HOW GOOD ARE FAST AND FRUGAL HEURISTICS?

Gerd Gigerenzer, Jean Czerlinski, Laura Martignon

Rationality and optimality are the guiding concepts of the probabilistic approach to cognition, but they are not the only reasonable guiding concepts. Two concepts from the other end of the spectrum, simplicity and frugality, have also inspired models of cognition. These fast and frugal models are justified by their psychological plausibility and adaptedness to natural environments. For example, the real world provides only scarce information, the real world forces us to rush when gathering and processing information and the real world does not cut itself up into variables whose errors are conveniently independently normally distributed, as many optimal models assume.

However, recent optimal models already address these constraints. There are many methods for dealing with missing information. Optimal models can also be extended to take into account the cost of acquiring information. Finally, variables with unusual distributions can be transformed into nearly normal distributions, and outliers can be excluded. So what's the big deal? Optimal models seem to have met the challenge of adapting to natural environments. And if people do not already use these models, then they would want to learn how to use them since they are, after all, optimal.

Thus it would seem that there is no need to turn to fast and frugal heuristics, which appear doomed to be both simplistic and inaccurate. Besides, there is an even

J. Shanteau et al. (eds.), *Decision Science and Technology*
© Kluwer Academic Publishers 1999

stronger reason to shun simplicity and frugality as the basis for human cognition. They deny some of the most striking self-images Homo sapiens has constructed of itself: from "l'homme éclairé" of the Enlightenment to Homo economicus of modern business schools (Gigerenzer et al., 1989).

These are the typical intuitive arguments in the debate between optimality and rationality on the one hand and simplicity and frugality on the other. But before you pass judgment on where you stand, move beyond these mere intuitions to consider the real substance of the two approaches and the actual relationship between them. This chapter provides some food for your thoughts on these issues in the form of a review of our recent findings on fast and frugal heuristics (Gigerenzer, Todd, & the ABC group, in press). How great is the advantage in terms of speed and simplicity? How large is the loss of accuracy? How robust are fast and frugal heuristics under a variety of conditions—and under which conditions should we avoid using them? We answer these questions by comparing fast and frugal heuristics with benchmark models from the optimality and rationality tradition. Our intention is not to rule out one set of guiding concepts or the other, forcing us to choose rationality and optimality *or* simplicity and frugality. Rather, we wish to explore how far we can get with simple heuristics that may be more realistic models of how humans make inferences under constraints of limited time and knowledge.

But first we have to understand the guiding concepts. The fundamental distinction in approaches to reasonableness is between unbounded rationality and bounded rationality (e.g., Simon, 1982, 1992). Unbounded rationality suggests building models that perform as well as possible with little or no regard for how time-consuming or informationally greedy these models may be. This approach includes Bayesian models and expected utility maximization models (e.g., Edwards, 1954, 1961). In contrast, bounded rationality suggests designing models specifically to reflect the peculiar properties and limits of the mind and the environment. The decision maker is bounded in time, knowledge and computational power. In addition, each environment has a variety of irregular informational structures, such as departures from normality.

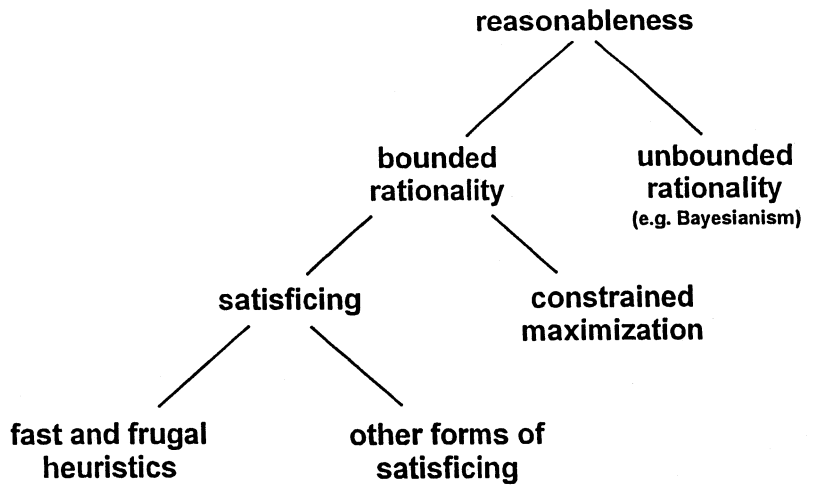

Figure 1. Visions of reasonableness.

There are, however, two approaches which compete for the title of bounded rationality: *constrained maximization* and *satisficing* (Figure 1). Constrained maximization means maximization under deliberation cost constraints. This demands even more knowledge and computation than unbounded rationality because the decision maker has to compute the optimal trade-off between accuracy and various costs, such as information search costs and opportunity costs. The paradoxical result is that "limited" minds are assumed to have the knowledge and computational ability of mathematically sophisticated econometricians and their statistical software packages (e.g., Sargent, 1993). The "father" of bounded rationality, Herbert Simon, has vehemently rejected this approach. In personal conversation, he once remarked in a mixture of anger and humor that he had thought of suing authors who misuse his concept of bounded rationality to construct ever more complicated models of human decision making.

Simon's view of bounded rationality is that of satisficing, which he contrasts to constrained maximization. In the satisficing interpretation, the two sides of bounded rationality, limited minds and structured environments, are not merely two additional complications to the optimality story. Rather, they form a happy and beneficial marriage: subtle environmental structures that were neglected by standard rational models are potentially exploitable by simple heuristics. (Egon Brunswik in particular has emphasized the interrelationship of cognition and environment, e.g., Brunswik, 1964). Satisficing asserts that our minds have evolved all sorts of nimble tricks to perform well in the quirky structures of the real world.

The types of models developed by the satisficing view are thus fairly simple, in stark contrast to those of the constrained maximization view. For instance, one of the best known examples of Simon's satisficing is to start with an aspiration level and then choose the first object encountered that satisfies this level (e.g., buy the first acceptable house). Still, satisficing can employ rather computationally expensive procedures (e.g., Simon, 1956). We use the term *fast and frugal heuristics* for a subset of satisficing strategies that work with a minimum of knowledge, time and computations. We call these heuristics "fast" because they process information in a relatively simple way, and we call them "frugal" because they use little information. The next section presents several examples of such heuristics.

1. Satisficing by fast and frugal heuristics

There are infinitely many kinds of tasks that heuristics can be designed to perform. This chapter focuses on the task of predicting or inferring which of two objects scores higher on a criterion. Which soccer team will win? Which of two cities has a higher homelessness rate? Which applicant will do a better job? To make such predictions, the heuristics use uncertain cues which indicate, with some probability, higher values on the criterion.

Consider, for example, the task of inferring which of two cities has a higher homelessness rate, using the data on 50 U.S. cities from Tucker (1987). An excerpt from this data including the values for Los Angeles, Chicago, New York and New Orleans on six cues and the criterion is shown in Table 1. One cue (rent control) is binary, and the other five have been dichotomized at the median. Unitary cue values ("1") indicate higher values on the criterion and zero cue values ("0") indicate lower values. For example, since cities with rent control more often have a higher

homelessness rate than cities without rent control, cities that have rent control are marked with a cue value of "1" for this cue. (In contrast, if cities without rent control more often had the higher homelessness rate, then having rent control would be marked by a "0.")

