

CHAPTER
33

Modeling Decision Heuristics

Konstantinos V. Katsikopoulos and Gerd Gigerenzer

Abstract

The engineering of human decisions should be based on models that (i) require realistic amounts of resources such as time, information, and computation; (ii) have high predictive accuracy; (iii) describe the decisions people make and the processes that lead to the decisions; and (iv) are easy to understand, apply, and explain. Fast and frugal heuristics are a family of simple algorithmic models of decision making that satisfy (i)–(iv). Theoretical analyses, data from laboratory experiments, and applications of the heuristics “in the wild” are presented. A future research direction is to develop a methodology for building technological interfaces based on fast and frugal heuristics.

Key Words: heuristics, lexicographic, fast and frugal trees, less is more, robustness

Introduction

Imagine that you are a pilot who spots another plane approaching. How can you determine whether you are heading for a collision or not? You could try to recall and apply the physics and aerodynamics you learned in college and flight school in order to calculate the trajectory of the other plane, predict its position in the next few moments, and decide whether you should dive or not. Is this what pilots really do? These calculations might not be completed before the impact occurs. Some flight instructors teach their students a simple heuristic, or rule of thumb, that is fast and accurate: Look for a scratch in your windshield and observe if the other plane moves relative to that scratch; if it does not, dive away quickly.

The study of heuristics in the cognitive sciences can be used to engineer human decision making in many areas besides cockpit decisions. For instance, how do environmentally minded citizens choose their electricity carrier? How do medical doctors categorize a patient as being at a high or low risk of having a serious heart condition? How do lay

investors allocate their wealth across the assets in a portfolio? There are some answers to such questions. For example, the naturalistic-decision-making research program (Klein, Orasanu, Calderwood, & Zsombok, 1993; also see Militello & Klein, this handbook) has studied professionals on the job and provided insights into how they make quick and accurate decisions in situations of real and high stakes, as in firefighting and military operations. To the extent, however, that this work has not developed computational models of decision making, it cannot support a precise and systematic approach to the engineering of decisions.

Models of decision making are most helpful for engineering decisions if they satisfy the following requirements. First, because people typically operate under conditions of *limited* time, information, and computational resources, the models should be able to lead to decisions when fed with limited amounts of resources. Traditional models of decision making often fail this requirement. For example, expected-utility-type models require, for each decision option, the knowledge of all values, utilities,

and probabilities, as well as their algebraic combination. Second, because the point of engineering is to bring about good outcomes, the decisions suggested by the models should be "accurate." It is not always possible to evaluate the accuracy of a decision. We consider situations in which it is possible to verify whether a decision was correct or not. For example, the decision that a patient was not at a high risk of having ischemic heart disease and thus was not allocated to the intensive care unit is incorrect if subsequently the patient suffered a heart attack. Engineering models of decision making should thus be accurate. Furthermore, it is critical that the models be accurate for decisions that have *not yet* been made. This type of accuracy is referred to as *predictive accuracy* or *robustness*.

Whereas the two requirements above had to do with the input and output of models, the next two refer to their use. The third requirement for an engineering model of decision making is that people actually do use it. People are much less willing to work with an interface that is based on a class of models they do not accept. For example, firefighters reject utility models, and it is difficult to get them to participate in utility-based decision analyses (Klein, Calderwood, & Clinton-Cirocco, 1986). Similarly, physicians typically reject logistic regression models to make a diagnosis (Corey & Merenstein, 1987; Pearson et al., 1994), while they accept fast and frugal trees, whose sequential structure corresponds to their own intuitive thinking (these trees are defined below). A decision maker is said to use a model when the model describes the decisions he or she makes as well as the process that led to the decisions. The investigation of the process is important because process data provide plenty of opportunities to reject a model. The fourth requirement is that the model should be easy to understand and apply, and to explain to other stakeholders such as colleagues and supervisors. This is so because it is a truism of cognitive engineering that people will not rely on an interface that is not usable and justifiable, and an interface inherits at least some of the usability and justifiability of the model on which it is based.

Fast and frugal heuristics are a family of models of decision making that satisfies the four requirements described above for the cognitive engineering of decisions. The rest of this chapter outlines the theory of some heuristics, illustrating some evidence for their accuracy and descriptive power, and presents the challenge of building interfaces based on heuristics. For other summaries of research on fast and frugal heuristics, see Todd and Gigerenzer

(2001) and the commentaries therein, Gigerenzer (2004, 2008), and Gigerenzer and Brighton (2009). For a collection of papers on fast and frugal heuristics, see Gigerenzer, Hertwig, and Pachur (2011).

