

Gigerenzer, G., Todd, P.M., & the ABC Research Group. (1999). *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

CHAPTER 4: Betting on One Good Reason: Take The Best and Its Relatives

Gerd Gigerenzer and Daniel G. Goldstein

Bounded rationality is what cognitive psychology is all about. And the study of bounded rationality is not the study of optimization in relation to task environments.

Herbert A. Simon, 1991

God, as John Locke (1690/1959) asserted, “has afforded us only the twilight of probability; suitable, I presume, to the state of mediocrity and probationership he has been pleased to place us in here ...” In the two preceding chapters, we argued that humans can make the best of this mediocre uncertainty. Ignorance about real-world environments, luckily, is often systematically rather than randomly distributed and thus allows organisms to navigate through the twilight with the recognition heuristic. In this chapter, we analyze heuristics which draw inferences from information beyond mere recognition. The source of this information can be direct observation, recall from memory, first-hand experience, or rumor. Darwin (1872), for instance, 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 or not to fight (Krebs & Davies, 1991). Inferences about the world are typically based on cues that are uncertain indicators: the eyes can deceive, and so can a medium-sized ethologist mimicking a large toad with a deep croak in the darkness. As Benjamin Franklin remarked in a letter in 1789: “In this world nothing is certain but death and taxes.”

How do people make inferences, predictions, and decisions from a bundle of imperfect cues and signals? The classical view of rational judgment under uncertainty is illustrated by Benjamin Franklin’s moral algebra. In an often cited letter to the British scientist Joseph Priestley, Franklin (1772/1987) explained how to decide which of two options to take, based on uncertain cues (which he calls “reasons”):

... [M]y Way is, to divide half a Sheet of Paper by a Line into two Columns, writing over the one *Pro*, and over the other *Con*. Then during three or four Days Consideration I put down under the different Heads short Hints of the different Motives that at different Times occur to me for or against the Measure. When I have thus got them all together in one View, I endeavor to estimate their respective Weights; and where I find two, one on each side, that seem equal, I strike them both out: If I find a Reason *pro* equal to some two Reasons *con*, I strike out the three. If I judge some two Reasons *con* equal to some three Reasons *pro*, I strike out the five; and thus proceeding I find at length where the Ballance lies; and if after a Day or two of farther Consideration nothing new that is of Importance occurs on either side, I come to a Determination accordingly.

And tho' the Weight or Reasons cannot be taken with the Precision of Algebraic Quantities, yet when each is thus considered separately and comparatively, and the whole lies before me, I think I can judge better, and am less likely to make a rash Step; and in fact I have found great Advantage from this kind of Equation, in what may be called *Moral* or *Prudential Algebra*.

Franklin's moral algebra, or what we will call Franklin's rule, is to search for all reasons, positive or negative, weigh each carefully, and add them up to see where the balance lies. This linear combination of reasons carries the moral sentiment of rational behavior: carefully look up every bit of information, weigh each bit in your hand, and combine them into a judgment. Franklin's method is a variant of the classical view of rationality which emerged in the Enlightenment (see Chapter 1), a view that is not bound to linear combinations of reasons. Classical rationality assumes that the laws of probability are the laws of human minds, at least of the educated ones (the *hommes éclairés*, see Daston, 1988). As Pierre Simon de Laplace (1814/1951, p. 196) put it, probability theory is "nothing more at bottom than good sense reduced to a calculus."

But in real-world situations with sufficient complexity, the knowledge, time, and computation necessary to realize the classical ideal of unbounded rationality can be prohibitive—too much for humble humans, and often also too much for the most powerful computers. For instance, if one updates Franklin's weighted linear combination of reasons into its modern and improved version, multiple linear regression, then a human would have to estimate the weights that minimize the error in the "least-squares" sense for all the reasons before combining them linearly—a task most of us could not do without a computer. If one were to further update Franklin's method to (non-linear) Bayesian networks, then the task could become too computationally complex to be solved by a computer.

Despite their psychological implausibility, the preferred models of cognitive processes since the cognitive revolution of the 1960s were those assuming demons: subjective expected utility maximizing models of choice, exemplar models of categorization, multiple regression models of judgment, Bayesian models of problem solving, and neural network models of almost everything. Demons that can perform amazing computations have not only swamped cognitive psychology, but also economics, optimal foraging theory, artificial intelligence, and other fields. Herbert Simon has countered "there is a complete lack of evidence that, in actual human choice situations of any complexity, these computations can be, or are in fact, performed" (1955a, p. 104).

Simon proposed to build models of bounded rationality rather than of optimizing. But how? What else could mental processes be, if not the latest statistical techniques?

Simple Stopping Rules

In this chapter, we deal with the same type of task as in Chapter 2: determining which of two objects scores higher on a criterion. This task is a special case of the more general problem of estimating which subclass of a class of objects has the highest values on a criterion (as in Chapter 3). Examples are treatment allocation (e.g., which of two patients to treat first in the emergency room, with life expectancy after treatment as the criterion), financial investment (e.g., which of two securities to buy, with profit as criterion), and demographic predictions (e.g., which of two cities has higher pollution, crime, mortality rates, and so on).

To illustrate the heuristics, consider the following two-alternative choice task:

Which of the two cities has a larger population?

- (a) Hannover
- (b) Bielefeld

Assume that a person has heard of both cities, so she cannot use the recognition heuristic. She needs to search for cues that indicate larger population. Search can be internal (in memory) or external (e.g., in libraries). Limited search is a central feature of fast and frugal heuristics: not all available information is looked up, and consequently, only a fraction of this information influences judgment. (In contrast, laboratory experiments in which the information is already conveniently packaged and laid out in front of the participants eliminate search, and in line with this experimental approach, many theories of cognitive processes do not even deal with search.)

[Figure 4-1]

Limited search implies a stopping rule. Fast and frugal heuristics use simple stopping rules. They do not follow the classical prescription to search as long as the perceived marginal benefits of acquiring additional information exceed the perceived marginal costs (Stigler, 1961). That minds could and would routinely calculate this optimal cost-benefit trade-off is a dominant, yet implausible, assumption in models of information search (see the epigram introducing this chapter).