Of course, people generally do not have such tables of information handy; they have to search for information, in their memories or in libraries. But how could one construct a heuristic that cheaply (rather than optimally) limits search and computations? Two examples of such heuristics are Minimalist and Take The Best, which are drawn from a family of fast and frugal heuristics (Gigerenzer & Goldstein, 1996; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Gigerenzer, Todd, & the ABC group, in press).

	Los Angeles	Chicago	New York	New Orleans
Homeless per million	10,526	6,618	5,024	2,671
Rent control (1 is yes)	1	0	1	0
Vacancy rate (1 is below median)	1	1	1	0
Temperature (1 is above median)	1	0	1	1
Unemployment (1 is above median)	1	1	1	1
Poverty (1 is above median)	1	1	1	1
Public housing (1 is below median)	1	1	0	0

Table 1. Cues for predicting homelessness in U.S. cities. Cues are ordered by validity, with rent control having the highest validity (.90). Further explanation in text. Source: Tucker (1987).

Minimalist. The minimal knowledge needed for cue-based inference is in which direction a cue "points." For instance, the heuristic needs to know whether warmer or cooler weather indicates a city with a higher rate of homelessness. In the 50 U.S. cities, warmer weather is indeed associated more often with higher homelessness rates than with lower rates, so a cue value of "1" is assigned to cities with warmer weather. Minimalist has only this minimal knowledge. Nothing is known about which cues have higher validity than others. The ignorant strategy of Minimalist is to look up cues in random order, choosing the city that has a cue value of "1" when the other city does not. Minimalist can be expressed in the following steps:

Step 1. Random search: Randomly select a cue and look up the cue values of the two objects.

Step 2. Stopping rule: If one object has a cue value of one ("1") and the other does not, then stop search. Otherwise go back to Step 1 and search for another cue. If no further cue is found, guess.

Step 3. Decision rule: Predict that the object with the cue value of one ("1") has the higher value in the criterion.

For instance, when inferring whether Chicago or New Orleans has a higher homelessness rate, the unemployment cue might be the first cue randomly selected, and the cue values are found to be one and one (Table 1). Search is continued, the public housing cue is randomly selected, and the cue values are one and zero.

Search is stopped and the inference is made that Chicago has a higher homelessness rate, as it indeed does.

So far, the only thing a person needs to estimate is which direction a cue points, that is, whether it indicates a higher or a lower value on the criterion. But there exist environments for which humans know not just the signs of cues, but roughly how predictive they are. If people can order cues according to their validities—whether or not this subjective order corresponds to the ecological order—then search can follow this order of cues. One of the heuristics that differs from the Minimalist in only this respect is called Take The Best; its motto is "Take the best, ignore the rest."

Take The Best. This heuristic is exactly like Minimalist except that the cue with the highest validity, rather than a random cue, is tried first. If this cue does not discriminate, the next best cue is tried, and so forth. Thus, Take The Best differs from Minimalist only in Step 1.

Step 1. Ordered search: Select the cue with the highest validity and look up the cue values of the two objects.

The *validity* v_i of cue i is the number of right (correct) inferences, R_i, divided by the number of right and wrong inferences, $R_i + W_i$, based on cue i alone, independent of the other cues. We count which inferences are right and wrong across all possible inferences in a reference class of objects. That is,

$$v_i = \frac{\text{right inferences}}{\text{right inferences + wrong inferences}} = \frac{R_i}{R_i + W_i} .$$

For example, since Los Angeles has a cue value of one for rent control while Chicago has a cue value of zero, the rent control cue suggests that Los Angeles has a higher homelessness rate; since Los Angeles does have a higher homelessness rate, this counts as a right inference. Between Chicago and New York, the rent control cue makes a wrong inference. And between Chicago and New Orleans, it does not discriminate—and cannot make an inference—because both cities have zero cue values for rent control. If we count the number of right and wrong inferences for all possible pairings of the 50 U.S. cities, we find that 90% of the inferences based on rent control are right; thus the cue validity of rent control is .90. Note that we only count as inferences the cases which are discriminated, that is, in which one object has a positive cue value and the other does not. Thus the sum of all right and wrong inferences in the denominator is equal to the number of pairs of cities on which the cue discriminates. In the simulations below we compute the validity from the actual, ecological cue values. But when Take The Best is used as a model of human inference, the validities are computed only from the cue values the person actually knows (or believes).

For instance, when inferring whether Chicago or New Orleans has a higher homelessness rate, Take The Best looks up first the cue values of the two cities for rent control, since it is the cue with the highest validity (.90). Unfortunately, this cue does not discriminate—both cities have cue values of zero (Table 1). So Take The Best looks up the second best cue, the vacancy rate cue (validity .73). This cue does discriminate, so search is stopped. Take The Best infers that Chicago, the city with the unitary cue value in contrast to New Orleans' zero, has the higher homelessness rate.

Take The Best and Minimalist are constructed from several building blocks of fast and frugal heuristics (Gigerenzer & Goldstein, 1996). These building blocks help us both in understanding the heuristics and in generating new heuristics.

The first building block is *step-by-step procedures,* that is, a cognitive strategy that searches for some information and checks whether this is sufficient to make a decision; if not, it searches for more information, checks whether this is sufficient, and so on (e.g., Miller, Galanter, & Pribram, 1960).

The second building block is *simple stopping rules,* which specify computationally simple conditions for halting the gathering of more cue information. There are a number of heuristics which use stopping rules, especially those that already use "attribute-based" rather than "alternative-based" information gathering (to use the terminology of Payne, Bettman, & Johnson, 1988). In the constrained maximization paradigm, for example, information search is halted when the marginal cost of another piece of information outweighs the marginal gain in accuracy expected. But calculating these marginals is a difficult game. In contrast, we propose stopping rules that do not need such cost-benefit computations. Take The Best and Minimalist stop gathering further cue information if one object has a unitary ("1") value for a cue and the other does not (i.e., has a zero, "0," or unknown value for that cue).

This simple stopping rule is in harmony with our third building block, *one-reason decision making.* Once search is stopped, a variety of computations could be performed on the information collected thus far. For example, multiple regression integrates all the cue values in a linear sum, and Bayesian models usually condition their probabilities on the values of several cues. But since Minimalist and Take The Best stop after the first piece of information that discriminates between the two objects, they base their decision only on this recent information, the last cue considered. Trade-offs between cues never surface. The vision behind such one-reason decision making is to avoid conflicts and avoid integrating information. Thus, the process underlying decisions is non-compensatory. Note that one-reason decision making could be employed with less simple stopping rules, such as gathering a larger number of cues (e.g., in a situation where one has to justify one's decision); the decision, however, is based on only one cue.

