In D. Koehler & N. Harvey (Eds.). (2004). *Blackwell handbook of judgment and decision making* (pp. 62–88). Oxford, UK: Blackwell

# 4

# *Fast and Frugal Heuristics: The Tools of Bounded Rationality*

**Gerd Gigerenzer**

## Introduction

If you open a book on judgment and decision making, chances are that you will stumble over the following moral: Good reasoning must adhere to the laws of logic, the calculus of probability, or the maximization of expected utility; if not, there must be a cognitive or motivational flaw. Don't be taken in by this fable. Logic and probability are mathematically beautiful and elegant systems. But they do not describe how actual people – including the authors of books on decision making – reason, as the subsequent story highlights. A decision theorist from Columbia University was struggling whether to accept an offer from a rival university or to stay. His colleague took him aside and said, "Just maximize your expected utility – you always write about doing this." Exasperated, the decision theorist responded, "Come on, this is serious."

I will introduce you to the study of cognitive heuristics: how people actually make judgments and decisions in everyday life, generally without calculating probabilities and utilities. The term *heuristic* is of Greek origin and means "serving to find out or discover." In the title of his Nobel Prize-winning paper of 1905, Albert Einstein used the term *heuristic* to indicate an idea that he considered incomplete, due to the limits of our knowledge, but useful (Holton, 1988). For the Stanford mathematician G. Polya (1954), heuristic thinking was as indispensable as analytical thinking for problems that cannot be solved by the calculus or probability theory – for instance, how to find a mathematical proof. The advent of computer programming gave heuristics a new prominence. It became clear that most problems of any importance are computationally intractable, that is, we do not know the optimal solution, nor a method for how to find it. This holds even for well-defined problems such as chess, the classic computer game Tetris, and the

traveling salesman problem (Michalewicz & Fogel, 2000). The same uncertainty holds for less well-structured problems, such as which job offer to accept, what stocks to invest in, and whom to marry. When optimal solutions are out of reach, we are not paralyzed to inaction or doomed to failure. We can use heuristics to discover good solutions.

## What is a Heuristic?

Imagine you want to build a robot that can catch balls – fly balls, as in baseball and cricket. (It's a thought experiment – no such robots exist yet.) For the sake of simplicity, consider situations where a ball is already high up in the air and will land in front of or behind the player. How would you build such a robot? One vision is *omniscience*: you aim at giving your robot a complete representation of its environment and the most sophisticated computational machinery. First, you might feed your robot the family of parabolas, because, in theory, balls have parabolic trajectories. In order to select the right parabola, the robot needs to be equipped with instruments that can measure the ball's initial distance, initial velocity, and projection angle. Yet in the real world, balls do not fly in parabolas, due to air resistance, wind, and spin. Thus, the robot would need further instruments that can measure the speed and direction of the wind at each point of the ball's flight, in order to compute the resulting path and the point where the ball will land, and to then run there. All this would have to be completed within a few seconds – the time a ball is in the air.

An alternative vision exists, which does not aim at complete representation and information. It poses the question: Is there a smart heuristic that can solve the problem? One way to discover heuristics is to study experienced players. Experimental studies have shown that players actually use several heuristics. One of these is the *gaze heuristic*. When a fly ball approaches, the player fixates the ball and starts running. The heuristic is to adjust the running speed so that the angle of gaze remains constant (or within a certain range; see McLeod & Dienes, 1996). The angle of gaze is the angle between the eye and the ball, relative to the ground. In our thought experiment, a robot that uses this heuristic does not need to measure wind, air resistance, spin, or the other causal variables. It can get away with ignoring every piece of causal information. All the relevant information is contained in one variable: the angle of gaze. Note that a player or robot using the gaze heuristic is not able to compute the point at which the ball will land. But the player will be there where the ball lands.

The gaze heuristic is a fast and frugal heuristic. It is fast because it can solve the problem within a few seconds, and it is frugal because it requires little information, just the angle of gaze. In general, a heuristic is a rule, such as "fixate the ball, start running, and adjust your running speed so that the angle of gaze remains constant." But a rule is not necessarily a heuristic, unless it embodies three qualities:

1 *Heuristics exploit evolved capacities.* A heuristic is *simple* relative to the evolved or learned capacities of an organism. For example, it is easy for humans to track a moving object against a noisy background; two-month-old babies can already hold

their gaze on moving targets (Rosander & Hofsten, 2002). Tracking objects, however, is difficult for a robot; a computer program that can solve this problem as well as a human mind can does not yet exist. (Similarly, in contrast to robots, humans are able to run.) Thus, the gaze heuristic is simple for humans but not for robots. Simplicity is not only a characteristic of beauty; it also allows making *fast*, *frugal*, *transparent*, and *robust* judgments. The gaze heuristic, like all heuristics, is transparent in the sense that it can be easily understood and taught to a novice, and the term *robust* refers to the ability of heuristics to generalize to new situations (see below). To summarize, a heuristic exploits hard-wired or learned cognitive and motor processes, and these features make it simple.

2 *Heuristics exploit structures of environments.* The rationality of heuristics is not logical, but ecological. Ecological rationality implies that a heuristic is not good or bad, rational or irrational per se, only relative to an environment. It can exploit certain structures of environments, or change an environment. For instance, the gaze heuristic transforms the complex trajectory of the ball in the environment into a straight line. All heuristics are to some degree domain-specific; they are designed to solve specific classes of problems. The gaze heuristic can solve problems that involve the interception of moving objects. If you learn to fly an airplane, you will be taught a version of it: When another plane is approaching, and you fear a collision, then look at a scratch in your windshield and observe whether the other plane moves relative to that scratch. If it does not, dive away quickly. For the pilot, the goal is to avoid a collision, whereas for the outfielder, the goal is to produce a collision. The nature of the heuristic is the same. To summarize, evolved capacities can make a heuristic simple, while the structure of the environment can make it smart.

3 *Heuristics are distinct from "as-if" optimization models.* The idea of calculating the ball's trajectory by solving differential equations is a form of optimization. When optimization is proposed to explain human behavior (as opposed to building artificial systems), this is called *as-if* optimization. In Richard Dawkins' (1976, p. 96) words: "When a man throws a ball high in the air and catches it again, he behaves as if he had solved a set of differential equations in predicting the trajectory of the ball." As-if optimization models are silent about the actual process, although it is sometimes suggested that the measurements and calculations might happen unconsciously. The gaze heuristic, however, illustrates that the logic of a heuristic, conscious or unconscious, can be strikingly distinct from as-if optimization. This yields an advantage. With a good model of a heuristic, one can deduce predictions that cannot be obtained from an as-if optimization model. The gaze heuristic, for instance predicts that players catch the ball while running, which follows from the fact that the player must move to keep the angle of gaze constant. Similarly, when the ball is thrown to the side of the player, one can predict that the player will run a slight arc, as can be observed in baseball outfielders and in dogs who catch Frisbees (e.g., Shaffer & McBeath, 2002).

In summary, a model of a heuristic is a rule whose purpose is to describe the actual process – not merely the outcome – of problem solving.

## What is Bounded Rationality?

In models of unbounded rationality, all relevant information is assumed to be freely available to everyone. In this framework the question is, if humans had perfect information and all eternity at their disposal, how would they behave? Models of bounded rationality put us back into our human skin and try to answer the question, how do humans who have little time and knowledge behave? However, Simon's (1955, 1956) term *bounded rationality* itself has become associated with three disparate programs: the study of optimization under constraints, the study of cognitive illusions, and the study of fast and frugal heuristics.

The term *bounded* can refer to constraints in the environment, such as information costs, and to constraints in the mind, such as limited memory (Todd, 2001). Adding one or more constraints to the program of as-if optimization results in optimization under constraints (Conlisk, 1996). For instance, the idea of measuring all causal variables that determine the trajectory of a ball's flight, subject to some constraint such as time, illustrates this program of as-if optimization. In personal conversation, Herb Simon once remarked with a mixture of humor and anger that he had considered suing authors who misused his concept of bounded rationality to construct even more unrealistic models of the human mind.

The study of cognitive illusions evolved in opposition to optimization, but also linked itself to the study of bounded rationality (e.g., Camerer, 1998; Kahneman, Slovic, & Tversky, 1982, p. xii). Its primary aim is to show that optimization is descriptively invalid, that is, to demonstrate that people's judgments do not actually follow the laws of probability or the maximization of expected utility. The result is a list of deviations from norms, which are interpreted as cognitive fallacies, emphasizing irrationality rather than rationality. The assumption is that these deviations can reveal the underlying cognitive processes (see Chapter 5, this volume).

How would this program approach the problem of how players catch a ball? Let us continue the thought experiment. One might try to demonstrate that players actually make systematic errors when computing the point where the ball will land. A player might be positioned on a fixed point in the field, a fly ball is thrown, and the player is asked to predict where the ball will hit the ground. If players cannot predict very well, such as underestimating their distance to the point where the ball will land, this error would be attributed to people's limited cognitive abilities. It might be labeled the *optimistic bias* in baseball, because underestimation suggests that players think they might catch the ball even when they can't. A debiasing training might be offered to players. In this thought experiment, the cognitive illusions program would correctly conclude that the optimizing model is descriptively disproved, but the optimistic bias would not lead to the discovery of the gaze heuristic or other heuristics that players use. Just like the optimization model, the cognitive illusions program would overlook that the actual goal of the player is not to predict where the ball will land, but to be there where the ball lands. The rationality of heuristics is not simply a means to a given end; the heuristic itself can define what the end is.