We demonstrate a simple stopping rule with Figure 4-1. This figure represents a person's knowledge about four objects a, b, c, and d (cities, for example) with respect to five cues (such as whether the city has a big-league soccer team, is a state capital, and so forth) and recognition (whether or not the person has heard of the city before). For instance, 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. Suppose we wish to decide which of city a and city b is larger. Both a and b are recognized so the recognition heuristic can not be used. Search for further knowledge in memory brings to mind information about Cue 1, the soccer team cue. City a has a soccer team in the major league, but city b does not (the positive and the negative values on Cue 1 in Figure 4-1). Therefore, the cue discriminates between the two cities. Search is terminated, and the inference is made that city a is the larger city. More generally, for binary (or dichotomous) cues, the simple stopping rule is:

Stopping rule: If one object has a positive cue value ("1") and the other does not (i.e., either "0" or unknown) then stop search.

For convenience, we use "1" for the cue value that indicates a higher value on the criterion (e.g., a larger population). If the condition of the stopping rule is not met, then search is continued for another cue, and so on. For instance, when deciding between objects b and c in Figure 4-1, Cue 1 does not discriminate, but Cue 2 does. Object b is inferred to be larger on the basis of this single cue. Note that limited search works in a step-by-step way; cues are looked up one-by-one, until the stopping rule is satisfied (similar to the Test Operate Test Exit procedures of Miller, Galanter, & Pribram, 1960). If no cue was found that satisfy the stopping rule, a random guess is made. No cost-benefit computations need to be performed to stop search. The following three heuristics—Minimalist, Take The Last, and Take The Best—use this simple stopping rule. They also use the same heuristic principle for decision, one-reason decision making, that is, they base an inference on only one reason or cue. They differ in how they search for cues.

Heuristics

The Minimalist. The minimal intuition needed for cue-based inference is the direction in which a cue points, for instance, whether having a soccer team in the major league indicates a large or a

small population. This direction can, for instance, be estimated from a small learning sample (and the estimated direction may sometimes be wrong, see below). The Minimalist has only this minimal intuition. Nothing more is known, for instance, about which cues are better predictors than others. Consequently, the heuristic for search that the Minimalist uses is to look up cues in random order. Whenever the Minimalist can, it will take advantage of the recognition heuristic (see Chapter 2). However, there are situations where the recognition heuristic can not be used, that is, when both objects are recognized, or when recognition is not correlated with the criterion.

The Minimalist heuristic can be expressed in the following steps:

Step 0. If applicable, use the recognition heuristic, that is, if only one object is recognized, predict that it has the higher value on the criterion. If neither is recognized, then guess. If both are recognized go on to Step 1.

Step 1. Random search: Draw a cue randomly (without replacement) and look up the cue values of the two objects.

Step 2. Stopping rule: If one object has a positive cue value and the other does not (i.e., either negative or unknown value) then stop search and go on to Step 3. Otherwise go back to Step 1 and search for another cue. If no further cue is found, then guess.

Step 3. Decision rule: Predict that the object with the positive cue value has the higher value on the criterion.

Take The Last. Like the Minimalist, Take The Last only has an intuition in which direction a cue points but not which cues are more valid than others. Take The Last differs from the Minimalist only in Step 1. It uses a heuristic principle for search that draws on a strategy known as an "Einstellung set." Karl Duncker and other Gestalt psychologists demonstrated that when people work on a series of problems, they tend to start with the strategy that worked on the last problem when faced with a new, similar-looking problem (Duncker 1935/1945; Luchins & Luchins, 1994), and thereby build up an Einstellung set of approaches to try. For the first problem, Take The Last tries cues randomly like the Minimalist, but from the second problem onward it starts with the cue that stopped search the last time. If this cue does not stop search, it tries the cue that stopped search the time before last, and so on. Because cues that recently stopped search tend to be cues that are more likely than others to stop search (i.e., they are cues with higher discrimination rates), Take The Last tends to search for fewer cues than the Minimalist. For instance, if the last decision was based on the soccer team cue, Take The Last would try the soccer team cue first on the next problem. In contrast to the Minimalist, Take The Last needs a memory for what cues discriminated in the past. Step 1 of the Take The Last is:

Step 1. Einstellung search: If there is a record of which cues stopped search on previous problems, choose the cue which stopped search on the most recent problem and has not yet been tried. Look up the cue values of the two objects. Otherwise try a random cue and build up such a record.

Take The Best. There are environments for which humans or animals know (rightly or wrongly) not just the signs of cues, but also which cues are better than others. An order of cues can be genetically prepared (e.g., cues for mate choice in many animal species) or learned by observation. In the case of learning, the order of cues can be estimated from the relative frequency with which it predicts the criterion. For example, the validity of the soccer team cue would be the relative frequency with which cities with soccer teams are larger than cities without teams. The validity is computed across all pairs in which one city has a team and the other does not. If people can order cues according to their perceived validities - whether or not this subjective order

corresponds to the ecological order- then search can follow this order of cues. Take The Best first tries the cue with the highest validity, and if it does not discriminate, the next best cue, and so on. Its motto is “take the best, ignore the rest.” Take The Best differs from the Minimalist only in Step 1, which becomes:

Step 1. Ordered search: Choose the cue with the highest validity that has not yet been tried for this choice task. Look up the cue values of the two objects.

Note that the order that Take The Best uses is not an “optimal” one - it is, rather, a frugal ordering. It does not attempt to grasp the dependencies between cues, that is, to construct an order from conditional probabilities or partial correlations (see Chapter 6). The frugal order can be estimated from a small sample of objects and cues (see Chapter 5).

To summarize, the three fast and frugal heuristics just presented embody the following properties: limited search using a step-by-step procedures, simple stopping rules, and one-reason decision making. One-reason decision making, basing inferences on just one cue, is implied by the specific stopping rule used here. It is not implied by all simple stopping rules. Furthermore, one-reason decision making does not necessarily imply the stopping rule used by the three heuristics. For instance, one could search for a large number of cues that discriminate between the two alternatives (such as in a situation where one has to justify one’s decision) but still base the decision on only one cue.