To summarize, Minimalist and Take The Best employ the following building blocks:

- Step-by-step procedures
- Search limited by simple stopping rules
- One-reason decision making

In the following sections we will see how these building blocks exploit certain structures of environments. We will not deal here with how they exploit a lack of knowledge (see Gigerenzer & Goldstein, 1996).

Some of these building blocks appear in other heuristics, which are related to Take The Best. Lexicographic strategies (e.g., Keeney & Raiffa, 1993; Payne, Bettman, & Johnson, 1993) are very close to Take The Best, but not Minimalist. The term "lexicographic" signifies that cues are looked up in a fixed order and the first discriminating cue makes the decision, like in the alphabetic ordering used to decide which of two words comes first in a dictionary. Take The Best can exploit a lack of knowledge by means of the recognition heuristic (when there is only limited knowledge, a case dealt with in Gigerenzer & Goldstein, 1996). A more distantly related strategy is Elimination By Aspects (Tversky, 1972), which is also an

attribute-based information processor and also has a stopping rule. Elimination By Aspects (EBA) differs from Take The Best in several respects; for instance, EBA chooses cues not according to the order of their validities but by another probabilistic criterion, and it deals with preference rather than inference. Another related strategy is classification and regression trees (CART), which deals with classification and estimation rather than two-alternative prediction tasks. The key difference is that in CARTs, heavy computation and optimizing are used to determine the trees and the stopping rules.

In Section 1 we have defined two fast and frugal heuristics. These heuristics violate two maxims of rational reasoning: they do not search for all available information and they do not integrate information. Thus Minimalist and Take The Best are fast and frugal, but at what price? How much more accurate are benchmark models that use and integrate all information when predicting unknown aspects of real environments?

This question was posed by Gigerenzer and Goldstein (1996), who studied the price of frugality in inferring city populations. The surprising result was that Take The Best made as many accurate inferences as standard linear models, including multiple regression, which uses both more computational power and more information. Minimalist generated only slightly fewer accurate inferences. In Section 2 we test whether these results generalize to other real-world environments and to situations in which the training set and the test set are different. For simplicity, we will only study the performance of the heuristics under complete knowledge of cue values, whereas Gigerenzer and Goldstein (1996) varied the degree of limited knowledge. In Section 3, we analyze the structure of information in real-world environments that fast and frugal heuristics can exploit, that is, their ecological rationality. Finally, in Section 4, we take up Ward Edwards' challenge to compare the performance of fast and frugal heuristics with more powerful strategies than multiple regression, in particular with Bayesian models.

2. Fast and frugal at what price?

Some psychologists propose multiple linear regression as a description of human judgment; others argue that it is too complex a model for humans to instinctively perform. Nevertheless, both camps often regard it as an approximation of the optimal strategy people should use, Bayesian models aside. A more psychologically plausible version of a linear strategy employs unit weights (rather than beta weights), as suggested by Robyn Dawes (e.g., 1979). This heuristic adds up the number of unitary ("1") cue values and subtracts the number of zero ("0") cue values. Thus it is fast (it does not involve much computation), but not frugal (it looks up all cues). For simplicity, we call this heuristic *Dawes' Rule.*[1]

In this section, we will compare the performance of fast and frugal heuristics against these standard linear models. We begin by describing a single task in detail: to predict which U.S. cities have higher homelessness rates. Thereafter, we present the full data—the average results of the contests in 20 empirical data sets. But performance isn't everything—we also want to know what price we must pay for our accuracy. For example, heuristic *A* might need twice as many cue values as heuristic *B* in order to make its inferences, but might be only a few percentage points more accurate. We will determine these accuracy-effort trade-offs for our heuristics, using measures of computational simplicity and frugality of cue use.

Predicting homelessness

The first contest deals with a problem prevalent in many cities, homelessness, and we challenge our heuristics to predict which cities have higher homelessness rates. As mentioned above, the data stem from an article by William Tucker (1987) exploring the causes of homelessness. He presents data for six possible factors for homelessness in 50 U.S. cities. Some possible factors have an obvious relationship to homelessness because they affect the ability of citizens to pay for housing, such as the unemployment rate and the percentage of inhabitants below the poverty line. Other possible causes affect the ability to find housing, such as high vacancy rates. When many apartments are vacant, tenants have more options of what to rent and landlords are forced to lower rents in order to get any tenants at all. Rent control is also believed to affect ability to find housing. It is usually instituted to make housing more affordable, but many economists believe landlords would rather have no rent than low rent. Thus less housing is available for rent and more people must live on the streets. The percentage of public housing also affects the ability to find housing because more public housing means that more cheap housing options are available. Finally, one possible cause does not relate directly to the landlord-tenant relationship. Average temperature in a city can affect how tolerable it is to sleep outside, leading to a number of possible effects, all of which suggest that warmer cities will have higher homelessness rates; warmer cities might attract the homeless from cooler cities, landlords might feel less guilty about throwing people out in warmer cities, and tenants might fight less adamantly against being thrown out in more tolerable climates.

We will ask our heuristics to use these six (dichotomized) cues to predict homelessness rates in the 50 cities.[2] The heuristics will be required to choose the city with more homelessness for all $50 \times 49/2 = 1225$ pairs of cities. Regression will use the matrix of cue values to derive optimal weightings of the cues (possible causes).[3] There will be two types of competitions. In the first competition, the test set is the same as the training set (from which a strategy learns the information it needs, such as the weights of the cues). In the second, more realistic competition, the test set is different from the training set (also known as cross-validation). The second competition can reveal how much a heuristic overfits the data. Only the first type of competition was studied in Gigerenzer and Goldstein (1996).

Performance: Test set = Training set

We begin with the case of learning the entire data set and trying to fit it as well as possible. Performance is measured by the percentage of the 1225 inferences that are correct (which city has higher homelessness?). Sometimes the heuristics must guess, for example between New Orleans and Miami, which have the same characteristics on the six cues (both are one on temperature, both are zero on rent control, both are one on poverty, etc.). When a heuristic guesses, it earns a score of 0.5 correct, on the grounds that half the time the heuristic will be correct in its guess.

Strategy	Average number of cues looked up	% correct when test set same as training set	% correct when test set different from training set
Minimalist	2.1	61	56
Take The Best	2.4	69	63
Dawes' Rule	6	66	58
Multiple regression	6	70	61

Table 2. Trade-off between accuracy and cues looked up in predicting homelessness, for two kinds of competition (test set = training set, and test set ≠ training set). The average number of cues looked up was about the same for both kinds of competition.

How well do the heuristics predict homelessness? Table 2 shows the results for the situation when the test set coincides with the training set. There are two surprises in these numbers. The first is that Take The Best, which uses only 2.4 cues on average, scores higher than Dawes' Rule, which uses all 6 of them. The second surprise is that Take The Best is almost as good as linear regression, which not only looks up all the cues but performs complicated calculations on them. So it seems that fast and frugal heuristics can be about as accurate as the more computationally expensive multiple regression! This confirms the findings of Gigerenzer and Goldstein (1996) in a task of predicting city populations.