**Table 4.1** Twelve examples of phenomena that were first interpreted as "cognitive illusions" (left) but later revalued as reasonable judgments given the environmental structure

| *Is a phenomenon due to a "cognitive illusion"...* | *... or to an environmental structure plus an unbiased mind?* |
|---|---|
| Overconfidence bias (defined as miscalibration) | "Miscalibration" can be deduced from an unbiased mind in an environment with unsystematic error, causing regression toward the mean (Dawes & Mulford, 1996; Erev et al., 1994) |
| Overconfidence bias (defined as mean confidence minus proportion correct) | "Overconfidence bias" can be deduced from an unbiased mind in an environment with unrepresentative sampling of questions; disappears largely with random sampling (Juslin, Winman, & Olsson, 2000) |
| Hard-easy effect | "Hard-easy effect" can be deduced from an unbiased mind in an environment with unsystematic error, causing regression toward the mean (Juslin et al., 2000) |
| Overestimation of low risks and underestimation of high risks | This classical phenomenon can be deduced from an unbiased mind in an environment with unsystematic error, causing regression toward the mean (Gigerenzer & Fiedler, 2004) |
| Contingency illusion | "Contingency illusion" can be deduced from an unbiased mind performing significance tests on samples with unequal sizes, such as minorities and majorities (Fiedler, Walther, & Nickel, 1999) |
| Most drivers say they drive safer than average | The distribution of the actual numbers of accidents is highly skewed, which results in the fact that most drivers (80% in one US study) have less accidents than the average number of accidents (Lopes, 1992; Gigerenzer, 2002) |
| Availability bias (letter "R" study) | "Availability bias" largely disappears when the stimuli (letters) are representatively sampled rather than selected (Sedlmeier, Hertwig, & Gigerenzer, 1998) |
| Preference reversals | Consistent social values (e.g., don't take the largest slice; don't be the first to cross a picket line) can create what look like preference reversals (Sen, 2002) |
| Probability matching | Probability matching is suboptimal for an individual studied in isolation, but not necessarily for individuals in an environment of social competition (Gallistel, 1990) |
| Conjunction fallacy | "Conjunction fallacy" can be deduced from the human capacity for semantic inference in social situations (Hertwig & Gigerenzer, 1999) |
| False consensus effect | This "egocentric bias" can be deduced from Bayes' rule for situations where a person has no knowledge about prior probabilities (Dawes & Mulford, 1996) |
| Violations of logical reasoning | A number of apparent "logical fallacies" can be deduced from Bayesian statistics for environments where the empirical distribution of the events (e.g., P, Q, and their negations) is highly skewed (McKenzie & Amin, 2002; Oaksford & Chater, 1994) and from the logic of social contracts (Cosmides & Tooby, 1992) |

The general argument is that an unbiased mind plus environmental structure (such as unsystematic error, unequal sample sizes, skewed distributions) is *sufficient* to produce the phenomenon. Note that other factors can also contribute to some of the phenomena. The moral is not that people would never err, but that in order to understand good and bad judgments, one needs to analyze the structure of the problem or of the natural environment.

Simon's vision of bounded rationality was neither optimization under constraints nor cognitive illusions (Gigerenzer, 2004). His notion of rationality was an ecological one: the match between mind and environment. This is best illustrated in an analogy he offered: "Human rational behavior is shaped by a scissors whose blades are the structure of task environments and the computational capabilities of the actor" (Simon, 1990, p. 7). If looking only at one blade, one cannot fully understand how the human mind works, just as one cannot understand how scissors cut.

For instance, as Table 4.1 illustrates, as soon as researchers began to study the structure of information in the environment, what looked like a dull cognitive illusion often turned out to be a sharp pair of scissors. In a series of experiments, for example, participants answered general-knowledge questions. The typical finding was that when participants were 100 percent confident of giving a correct answer, the average number correct was lower, such as 80 percent. This phenomenon was labeled *overconfidence bias* or *miscalibration* and interpreted as a cognitive illusion. A glance at the environmental structure, however, reveals a large unsystematic error, which *in the absence of any cognitive bias* leads to regression towards the mean, that is, the average number correct is always lower than a high confidence level. When one plots the data the other way round, the unsystematic error produces a pattern that looks like *underconfidence*: When participants answered 100 percent correctly, their mean confidence was lower, such as 80 percent. Rather than being a cognitive illusion, the phenomenon seems largely a consequence of environments with substantial unsystematic error (Erev, Wallsten, & Budescu, 1994; but see Chapter 9, this volume, for a different view).

## Models of Heuristics

A model of a heuristic specifies: (1) a process rule; (2) the capacities that the rule exploits to be simple; and (3) the kinds of problems the heuristic can solve, that is, the structures of environments in which it is successful. The latter two are Simon's blades. Models of heuristics need to be distinguished from mere labels. For instance, terms such as *representativeness* and *availability* are common-sense labels without specification of a process and the conditions under which a heuristic succeeds and fails. These need to be developed into testable models; otherwise they can post hoc account for almost everything (see Gigerenzer, 1996, 2000; Gigerenzer & Murray, 1987, Chapter 5; Gigerenzer & Regier, 1996; Kahneman & Tversky, 1996).

There do already exist a number of testable models for heuristics, such as satisficing (Selten, 2001; Simon, 1982), elimination by aspect (Tversky, 1972), and various heuristics for multiattribute choice discussed in Payne, Bettman, & Johnson (1993), and Chapter 6, this volume. Much of this earlier work addressed heuristics for preferences, not for inferences, that is, for problems where no external criterion of success exists. Criteria for the accuracy of heuristics were typically internal, such as whether they used all of the information or how closely they mimicked the gold standard of a weighted additive model. Because there were no external criteria for accuracy, the true power of heuristics could not be demonstrated. Some concluded that heuristics generally lead to

irrational judgments, at best to second-best choices. For instance, when Keeney and Raiffa (1993) discussed lexicographic heuristics, they repeatedly inserted warnings that such a 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. 77–8). But the authors failed to report such a test – preferences alone cannot reveal how accurate heuristics really are. In this chapter I will report such tests.

I will focus on heuristics for inferences – such as comparative judgments, classification, and estimation. From the seminal work on heuristics with simple unit weights (such as +1 and −1; see Dawes, 1979), we know that the predictive accuracy of simple heuristics can be as high as or higher than that of the gold standard of weighing and adding. For instance, unit weights predicted the academic performance of students as well as or better than multiple regression (Dawes & Corrigan, 1974), and the Take The Best heuristic predicted the outcomes of the basketball games in the 1996 NBA season as well as Bayes's rule, but did so faster and with less information (Todorov, 2002). Models of heuristics for classification, estimation, comparative judgments, and choice are discussed in Gigerenzer, Todd, and the ABC Research Group (1999), Gigerenzer and Selten (2001), and Todd and Gigerenzer (2000). In what follows, I will select a few heuristics and discuss the ecological rationality and the empirical evidence.

## Recognition Heuristic

Imagine you are a contestant in a TV game show and face the $1 million question: "Which city has more inhabitants: San Diego or San Antonio?" What is your answer? If you are American, then your chances of finding the right answer, San Diego, are not bad. Some two thirds of undergraduates at the University of Chicago did (Goldstein & Gigerenzer, 2002). If, however, you are German, your prospects look dismal because most Germans know little about San Diego, and many have not even heard of San Antonio. How many correct inferences did the less knowledgeable German group that we tested achieve? Despite a considerable lack of knowledge, 100 percent of the Germans answered the question correctly. How can people who know less about a subject nevertheless make more correct inferences? The answer is that the Germans used a fast and frugal heuristic, the recognition heuristic: If you recognize the name of one city but not the other, then infer that the recognized city has the larger population. The Americans could not use the heuristic, because they had heard of both cities. They knew too much.

The recognition heuristic is useful when there is a strong correlation – in either direction – between recognition and criterion. For simplicity, we assume that the correlation is positive. For two-alternative choice tasks, the heuristic can be stated as follows:

> *Recognition heuristic*: If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value with respect to the criterion.

The recognition heuristic builds on an evolved capacity for recognition – such as face, voice, and name recognition. No computer program yet exists that can perform face recognition as well as a human child does. Note that the capacity for recognition is

different from that for recall. For instance, one may recognize a face but not recall anything about who that person is (Craik & McDowd, 1987).

> *Ecological rationality*: The recognition heuristic is successful when ignorance is systematic rather than random, that is, when recognition is strongly correlated with the criterion.

The direction of the correlation between recognition and the criterion can be learned from experience, or it can be genetically coded. Substantial correlations exist in competitive situations, such as between name recognition and the excellence of colleges, the value of the products of companies, and the quality of sports teams. Consider forecasting the outcomes of the 32 English FA Cup third-round soccer matches, such as Manchester United versus Shrewsbury Town. Ayton and Önkal (1997) tested 50 Turkish students and 54 British students. The Turkish participants had very little knowledge about (or interest in) English soccer teams, while the British participants knew quite a bit. Nevertheless, the Turkish forecasters were nearly as accurate as the English ones (63 percent versus 66 percent correct). Their predictions were consistent with the recognition heuristic in 627 out of 662 cases (95 percent). Experimental studies by Goldstein and Gigerenzer (2002) indicate that in situations where the recognition heuristic is ecologically rational, people rely on it in about 90 percent of all cases.