Compare the spirit of these simple heuristics to Franklin’s rule. One striking difference is that all three heuristics practice one-reason decision making. Franklin’s moral algebra, in contrast, advises us to search for all reasons—at least during several days consideration—and to carefully weigh each reason and add them all up to see where the balance lies. The three heuristics avoid conflicts between cues that may point in opposite directions. Avoiding conflicts makes the heuristics non-compensatory: no amount of contrary evidence from later (unseen) cues can compensate for or counteract the decision made by an earlier cue. An example is the inference that a is larger than b in Figure 4-1; neither the two positive values for b nor the negative value for a can reverse this inference. Basing an entire decision on just one reason is certainly bold, but is it smart?

Psychologically Plausible But Dumb?

Consider first a species that practices one-reason decision making closely resembling Take The Best. In populations of guppies, the important adaptive task of mate choice is undertaken by the females, which respond to both physical and social cues (Dugatkin, 1996). Among the physical cues they value are large body size and bright orange body color. The main social cue they use is whether or not they have observed the male in question mating with another female. The cues seem to be organized in a hierarchy, with the orange-color cue dominating the social cue. If a female has a choice between two males, one of which is much more orange than the other, she will choose the more orange one. If the males are close in orangeness, she prefers the one she has seen mating with another female. She prefers this one even if he has slightly less orange color. The stopping rule for the orangeness cue is that one male must be much (about 40%) more orange than the other. Mate choice in female guppies illustrates limited search, simple stopping rules, and one-reason decision making.

People, not just lower animals, often look up only one or two relevant cues, avoid searching for conflicting evidence, and use non-compensatory strategies (e.g. Einhorn, 1970; Einhorn & Hogarth, 1981, p. 71; Fishburn, 1988; Hogarth, 1987; Payne et al., 1993; Shepard, 1967). For

instance, Take The Best (unlike the Minimalist and Take the Last) is related to lexicographic strategies. The term “lexicographic” signifies that the cues are looked up in a fixed order of validity, like the alphabetic order used to arrange words in a dictionary. The Arabic (base 10) and Babylonian (base 12) number systems are lexicographic. To see which of two numbers with equal numbers of digits is larger, one has to look at the first digit: if this digit is larger, the whole number is larger. If they are equal, one has to look at the second digit, and so on. This simple method is not possible for Roman numbers, which are not lexicographic. In experimental studies, lexicographic strategies seem to be favored under time constraints (Payne et al., 1993, see also Chapter 7). In addition, Take The Best and the more general framework of probabilistic mental models (Gigerenzer, Hoffrage, & Kleinbölting, 1991) have been successful in integrating various empirical phenomena (Gigerenzer et al., 1991; Juslin, 1993; DiFonzo, 1994; McClelland and Bolger, 1994).

However, simple heuristics that embody one-reason decision making, avoid conflicts, and are non-compensatory were often discredited as irrational, because they look stupid in comparison to traditional norms of rationality which focus on coherence rather than on performance in real-world environments. For instance, when Keeney and Raiffa (1993) discuss lexicographic strategies, they repeatedly insert warnings that this strategy “is more widely adopted in practice than it deserves to be” because “it is naively simple” and “will rarely pass a test of ‘reasonableness’”(pp. 77-78). They did not actually perform such a test. We shall.

Can Fast and Frugal Heuristics Be Accurate?

Heuristics are often evaluated by principles of internal coherence, rather than by criteria that measure their performance in the external world: accuracy, frugality and speed, among others. The major exception in judgment and decision-making research is the work by Payne et al. (1993), who have systematically compared the “accuracy-effort” trade-off of simple strategies to the performance of the weighted additive rule (Franklin’s rule), which is often taken as normative for preferences (see also Beach & Mitchell, 1978; Beach et al., 1986). In contrast to our research, Payne and his colleagues studied preferences in artificial problems rather than inferences about the real world. One consequence of this is that there is no external criterion for accuracy (for example the actual population of a city), so norms must be constructed. In their studies, the weighted additive rule is taken as the gold standard, and accuracy is defined as how close a strategy comes to this rule. Therefore, no strategy can ever be more accurate than the norm. When making inferences about the real world, however, it does not necessarily hold that the weighted additive rule is the best one can do.

How accurate can heuristics be that violate the following two commandments that are often taken as characteristic of rational judgment?

Complete search. Thou shalt find all the information available. If thou cannot because of time or computational constraints, then compute the point where the cost of further searching exceeds the benefits of doing so, and search until this point.

Compensation. Thou shalt combine all pieces of information. Thou shalt not rely on just one piece.

While Franklin’s rule respects both commandments, the Minimalist, Take The Last, and Take The Best heuristics violate them. They do not look up all cue values (limited search) and do use

a simple stopping rule. They do not combine cue values (non-compensation). The Minimalist, in addition, can violate transitivity, a sacred principle of internal coherence.¹

To answer the question of how accurate fast and frugal heuristics are, we evaluated their performance in a competition that pitted three standard statistical strategies against the three fast and frugal heuristics introduced above. The goal was to see which strategy would yield the most accurate inferences while looking up the fewest cue values.

The Competitors

To provide standards of comparison, we introduce three competitors that do not violate these commandments of rational judgment. The first is a weighted linear combination of cues, which we call Franklin's rule, because it applies Franklin's principles to the two-alternative choice tasks considered here. It is actually a more sophisticated method than Franklin's original moral algebra because the weights are not subjective but computed from the data. In the present simulation, the cue weights are ecological validities, to be defined shortly. Franklin's rule multiplies each cue value by its weight and sums the total, inferring that the object with the larger sum is the larger object. In the simulation, positive and negative cue values are coded as 1 and 0, respectively.