Although Take The Best's accuracy is very close to that of regression, its absolute value does not seem to be very high. What is the upper limit on performance? The upper limit is not 100%, but 82%. This would be obtained by an individual who could memorize the whole table of cue values and, for each pair of cue profiles, memorize which one has the higher homelessness rate (but for the purpose of the test forgets the city names). If a pair of cue profiles appears more than once, this Profile Memorization Method goes for the profile that leads to the right answer more often.[4] The Profile Memorization Method results in 82% correct inferences for the homelessness data (see Section 4).

Performance: Test set ≠ Training set

The prediction task we have considered thus far is limited to static situations, when we are merely trying to "fit" a phenomenon about which we already have all information. How well do the heuristics perform if the test set is different from the training set? This situation is a version of one-step learning and prediction. The data set is broken into two halves, with random assignment of cities to one or the other half. The heuristics are allowed to use one half to build their models (calculate regression weights, get cue orders, determine cue direction); then they must make predictions on the other half, using the parameters estimated on the first half, and their accuracy is scored. This process is repeated 1000 times, with 1000 random ways of breaking the data into two halves in order to average out any particularly helpful or hurtful ways of halving the data.

Training might not seem to affect Dawes' Rule and Minimalist, but in fact it does. Both strategies use the first half of the data set to estimate the direction of the cue (whether a higher or a lower cue value signals a higher criterion value). When the test set was different from the training set, the performance of Minimalist dropped from 61% correct to 56%, and that of Dawes' Rule from 66% to 58% (Table 2). Take The Best needs to learn more than merely the direction of the cues; it must also learn the order of the cue validities. With this slight additional

knowledge, Take The Best scores 63% correct, down from 69%. Regression needs to learn not only the direction of the cues but also their interrelationship in order to determine the best linear weighting scheme. Despite all this knowledge, regression's performance falls more than that of Take The Best. While regression scored 70% correct when it merely had to fit the data, it scores only 61% correct in the cross-validated case, falling to second place.

In summary, when heuristics built their models on half of the data and inferred on the other half, Take The Best was the most accurate strategy for predicting homelessness, followed closely by regression. This seems counterintuitive since Take The Best looks up only 2.4 of the 6 cues and (as we will soon see) is computationally simpler.

Note that we no longer determine the upper limit by the Profile Memorization Method. This method cannot be used if cue profiles that were not present in the first half are present in the second half.

Generalization[5]

How well do these results generalize to making predictions about other environments? We now consider results across 20 data sets. These data sets have real-world structure in them rather than artificial, multivariate normal structures. In order to make our conclusions as robust as possible, we also tried to choose as wide a range of empirical environments as possible. So they range from having 17 objects to 395 objects, and from 3 cues (the minimum to distinguish between the heuristics) to 19 cues. Their content ranges from social topics like high school drop-out rates, to chemical ones such as the amount of oxidant produced from a reaction, to biological ones like the number of eggs found in individual fish.

Strategy	Average number of cues looked up	% correct when test set same as training set	% correct when test set different from training set
Minimalist	2.2	70	65
Take The Best	2.4	76	71
Dawes' Rule	7.4	73	70
Multiple regression	7.4	78	67

Table 3. Trade-off between accuracy and cues looked up, averaged across 20 data sets, for two kinds of competitions (test set = training set, and test set ≠ training set). The average number of cues looked up was about the same for both kinds of competition.

Table 3 shows the performance of the heuristics averaged across the 20 data sets. When the task is merely fitting the given data (test set same as training set), multiple linear regression is the most accurate strategy, by two percentage points, followed by Take The Best. But when the task is to generalize from a training set to a test set, Take The Best is the most accurate. It outperforms multiple regression by four percentage points. Note that multiple regression has all the information Take The Best uses, and more. Dawes' Rule lives up to its reputation for robustness in the literature (Dawes, 1979) by taking second place and beating regression by three percentage points. Finally, Minimalist performs surprisingly well, only two percentage points behind regression. In short, Dawes' Rule is not the only robust yet simple model; Take The Best and Minimalist are also fairly accurate and robust under a broad variety of conditions. In fact, Take the Best is

even slightly more accurate than Dawes' Rule although it is more frugal. In Section 3, we will explore how this is possible—how fast and frugal heuristics can also be accurate.

How much information processing is performed?

We established empirically that Take The Best and Minimalist are frugal—on average, they stopped searching and made a prediction after having looked up fewer than one-third of the cues. But are the heuristics also fast, that is, simple in their computations? Given that Take The Best performs so well, it must be doing some work, perhaps hidden in the training phase of the cross-validation if not in the test phase. Thus, we now wish to be more precise about measuring how fast (computationally simple) our heuristics are, both in the training and test phase.

Let us begin by measuring the amount of learning required by the heuristics to build their models in order to perform their predictions later. We can use the suggestion of Newell and Simon (1972) and Payne, Bettman, and Johnson (1990) to count the number of elementary information processing (EIP) units necessary for the training phase. These EIPs include addition, subtraction, multiplication, division, comparison of two numbers, reading a number, writing a number, and so on. For each such operation, we count one unit. These elementary processing units are easy to count, and Payne, Bettman, and Johnson (1990) present experimental evidence that they are a reasonable estimate of the cognitive effort involved in executing a particular choice strategy in a specific task environment. For our tasks, the number of EIPs required depends on N, the number of objects, and M, the number of cues in the data set. Table 4 specifies both the approximate number of EIPs used, for any values of N and M that a data set has, and the number of EIPs for the specific case of predicting homelessness, with $N = 50$ and $M = 6$.

Strategy	Knowledge about cues obtained in training phase	Approximate number of EIPs used in training phase for any N, M	Number of EIPs in training for homelessness ($N = 50$, $M = 6$)
Minimalist	Direction	$\approx 10NM$	3,398
Take The Best	Direction + order	$\approx 10NM$	3,448
Dawes' Rule	Direction	$\approx 10NM$	3,398
Multiple regression	Beta weights	$\approx 10NM^2$	20,020

Table 4. Approximate number of Elementary Information Processing units (EIPs) needed for the training phase of each strategy. The task is predicting homelessness. N = number of objects; M = number of cues (for details see Czerlinski, 1997).

Fast and frugal heuristics require significantly less calculation in the training phase than multiple regression. This is the case even though in calculating the number of EIPs in regression, we neglected the usual invertibility and computer overflow checks, so 20,000 EIPs is really a lower bound. In practical applications, fast and frugal heuristics might be as much as 1/100 simpler. Note that we differ from earlier theorists such as Dawes (1979) in including learning the direction of cues as a real problem; other theorists have assumed this is known already, making fast and frugal heuristics even simpler.

Strategy	Process of inference	Number of EIPs used in test phase for any N, M, M_a	Number of EIPs in inferring homelessness ($N = 50$, $M = 6$)
Minimalist	Search through cues randomly until decision possible	$3M_a$	6.2 ($M_a = 2.1$)
Take The Best	Search through ordered cues until decision possible	$3M_a$	7.2 ($M_a = 2.4$)
Dawes' Rule	Count number of unitary and zero cue values; compare	$8M-7$	41
Multiple regression	Linearly use beta weights to estimate criterion; compare	$16M-7$	89

Table 5. Number of Elementary Information Processing units (EIPs) needed for the test phase of each strategy. The task is predicting homelessness. N = number of objects; M = number of cues; M_a = average number of cues used (for details see Czerlinski, 1997).