One way to measure the degree of ecological rationality of the recognition heuristic (the correlation between recognition and criterion) is the *recognition validity* $\alpha$, which is the proportion of times a recognized object has a higher criterion value than an unrecognized object in a reference class, such as cities, companies, or sports teams:

$$\alpha = R/(R + W) \tag{4.1}$$

where $R$ is the number of correct (right) inferences the recognition heuristic would achieve, computed across all pairs in which one object is recognized and the other is not, and $W$ is the number of incorrect (wrong) inferences under the same circumstances.

The recognition heuristic should not be confused with labels such as *availability* (Tversky & Kahneman, 1974) or *familiarity* (Griggs & Cox, 1982). Availability refers to ease of recall, not recognition. The recognition heuristic implies several counterintuitive phenomena that cannot be deduced from any other theory I am aware of. For instance, recognition information tends to dominate further knowledge, in rats as well as in people, even if there is conflicting evidence (Goldstein & Gigerenzer, 2002). Next, I will deduce a counterintuitive phenomenon, the *less-is-more effect*, and the conditions under which it will occur.

### *The less-is-more effect*

Equation 4.2 specifies the proportion of correct answers $c$ on an exhaustive test of all pairs of $N$ objects (such as cities, soccer teams) for a person that recognizes $n$ of these objects:

$$c = \frac{2n(N-n)}{N(N-1)}\alpha + \frac{(N-n)(N-n-1)}{N(N-1)}\frac{1}{2} + \frac{n(n-1)}{N(N-1)}\beta \tag{4.2}$$

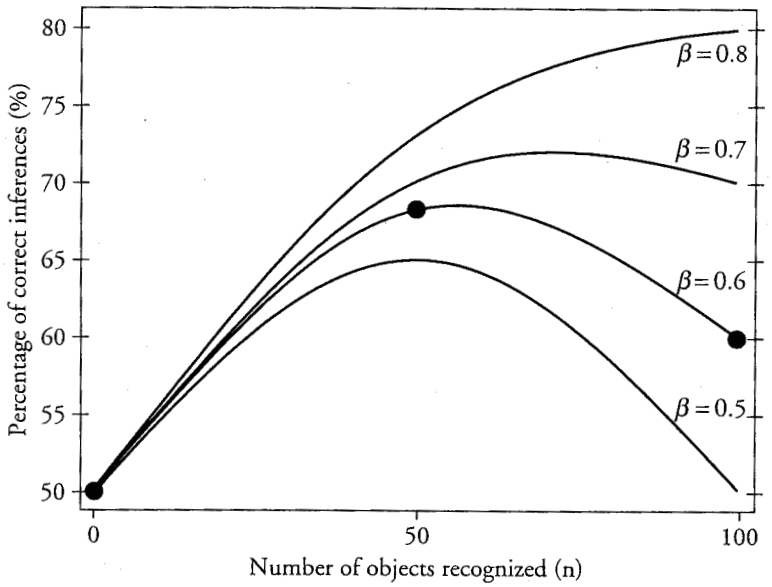


**Figure 4.1** The less-is-more effect is a consequence of the recognition heuristic

The effect occurs when the recognition validity $\alpha$ is larger than the knowledge validity $\beta$. The curves shown are for $\alpha = 0.8$. A less-is-more effect can occur between people with the same $\beta$, as shown by the middle and right-hand points. It can also occur between people with different knowledge validities. For instance, a person who recognizes only half of the objects ($n = 50$) and has no useful knowledge ($\beta = 0.5$) will make more correct inferences than a person who recognizes all objects ($n = 100$) and has useful knowledge ($\beta = 0.6$)

The three terms on the right side of the equation correspond to the three possibilities: a person recognizes one of the two objects, none, or both. The first term accounts for the correct inferences made by the recognition heuristic, the second term for guessing, and the third term equals the proportion of correct inferences made when knowledge beyond recognition is used. The *knowledge validity* $\beta$ is the relative frequency of getting a correct answer when both objects are recognized, which is computed like the recognition validity. All parameters in Equation 4.2 can be independently measured.

When one plots Equation 4.2, a counterintuitive implication can be seen (Figure 4.1). Consider first the curve for $\beta = 0.5$, that is, for people who have no predictive knowledge beyond recognition. A person that has heard of none of the objects will perform at chance level (50 percent, left side). A person that has heard of all objects will also perform at chance level (50 percent, right side). Only a person who has heard of some but not all objects can use the recognition heuristic, and her accuracy will first increase with $n$ but then decrease again. The reason is that the recognition heuristic can be used most often when about half of the objects are recognized, in comparison to when all or none are recognized. When half of the objects are recognized, a person can use the recognition heuristic about half of the time, which results in some 65 percent (40 percent for $\alpha = 0.8$ plus 25 percent for guessing) correct inferences, as can be calculated from Equation 4.2. The next curve with three dots shows a less-is-more effect in the presence

of knowledge beyond mere recognition, for $\beta = 0.6$. The left dot represents a person who has not heard of any objects, while the dot on the right represents someone who has heard of all objects and has recall knowledge that does better than chance. The middle dot represents a person who recognizes less objects but gets more correct inferences. In general, assuming that $\alpha$ and $\beta$ are constant, the following result can be proven (Goldstein & Gigerenzer, 2002):

*Less-is-more effect*: The recognition heuristic will yield a less-is-more effect if $\alpha > \beta$.

A less-is-more effect can emerge in at least three different situations. First, it can occur between two groups of people, when a more knowledgeable group makes worse inferences than a less knowledgeable group in a given domain. An example is the performance of the American and German students on the question of whether San Diego or San Antonio is larger. Second, a less-is-more effect can occur between domains, that is, when the same group of people achieve higher accuracy in a domain in which they know little than in a domain in which they know a lot. For instance, when American students were tested on the 22 largest American cities (such as New York versus Chicago) and on the 22 largest German cities (such as Cologne versus Frankfurt), they scored a median 71.0 percent (mean 71.1 percent) correct on their own cities but slightly higher on the less familiar German cities, with a median of 73.0 percent correct (mean 71.4 percent). This effect was obtained despite a handicap: Many Americans already knew the three largest US cities in order, and did not have to make any inferences (Goldstein & Gigerenzer, 2002). A similar less-is-more effect was demonstrated with Austrian students, whose scores for correct answers were slightly higher for the 75 largest American cities than for the 75 largest German cities (Hoffrage, 1995; see also Gigerenzer, 1993). Third, a less-is-more effect can occur during knowledge acquisition, that is, when an individual's performance curve first increases but then decreases again.

### *Less-is-more in groups*

Consider now group decision making. Three people sit in front of a computer screen on which questions such as "Which city has more inhabitants: San Diego or San Antonio?" are displayed. The task of the group is to find the correct answer through discussion, and they are free to use whatever means. In this task, the correct solution is difficult to "prove" by an individual group member; thus, one might expect that the majority determines the group decision (the *majority rule*; see Gigone & Hastie, 1997). Consider now the following conflict. Two group members have heard of both cities and each concluded independently that city *A* is larger. But the third group member has not heard of *A*, only of *B*, and concludes that *B* is larger (relying on the recognition heuristic). After the three members finished their negotiation, what will their consensus be? Given that two members have at least some knowledge about both cities, one might expect that the consensus is always *A*, which is also what the majority rule predicts. In fact, in more than half of all cases (59 percent), the group voted for *B* (Reimer & Katsikopoulos, 2003). This number rose to 76 percent when two members relied on mere recognition.

That group members let their knowledge be dominated by others' lack of recognition may seem odd. But in fact this apparently irrational decision increased the overall accuracy of the group. This result can be analytically deduced (Reimer & Katsikopoulos, 2003) and intuitively seen from Figure 4.1. When the recognition heuristic is used in group decisions, then a less-is-more effect results if $\alpha > \beta$, just as in Figure 4.1, but more strongly. Consistent with the theory, Reimer and Katsikopoulos (2003) observed when two groups had the same average $\alpha$ and $\beta$, the group who recognized *fewer* cities (smaller $n$) typically had *more* correct answers. For instance, the members of one group recognized on average only 60 percent of the cities, and those in a second group 80 percent; but the first group got 83 percent answers correct in a series of over 100 questions, whereas the second only 75 percent. Thus, group members seem to intuitively trust the recognition heuristic, which can improve accuracy and lead to the counterintuitive less-is-more effect between groups.

## Heuristics Based on Reasons

When recognition is not valid, or people know too much, heuristics can involve search for reasons or cues. A few years after his voyage on the *Beagle*, the 29-year-old Charles Darwin divided a scrap of paper (titled, "This is the Question") into two columns with the headings "Marry" and "Not Marry" and listed supporting reasons for each of the two possible courses of action, such as "nice soft wife on a sofa with good fire" opposed to "conversation of clever men at clubs." Darwin concluded that he should marry, writing "Marry – Marry – Marry Q.E.D" decisively beneath the first column (Darwin, 1969[1887], pp. 232–3). The following year, Darwin married his cousin, Emma Wedgwood, with whom he eventually had 10 children. How did Darwin decide to marry, based on the possible consequences he envisioned – children, loss of time, a constant companion? He did not tell us. But we can use his "Question" as a thought experiment to illustrate various visions of decision making.