The other two competitors are linear combinations of cues like Franklin's rule. One of them demands considerably more knowledge and computation, and one demands less. The more demanding algorithm is multiple linear regression. Multiple regression takes care of the dependencies between cues by calculating weights that minimize the error in the "least-squares" sense. Variants of weighted linear models have been proposed as descriptive or prescriptive models of cognitive processes, for instance, in N. H. Anderson's (e.g., 1981) information integration theory, and in social judgment research (Brehmer, 1994; Brunswik, 1955). As descriptions of psychological processes, weighted linear models, and particularly multiple linear regression are questionable, given the complex computations they assume (Brehmer & Brehmer, 1988; Einhorn & Hogarth, 1975; Hogarth, 1987). A more psychologically plausible version of a linear strategy employs unit weights, as suggested by Robyn Dawes (e.g., 1979). This strategy simply adds up the number of positive cue values (or ones) and subtracts the number of negative cue values (or zeroes). Thus it is fast (it does not involve much computation), but not frugal (it looks up all cues). For short, we call this strategy Dawes's rule.

In the simulations we report, these three linear models serve as benchmarks against which to evaluate the performance of the fast and frugal heuristics. Note that Franklin's rule and multiple linear regression use all the information the three heuristics use, and more. They also carry out more sophisticated computations on this information.

The Environment

After Germany was reunified in 1990, the country had 83 cities with more than 100,000 inhabitants. These cities and nine cues for population size constituted the environment for the

¹ Intransitivity can result from the fact that the Minimalist picks cues in random order, as is illustrated by Figure 4-1. For instance, if Cue 1 happens to be applied to objects a and b, Cue 2 to b and c, and Cue 3 to a and c, we get the intransitive judgment a>b, b>c, and c>a. Intransitivity can result from the fact that the Minimalist picks cues in random order, as is illustrated by Figure 4-1. For instance, if Cue 1 happens to be applied to objects a and b, Cue 2 to b and c, and Cue 3 to a and c, we get the intransitive judgment a>b, b>c, and c>a.

simulation. The cues were chosen from people's reported cues in experiments (Gigerenzer et al., 1991; Gigerenzer & Goldstein, 1996a). The task was to infer which of two cities has a larger population. Each cue has two important characteristics: its ecological validity and its discrimination rate. The ecological validity of a cue is the relative frequency with which the cue correctly predicts the criterion, defined with respect to the reference class (here, 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 about 87% of these cases the city with the team also has the higher population. This .87 value is the ecological validity of the soccer team cue. In general, the ecological validity v_i of the i th cue is:

$$v_i = \text{number of correct predictions/number of predictions}$$

where the number of predictions is the number of pairs in which one object has a positive and the other a negative value. The ecological validities of the cues varied over the whole range (Table 4-1).

A cue with a high ecological validity, however, is not very useful if its discrimination rate is small. The discrimination rate of a cue is the relative frequency with which a cue discriminates between pairs of 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 i th 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 $2x_i y_i$.

The larger the ecological validity of a cue, the better the inferences. The larger the discrimination rate, the more often a cue can be used to make an inference. The pairwise correlations between the 9 cues ranged between -.25 and .54, with an average absolute value of .19.

Different strategies extract different information from the environment. The Minimalist, for instance, does not extract information about which cues are better than others, it only needs to estimate in which direction a cue points. Take The Best extracts information about the order in which cues should be tried. All competitors made use of the actual cue values from the complete environment to calculate parameters such as ecological validities or regression coefficients.

[Table 4-1]

Limited Knowledge

We simulated subjects with varying degrees of knowledge about this environment. Limited knowledge can take two forms. One is limited recognition of objects. The other is limited knowledge about the cue values of recognized objects. To model limited recognition knowledge, we simulated subjects who recognized between 0 and all (83) German cities (i.e., 84 different levels of recognition). To model limited knowledge of cue values, we simulated 6 classes of subjects, 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 x 84 types of subjects each having different degrees and kinds of limited knowledge. For each type of subject, we created 500

simulated individuals, who differed randomly from one another in the particular objects and cue values they knew.

The simulation needed to be realistic in the sense that the simulated subjects should be able to invoke the recognition heuristic. Therefore, the sets of cities the simulated subjects recognized 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 relationship between the recognition of cities and city populations. In a survey of 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 the cases. We incorporated this value into our simulations by choosing sets of cities (for each knowledge state, that is, 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 subjects had the same relationship between recognition and population as did those of the human subjects. For details of the simulation see Gigerenzer and Goldstein (1996a).

Each simulated subject made inferences about which of two cities is larger, using each of six strategies: the three fast and frugal heuristics (Take The Best, Take The Last, and Minimalist) and the three linear methods (regression, Franklin's rule, and Dawes's rule). The question of how well a fast and frugal heuristic performs in a real-world environment has rarely been posed in research on inductive inference. If the simple heuristics are adapted to environmental structures, then they should not fail outright.

How Frugal Are the Heuristics?

We measure frugality by the number of cues a heuristic looks up. The three linear models always look up and integrate all 10 cues (9 ecological cues plus recognition). Across all states of limited knowledge, Take The Last looked up on average only 2.6 cues, the Minimalist 2.8 cues, and Take The Best 3.0 cues (Table 4-2). Take The Last owes its frugality to the "Einstellung set" which tends to collect the cues which discriminate most often. The reason why the Minimalist looked up fewer cues than Take The Best is that cue validities and cue discrimination rates are negatively correlated (Table 4-1). Therefore, randomly chosen cues tend to have higher discrimination rates than cues chosen by cue validity. All in all, the three heuristics look up less than a third of the cues used by the linear models, on average.

[Table 4-2]

How Accurate Are the Heuristics?

How accurate are the three heuristics, given that they look up only a fraction of the available information? Recall that the Minimalist looks up on average only 2.8 cues, uses one-reason-decision making, does not know which cues are better than others, and can violate transitivity. It must be doomed to fail. Table 4-2, however, shows that the Minimalist achieves an average accuracy of 64.7%. This is slightly higher than Take The Last, but lower than Take The Best with 65.8%. But how much more accurate are Dawes's rule, Franklin's rule, and multiple regression, which use all cues' values and combine them? The result in Table 4-2 is surprising. Dawes's rule is outperformed by each of the three heuristics, although Dawes's rule has all the information that the Minimalist and Take The Last have (only Take The Best knows about the order of cues, which is not available to

Dawes's rule). Franklin's rule has all the information that each of the three heuristics has, and more. Still, it is outperformed by even the most frugal of the simple heuristics.