Of course, learning a model of the data is only the first step. Implementing the heuristic has a cost, too. Table 5 specifies the number of EIPs in the test phase. Fast and frugal heuristics are always at least as efficient as the others because they look up fewer cues and perform fewer calculations on those cues. Since fast and frugal heuristics generally do not use all of the available cues, we also need to consider the "actual" number of cues looked up, M_a. For example, Take The Best uses on average only 2.4 cues for predicting homelessness.

Table 5 clearly shows that the cue-based predictions of Minimalist and Take The Best are highly efficient, about 5 times simpler than the simplest linear model, Dawes' Rule, and about 10 times simpler than multiple regression. We now have a measure of how "fast" (computationally simple) the heuristics are, and we have shown that fast and frugal heuristics can be from 5 to 10 times faster theoretically than regression (and practically even more). The calculation of EIPs does not have to assume serial processing; if the brain implements certain aspects of the calculation in parallel, then the total number of calculations would be the same, but they would be completed more quickly. For example, if we could compute the validity of all cues in parallel, we would effectively have $M = 1$, and this could be plugged in to the formulae above. However, even under such conditions, fast heuristics could not be slower than regression and could still be faster, just not as much faster as they are under the assumption of serial processing. And, of course, they would still be more frugal.

In summary, our fast and frugal heuristics learn with less information, perform fewer computations while learning, look up less information in the test phase, and perform fewer computations when predicting. Nevertheless, fast and frugal heuristics can be almost as accurate as multiple regression when fitting data. Even more counterintuitively, one of these fast and frugal heuristics, Take The Best, was on average more accurate than regression in the more realistic situation where the training set and test set were not the same (cross-validation). How is this possible?

3. Ecological rationality: Why and when are fast and frugal heuristics good?

Note first that these data sets have been collected from "real-world" situations. What are the characteristics of information in real-world environments that make Take The Best a better predictor than other strategies, and where will it fail? When we talk of properties of information, we mean the information about an environment known to a decision maker. We discuss three properties. The first of these properties is one that characterizes many real-world situations: the available information is *scarce.* Take The Best is smarter than its competitors when information is scarce.

Scarce information

In order to illustrate the concept of scarce information, let us recall an important fact from information theory: a class of N objects contains log$N$ bits of information. This means that if we were to encode each object in the class by means of binary cue profiles of the same length, this length should be at least log$N$ if each object is to have a unique profile. The example in Table 6 illustrates this relation for $N = 8$ objects. The eight objects are perfectly predictable by the three (log8 = 3) binary cues. If there were only two cues, perfect predictability could simply not be achieved.

Theorem: If the number of cues is less than log$N$ the Profile Memorization Method will never achieve 100% correct inferences. Thus no other strategy will either.

This theorem motivates the following definition:

Definition: A set of M cues provides *scarce* information for a reference class of N objects if $M \leq \log N$.

Objects	First cue	Second cue	Third cue
A	1	1	1
B	1	1	0
C	1	0	1
D	1	0	0
E	0	1	1
F	0	1	0
G	0	0	1
H	0	0	0

Table 6. Illustration of the fact that 8 objects can be perfectly predicted by log8 = 3 binary cues.

We can now formulate a theorem that relates the performance of Take The Best to that of Dawes' Rule.

Theorem: In the case of scarce information, Take The Best is on average more accurate than Dawes' Rule.

The proof is in Martignon, Hoffrage, and Kriegeskorte (1997). The phrase "on average" means across all possible environments, that is, all combinations of binary entries for $N \times M$ matrices. The intuition underlying the theorem is the following: in scarce environments, Dawes' Rule can take little advantage of its strongest property, namely, compensation. If in a scarce environment cues are *redundant,* that is, if a subset of these cues does not add new information, things will be even worse

for Dawes' Rule. Take The Best suffers less from redundancy because decisions are taken at a very early stage.

Abundant information

Adding cues to a scarce environment will do little for Take The Best, if the best cues in the original environment are already highly valid, but it may compensate for various mistakes Dawes' Rule would have made based on the first cues. In fact, by adding and adding cues we can make Dawes' Rule achieve perfection. This is true even if all cues are favorable (i.e., their validity is > 0.5) but uncertain (i.e., their validity is < 1).

Theorem: Assume that the environment consists of $N \geq 5$ objects. If an environment consists of all possible uncertain but favorable cues, Dawes' Rule will discriminate among all objects and make only correct inferences.

The proof is given in Martignon et al. (1997). Note that we are using the term cue to denote a binary-valued function on the reference class. Therefore, the number of different cues on a finite reference class is finite. The theorem can be generalized to linear models that use cue validities as weights rather than unit weights. As a consequence, Take The Best will be outperformed on average by linear models in abundant environments.

Non-compensatory information

Environments may be compensatory or non-compensatory. Among the 20 environments studied in Section 2, we found 4 in which the weights for the linear models were non-compensatory (i.e., each weight is larger than the sum of all other weights to come, such as 1/2, 1/4, 1/8, ...). The following theorem states an important property of non-compensatory models and is easily proved (Martignon et al., 1997).

Theorem: Take The Best is equivalent—in performance—to a weighted linear model whose weights form a non-compensatory set.

If multiple regression happens to have a non-compensatory set of weights (where the order of this set corresponds to the order of cue validities), then its accuracy is equivalent to Take The Best.

Why is Take The Best so robust?

The answer is simple: Take The Best uses few cues (only 2.4 cues on average in the data sets presented here). Thus its performance depends on very few parameters. The top cues usually have high validity. In general, highly valid cues will remain highly valid across different subsets of the same class of objects. Even the order of their cue validities tends to be fairly stable. The stability of highly valid cues is a main factor for the robustness of Take The Best, in cross-validation as well as in other forms of incremental learning.

Strategies that use all cues must estimate a number of parameters larger than or equal to the number of cues. Some, like multiple regression, use a huge number of parameters. Thus they suffer from overfitting, in particular with small data sets.

To conclude, scarceness and redundancy of information are characteristics of information gathered by humans. Humans are not always good at finding large numbers of cues for making predictions. The magic number 7 ± 2 seems to represent the basic information capacity human minds work with in a short time interval. Further, humans are not always good at detecting redundancies between

cues, and quantitatively estimating the degree of these redundancies. Fast and frugal Take The Best is a heuristic that works well with scarce information and does not even try to estimate redundancies and cue intercorrelations. In this way, it compensates for the limits in human information processing. If the structure of the information available to an organism is scarce or non-compensatory, then Take The Best will be not only fast and frugal, but also fairly accurate, even relative to more computationally expensive strategies.