Fast and Frugal Heuristics

An approach to modeling decisions made under conditions of limited time, information, and computational resources is the research program on fast and frugal heuristics (Gigerenzer, Todd, & ABC Research Group, 1999; Todd, Gigerenzer, & ABC Research Group, 2012). These heuristics are rules of thumb that can be formulated as simple algorithms. Their objective is not to satisfy logical requirements of consistency (such as transitivity) or to find a mathematical solution to an optimization problem, but to achieve success in the ecology, as measured by criteria such as speed, frugality, accuracy, and robustness (Gigerenzer & Brighton, 2009).

A good way of understanding fast and frugal heuristics is to take a Darwinian perspective. Because evolution does not follow a grand plan, there is a patchwork of heuristics, tailored to particular jobs. Just as evolution produces adaptations that are bound to a particular ecological niche, the heuristics are neither good nor bad, rational nor irrational, *per se*, but only relative to their niches. Finally, fast and frugal heuristics exploit both core psychological capacities (e.g., recognition, recall, or imitation) and the structure of the environment (e.g., non-compensatory cue validities or the trade-off between an algorithm's bias and variance).

Before introducing the basics of fast and frugal heuristics, we point out that there are multiple interpretations of the meaning of the term "heuristics" and multiple approaches to the study of heuristics. In psychology and behavioral economics, heuristics are often linked with biases (Tversky & Kahneman, 1974, p. 1124): "in general, . . . heuristics are quite useful, but sometimes they lead to severe and systematic errors." A critique of this approach is that it has not been specified precisely what "sometimes" means; instead, answers are typically cast as a litany of experimental manipulations that increase or decrease the accuracy of a heuristic. Furthermore, heuristics such as "availability" are mere verbal labels instead of testable formal models. In applied mathematics and operations research, heuristics have a positive meaning and are understood as computational shortcuts that allow "quickly finding good feasible solutions" (Hillier & Lieberman, 2001, p. 624). For the purpose of designing interfaces to be used by people, a limitation of the heuristics

developed in operations research is that they lack a psychological basis. In engineering, some authors (Magee & Frey, 2006) have also discussed practitioners' use of shortcuts and called them heuristics. These heuristics are somewhat different from what we mean by fast and frugal heuristics in that they apply to problems of experimentation and design rather than to problems of reasoning, choice, and inference.

Ignorance-Based Decision Making

Imagine that you are a contestant in a TV game show and face the \$1 million question: Which city has more inhabitants: Detroit or Milwaukee?

What is your answer? If you are American, then your chances of finding the right answer, Detroit, are not bad—60% of undergraduates at the University of Chicago did. If, however, you are German, your prospects look dismal because most Germans know little about Detroit, and many have not even heard of Milwaukee. How many correct inferences did the less knowledgeable German group achieve? Ninety percent of the Germans answered the question correctly (Goldstein & Gigerenzer, 2002). How can people who know less about a subject nevertheless make more correct inferences? The answer seems to be that the Germans used a 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 paired-comparison tasks, where the goal is to infer which one of two objects (e.g., cities) has the higher value on a numerical criterion (e.g., population), the heuristic is 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 on the criterion.

The recognition heuristic builds on the core capacity of 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.

Intuitively, the recognition heuristic is successful when ignorance is systematic rather than random, that is, when recognition is strongly correlated with the criterion. 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. A number of experimental studies have found that if the accuracy of the recognition heuristic is substantial (i.e., it exceeds 0.7), people use the recognition heuristic in about 90% of all cases (Pachur & Hertwig, 2006).

A strong prediction of the recognition heuristic is that no other pieces of information can change the decision to which recognition points. For example, suppose that a person (i) recognizes Detroit and not Milwaukee and (ii) recalls that the automobile industry in Detroit has been hit for long time by a recession. The prediction of the recognition heuristic is that he or she will infer that Detroit is more populous despite Detroit's recession. In other words, recognition is predicted to be used in a *non-compensatory* fashion. For instance, Pachur, Broeder, and Marewski (2008) reported that an amazing 50% of the participants in their study chose the recognized object consistently, that is, in every single trial, even when they had knowledge of three cues indicating that the recognized object should have a low criterion value.

The Less-Is-More Effect

Beyond the non-compensatory use of recognition, the recognition heuristic leads to a counterintuitive prediction, one that has to do with accuracy. This is the *less-is-more* effect, where less information leads to more accuracy (Goldstein & Gigerenzer, 2002). The effect can be viewed as a violation of the effort-accuracy trade-off where effort is measured by the amount of information used to make a decision. Below, we present a theoretical analysis of the less-is-more effect.