How do the heuristics compare to a more powerful competitor? Multiple regression calculates a set of weights considered optimal for linear prediction, and arriving at these weights requires considerable computational might. Though it makes more accurate inferences than both the Minimalist and Take The Last, regression is matched in accuracy by the fast and frugal Take The Best.

Figure 4-2 shows the accuracy of the six competitors as a function of the number of cities recognized. Here, the situation where all competitors perform best is shown, namely when knowledge of cue values is 100%. The figure shows that the Minimalist and Take The Last can compete well with the other algorithms in accuracy when the number of objects recognized is limited, but take a loss when all are known, that is, when complete information is available. Franklin's rule and Dawes's rule match Take The Best when no or all objects are recognized, but suffer with intermediate levels of recognition. Why is this? The reason is that these two strategies violate the wisdom of the recognition heuristic. They sometimes choose unrecognized cities as larger than recognized ones. In this environment, most cities have more negative cue values than positive ones, for example the average city is not a state capital, does not have a major league soccer team, and so on. Dawes's rule, which subtracts the number of negative cue values from the number of positive ones, often arrives at a negative total for a recognized city that exceeds that of an unrecognized city (which is always -1, because of one negative reason: no recognition). The same holds for Franklin's rule, which weights the reasons (Gigerenzer & Goldstein, 1996a). Therefore, an unrecognized city is often inferred to be larger than a recognized one, which turns out to be a bad idea in this environment where the recognized cities were larger than the unrecognized cities 80% of the time. When one helps the linear strategies by endowing them with the recognition heuristic, their performance roughly matches that of Take The Best and multiple regression.

[Figure 4-2]

Figure 4-2 also illustrates a less-is-more effect (see Chapter 3) in four of the six strategies. In contrast to Figure 3-4, which shows a noisy less-is-more effect obtained by Take The Best in a simulation where the recognition validity was determined empirically at each level of recognition, here we see it in a smooth, refined form—a result of holding the recognition validity constant at our estimate of its empirical average.

Trade-off Between Accuracy and Frugality

Within the three heuristics, the expected trade-off holds: the more frugal (the fewer cue values looked up), the less accurate. However, when we compare the family of heuristics to the three linear strategies, then things get very interesting. Compared to multiple regression, Take The Best did not sacrifice accuracy for frugality, it achieved both. Compared to Dawes's and Franklin's rules, all three heuristics managed to be more accurate and yet more frugal at the same time.

When we first obtained these results, we could not believe them. We hired independent programmers in the US and Germany to rerun the simulations to exclude possible wishful thinking on our part. When we finally published the results, we also included the data on the environment so that everyone could perform their own replications, and many did (Gigerenzer & Goldstein, 1996a). Fast and frugal heuristics do not necessarily have to trade off accuracy for simplicity.

Can Frugality and Accuracy Both Be Possible?

Fast and frugal heuristics can make accurate inferences about unknown properties of the world, that is, inferences that are equal to or more accurate than the three linear strategies. In designing these simulations, we wondered if the heuristics would fail dismally. Before we reported the results, three eminent researchers in judgment and decision-making predicted that Take The Best might perform 10 or 5 percentage points worse than the linear strategies. Each of the three heuristics, however, exceeded these expectations, and even outperformed some of the linear strategies. Take The Best matched or outperformed them all. At that juncture we did not understand how the competition could come out that way. The answer—in the form of what we call ecological rationality—was only found after some further struggling, and will be developed in Chapter 7. Here we summarize a few insights.

The observation of a flat maximum for linear models is one insight. If many sets of 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 Robyn Dawes and others (e.g., Dawes & Corrigan, 1974) made this phenomenon known to decision researchers, but has actually been known for longer. Since Wilks (1938) wrote about the robustness of equal weights, many have argued that weights are irrelevant both for making predictions by an artificial system (such as an IQ test) and for describing actual human inferences. In psychometrics, weighting the components of a test battery is rare because various weighting schemes result in surprisingly similar composite scores, that is, in flat maxima (e.g., Gulliksen, 1950). Flat maxima seem to occur when cues are strongly positively correlated. The performance of fast and frugal heuristics indicates that a flat maximum can extend beyond the issue of weights to decision strategies themselves: inferences based solely on the best cue can be as accurate as those based on a weighted linear combination of all cues.

There is also scattered earlier evidence that simple, non-compensatory heuristics can perform well. However, because much of the earlier work concentrated on preferences (rather than inferences) and on artificial stimuli (rather than real-world environments), external criteria of performance were often hard to come by. As mentioned before, the closest relatives of Take The Best are lexicographic strategies. Payne, Bettman, and Johnson (1993) showed that lexicographic judgments can sometimes be close those of a weighted linear model, but they had no external criteria for accuracy. A second class of close relatives are simple algorithms in machine learning, which can perform highly accurate classifications (Holte, 1996; Rivest, 1996). A more distant relative to Take The Best is Elimination By Aspects (Tversky, 1972), which also employs limited search and a stopping rule, but deals with preference rather than inference, does not use the order of cues but a probabilistic criterion for search which requires knowledge of the quantitative validities of each cue, has no recognition heuristic built in, and does not employ one-reason decision making. Another more distant class of relatives are classification and regression trees (CART), which use a simple decision tree and one-reason decision making, but differ in the knowledge and computational power they use for setting up the simple tree. For instance, Breiman et al. (1993) reported a simple CART algorithm with only three binary, ordered cues that classified heart-attack patients into “high” and “low” risk groups. This non-compensatory tree was more accurate than standard statistical classification methods, which used up to 19 variables (see Chapter 1). The practical relevance 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 since there is no integration. For theories that postulate

mechanism that resemble Take The Best see Relevance Theory (Sperber, Cara, & Girotto, 1995) and optimality theory (Prince & Smolensky, 1991; Legendre, Raymond, & Smolensky, 1993).