Darwin searched in his memory for reasons. There are two visions of search: optimizing search and heuristic search. Following Wald's (1950) optimizing models of sequential analysis, several psychological theories postulated versions of sequential search and stopping rules (e.g., Busemeyer & Townsend, 1993; see Chapter 7, this volume). In the case of a binary hypothesis (such as to marry or not marry), the basic idea of most sequential models is the following: A threshold is calculated for accepting one of the two hypotheses, based on the costs of the two possible errors, such as wrongly deciding that to marry is the better option. Each reason or observation is then weighted and the evidence is accumulated until the threshold for one hypothesis is met, at which point search is stopped, and the hypothesis is accepted. If Darwin had followed this procedure, he would have had to estimate, consciously or unconsciously, how many conversations with clever friends are equivalent to having one child, and how many hours in a smoky abode can be traded against a lifetime of soft moments on the sofa. Weighting and adding is a mathematically convenient assumption, but it assumes that there is a common currency for all beliefs and desires in terms of quantitative probabilities and utilities. These models

are often presented as *as-if* models, whose task is to predict the outcome rather than the process of decision making, although it has been suggested that the calculations might be performed unconsciously using the common currency of neural activation.

The second vision of search is that people use heuristics – either social heuristics or reason-based heuristics – that exploit some evolved capacities. Social heuristics exploit the capacity of humans for social learning and imitation (imitation need not result in learning), which is unmatched among the animal species. For instance, the following heuristic generates social facilitation (Laland, 2001):

> *Do-what-the-majority-do heuristic*: If you see the majority of your peers display a behavior, engage in the same behavior.

For the marriage problem, this heuristic makes a man start thinking of marriage at a time when most other men in one's social group do, say, around age 30. It is a most frugal heuristic, for one does not even have to think of pros and cons. Do-what-the-majority-do tends to be ecologically rational when (1) the observer and the demonstrators of the behavior are exposed to similar environments that (2) are stable rather than changing, and (3) noisy, that is, where it is hard to see what the immediate consequence of one's action is (Boyd & Richerson, 1985; Goldstein, Gigerenzer, Hogarth, et al., 2001).

Darwin, however, seems to have based his decision on reasons. I will describe two classes of heuristics that search for reasons. Unlike optimizing models, they do not both weight and add cues. One class of heuristics dispenses with adding, and searches cues in order (a simple form of weighing). I will refer to this class as *one-reason decision making*. The second class dispenses with weighing, and adds up cues until a threshold is met. I will refer to the second class as *tallying* heuristics. Each of the heuristics consists of three building blocks: a rule for search, stopping, and decision making. I will specify some of the conditions under which each class of heuristics will be successful, and in order to do this, I will turn to inference rather than preference.

## *Take The Best and tallying*

Consider the task of predicting which alternative, $a$ or $b$, has the higher value on a criterion, where $a$ and $b$ are elements of a set of $N$ alternatives (which can be actions, objects, events). The prediction can be based on $M$ binary cues ($1, 2, \ldots, i, \ldots, M$), where the cue values 1 and 0 indicate higher and lower criterion values, respectively. To illustrate, consider an experiment by Newell, Weston, & Shanks (2003). The participants were presented with a series of choices between the shares of two fictional companies. In each trial, two companies were presented on a computer screen, and the participants were asked to infer which share would prove to be more profitable. To help find the more profitable share, participants could acquire information concerning six cues, such as: "Does the company invest in new projects?" and "Does the company have financial reserves?" The cost of information about each cue was 1p (penny). After participants had bought as many cues as they desired, they made their choice, and feedback was given whether the answer was correct. When the answer was correct, the participants received

7p minus the amount they had spent searching for information. How do people make an inference when they have to search for information?

One hypothesis about how people make inferences is the *Take The Best* heuristic (Gigerenzer & Goldstein, 1999), which is a form of one-reason decision making. It consists of three building blocks: a search rule, a stopping rule, and a decision rule:

Take The Best:

1 Search by validity: Search through cues in order of their validity. Look up the cue values of the cue with the highest validity first.
2 One-reason stopping rule: If one object has a positive cue value (1) and the other does not (0 or unknown), then stop search and proceed to Step 3. Otherwise exclude this cue and return to Step 1. If no more cues are found, guess.
3 One-reason decision making: Predict that the object with the positive cue value (1) has the higher value on the criterion.

The validity of a cue $i$ is defined as $v_i = R_i/P_i$, where $R_i$ = number of correct predictions by cue $i$, and $P_i$ = number of pairs where the values of cue $i$ differ between objects. In the Newell, Weston and Shanks (2003) task, for example, the participant would start by looking up the most valid cue for predicting profitability, and see if the two companies differed with respect to that cue. If they did, the participant would stop search and choose accordingly; if not, the participant would look up the next most valid cue, and repeat the process until a choice is made. By using this stopping rule, participants can draw inferences without having to look up all of the available cue values.

Now consider an example for a tallying heuristic, which relies on adding but not on weighing (or order):

Tallying:

1 Random search: Search through cues in random order. Look up the cue values.
2 Stopping rule: After $m$ ($1 < m \leq M$) cues, stop search and determine which object has more positive cue values (1), and proceed to Step 3. If the number is equal, return to Step 1 and search for another cue. If no more cues are found, guess.
3 Tallying rule: Predict that the object with the higher number of positive cue values (1) has the higher value on the criterion.

Versions of tallying have been discussed in the literature, such as unit-weight models in which all cues ($m = M$), or the $m$ significant cues are looked up (Dawes, 1979). Unlike as-if models, which predict outcomes only, these models of heuristics predict process and outcome, and can be subjected to a stronger test. In Newell, Weston, and Shanks's study (2003), each of the three building blocks was tested independently.

*Search rule.* In theory, participants can search through cues in many different ways. If they looked up all six cues (which is unlikely, given the pay-off function), there would be 6! = 720 different orders. The search rule of the tallying heuristic does not predict a specific order, but the search rule from the Take The Best heuristic makes a strong prediction. People will search by one of these orders, the one defined by $v_i$. In order to learn the validities, Newell, Weston, and Shanks (2003) exposed each participant to 120 learning trials, with feedback (correct/incorrect) given after each response. The six cues varied in their validity. The learning phase was followed by a test phase with 60 trials.

During the test phase, 75 percent of the participants followed the search rule of Take The Best. When there were only two cues, this number increased to 92 percent. Thus, the great majority of participants did not search randomly, but in order of validity.

*Stopping rule.* The logical possibilities for stopping search are fewer than those for search. There are six possibilities, after the first, second, . . ., sixth cue (not counting the possibility that people would not search but simply guess). Tallying postulates that participants add up more than one cue, but leaves open how many (i.e., the number $m$ must be independently estimated). In contrast, Take The Best postulates that search is stopped immediately after the first discriminating cue is found, not beforehand and not later. Note that each stopping rule can be valid independent of the results for the search rule. For instance, people can search in one of the 719 orders not consistent with $v_i$ but stop after the first discriminating cue is found, or search can follow validity but is only stopped after all cues have been looked up. Thus, the empirical result for the search rule does not constrain the stopping behavior. Newell, Weston, and Shanks (2003) reported that in 80 percent of all cases (where participants bought any information at all), participants did not continue beyond a single discriminating cue, and this number increased to 89 percent when there were only two cues. This means that the great majority stopped search immediately after they found the first cue that made a difference.

*Decision rule.* In theory, participants can use infinite ways to combine the information concerning six cues. This includes linear models, weighted or unweighted. If a person follows the one-reason stopping rule, this constrains the ways to arrive at a decision (whereas, as mentioned before, the search rules impose no constraints on the stopping and decision rules). If only one piece of discriminating information is obtained, it seems that the only reasonable decision rules left are forms of one-reason decision making. The multiple-reason stopping rule, in contrast, would not constrain possible decision rules. Newell, Weston, and Shanks (2003) report that the decision rule of Take The Best was followed by their participants in 89 percent of trials, both for six and two cues.

There are now a substantial number of experiments that have analyzed under what conditions people use Take The Best (e.g., Bröder, 2000, 2003; Newell & Shanks, 2003; Newell, Weston, and Shanks, 2003) and where Take The Best was compared with other heuristics or optimizing models in the same task (Bröder, 2000, 2002; Bröder & Schiffer, 2003a, 2003b; Lee & Cummins, in press; Newell, Rakow, Weston, & Shanks, in press; Rieskamp & Hoffrage, 1999; Rieskamp & Otto, 2004; Todorov, 2002). Comparatively little experimental work has examined tallying (Bröder, 2000; Rieskamp & Hoffrage, 1999). One-reason decision making has been observed in high-stake decisions. British magistrates tend to make bail decisions on the basis of one good reason only (Dhami, 2003; Dhami & Ayton, 2001), and so do British general practitioners when they prescribe lipid-lowering drugs (Dhami & Harries, 2001). Many parents rely on one reason to decide which doctor to drive to in the night when their child becomes seriously ill (Scott, 2002).

