competition communities. *The American Naturalist, 109*, 51–60.

Goldstein, D. G. (1994). *The less-is-more effect in inference.* Unpublished master's thesis, University of Chicago.

Goldstein, D. G., & Gigerenzer, G. (1996). *Reasoning by recognition alone: How to exploit a lack of knowledge.* Unpublished manuscript.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411–435.

Hammond, K. R. (1966). *The psychology of Egon Brunswik.* New York: Holt, Rinehart & Winston.

Hammond, K. R. (1990). Functionalism and illusionism: Can integration be usefully achieved? In R. M. Hogarth (Ed.), *Insights in decision making* (pp. 227–261). Chicago: University of Chicago Press.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71*, 438–456.

Hammond, K. R., & Summers, D. A. (1965). Cognitive dependence on linear and nonlinear cues. *Psychological Review, 72*, 215–244.

Hammond, K. R., & Wascoe, N. E. (Eds.). (1980). *Realizations of Brunswik's representative design: New directions for methodology of social and behavioral science.* San Francisco: Jossey-Bass.

Hertwig, R., Gigerenzer, G., & Hoffrage, U. (in press). The reiteration effect in hindsight bias. *Psychological Review.*

Hoffrage, U. (1994). *Zur Angemessenheit subjektiver Sicherheits-Urteile: Eine Exploration der Theorie der probabilistischen mentalen Modelle* [On the validity of confidence judgments: A study of the theory of probabilistic mental models]. Unpublished doctoral dissertation, Universität Salzburg, Salzburg, Austria.

Huber, O. (1989). Information-processing operators in decision making. In H. Montgomery & O. Svenson (Eds.), *Process and structure in human decision making* (pp. 3–21). New York: Wiley.

Huttenlocher, J., Hedges, L., & Prohaska, V. (1988). Hierarchical organization in ordered domains: Estimating the dates of events. *Psychological Review, 95*, 471–484.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Juslin, P. (1993). An explanation of the hard-easy effect in studies of realism of confidence in one's general knowledge. *European Journal of Cognitive Psychology, 5*, 55–71.

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57*, 226–246.

Juslin, P., Winman, A., & Persson, T. (1995). Can overconfidence be used as an indicator of reconstructive rather than retrieval processes? *Cognition, 54*, 99–130.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, England: Cambridge University Press.

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives.* Cambridge, England: Cambridge University Press.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107–118.

Krebs, J. R., & Davies, N. B. (1987). *An introduction to behavioral ecology* (2nd ed.). Oxford: Blackwell.

Legendre, G., Raymond, W., & Smolensky, P. (1993). Analytic typology of case marking and grammatical voice. *Proceedings of the Berkeley Linguistics Society, 19*, 464–478.

Lewontin, R. C. (1968). Evolution of complex genetic systems. In M. Gerstenhaber (Ed.), *Some mathematical questions in biology.* Providence, RI: American Mathematical Society.

Lopes, L. L. (1992). Three misleading assumptions in the customary rhetoric of the bias literature. *Theory and Psychology, 2*, 231–236.

Lopes, L. L. (1994). Psychology and economics: Perspectives on risk, cooperation, and the marketplace. *Annual Review of Psychology, 45*, 197–227.

Lopes, L. L. (1995). Algebra and process in the modeling of risky choice. In J. R. Busemeyer, R. Hastie, and D. Medin (Eds.), *Decision making from the perspective of cognitive psychology* (pp. 177–220). New York: Academic Press.

Lovie, A. D., & Lovie, P. (1986). The flat maximum effect and linear scoring models for prediction. *Journal of Forecasting, 5*, 159–168.

Luchins, A. S., & Luchins, E. H. (1994). The water jar experiments and Einstellung effects: I. Early history and surveys of textbook citations. *Gestalt Theory, 16*, 101–121.

McClelland, A. G. R., & Bolger, F. (1994). The calibration of subjective probabilities: Theories and models 1980–1994. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 453–482). Chichester, England: Wiley.

McClennen, E. F. (1990). *Rationality and dynamic choice.* Cambridge, England: Cambridge University Press.

McCloskey, D. N. (1985). *The rhetoric of economics.* Madison: University of Wisconsin Press.