All in all, the observation of flat maxima, the performance of simple machine learning rules and CART trees, and the work by Payne, Bettman, and Johnson, gave us hope that there was something larger to discover behind this first surprising finding.

Matching Stopping Rules To Environments

What structures of information in real-world environments can fast and frugal heuristics exploit in order to perform as accurately as they did? Where would they fail? Chapters 6 and 7 will address these questions. Here, we will illustrate this idea of ecological rationality—the match between mind and environment—by the positive bias of the stopping rule. Recall that the combination of a positive value and an unknown value stop search, but a negative and an unknown value do not. This asymmetry is what we mean by a positive bias. Positive biases of various kinds have been observed in humans (e.g., Klayman & Ha, 1987), and can result in both more frugal and more accurate inferences than an unbiased stopping rule. Consider first an unbiased stopping rule that demands a positive and a negative cue value (as proposed by Gigerenzer et al., 1991). This stopping rule would be less frugal, because search would take longer when there is limited knowledge (i.e., unknown cue values) than it would with a positive bias. Now consider a faster, unbiased stopping rule that always terminates search when the positive bias rule does, but in addition when a negative and an unknown value are obtained. Compared to this second unbiased stopping rule, a positive bias can be shown to achieve more accurate judgments in environments where negative cue values are more frequent than positive ones. The intuition for this result is that the unknown value is most likely a negative value. If the unknown value is negative, however, this will lead to fewer accurate judgments when one stops with a negative and an unknown value, because this would often mean that there were actually two negative values. Thus, a stopping rule with positive bias is ecologically rational in environments where negative cue values outnumber positive ones. An example is the environment studied in this chapter, where only relatively few cities have soccer teams in the major league, and only a few are state capitals (other examples lead to the “rarity” assumption of Oaksford & Chater, 1994). More generally, in environments where positive indicators are few and scattered—a rare symptom that signals a disease, an unusual feature that hints competence—a stopping rule with positive will prove ecologically rational.

Generalization

How does Take The Best estimate the order of cues? How do Take The Last and the Minimalist learn in which direction a cue points? There are several ways cues and their ranking may be learned. Cues, or the preparedness to learn cues, may be genetically coded through evolution. Cues for distance perception, mate choice, and food avoidance have been proposed as examples (e.g., Buss, 1992). Cues can also be learned through cultural transmission. For example, the cues needed for expertise can be learned from apprenticeship and the exchange of trade secrets. Finally, cues can be learned from direct observation. For instance, a person who knows some cue values for just 10 German cities, and knows for some pairs of these cities which has a higher population, could use this knowledge to estimate the rank order and direction of cues for the entire set. In contrast, in

the simulations reported in this chapter, each strategy computed the parameters needed (direction of cue, cue order, cue validities, regression coefficients) from the entire data set.

How well would Take The Best do if it were to learn cues from a small sample? Recall that Take The Best extracts from a learning sample only the order and sign of the cues, a very small amount of information compared to the real-valued weights, regression coefficients or conditional probabilities extracted by more complex statistical procedures. Thus, in a learning situation, Take The Best takes away only a small amount of information from a small sample. Regression, in contrast, extracts considerably more information from a small sample. Which is the better policy?

Figure 4-3 shows Take The Best, Take The Last, and the Minimalist competing with multiple regression at making generalizations from a training set to a test set. Each strategy estimated its respective parameters from a proportion (between 10% and 90%) of the German cities, and made predictions about the complement. The process of dividing the environment into training and test sets, learning the parameters from the training set, and making predictions about the test set was repeated 500 times. In these simulations recognition was not a factor, that is, all objects were assumed to be recognized. Let us first consider the situation in which all cue values are known for all objects in the training and test sets (Figure 4-3a). At the point where the training set is 50% of the total environment, for instance, Take The Best reaches 72% correct predictions, whereas Multiple Regression achieves 71%. More generally, throughout the entire range of training set sizes, Take The Best outperforms Multiple Regression, especially when the training set is small. Figure 4-3b shows a more difficult situation where half of the cue values were eliminated from the environment before the training and test sets were created. Here, the advantage of Take The Best is slightly more pronounced. Furthermore, when the training set is very small, the two most simple heuristics, Take The Last and Minimalist, perform as well as or better than the other strategies.. These results indicate that Take The Best is more robust than Multiple Regression on this data set, and less prone to overfit a training set. Under situations of limited knowledge, simpler strategies may be more robust.

[Figure 4-3]

What about the generalization ability of more computationally expensive strategies than multiple regression? Using the German cities environment, Chater et al. (1997) tested Take The Best against complex strategies, including neural networks and exemplar models of categorization. Like multiple regression, none of these strategies has a stopping rule, but rather use all available cues. When the training set was less than 40% of the test set, Take The Best outperformed all other competitors. This advantage was largest (ten percentage points) when the size of the training set was smallest. Only when the training set grew beyond 40% of the German cities environment (which is actually more knowledge than most anybody has about German demographics, Germans included) then the competitors' performance increased above that of Take The Best, at most attaining a margin of about five percentage points. Note however that the simulations of Chater et al. have only dealt with the case where there were not any unknown cue values (as represented by the question marks in Figure 4-1).

These results, which came as a surprise to us, show how very simple heuristics can excel in situations where knowledge is limited, and where generalizations must be made from one sample to another. Chapters 6 and 7 will address the robustness of fast and frugal heuristics in more detail.

The Adaptive Toolbox

The Minimalist, Take The Last, and Take The Best are candidates for the collection of heuristics in what we call the adaptive toolbox. The emphasis is on “collection.” None of these three strategies can perform all possible inferences under uncertainty—for instance, all three are designed to make estimates about which of two objects is larger, more effective, more dangerous, and so on. They cannot, for instance, estimate the quantitative values of one object. However, some of the building blocks—simple stopping rules, one-reason decision making—can be recombined to make heuristics for quantitative estimation, classification, and other tasks, as we will see in later chapters.