Take The Best and tallying have been proposed and tested as components of a number of judgmental processes, such as in probabilistic mental models theory (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Slegers, Brake, & Doherty, 2000) and RAFT, the first process model for the hindsight bias (Hoffrage, Hertwig, & Gigerenzer, 2000). We know from many studies that hindsight bias sometimes occurs and sometimes does not. The process model can predict for each participant and question whether hindsight bias

Hindsight Bias: Confidence

***Question: Which has more cholesterol, cake or pie?***

| Cues | Original | Feedback | Recall |
|---|---|---|---|
| Saturated fat (80%) | cake ? pie | "cake" | cake > pie |
| Calories (70%) | cake > pie | | stop search |
| Protein (60%) | stop search | | |
| Choice | cake | | cake |
| Confidence | 70% | | 80% |

**Figure 4.2** A process model of hindsight bias

Participants learn cues (saturated fat, calories, protein) and their validities (in parentheses) in order to judge which of two supermarket food items has more cholesterol. There are three points in time: original judgment, feedback, and recall of the original judgment. The original judgment is generated by Take The Best, which implies that cues are looked up in memory in the order of validity. In the example given, the first cue, saturated fat, does not stop search, because the participant is ignorant of whether cake or pie has a higher value (indicated by "?"). The second cue stops search because the participant learned that cake has more calories than pie (indicated by ">"). The answer is *cake* and the confidence is 70 percent, that is, the validity of the cue. At the second point in time, feedback is given that *cake* was correct. Feedback is automatically used to update missing information in memory about cues (question marks). Thus, at the time of the recall, the "?" for saturated fat is likely to have changed into ">", following the direction of feedback. Recall of the original judgment again follows Take The Best, but now the first cue stops search and the recalled answer is *cake* as before, whereas the recalled confidence is 80 percent, which is incorrect and known as hindsight bias. By manipulating what participants know and don't know, one can predict for each question whether hindsight bias will occur or not. *Source*: Hoffrage et al. (2000)

will or will not occur (Figure 4.2). The bias itself seems to be a byproduct of an adaptive memory updating process.

## Ecological Rationality

What structures of environments can each of the two heuristics exploit? Consider a situation with five binary cues, as in Figure 4.3 (left), where the weights correspond to the order of cues in Take The Best. In an environment where the weights of the cues (e.g., beta weights) decrease exponentially, such as $^1/_2$, $^1/_4$, $^1/_8$, and so on, no linear model, including multiple regression, can outperform the faster and more frugal Take The Best. The proof is in Martignon & Hoffrage (1999, 2002). One can see this result intuitively because the sum of all cue weights to the right of a cue can never be larger than this cue's weight – they cannot compensate for the cues with higher weights. This type of environment is structured by *noncompensatory information*. Here, relying on one reason and ignoring the rest is as accurate as integrating all reasons by any linear method. Given the superior robustness of frugal heuristics (see below), Take The Best is actually likely to be more accurate.

Tallying will not do well with noncompensatory information. It can exploit environments where the cue weights do not differ much. In the extreme case shown in

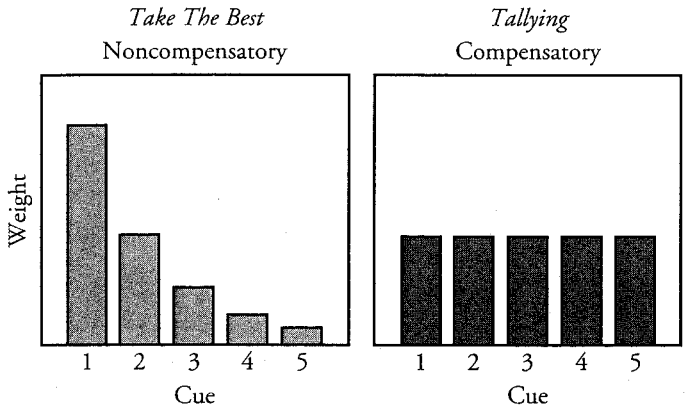


**Figure 4.3** Ecological rationality of Take The Best and tallying

The environment on the left side has perfectly noncompensatory information (the weights of five binary cues decrease exponentially); the one on the right side has perfectly compensatory information (all weights are equal). Take The Best (but not tallying) can exploit noncompensatory information, and tallying (but not Take The Best) can exploit compensatory information.
*Source*: Martignon & Hoffrage (1999)

Figure 4.3 (right), where all cues have the same weights, one can see that a tallying heuristic that tallies all cues ($m = M$) will lead to the same accuracy as any linear model.

There are further structures that these two heuristics can exploit (Forster, Martignon, Vitouch, & Gigerenzer, 2003; Martignon & Hoffrage, 1999, 2002). For instance, when information is scarce, that is, the data points are few compared to the number of cues, then Take The Best and other simple heuristics will generally be of advantage compared to multiple regression and other statistical models that need large learning samples (Gigerenzer et al., 1999; Chater, Oaksford, Nakisa, & Redington, 2003).

## How Do People Know Which Heuristic to Use?

Research suggests that people hardly ever make conscious decisions about which heuristic to use, but that they quickly and unconsciously tend to adapt heuristics to changing environments, provided there is feedback (Payne et al., 1993). This adaptive process is illustrated by an experiment by Rieskamp and Otto (2004). Participants took on the role of bank consultants with the task of evaluating which of two companies applying for a loan was more creditworthy on the basis of six cues, such as qualification of employees and profitability (similar to the experiment by Newell et al., 2003, except that there were no costs for looking up cue values). For the first 24 pairs of companies, no feedback was provided as to the correctness of the participant's inference. Participants followed Take The Best in only about 30 percent of the cases, which is not unusual for situations where information is free. In the following trials, feedback was given. For one group of participants, the environment was noncompensatory (see Figure 4.3), that is, the company that was more creditworthy was determined by the cue with the highest validity (on which

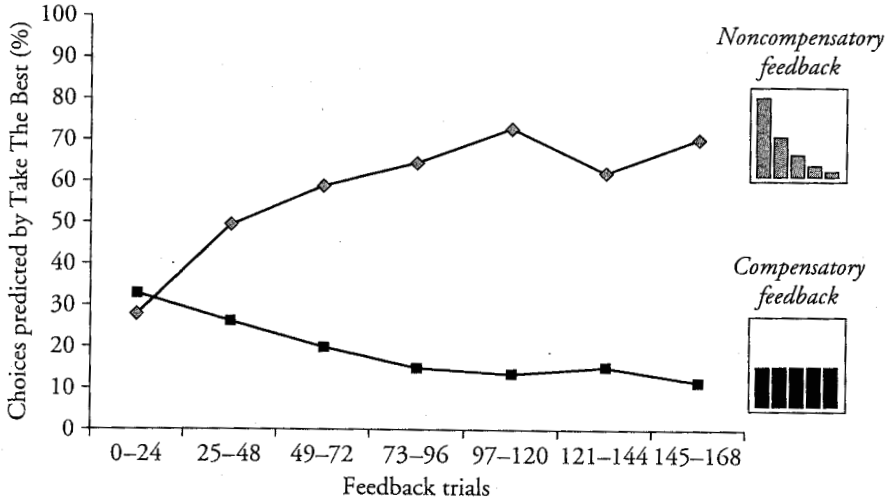


**Figure 4.4** How people adapt their heuristics to the structure of environment
*Source*: Based on Rieskamp & Otto (2004)

the two companies differed) in about 90 percent of the cases. For the second group, the environment was compensatory, that is, feedback was determined by the weighted additive rule in about 90 percent of the cases. Did people intuitively adapt their heuristics to the structure of the environment? As can be seen in Figure 4.4, feedback about the structure of the environment changed the frequency of using Take The Best. People learned without instruction that different heuristics are successful in different environments. Bröder (2003) reported that people with higher IQs are better at detecting the structure of the environment, and consequently in knowing which heuristic to use. While individual correlates of strategy use are difficult to find, individual correlates of strategy adaptation seem to be easier to demonstrate.

This experiment illustrates individual learning by feedback. Which heuristic to use for which problem can also be learned by evolutionary and cultural learning. For instance, a female guppy comes already equipped with a heuristic for mate choice, which resembles Take The Best (Dugatkin, 1996). When she has to decide between two potential mates, the most important cue seems to be the extent of orange color. If one male is noticeably more orange than the other, this cue is sufficient to stop search and decide in favor of him. Evolutionary learning is slowest, while social learning is the fastest way to learn what heuristic to use when. A novice baseball outfielder, pilot, or sailor can be taught the gaze heuristic in a few minutes.

## Robustness

A good heuristic needs to be robust. *Robustness* is the ability to make predictions about the future or new events, whereas *fitting* refers to the ability to fit the past or already known

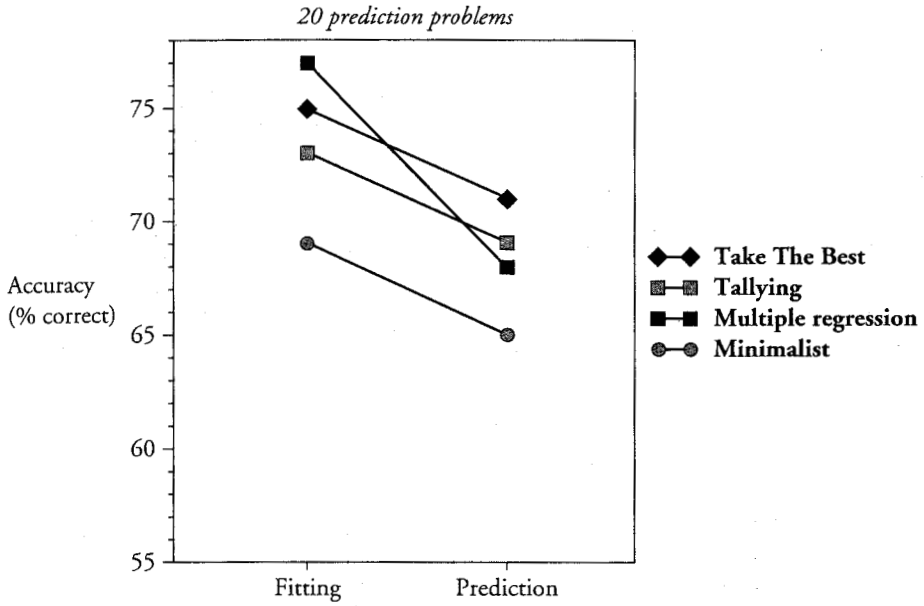


**Figure 4.5** Simplicity can lead to higher predictive accuracy

The predictive power of three heuristics is compared to that of multiple regression in 20 problems. Two of the heuristics (Take The Best and minimalist) are from the one-reason decision-making family; the third is from the tallying family (unit weight linear model). The 20 prediction problems include psychological, economic, environmental, biological, and health problems. Most were taken from statistical textbooks, where they served as good examples for the application of multiple regression. The number of cues varied between 3 and 19, and these were binary or dichotomized at the median. For each of the 20 problems and each of the four strategies, the 95 percent confidence intervals were $\leq 0.42$ percentage points. Although multiple regression has the best fit, two of the heuristics have higher predictive accuracy.
*Source*: Czerlinski et al. (1999)

data. An excellent fit can mean little more than *overfitting* (Mitchell, 1997; Roberts & Pashler, 2000). To define overfitting, we need to distinguish between a learning sample from which a model estimated its parameters and the test sample on which it is tested. Both samples are randomly drawn from the same population.