4. How does Take The Best compare with good Bayesian models?

It happened that Ward Edwards was a reviewer of one of our group's first papers on fast and frugal heuristics (Goldstein & Gigerenzer, 1996). Ward sent us a personal copy of his review, as he always does. No surprise, his first point was "specify how a truly optimal Bayesian model would operate." But Ward did not tell us which Bayesian model of the task (to predict the populations of cities) he would consider truly optimal.

In this section, we present a possible Bayesian approach to the type of task discussed in the previous sections. We do not see Bayesian models and fast and frugal heuristics as incompatible, or even opposed. On the contrary, considering the computational complexity Bayesian models require, and the fact (as we will see) that fast and frugal heuristics do not fall too far behind in accuracy, one can be a satisficer when one has limited time and knowledge, and a Bayesian when one is in no hurry and has a computer at hand. A Bayesian can decide when it is safe and profitable to be a satisficer.

Bayesian networks[6]

If training set and test set coincide, the Bayesian knows what she will do: she will use the Profile Memorization Method if she has perfect memory. If training set and test set are different the Bayesian has to construct a good model. Regression is not necessarily the first model that would come to her mind. Given the kind of task, she may tend to choose from the flexible family of Bayesian networks. Another possibility is a Bayesian CART and a third is a mixture of these two.

The task is to infer which of two objects A or B scores higher on a criterion, based on the values of a set of binary cues. Assume, furthermore, that the decision maker has nine cues at her disposal and she has full knowledge of the values these cues take on A and B. To work out a concrete example, let A and B have the cue profiles (100101010) and (011000011) respectively. The Bayesian asks herself: What is the probability that an object with cue profile (100101010) scores higher than an object with cue profile (011000011) on the established criterion? In symbols:

$$\text{Prob}(A > B \mid A \cong (100101010),\ B \cong (011000011)) = ? \qquad (*)$$

Here the symbol $\cong$ is used to signify "has the cue profile." As a concrete example, let us discuss the task investigated in Gigerenzer and Goldstein (1996), where pairs of German cities were compared as to which had a larger population. There were nine cues: "Is the city the national capital?" (NC); "Is the city a state capital?" (SC); "Does the city have a soccer team in the major national league?" (SO); "Was the city once an exposition site?" (EX); "Is the city on the Intercity train line?" (IT); "Is the abbreviation of the city on the license plate only one letter long?"

(LP); "Is the city home to a university?" (UN); "Is the city in the industrial belt?" (IB); "Is the city in former West Germany?" (WG).

A network for our type of task considers pairs of objects (A,B) and the possible states of the cues, which are the four pairs of binary values (0,0), (0,1), (1,0), (1,1) on pairs of objects. A very simple Bayesian network would neglect all interdependencies between cues. This is known as Idiot Bayes. It computes (*) from the product of the different probabilities of success of all cues. Forced to a deterministic answer, Idiot Bayes will predict that A scores higher than B on the criterion, if the probability of *"A larger than B"* computed in terms of this product is larger than the probability of *"B larger than A."* Due to its simplicity, Idiot Bayes is sometimes used as a crude estimate of probability distributions. This is not the procedure the Bayesian will use if she wants accuracy.

The other extreme in the family of Bayesian networks is the fully connected network, where each pair of nodes is connected both ways. Computing (*) in terms of this network when training and test set coincide amounts to using the Profile Memorization Method. Both these extremes, namely Idiot Bayes and the fully connected network are far from being optimal when training set and test set differ. A more accurate Bayesian network has to concentrate on the important conditional dependencies between cues, as some dependencies are more relevant than others. Some may be so weak that it is convenient to neglect them, in order to avoid overfitting. The Bayesian needs a Bayesian strategy to decide which are the relevant links that should remain and to prune all the irrelevant ones. She needs a strategy to search through the possible networks and evaluate each network in terms of its performance. Bayesian techniques for performing this type of search in a smart, efficient way have been developed both in statistics and artificial intelligence. These methods are efficient in learning both structure and parameters. Nir Friedman and Leo Goldszmit (1996), for instance, have devised software[7] for searching over networks and finding a good fit for a given set of data in a classification task. Since our task is basically a classification task (we are determining whether a pair of objects is rightly ordered or not), we are able to make use of Friedman and Goldszmit's network. But a smart Bayesian network is often too complex to be computed. The following theorem offers a way to reduce the huge number of computations that would be, at first glance, necessary for computing (*) based on a Bayesian network. In a Bayesian network the nodes with arrows pointing to a fixed node are called the *parents* of that node. The node itself is called a *child* of its parents. What follows is a fundamental rule for operating with Bayesian networks.

Theorem: The conditional probability of a node j being in a certain state given knowledge on the state of all other nodes in the network is proportional to the product of the conditional probability of the node given its parents times the conditional probability of each of its children given its parents.

In symbols:

$$\text{Prob(node } j \mid \text{other nodes)} =$$
$$K \times \text{Prob(node } j \mid \text{parents of } j) \times \prod \text{Prob(child } k \text{ of } j \mid \text{parents of } k) .$$

Here K is a normalizing constant. The set consisting of a node, its parents, its children and the other parents of its children is called the *Markov Blanket* of that node. What the theorem states is that the Markov Blanket of a node determines the state of the node regardless of the state of all other nodes not in the Blanket.

The theorem just stated, based essentially on Bayes' rule, represents an enormous computational reduction in the calculation of probability distributions. It is precisely due to this type of reduction of computational complexity that Bayesian networks have become a popular tool both in statistics and in artificial intelligence in the last decade.

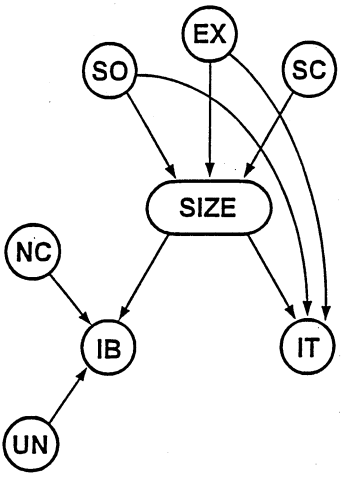

Figure 2. A Bayesian network for predicting population size (which of two German cities *A* or *B* is larger). The cues are SO = soccer team; EX = exposition site; SC = state capital; IB = industrial belt; NC = national capital; UN = university; IT = intercity train.

Figure 2 shows the Bayesian network obtained with Friedman's search method, for the task of comparing German cities according to their population, as in Gigerenzer and Goldstein (1996). In that paper, the reference class of the 83 German cities with more than 100,000 inhabitants was analyzed. The Bayesian network reveals that two of the nine cues, LP and WG, are irrelevant when the other seven cues are known. Figure 2 illustrates the Markov Blanket of the node size, which represents the hypothesis "city *A* has more inhabitants than city *B*" and obviously can be in two states (the other state is "city *B* has more inhabitants than city *A*"). According to the theorem specified above:

$$\text{Prob(size} \mid \text{UN, NC, IB, SO, EX, SC, IT)} = K \times \text{Prob(size} \mid \text{SO, EX, SC)} \times$$
$$\text{Prob(IB} \mid \text{size, UN, NC)} \times \text{Prob(IT} \mid \text{size),}$$

where K is a constant. In order to determine each of the probabilities on the right-hand side of the equation the program produces simple decision trees (actually CARTs), as illustrated in Figure 3 for Prob(size | SO, EX, SC). The program searches among all possible trees for the one that fits the data best, pruning all irrelevant branches. That is, this approach combines a Bayesian network with a CART step at the end. CART models were popularized in the statistical community by the seminal book by Breiman, Friedman, Olshen, and Stone (1984).