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

Prince, A., & Smolensky, P. (1991). *Notes on connectionism and harmony theory in linguistics* (Tech. Rep. No. CU-CS-533-91). Boulder: University of Colorado, Department of Computer Science.

Shepard, R. N. (1967). On subjectively optimum selections among multi-attribute alternatives. In W. Edwards & A. Tversky (Eds.), *Decision making* (pp. 257–283). Baltimore: Penguin Books.

Simon, H. A. (1945). *Administrative behavior: A study of decision-making processes in administrative organization.* New York: Free Press.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129–138.

Simon, H. A. (1982). *Models of bounded rationality.* Cambridge, MA: MIT Press.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*, 1–19.

Simon, H. A. (1992). *Economics, bounded rationality, and the cognitive revolution.* Aldershot Hants, England: Elgar.

Sniezek, J. A., & Buckley, T. (1993). Becoming more or less uncertain. In N. J. Castellan (Ed.), *Individual and group decision making* (pp. 87–108). Hillsdale, NJ: Erlbaum.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 57*, 31–95.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory.* Princeton, NJ: Princeton University Press.

Stone, G. O. (1986). An analysis of the delta rule and the learning of statistical associations. In D. Rumelhart, J. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 444–459). Cambridge, MA: MIT Press.

Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 331–376). New York: Academic Press.

Tweney, R. D., & Walker, B. J. (1990). Science education and the cognitive psychology of science. In B. F. Jones & L. Idol (Eds.), *Dimen-*
sions of thinking and cognitive instruction (pp. 291–310). Hillsdale, NJ: Erlbaum.

von Winterfeldt, D., & Edwards, W. (1982). Costs and payoffs in perceptual research. *Psychological Bulletin, 91,* 609–622.

Weber, U., Böckenholt, U., Hilton, D. J., & Wallace, B. (1993). Determinants of diagnostic hypothesis generation: Effects of information, base rates, and experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1151–1164.

Wimsatt, W. C. (1976). Reductionism, levels of organization, and the mind–body problem. In G. G. Globus, G. Maxwell, & I. Savodnik (Eds.), *Consciousness and the brain: A scientific and philosophical inquiry* (pp. 199–267). New York: Plenum.

Appendix

The Environment

City	Population	Soccer team	State capital	Former East Germany	Industrial belt	Licence plate	Intercity trainline	Exposition site	National capital	University
Berlin	3,433,695	−	+	−	−	+	+	+	+	+
Hamburg	1,652,363	+	+	−	−	−	+	+	−	+
Munich	1,229,026	+	+	−	−	+	+	+	−	+
Cologne	953,551	+	−	−	−	+	+	+	−	+
Frankfurt	644,865	+	−	−	−	+	+	+	−	+
Essen	626,973	−	−	−	+	+	+	+	−	+
Dortmund	599,055	+	−	−	+	−	+	+	−	+
Stuttgart	579,988	+	+	−	−	+	+	+	−	+
Düsseldorf	575,794	−	+	−	−	+	+	+	−	+
Bremen	551,219	+	+	−	−	−	+	−	−	+
Duisburg	535,447	−	−	−	+	−	+	−	−	+
Hannover	513,010	−	+	−	−	+	+	+	−	+
Leipzig	511,079	−	−	+	−	+	+	+	−	+
Nuremberg	493,692	+	−	−	−	+	+	+	−	+
Dresden	490,571	+	−*	+	−	−	+	−	−	+
Bochum	396,486	+	−	−	+	−	+	−	−	+
Wuppertal	383,660	−	−	−	+	+	+	−	−	+
Bielefeld	319,037	−	−	−	−	−	+	−	−	+
Mannheim	310,411	−	−	−	−	−	+	−	−	+
Halle	310,234	−	−	+	−	−	+	−	−	−
Chemnitz	294,244	−	−	+	−	+	−	−	−	−
Gelsenkirchen	293,714	+	−	−	+	−	+	−	−	−
Bonn	292,234	−	−	−	−	−	+	−	−	+
Magdeburg	278,807	−	+	+	−	−	+	−	−	−
Karlsruhe	275,061	+	−	−	−	−	+	−	−	−
Wiesbaden	260,301	−	+	−	−	−	+	−	−	−
Münster	259,438	−	−	−	−	−	+	−	−	+
Mönchengladbach	259,436	+	−	−	−	−	−	−	−	−
Braunschweig	258,833	−	−	−	−	−	+	−	−	+
Augsburg	256,877	−	−	−	−	+	+	−	−	+
Rostock	248,088	−	−	+	−	−	+	−	−	−
Kiel	245,567	−	+	−	−	−	+	−	−	+
Krefeld	244,020	−*	−	−	−	−	−	−	−	−
Aachen	241,961	−	−	−	−	−	+	−	−	+
Oberhausen	223,840	−	−	−	+	−	+	−	−	−
Lübeck	214,758	−	−	−	−	−	+	−	−	−
Hagen	214,449	−	−	−	+	−	+	−	−	−
Erfurt	208,989	−	+	+	−	−	+	−	−	−
Kassel	194,268	−	−	−	−	−	+	−	−	+
Saarbrücken	191,694	+	+	−	−	−	+	+	−	+