> *Definition*: A model $O$ *overfits* the learning sample if there exists an alternative model $O'$ such that $O$ has a smaller error than $O'$ in the learning sample but a larger error in the test sample. In this case, $O'$ is called the more *robust* model.

Consider Figure 4.5, which shows the accuracy of three heuristics compared to multiple regression, averaged across 20 real-world problems (Czerlinski, Gigerenzer, & Goldstein, 1999). In each problem, the task was to predict which of two objects scores higher on a criterion. For instance, one problem was to predict which Chicago public high school has the higher dropout rate. The cues included the attendance rates of the students, the socioeconomic and ethnic compositions of the student bodies, the sizes of the classes, and the scores of the students on various standardized tests. Other problems involved

the prediction of people's attractiveness judgments, of homelessness rates, of professors' salaries, and of adolescents' obesity at age 18. The three heuristics were Take The Best, minimalist (which is like take the best but searches cues in random order), and a tallying heuristic that looks up all cues ($m = M$), that is, a unit-weight linear rule. Take The Best and minimalist were most frugal; they looked up, on average, only 2.4 and 2.2 cues before they stopped search. Tallying and multiple regression looked up all cue information (exhaustive search), which amounted to an average of 7.7 cues. How accurate were the heuristics?

The important point is to distinguish between data fitting and prediction. In data fitting, the test set is the same as the training set, and here it is a mathematical truism that models with more adjustable parameters generally do better. Consequently, multiple regression had the best fit. However, the true test of a model concerns its predictive accuracy, which was tested by cross-validation, that is, the four models learned their parameters on half of the data, and were tested on the other half. The predictive accuracy of Take The Best and tallying was, on average, higher than that of multiple regression. This result may sound paradoxical because multiple regression processed all the information and more than each of the heuristics did.

Figure 4.5 shows that multiple regression overfitted the data relative to both Take The Best and tallying (see also Dawes, 1979). An intuitive way to understand overfitting is the following: a set of observations consists of information that generalizes to the other samples, and of information that does not (e.g., noise). If one extracts too much information from the data, one will get a better fit (a higher *explained variance*), but one will mistake more noise for predictive information. The result can be a substantial decrease in one's predictive power. Note that both forms of simplifying – dispensing either with adding or with weighting – resulted in greater robustness. Minimalist, however, which dispensed with both weighting and adding, extracted too little information from the data.

In general, the predictive accuracy of a model increases with its fit, decreases with its number of adjustable parameters, and the difference between fit and predictive accuracy gets smaller with larger number of data points (Akaike, 1973; Forster & Sober, 1994). The general lesson is that in judgments under uncertainty, one has to ignore information in order to make good predictions. The art is to ignore the right kind. Heuristics that promote simplicity, such as using the best reason that allows one to make a decision and ignore the rest, have a good chance of focusing on the information that generalizes.

These results may appear counterintuitive. More information is always better; more choice is always better – so the story goes. This cultural bias makes contrary findings look like weird oddities (Hertwig & Todd, 2003). Yet experts base their judgments on surprisingly few pieces of information (Shanteau, 1992), and professional handball players make better decisions when they have less time (Johnson & Raab, 2003). People can form reliable impressions of strangers from video clips lasting half a minute (Ambady & Rosenthal, 1993), shoppers buy more when there are fewer varieties (Iyengar & Lepper, 2000), and zero-intelligence traders make as much profit as intelligent people do in experimental markets (Gode & Sunder, 1993). Last but not least, satisficers are reported to be more optimistic and have higher self-esteem and life satisfaction, whereas maximizers excel in depression, perfectionism, regret, and self-blame (Schwartz, Ward, Monterosso, Lyubomirsky, White, & Lehman, 2002). Less can be more.

## The Building Blocks of Heuristics

One way to think of the relation between heuristics and their building blocks is the periodic table in chemistry, where there are many elements but only few particles. Just as the same particles combine to new chemical elements, the same building blocks can construct new heuristics to deal with new tasks. Consider the following problem.

A man is rushed to the hospital with serious chest pains. The doctors suspect acute ischemic heart disease (myocardial infarction) and need to make a decision, and they need to make it quickly: Should the patient be assigned to the coronary care unit or to a regular nursing bed with ECG telemetry? In a Michigan hospital, doctors sent 90 percent of their patients to the coronary care unit. This defensive decision making led to overcrowding, decreased the quality of care provided, and became a health risk for patients who should not have been in the unit. An expert system with some 50 probabilities and a logistic regression did better than the physicians, but physicians do not like to use these systems because they are not transparent, that is, they don't understand them. To find a solution, researchers at the University of Michigan Hospital (Green & Mehr, 1997) used the building blocks of Take The Best to design a classification heuristic in the form of a fast and frugal tree (Figure 4.6). If a patient has a certain anomaly in his electrocardiogram (the so-called ST segment), he is immediately admitted to the coronary

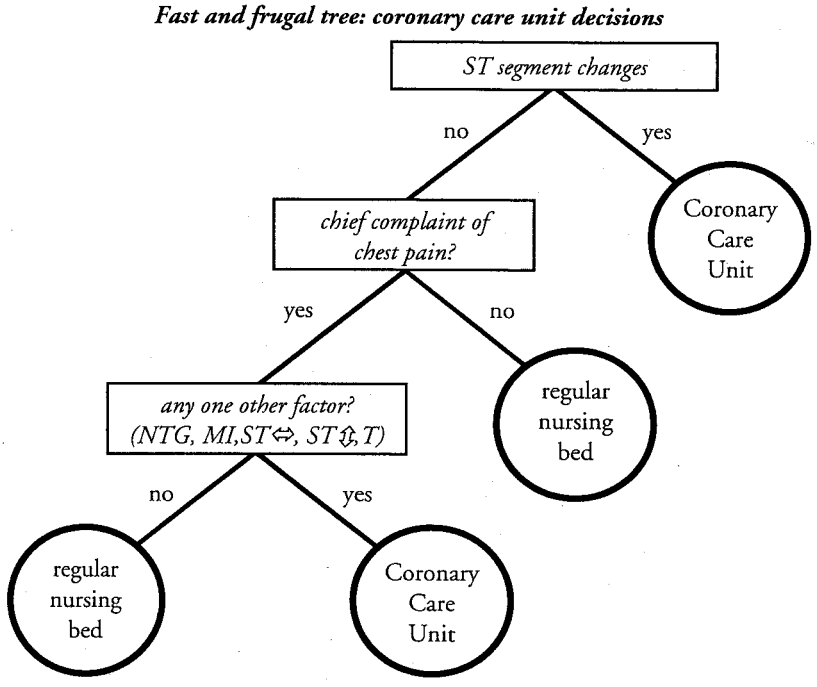


**Figure 4.6** A heuristic for coronary care unit allocation

*Source*: Based on Green & Mehr (1997)

care unit. No other information is required. If that is not the case, a second variable is considered: whether the patient's primary complaint is chest pain. If this is not the case, he is immediately classified as low risk and assigned to a regular nursing bed. No further information is considered. If the answer is yes, then a third, composite question is asked to finally classify the patient.

Green and Mehr (1997) report that the fast and frugal tree was more accurate (as measured by the occurrence of myocardial infarction) than physician's decisions. It had a higher sensitivity and a smaller false positive rate. The heuristic was also more accurate than the expert system that had all the information that the fast and frugal tree had, and more.

A fast and frugal tree is defined as a classification tree that allows for a classification at each level of the tree. It has $M + 1$ exits or end nodes ($M$ is the number of variables or cues). In contrast, the number $2^M$ of end nodes of a complete tree increases exponentially, which makes complete trees computationally intractable for large numbers of variables. A fast and frugal tree has the same building blocks as Take The Best: ordered search, one-reason stopping rule, and one-reason decision making. When the clinical cues follow the structure of Figure 4.3 (left), one can show that no logistic regression can be more accurate than the heuristic. When the clinical cues follow a more compensatory structure (Figure 4.3, right), then a tallying heuristic will be more accurate. A tallying heuristic for the care unit allocation problem is as follows (Forster et al. 2003):

> *Tally Three*: If three positive cue values are present, stop search and send patient to the coronary care unit.