This method, a mixture of a Bayesian network and CART, is much more computationally intensive than multiple regression, not to speak of Take The Best.

In fact, if we were to compute its EIPs as we did in the previous section, we would clearly reach a function of M and N containing an exponential term in M.

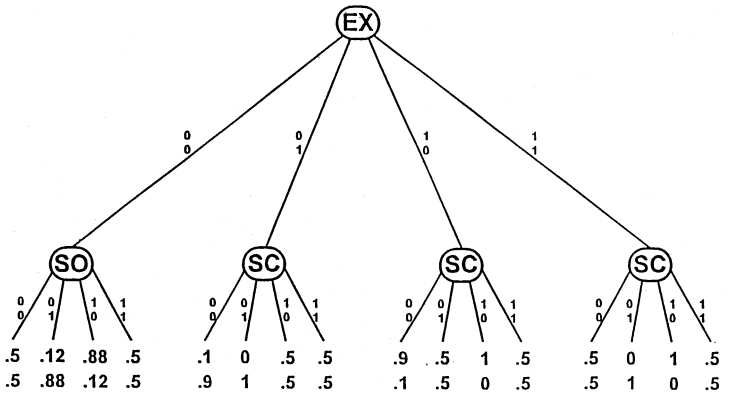

Figure 3. CART (*C*lassification *A*nd *R*egression *T*ree) for quick computation of Prob(size | SO, EX, SC). For instance, if neither of the two cities A and B is an exposition site (symbolized by the two zeros in the left branch), then the only relevant cue is SO, that is, whether a city has a soccer team in the major league (SC is irrelevant). If A has a soccer team but B does not ("1" and "0"), then Prob($A > B$ | SO, EX, SC) = .88, and Prob($A < B$ | SO, EX, SC) = .12. "$A > B$" stands for "A has a larger population than B."

How much more accurate is such a computationally complex Bayesian network than the simple Take The Best? Table 7 shows the performance of the Bayesian network and the Profile Memorization Method (the upper limit) when training and test set coincide. Performance was tested in five environments: Which of two German cities has the higher population? Which of two U.S. cities has a higher homelessness rate? Which of two individual Arctic female charr fish produces more eggs? Which of two professors at a Midwestern college has a higher salary?

	Take The Best	Multiple regression	Bayesian network	Profile Memorization Method
City population	74	74	76	80
Homelessness	69	70	77	82
Fish fertility	73	75	75	75
Professors' salaries	80	83	84	87

Table 7. Percentage of correct inferences when test set = training set.

For predicting city populations, the Bayesian network gets 2 percentage points more correct answers than Take The Best. The upper limit of correct predictions can be computed by the Profile Memorization Method as 80%, which is four percentage points above the performance of the Bayesian network. When the test set is different from the training set (Table 8), then multiple regression takes a slightly larger loss than Take The Best and the Bayesian network. Recall that the upper limit cannot be calculated by the Profile Memorization Method when the test set is different from the training set.

	Take The Best	Multiple regression	Bayesian network
City population	72	71	74
Homelessness	63	61	65
Fish fertility	73	75	75
Professors' salaries	80	80	81

Table 8. Percentage of correct inferences when test set is different from training set (cross-validation).

When predicting homelessness, the Bayesian network performs 8 percentage points better than Take The Best (Table 7). This difference is reduced to 2 percentage points when the test set is different from the training set (Table 8). Here, Take The Best is the most robust heuristic under cross-validation.

The fish fertility data set is of particular interest, because it contains a large set of objects (395 individual fish). The cues for the criterion (numbers of eggs found in a given fish) were weight of fish, age of fish, and average weight of her eggs. Here, as one would expect for a reasonably large data set, all results are quite stable when one cross validates.

The next problem is to predict which of two professors at a Midwestern college has a higher salary. The cues are gender, his or her current rank, the number of years in current rank, the highest degree earned, and the number of years since highest degree earned. When the test set is the same as the training set, Take The Best makes 4 percentage points fewer accurate inferences than the Bayesian network. However, when the test set is different from the training set, then Take The Best almost matches the Bayesian network.

Across these four environments, the following generalizations emerge:

1. When the test set is the same as the training set, the Bayesian network is considerably more accurate than Take The Best. On average, it was only 3 points behind the Profile Memorization Method, which attains maximal accuracy. However, when the test set is different from the training set, the accuracy of Take The Best is, on average, only 1 to 2 percentage points less than that of the Bayesian network. This result is noteworthy given the simplicity and frugality of Take The Best compared with the computational complexity of the Bayesian network.

2. Take The Best is more robust—measured in loss of accuracy from Table 7 to Table 8—than both multiple regression and the Bayesian network.

What is extraordinary about fast and frugal Take The Best is that it does not fall too far behind the complex Bayesian network. And it can easily compete in 20 different environments (Section 2) with Dawes' Rule and multiple regression.

5. Conclusions

L. J. Savage wrote that the only decision we have to make in our lives is how to live our lives (1954, p. 83). But "how to live our lives" means basically "how to make decisions." Are we going to adopt Bayesian decision making or use some simple heuristics, like the satisficing ones presented in this chapter? This might not be an exclusive "or": fast and frugal heuristics can have their place in everyday affairs where time is limited and knowledge is scarce, and Bayesian tools can be the choice for someone who is in no hurry and has a computer in her bag (von Winterfeldt & Edwards, 1986). A Bayesian who tries to maximize under deliberation constraints must choose a strategy under a combination of criteria, such

as computational cost, frugality, accuracy, and perhaps even transparency. Thus, it may happen that a Bayesian herself may choose Take The Best, or another fast and frugal heuristic, over expensive but less robust Bayesian networks in some situations. Bayesian reasoning itself may tell us when to satisfice.

The major results summarized in this chapter are the following. First, across 20 real-world environments, the fast and frugal Take The Best outperformed multiple regression in situations with learning (test set ≠ training set), while even the simpler Minimalist came within 2 percentage points of it. Second, we specified which characteristics of information in real-world environments enable Take The Best to match or outperform linear models. Third, we showed that sophisticated Bayesian networks were only slightly more accurate than Take The Best.