One may think of a collection of heuristics as a body made up of organs that have evolved over time rather than being designed in a grand plan. Thus, the adaptive toolbox may have evolved by adding features to already existing tools, rather than by replacing one generation of tools with a completely new generation. The three heuristics studied in this chapter, for instance, are built around the recognition heuristic. If the recognition heuristic can be used, search for further knowledge is not needed. If it cannot, the inference is made by the additional tools. Here, the order in which these two layers of heuristics are invoked follows their likely developmental and evolutionary order: Recognition and recognition memory are the more fundamental adaptive functions, less able to be damaged by age and brain injury (see Chapter 3) than the recall memory used by Take The Best and its relatives.

The single most important result in this chapter is: Fast and frugal heuristics that embody simple psychological mechanisms can yield inferences about a real-world environment that are at least as accurate as standard linear statistical strategies embodying classical properties of rational judgment. This result liberates us from the widespread view that only “rational” algorithms, from Franklin’s rule to multiple regression, can be accurate. Human inference does not have to forsake accuracy for simplicity. The mind can have it both ways.

When we concluded our first report of these results (Gigerenzer & Goldstein, 1996a) with the previous sentence, deep in our hearts we still had nagging doubts. Can heuristics really be fast, frugal, and accurate at the same time? Maybe there is something peculiar to city populations, or to German cities? Does the power of these heuristics to combine simplicity and frugality with accuracy generalize to other domains? What structures of information in natural environments do these heuristics exploit? Where do they break down? The following chapters tell what we have learned, so far. More surprises are to come.

Cue	Ecological Validity	Discrimination Rate
National Capital (is the city the national capital?)	1.0	.02
Exposition Site (was the city once an exposition site?)	.91	.25
Soccer Team (does the city have a team in the major leagues?)	.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

Table 4-1: Cues, ecological validities, and discrimination rates

Strategy	Knowledge about cues	Frugality (number of cues looked up)	Accuracy (%)
Take The Last	direction	2.6	64.5
Minimalist	direction	2.8	64.7
Take The Best	order	3.0	65.8
Dawes's rule	direction	10.0	62.1
Franklin's rule	validities	10.0	62.3
Multiple regression	beta weights	10.0	65.7

Table 4-2: A tournament between three fast and frugal heuristics (Minimalist, Take The Last, Take The Best) and three linear models (Dawes's rule, Franklin's rule, and multiple regression). Results are averaged across all levels of limited knowledge, that is, limited recognition and limited number of cue values known (see text). For instance, the Minimalist looked up only 2.8 cues on the average and made 64.7% correct inferences.

Take The Best (one-reason decision making)

	a	b	c	d
Recognition	+	+	· · +	-
Cue 1	1	0	· · ?	?
Cue 2	?	· 1 · · ?	· · ?	?
Cue 3	0	1	1	?
Cue 4	?	0	0	?
Cue 5	?	?	0	?
⋮	⋮	⋮	⋮	⋮

Figure 4-1: Illustration of bounded search through limited knowledge. Objects a, b, and c are recognized, d is not. Cue values are positive (+), or negative (-); missing knowledge is shown by a question mark. For instance, to infer whether $a > b$, Take The Best looks up only the 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 and so are shown within the diagram as shrouded in the fog of memory.

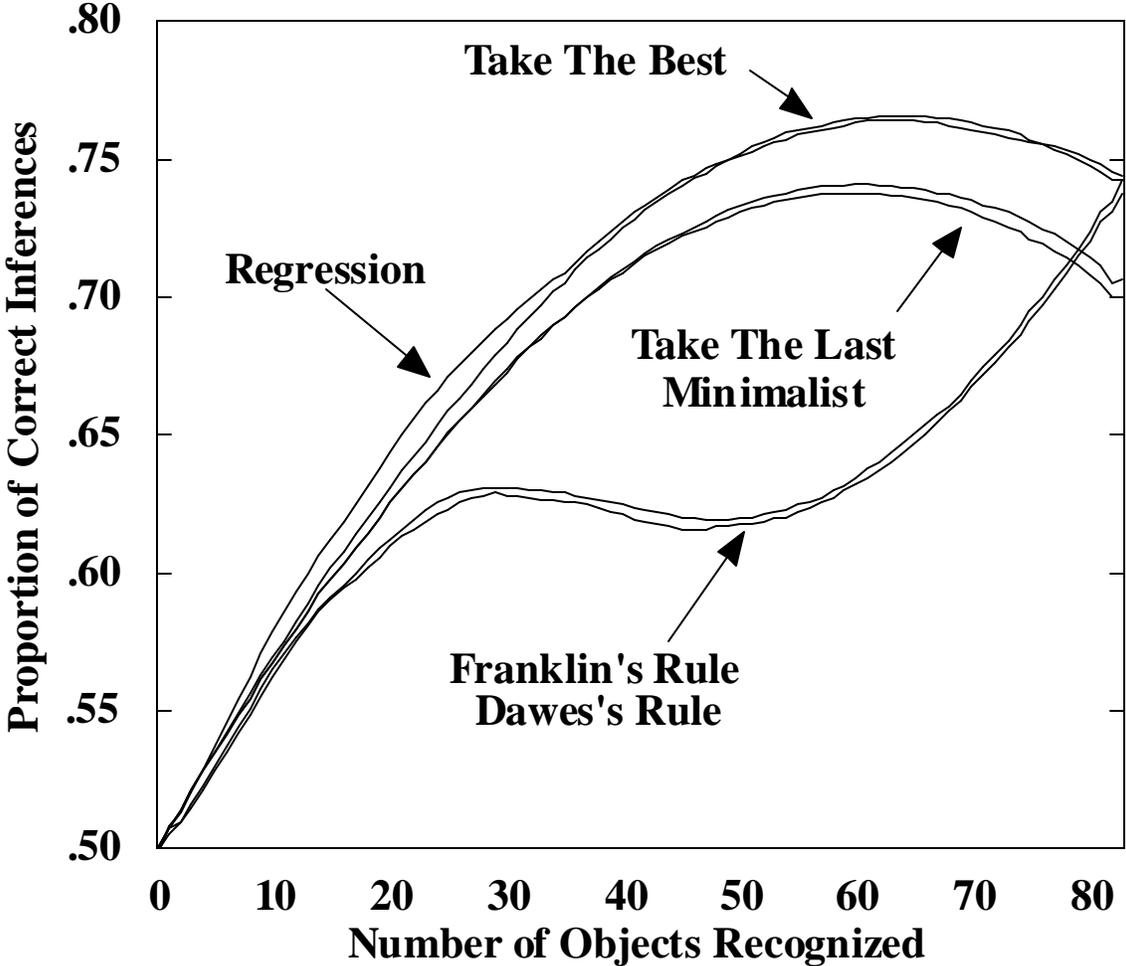


Figure 4-2: Results of the competition when knowledge of cue values is 100%.