Both heuristics for classification are transparent, that is, physicians can understand their logic easily and therefore are more willing to accept and actually use them than a logistic regression. The predictive accuracy of these heuristics is high: (1) if they can exploit the structure of the environment (see Figure 4.3), and (2) because simplicity tends to promote robustness. Note that the logistic regression may have had an excellent fit in the clinical population where it was originally validated, but it is now applied in a Michigan hospital with a patient population that differs to an unknown extent. That is, unlike the situation in Figure 4.5, where the training and test samples were drawn from the same population, robustness here refers to an unknown population.

Systematic teaching of fast and frugal decision making is currently being introduced into medicine as an alternative to classical decision theory (Elwyn, Edwards, Eccles, & Rovner, 2001), and also as an explication of the superb intuitions of master clinicians (Naylor, 2001; see also Chapter 15, this volume). Fast and frugal trees have been proposed as normative guidelines, such as when to prescribe antibiotics to young children (Fischer et al., 2002).

For classification problems with a larger number of categories, heuristics can rely on *elimination* rather than one-reason decision making or tallying. This building block is part of *categorization by elimination* (Berretty, Todd, & Martignon, 1999) and of a heuristic that can solve problems that involve quantitative estimation, *QuickEst* (Hertwig, Hoffrage, & Martignon, 1999). It was used earlier for preferential choice in *elimination by aspects* (Tversky, 1972). Building blocks for heuristics not covered in this chapter

include aspiration levels (as in mate choice; see Todd & Miller, 1999), social notions of equity (as in parental investment; see Hertwig, Davis, & Sulloway, 2002), social norms (as in conformism; see Boyd & Richerson, 2001), and emotional processes, whose function can be analog to that of cognitive building blocks, yet stronger and longer lasting (see the chapters in Gigerenzer & Selten, 2001). For instance, falling in love can be seen as a powerful stopping rule that ends search for a partner and fosters commitment. Disgust can limit the choice set, and social norms can free us from having to think about how to make decisions all the time.

## The Adaptive Toolbox

The study of smart heuristics is concerned with identifying (1) the building blocks of heuristics, and (2) the structures of environments that a given heuristic can exploit, that is, the kind of problems it can solve. In other words, its first objective is the study of the adaptive toolbox, and the second that of ecological rationality, with aims that are both descriptive and prescriptive. The study of the adaptive toolbox aims at description, including individual differences in the use of heuristics, and the change in the adaptive toolbox over the life course (Gigerenzer, 2003). Models of heuristics allow for qualitative predictions, such as whether players will catch a ball while running or the conditions in which hindsight bias will and will not occur. They also allow for quantitative predictions, such as the proportion of correct answers when using the recognition heuristic. Consistent with Bayesian model testing (MacKay, 1992), the strongest tests can be obtained from: (1) counterintuitive predictions, such as the less-is-more effect in individual and group decision making, and (2) models with zero adjustable parameters, such as the search, stopping, and decision rules of Take The Best.

The study of ecological rationality, in contrast, is descriptive and prescriptive. Its results concerning the match between heuristics and structures of environments can be used to derive hypotheses about people's adaptive use of heuristics. These results also carry prescriptive force. For instance, when the available information is noncompensatory, we can recommend a fast and frugal tree for classification, or Take The Best for paired comparison, because these heuristics will predict as well as any linear model, yet faster, more frugally, and more transparently. If in addition the available information is scarce, we expect that the heuristics will be more accurate, because they tend to be robust. Most problems that worry our minds and hearts are computationally intractable – no machine or mind can find the optimal solution. The systematic study of fast and frugal heuristics can provide normative recommendations on an empirical basis, even when we can never know the best solution.

In this chapter, I invited you to a journey into a land of rationality that is different from the familiar one we know where the sun of enlightenment shines down in beams of logic and probability. The new land of rationality we set out to explore is shrouded in a dim mist of uncertainty, populated with people who have limited time and knowledge, but with smart heuristics at their disposal. Welcome, and I hope you feel at home in this world.

## Acknowledgment

I am grateful for comments by Jerome R. Busemeyer, Anja Dieckmann, Barbara Fasolo, William Goldstein, Nigel Harvey, Ralph Hertwig, Ulrich Hoffrage, Oswald Huber, Derek Koehler, Craig McKenzie, Ben R. Newell, Torsten Reimer, Jörg Rieskamp, Lael Schooler, and Peter Todd.

## References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (eds.), *2nd International Symposium on Information Theory* (pp. 267–81). Budapest: Akademiai Kiado.

Ambady, N. & Rosenthal, R. (1993) Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness, *Journal of Personality and Social Psychology*, 64, 431–41.

Ayton, P. & Önkal, D. (1997) Forecasting football fixtures: Confidence and judged proportion correct. Unpublished manuscript.

Berretty, P. M., Todd, P. M., & Martignon, L. (1999) Categorization by elimination: Using few cues to choose. In G. Gigerenzer, P. M. Todd, and the ABC Research Group, *Simple Heuristics that Make Us Smart* (pp. 235–57). New York: Oxford University Press.

Boyd, R. & Richerson, P. J. (1985) *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Boyd, R. & Richerson, P. J. (2001) Norms and bounded rationality. In G. Gigerenzer and R. Selten (eds.), *Bounded Rationality: The Adaptive Toolbox* (pp. 281–96). Cambridge, MA: MIT Press.

Bröder, A. (2000) Assessing the empirical validity of the "Take The Best"-heuristic as a model of human probabilistic inference, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1332–46.

Bröder, A. (2002) Take The Best, Dawes' Rule, and compensatory decision strategies: A regression-based classification method, *Quality & Quantity*, 36, 219–38.

Bröder, A. (2003) Decision making with the "adaptive toolbox": Influence of environmental structure, intelligence, and working memory load, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 611–25.

Bröder, A. & Schiffer, S. (2003a) Bayesian strategy assessment in multi-attribute decision making, *Journal of Behavioral Decision Making*, 16, 193–213.

Bröder, A. & Schiffer, S. (2003b) Take The Best versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format, *Journal of Experimental Psychology: General*, 132, 277–93.

Busemeyer, J. R. & Townsend, J. T. (1993) Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment, *Psychological Review*, 100, 432–59.

Camerer, C. F. (1998) Bounded rationality in individual decision making, *Experimental Economics*, 1, 163–83.

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

Conlisk, J. (1996) Why bounded rationality? *Journal of Economic Literature*, 34, 669–700.

Cosmides, L. & Tooby, J. (1992) Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, and J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 163–228). Oxford: Oxford University Press.

Craik, F. I. M. & McDowd, M. (1987) Age differences in recall and recognition, *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 474–79.

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

Darwin, C. (1969) *The Autobiography of Charles Darwin, 1809–1882.* New York: Norton. (Original work published 1887.)

Dawes, R. M. (1979) The robust beauty of improper linear models in decision making, *American Psychologist*, 34, 571–82.

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

Dawes, R. M. & Mulford, M. (1996) The false consensus effect and overconfidence: Flaws in judgment, or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes*, 65, 201–11.

Dawkins, R. (1976) *The Selfish Gene.* Oxford: Oxford University Press.

Dhami, M. K. (2003) Psychological models of professional decision-making, *Psychological Science*, 14, 175–80.

Dhami, M. K. & Ayton, P. (2001) Bailing and jailing the fast and frugal way, *Journal of Behavioral Decision Making*, 14, 141–68.

Dhami, M. K. & Harries, C. (2001) Fast and frugal versus regression models in human judgement, *Thinking & Reasoning*, 7, 5–27.

Dugatkin, L. A. (1996) Interface between culturally based preferences and genetic preferences: Female mate choice in Poecilia reticulata, *Proceedings of the National Academy of Sciences, USA*, 93, 2770–3.

Elwyn, G., Edwards, A., Eccles, M., & Rovner, D. (2001) Decision analysis in patient care, *The Lancet*, 358, 571–4.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994) Simultaneous over- and underconfidence: The role of error in judgment processes, *Psychological Review*, 101, 519–27.

Fiedler, K., Walther, E., & Nickel, S. (1999) Covariation-based attribution: On the ability to assess multiple covariates of an effect, *Personality & Social Psychology Bulletin*, 25, 607–22.

Fischer, J. E., Steiner, F., Zucol, F., Berger, C., Martignon, L., Bossart, W., et al. (2002) Use of simple heuristics to target macrolide prescription in children with community-acquired pneumonia, *Archives of Pediatrics and Adolescent Medicine*, 156, 1005–8.

Forster, M., Martignon, L., Vitouch, O., & Gigerenzer, G. (2003, forthcoming) Simple heuristics versus complex predictive instruments: Which is better and why? Manuscript submitted for publication.

Forster, M. & Sober, E. (1994). How to tell when simpler, more unified, and less *ad hoc* theories will provide more accurate predictions, *British Journal of the Philosophy of Science*, 45, 1–35.

Gallistel, C. R. (1990) *The Organization of Learning.* Cambridge, MA: MIT Press.

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

Gigerenzer, G. (1996) On narrow norms and vague heuristics: A rebuttal to Kahneman and Tversky (1996), *Psychological Review*, 103, 592–6.

Gigerenzer, G. (2000) *Adaptive Thinking: Rationality in the Real World.* New York Oxford University Press.

Gigerenzer, G. (2002) *Calculated Risks: How to Know When Numbers Deceive You.* New York: Simon & Schuster. (UK edition: *Reckoning with Risk. Learning to Live with Uncertainty.* London: Penguin.)