Assume that there exist N objects (e.g., cities) and that a person performs all $N(N-1)/2$ paired comparisons according to a numerical criterion involving two of these objects (e.g., compare two city populations). The amount of information a person uses is measured by the number of objects the person recognizes, n . The question is if and under what conditions a smaller n can lead to higher accuracy than a larger n . The probability of being able to use the heuristic for a paired comparison equals the probability of exactly one object in the pair being recognized, or

$$r(n) = 2n(N-n) / [N(N-1)] \quad (1)$$

Similarly, the probability that both objects are recognized, and thus other knowledge beyond recognition must be used, equals

$$k(n) = n(n-1) / [N(N-1)] \quad (2)$$

Finally, the probability that neither object is recognized, which means that the person has to guess, equals

$$g(n) = (N-n)(N-n-1) / [N(N-1)] \quad (3)$$

Let α be the accuracy of the recognition heuristic and β the accuracy when both objects are recognized and other knowledge is used (where $\alpha, \beta > 1/2$). We also assume that accuracy equals $1/2$ when none of the objects is recognized. The overall accuracy of a person who recognizes n objects equals

$$f(n) = r(n)\alpha + k(n)\beta + g(n)(1/2) \quad (4)$$

It is a mathematical fact that $f(n)$ is a non-increasing function of n when $\alpha > \beta$, assuming that both are constant across n (Goldstein & Gigerenzer, 2002). In other words, if $\alpha > \beta$, a less-is-more effect is predicted. For example, assume that there are three sisters who study for a geography quiz with $N = 100$ cities. The three sisters have the same $\alpha = 0.8$ and $\beta = 0.6$, but differ in the number of cities n they recognize. The little sister, who recognizes zero cities, has an accuracy of $f(0) = 0.50$. The eldest sister, who recognizes all 100 cities, has an accuracy of $f(100) = 0.60$. The middle sister, who recognizes 50 cities, is the most accurate of the three sisters, as $f(50) = 0.68$.

At a first glance, the less-is-more effect might appear paradoxical. But it is not because less recognition information may simply enable more accurate cognitive processing via the use of the recognition heuristic. This idea is formalized by the condition $\alpha > \beta$.

In Figure 33.1, accuracy for all n from 0 to 100, when $\alpha = 0.8$ and $\beta = 0.6$, is graphed by a solid curve. The dashed curve presents the accuracy of groups with three sisters in each group (where all sisters in the group have the same n , and all sisters in all groups have $\alpha = 0.8$ and $\beta = 0.6$). All groups use the simple majority rule to decide: If two or three group members infer that city A is more populous than city B , then the group infers that A is more populous. The middle triplet, where each of the three sisters recognizes 50 cities, is more accurate than the eldest triplet, where each sister recognizes all cities. It is a mathematical fact that if $\alpha > \beta$, a less-is-more effect is predicted for groups who use a variety of majority rules (Reimer & Katsikopoulos, 2004).

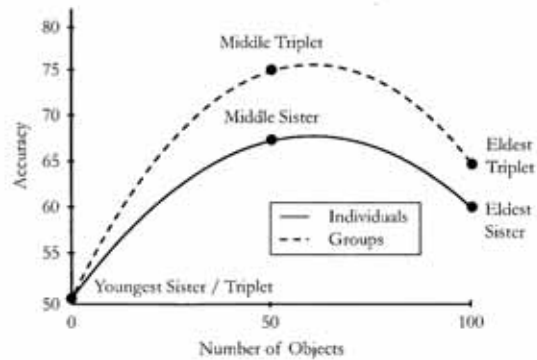


Figure 33.1 Predicted accuracy as a function of the number of recognized objects for $\alpha = 0.8$ and $\beta = 0.6$, for individuals (solid curve) and three-member groups that use the simple majority rule (dashed curve).

Reimer and Katsikopoulos (2004) investigated how groups of people decided together. In this study, three people sat in front of a computer screen on which questions such as "Which city has more inhabitants: San Diego or San Antonio?" were displayed. The task of the group was to find the correct answer through discussion. Consider the following conflict: Two group members have heard of both cities, and each concluded independently that city A is more populous. But the third group member has not heard of A , only of B , and concludes that B is more populous (relying on the recognition heuristic). 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%), the group voted for B (Reimer & Katsikopoulos, 2004). This percentage rose to 76% if two members used recognition.

Group members letting their knowledge be dominated by others' lack of recognition may seem odd. But this apparently irrational decision increased the overall accuracy of the group. Consistent with the mathematical theory, Reimer and Katsikopoulos (2004) observed that when two groups had approximately equal average α and β (that were such that $\alpha > \beta$), the group that recognized fewer cities (smaller n) typically had more correct answers. For instance, the members of one group recognized on average only 60% of the cities while those in a second group recognized 80%, but, in a series of over 100 questions, the first group got 83% of the answers correct, whereas the second got only 75%.

We next turn to situations where the recognition heuristic cannot be used.