Figure 4-3a

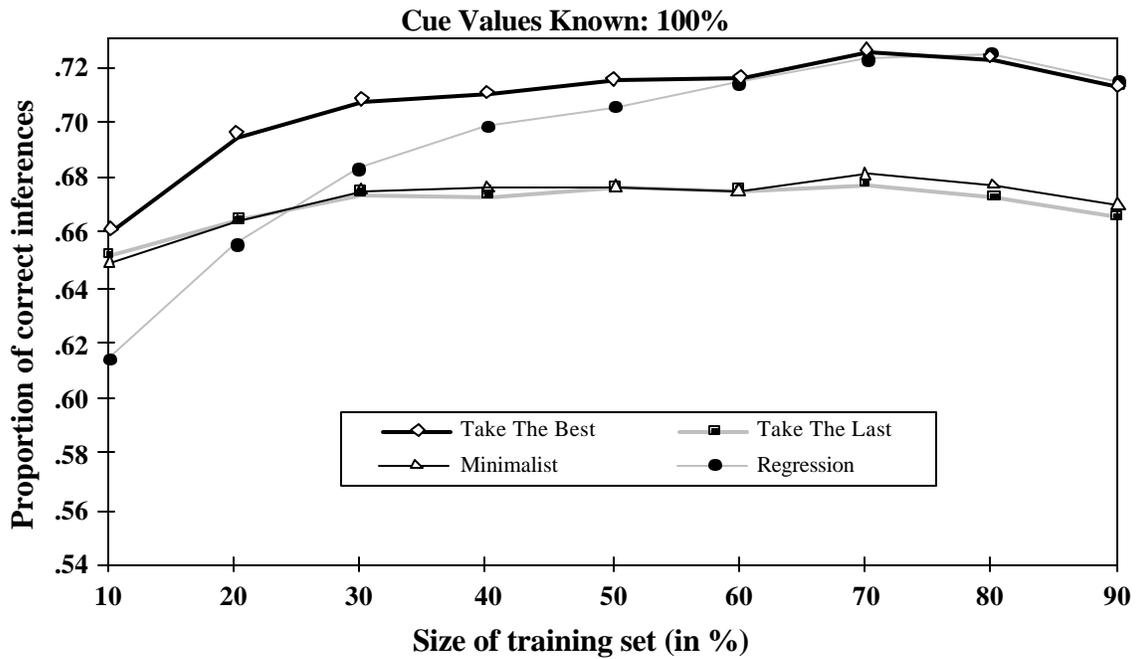


Figure 4-3b

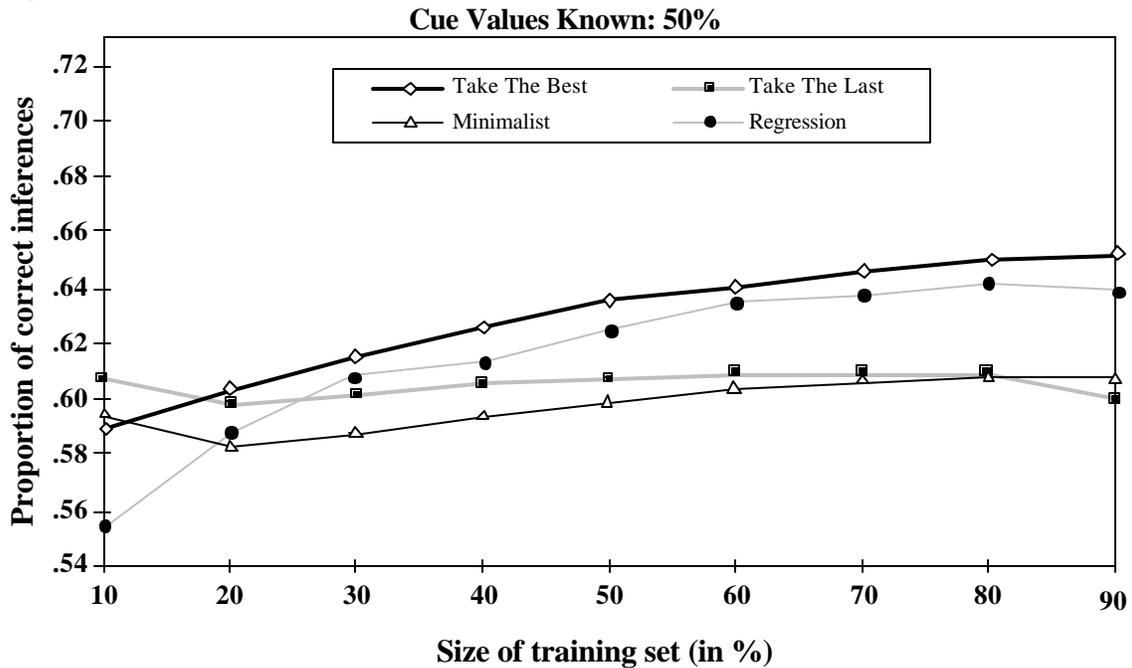


Figure 4-3: Generalizing from a training set to a test set. Results are shown for training sets with between 10% and 90% of the objects. Competitors were tested on the complement of the training set, so in the 10% condition, the test set included the remaining 90% of the objects. We also varied amount of missing knowledge in the environment. Figure 4-3a shows the cases where the training and test sets had no missing cue values. Figure 4-3b shows the case where 50% of the cue values, selected at random, were

eliminated (replaced with question marks) from the overall environment before dividing it into training and test sets.