Gigerenzer, G. (2003) The adaptive toolbox and lifespan development: Common questions? In U. M. Staudinger and U. E. R. Lindenberger (eds.), *Interactive Minds: Life-span Perspectives on the Social Foundation of Cognition* (pp. 319–46). Cambridge: Cambridge University Press.

Gigerenzer, G. (2004) Striking a blow for sanity in theories of rationality. In M. Augier and J. G. March (eds.), *Models of a Man: Essays in Memory of Herbert A. Simon* (pp. 389–409). Cambridge, MA: MIT Press.

Gigerenzer, G. & Fiedler, K. (2004) Minds in environments: The potential of an ecological approach to cognition. Manuscript submitted for publication.

Gigerenzer, G. & Goldstein, D. G. (1999) Betting on one good reason: The Take The Best heuristic. In G. Gigerenzer, P. M. Todd, and the ABC Research Group, *Simple Heuristics that Make Us Smart* (pp. 75–95). New York: Oxford University Press.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991) Probabilistic mental models: A Brunswikian theory of confidence, *Psychological Review*, 98, 506–28.

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

Gigerenzer, G. & Regier, T. (1996) How do we tell an association from a rule? Comment on Sloman (1996), *Psychological Bulletin*, 119, 23–6.

Gigerenzer, G. & Selten, R. (eds.) (2001) *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press.

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

Gigone, D. & Hastie, R. (1997) The impact of information on small group choice, *Journal of Personality and Social Pscyhology*, 72, 132–40.

Gode, D. K. & Sunder, S. (1993) Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality, *Journal of Political Economy*, 101, 119–37.

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

Goldstein, D. G., Gigerenzer, G., Hogarth, R. M., Kacelnik, A., Kareev, Y., Klein, G., et al. (2001) Group report: Why and when do simple heuristics work? In G. Gigerenzer and R. Selten (eds.), *Bounded Rationality: The Adaptive Toolbox* (pp. 173–90). Cambridge, MA: MIT Press.

Green, L. & Mehr, D. R. (1997) What alters physicians' decisions to admit to the coronary care unit? *The Journal of Family Practice*, 45, 219–26.

Griggs, R. A. & Cox, J. R. (1982) The elusive thematic-materials effect in Wason's selection task, *British Journal of Psychology*, 73, 407–20.

Hertwig, R., Davis, J. N., & Sulloway, F. J. (2002) Parental investment: How an equity motive can produce inequality, *Psychological Bulletin*, 128, 728–45.

Hertwig, R. & Gigerenzer, G. (1999) The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors, *Journal of Behavioral Decision Making*, 12, 275–305.

Hertwig, R., Hoffrage, U., & Martignon, L. (1999) Quick estimation: Letting the environment do the work. In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple Heuristics that Make Us Smart* (pp. 209–34). New York: Oxford University Press.

Hertwig, R. & Todd, P. M. (2003). More is not always better: The benefits of cognitive limits. In D. Hardman and L. Macchi (eds.), *Reasoning and Decision Making: A Handbook* (pp. 213–31). Chichester: Wiley.

Hoffrage, U. (1995) Zur Angemessenheit subjektiver Sicherheits-Urteile: Eine Exploration der Theorie der probabilistischen mentalen Modelle [The adequacy of subjective confidence judgments: Studies concerning the theory of probabilistic mental models]. Unpublished doctoral dissertation, University of Salzburg, Austria.

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000) Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 566–81.

Simon's vision of bounded rationality was neither optimization under constraints nor cognitive illusions (Gigerenzer, 2004). His notion of rationality was an ecological one: the match between mind and environment. This is best illustrated in an analogy he offered: "Human rational behavior is shaped by a scissors whose blades are the structure of task environments and the computational capabilities of the actor" (Simon, 1990, p. 7). If looking only at one blade, one cannot fully understand how the human mind works, just as one cannot understand how scissors cut.

For instance, as Table 4.1 illustrates, as soon as researchers began to study the structure of information in the environment, what looked like a dull cognitive illusion often turned out to be a sharp pair of scissors. In a series of experiments, for example, participants answered general-knowledge questions. The typical finding was that when participants were 100 percent confident of giving a correct answer, the average number correct was lower, such as 80 percent. This phenomenon was labeled *overconfidence bias* or *miscalibration* and interpreted as a cognitive illusion. A glance at the environmental structure, however, reveals a large unsystematic error, which *in the absence of any cognitive bias* leads to regression towards the mean, that is, the average number correct is always lower than a high confidence level. When one plots the data the other way round, the unsystematic error produces a pattern that looks like *underconfidence*: When participants answered 100 percent correctly, their mean confidence was lower, such as 80 percent. Rather than being a cognitive illusion, the phenomenon seems largely a consequence of environments with substantial unsystematic error (Erev, Wallsten, & Budescu, 1994; but see Chapter 9, this volume, for a different view).

## Models of Heuristics

A model of a heuristic specifies: (1) a process rule; (2) the capacities that the rule exploits to be simple; and (3) the kinds of problems the heuristic can solve, that is, the structures of environments in which it is successful. The latter two are Simon's blades. Models of heuristics need to be distinguished from mere labels. For instance, terms such as *representativeness* and *availability* are common-sense labels without specification of a process and the conditions under which a heuristic succeeds and fails. These need to be developed into testable models; otherwise they can post hoc account for almost everything (see Gigerenzer, 1996, 2000; Gigerenzer & Murray, 1987, Chapter 5; Gigerenzer & Regier, 1996; Kahneman & Tversky, 1996).

There do already exist a number of testable models for heuristics, such as satisficing (Selten, 2001; Simon, 1982), elimination by aspect (Tversky, 1972), and various heuristics for multiattribute choice discussed in Payne, Bettman, & Johnson (1993), and Chapter 6, this volume. Much of this earlier work addressed heuristics for preferences, not for inferences, that is, for problems where no external criterion of success exists. Criteria for the accuracy of heuristics were typically internal, such as whether they used all of the information or how closely they mimicked the gold standard of a weighted additive model. Because there were no external criteria for accuracy, the true power of heuristics could not be demonstrated. Some concluded that heuristics generally lead to

Reimer, T. & Katsikopoulos, K. (2003) Group decision-making, recognition heuristic, and less-is-more effect. Manuscript submitted for publication.

Rieskamp, J. & Hoffrage, U. (1999) When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. M. Todd, and the ABC Research Group, *Simple Heuristics that Make Us Smart* (pp. 141–67). New York: Oxford University Press.

Rieskamp, J. & Otto, P. E. (2004) Adaptive strategy selection in decision making: The learning rule theory. Manuscript submitted for publication.

Roberts, S. & Pashler, H. (2000) How persuasive is a good fit? A comment on theory testing, *Psychological Review*, 107, 358–67.

Rosander, K. & Hofsten, C. von (2002) Development of gaze tracking of small and large objects, *Experimental Brain Research*, 146, 257–64.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002) Maximizing versus satisficing: Happiness is a matter of choice, *Journal of Personality and Social Psychology*, 83, 1178–97.

Scott, A. (2002). Identifying and analysing dominant preferences in discrete choice experiments: An application in health care, *Journal of Economic Psychology*, 23, 383–98.

Sedlmeier, P., Hertwig, R., & Gigerenzer, G. (1998) Are judgments of the positional frequencies of letters systematical biased due to availability? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 754–70.

Selten, R. (2001) What is bounded rationality? In G. Gigerenzer and R. Selten (eds.), *Bounded Rationality: the Adaptive Toolbox* (pp. 13–36). Cambridge, MA: MIT Press.

Sen, A. (2002) *Rationality and Freedom*. Cambridge, MA: Harvard University Press.

Shaffer, D. M. & McBeath, M. K. (2002) Baseball outfielders maintain a linear optical trajectory when tracking uncatchable fly balls, *Journal of Experimental Psychology: Human Perception and Performance*, 28, 335–48.

Shanteau, J. (1992) How much information does an expert use? Is it relevant? *Acta Psychologica*, 81, 75–86.

Simon, H. A. (1955) A behavioral model of rational choice, *Quarterly Journal of Economics*, 69, 99–118.

Simon, H. A. (1956) Rational choice and the structure of environments, *Psychological Review*, 63, 129–38.

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

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

Slegers, D. W., Brake, G. L., & Doherty, M. E. (2000) Probabilistic mental models with continuous predictors, *Organizational Behavior and Human Decision Processes*, 81, 98–114.

Todd, P. (2001) Fast and frugal heuristics for environmentally bounded minds. In G. Gigerenzer and R. Selten (eds.), *Bounded Rationality. The Adaptive Toolbox* (pp. 51–70). Cambridge, MA: MIT Press.

Todd, P. M. & Gigerenzer, G. (2000) Precis of *Simple Heuristics that Make Us Smart*, *Behavioral and Brain Sciences*, 23, 727–80.

Todd, P. M. & Miller, G. F. (1999) From pride to prejudice to persuasion: Satisficing in mate search. In G. Gigerenzer, P. M. Todd, and the ABC Research Group, *Simple Heuristics that Make Us Smart* (pp. 287–308). New York: Oxford University Press.

Todorov, A. (2002) Predicting real outcomes: When heuristics are as smart as statistical models. Unpublished manuscript.

Tversky, A. (1972) Elimination by aspects: A theory of choice, *Psychological Review*, 79, 281–99.

Tversky, A. & Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases, *Science*, 185, 1124–31.

Wald, A. (1950) *Statistical Decision Functions*. New York: Wiley.