City	Population	Soccer team	State capital	Former East Germany	Industrial belt	Licence plate	Intercity trainline	Exposition site	National capital	University
Freiburg	191,029	−	−	−	−	−	+	−	−	+
Hamm	179,639	−−	−	−	+	−	+	−	−	−
Mainz	179,486	−	+	−	−	−	+	−	−	+
Herne	178,132	−	−	−	+	−	−	−	−	−
Mülheim	177,681	−	−	−	+	−	−	−	−	−
Solingen	165,401	−	−	−	−	−	+	−	−	−
Osnabrück	163,168	−	−	−	−	−	+	−	−	+
Ludwigshafen	162,173	−	−	−	−	−	+	−	−	−
Leverkusen	160,919	+	−	−	−	−	−	−	−	−
Neuss	147,019	−	−	−	−	−	−	−	−	−
Oldenburg	143,131	−	−	−	−	−	+	−	−	+
Potsdam	139,794	−	+	+	−	+	+	−	−	−
Darmstadt	138,920	−	−	−	−	−	+	−	−	+
Heidelberg	136,796	−	−	−	−	−	+	−	−	+
Bremerhaven	130,446	−	−	−	−	−	+	−	−	−
Gera	129,037	−	−	+	−	+	+	−	−	−
Wolfsburg	128,510	−	−	−	−	−	−	−	−	−
Würzburg	127,777	−	−	−	−	−	+	−	−	+
Schwerin	127,447	−	+	+	−	−	+	−	−	−
Cottbus	125,891	−	−	+	−	−	−	−	−	−
Recklinghausen	125,060	−	−	−	+	−	+	−	−	−
Remscheid	123,155	−	−	−	−	−	−	−	−	−
Göttingen	121,831	−	−	−	−	−	+	−	−	+
Regensburg	121,691	−	−	−	−	+	+	−	−	+
Paderborn	120,680	−	−	−	−	−	−	−	−	+
Bottrop	118,936	−	−	−	+	−	−	−	−	−
Heilbronn	115,843	−	−	−	−	−	−	−	−	−
Offenbach	114,992	−	−	−	−	−	−	+	−	−
Zwickau	114,636	−	−	+	−	+	−	−	−	−
Salzgitter	114,355	−	−	−	−	−	−	−	−	−
Pforzheim	112,944	−	−	−	−	−	+	−	−	−
Ulm	110,529	−	−	−	−	−	+	−	−	+
Siegen	109,174	−	−	−	−	−	−	−	−	+
Koblenz	108,733	−	−	−	−	−	+	−	−	+
Jena	105,518	−	−	+	−	+	+	−	−	+
Ingolstadt	105,489	−	−	−	−	−	+	−	−	−
Witten	105,403	−	−	−	+	−	−	−	−	−
Hildesheim	105,291	−	−	−	−	−	+	−	−	+
Moers	104,595	−	−	−	+	−	−	−	−	−
Bergisch Gladbach	104,037	−	−	−	−	−	−	−	−	−
Reutlingen	103,687	−	−	−	−	−	−	−	−	−
Fürth	103,362	−	−	−	−	−	+	−	−	−
Erlangen	102,440	−	−	−	−	−	+	−	−	+

Note. City populations were taken from *Fischer Welt Almanach* (1993).

* The two starred minus values are, in reality, plus values. Because of transcription errors, we ran all simulations with these two minus values. These do not affect the rank order of cue validities, should not have any noticeable effect on the results, and are irrelevant for our theoretical argument.