The results reported in this chapter were obtained with real-world data but must be evaluated with respect to the conditions used, which include the following. First, we studied inferences only under complete knowledge, unlike Gigerenzer and Goldstein (1996), who studied the performance of heuristics under limited knowledge. Limited knowledge (e.g., knowing only a fraction of all cue values) is a realistic condition that applies to many situations in which predictions must be made. In the simulations reported by Gigerenzer and Goldstein, the major result was that the more limited the knowledge, the smaller the discrepancy between Minimalist and other heuristics becomes. Thus Minimalist, whose respectable scores were nevertheless always the lowest, really flourishes when there is only limited knowledge. Gigerenzer and Goldstein (1996) also develop circumstances under which the counterintuitive less-is-more effect is possible: when knowing less information can lead to better performance than knowing more information.

Other conditions of the studies reported here include the use of binary and dichotomized data, which can be a disadvantage to multiple regression and Bayesian networks. Finally, we have used only correct data, and not studied predictions under the realistic assumption that some of the information is wrong.

Some of the results obtained are reminiscent of the phenomenon of flat maxima. If many sets of weights, even unit weights, can perform about as well as the optimal set of weights in a linear model, this is called a flat maximum. The work by Dawes and others (e.g., Dawes & Corrigan, 1974) made this phenomenon known to decision researchers, but it is actually much older (see John, Edwards, & von Winterfeldt, n.d.). The performance of fast and frugal heuristics in some of the environments indicates that a flat maximum can extend beyond the issue of weights: inferences based solely on the best cue can be as accurate as those based on any weighted linear combination of all cues. The results in Section 3, in particular the theorem on non-compensatory information, explain conditions under which we can predict flat maxima.

The success of fast and frugal heuristics emphasizes the importance of studying the structure of the information in the environment. Such a program is a Brunswikian program, but it is one that dispenses with multiple regression as *the* tool for describing both the processes of the mind and the structure of the environment. Fast and frugal heuristics can be ecologically rational in the sense that they exploit specific and possibly recurrent characteristics of the environment's structure (Tooby & Cosmides, in press). Models of reasonable judgment should look outside of the mind, to its environment. And models of reasonableness do not have to forsake accuracy for simplicity. The mind can have it both ways.

Acknowledgement

We know Ward Edwards as a poet of limericks, as the ghost writer who jazzes up the boring titles of our talks at the annual Bayesian meetings, as the rare reviewer who sends his reviews directly to the authors, and as the man who envisions the 21st century as "the century of Bayes." In research Ward has found a calling. Rather than promoting himself, he has chosen to promote the truth. For instance, he was strong enough to set the popular cognitive illusions program in motion and then jump off his own bandwagon, knowing that staying on it would have boosted his career. Ward is always willing to criticize his own thinking and reconsider his past views. The only possible exception is his dedicated Bayesianism. A great physicist once said of Max Planck: "You can certainly be of a different opinion from Planck's, but you can only doubt his upright, honorable character if you have none yourself." This statement could just as well have been made about Ward Edwards.

Footnotes

1. Dawes and Corrigan (1974) write, "The whole trick is to decide what variables to look at and then to know how to add" (p. 104). The problem of what variables to look at is, however, not defined; it is the job of the expert to determine both the cues and their directional relationship with the criterion (Dawes, 1979). In our simulations, we will use the full set of cues and simply calculate the actual direction of the cues (rather than asking an expert).

2. In contrast to Gigerenzer and Goldstein (1996), we always provide full information for the algorithms (no unknown cue values).

3. Note that if the optimal weight is negative, then regression says the cue points in the opposite direction from that indicated by the ones and zeros. This can happen because the ones and zeros are calculated for each cue independently while regression operates on all cues simultaneously, taking their interrelationship into account.

4. The Profile Memorization Method is essentially a Bayesian method. If there are several pairs of objects with the same one pair of cue profiles, the Profile Memorization Method looks at all such pairs and determines the frequency with which a city with the first cue profile has more homeless than a city with the second cue profile. This proportion is the probability that the first city scores higher on this criterion. If forced to give a deterministic answer, and if the penalty for incorrectly guessing city 1 is the same as the penalty for incorrectly guessing city 2, the method picks the object that has the highest probability of a high value on the criterion (e.g., a higher homelessness rate). Thus, in this situation the Bayesian becomes a frequentist making optimal use of every bit of information.

5. The results that follow are explained in detail in Czerlinski, Gigerenzer, and Goldstein (in press).

6. The results that follow are explained in detail in Martignon and Laskey (in press).

7. The software for this procedure has been kindly put to the disposition of Kathy Laskey and Laura Martignon by Nir Friedman.

References

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

Brunswik, E. (1964). Scope and aspects of the cognitive problem. In J. S. Bruner et al. (Eds.), *Contemporary approaches to cognition* (pp. 5–31). Cambridge, MA: Harvard University Press.

Czerlinski, J. (1997). *Algorithm calculation costs measured by EIPs.* Manuscript. Max Planck Institute for Psychological Research, Munich.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. (in press). Pushing fast and frugal heuristics to the limits. In G. Gigerenzer, P. Todd, & the ABC group, *Simple heuristics that make us smart.* New York: Oxford University Press.

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

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

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

Edwards, W. (1961). Behavioral decision theory. *Annual Review of Psychology, 12,* 473–498.

Friedman, N., & Goldszmit, L. (1996). A software for learning Bayesian networks. (Not released for public use.)

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

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

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

Gigerenzer, G., Todd, P., & the ABC group (in press). *Simple heuristics that make us smart.* New York: Oxford University Press.

Goldstein, D. G., & Gigerenzer, G. (1996). Satisficing inference and the perks of ignorance. In G. W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 137–141). Mahwah, NJ: Erlbaum.

John, R. S., Edwards, W., & Winterfeldt, D. von (n.d.). *Equal weights, flat maxima, and trivial decisions.* Research Report 80–2. Social Science Research Institute, University of Southern California.

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

Martignon, L., Hoffrage, U., & Kriegeskorte, N. (1997). *Lexicographic comparison under uncertainty: A satisficing cognitive algorithm.* Submitted for publication.

Martignon, L., & Laskey, K. (in press). Laplace's Demon meets Simon: The role of rationality in a world of bounded resources. In G. Gigerenzer, P. Todd, & the

ABC group, *Simple heuristics that make us smart.* New York: Oxford University Press.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart & Winston.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

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

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1990). The adaptive decision maker: Effort and accuracy in choice. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 129–153). Chicago: The University of Chicago Press.

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

Sargent, T. J. (1993). *Bounded rationality in macroeconomics.* Oxford: Clarendon Press.

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

Simon, H. A. (1956). Dynamic programming under uncertainty with a quadratic criterion function. *Econometrica, 24,* 19–33.

Simon, H. A. (1982). *Models of bounded rationality.* 2 vols. Cambridge, MA: MIT Press.

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

Tooby, J., & Cosmides, L. (in press). Ecological rationality and the multimodular mind. Grounding normative theories in adaptive problems. In J. Tooby & L. Cosmides, *Evolutionary psychology: Foundational papers*, Cambridge, MA: MIT Press.

Tucker, W. (1987). Where do the homeless come from? *National Review,* Sept. 25, pp. 34–44.

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

Winterfeldt, D. von, & Edwards, W. (1986). *Decision analysis and behavioral research.* Cambridge: Cambridge University Press.