Social Heuristics

When recognition is not valid, or when people recognize all objects, fast and frugal heuristics can involve the search for reasons. 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. 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. One possibility is that he arrived at his decision by relying on heuristics that process reasons, such as take-the-best and tallying (see below for definitions of these heuristics), which exploit core capacities such as recall memory. Another possibility is social heuristics that exploit the social capacities of people, such as imitation, which is unmatched among animal species. For instance, consider the following (Laland, 2001):

Do-what-the-majority-do heuristic: If the majority of your peers display a behavior, engage in it as well.

For the marriage problem, this heuristic makes a man start thinking of marriage at a time when most other men in his 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. The do-what-the-majority-do heuristic tends to perform well when (i) the observer and the demonstrators of the behavior are exposed to similar environments that (ii) are stable rather than changing and (iii) noisy, that is, where it is hard to see what the immediate consequence of one’s action is (Boyd & Richerson, 2005).

Social heuristics appear to guide many of our decisions, and do-what-the-majority-do is only one such heuristic in the adaptive toolbox of humans. Consider deciding about green versus gray energy. Assume you have moved into a new apartment, and you need to choose providers for the basic utilities. In the United States, the United Kingdom, and many countries in Europe, 50% to 90% of the people asked say that they would favor a “green” electricity carrier and are even willing to pay a small premium for it. But, unfortunately, these statements do not

reflect behavior: The percentage of people who consume green electricity is marginal; for example, 2% in Germany and 0.5% in the United Kingdom. This discrepancy between what people say and what they do can be explained by the use of a social heuristic (Pichert & Katsikopoulos, 2008): When one moves into a new apartment, there is typically an electricity carrier that provides a default (the carrier that was used by the previous tenant or the carrier that the landlord has chosen). The new tenants typically take no action, and the default is used.

Default heuristic: If a decision is set as the default, do not change it.

The default heuristic can explain a flurry of phenomena such as peoples’ retirement plans and whether people are organ donors or not (Johnson & Goldstein, 2003). Darwin seems to have reached his decision not by consulting defaults or peers, but by his own thinking based on pieces of information or, in psychological jargon, *cues* (Brunswik, 1955; Kirlik, 2006). We next describe a class of cue-based heuristics that uses a minimum number of cues in order to make a decision.

Lexicographic Heuristics

Consider a paired comparison in which both objects are recognized. The decision is made on the basis of binary cues: For example, if the task is to infer which of two professors at a Midwestern college gets the higher salary, gender is a highly predictive cue. Cues such as gender are naturally binary; if a cue is continuous (e.g., years since doctorate degree), it can be dichotomized. The standard way of comparing two objects is the class of linear rules (as in the usual application of Brunswik’s lens model; see Hammond & Stewart, 2001), in which the weighted sum of cue values for each object is computed, and the object with the higher sum is inferred to have the higher criterion value (if the sums are equal, one object is picked randomly). The weight w_i for cue c_i can be computed in a number of ways as, for example, by minimizing the sum of squared differences between the real criterion values in the ecology and the criterion values estimated by the linear rule.

For making inferences, as in predicting salaries or the outcomes of football matches, the linear rule is the analog of the expected utility rule for making choices: The relevant pieces of information (cues) are weighted and added. The class of *lexicographic* heuristics dispenses with adding, and instead orders cues (a simple form of weighing). In the *take-the-best* heuristic (Gigerenzer & Goldstein, 1996), cues are

ordered by decreasing validity. The validity v_i of cue c_i is a measure of cue goodness. It is the conditional probability that the cue points to the object with the higher criterion value ($c_i = 1$ on the object with the higher criterion value, and $c_i = 0$ on the other object) given that the cue has different values on the two objects. After cues are ordered, the decision maker inspects the first cue. If this cue points to one of the objects, then this object is inferred to have the higher criterion value. If the cue does not discriminate between the objects, then the second cue is inspected and so on until a discriminating cue is found; if no such cue exists, an object is picked at random.

Note that, like the recognition heuristic, take-the-best is non-compensatory. Furthermore, unlike linear rules, lexicographic heuristics specify the processes by which people make inferences. In take-the-best, it is specified how people search for information (by ordering cues by validity), how they decide to stop the search (as soon as one cue discriminates between the objects and allows making a decision), and how they decide based on the available information (by using the discriminating cue). There have been a number of laboratory tests of these processes (as well as of the decision outcomes predicted by take-the-best), and this research is summarized in Broeder and Newell (2008; see also the other articles in this journal special issue). Overall, whether people use heuristics such as take-the-best depends on the characteristics of the decision environment, as, for example, whether there is time pressure or not, and how skewed the distribution of cue validities is.

In engineering, Rothrock and Kirlik (2003) claimed that, if time and information are limited, human judgments are not described best by linear models, but by fast and frugal heuristics. To capture the outcomes of the heuristics, Rothrock and Kirlik proposed a lens model that is nonlinear and non-compensatory (see also Gigerenzer & Kurz, 2001). Katsikopoulos (2009) showed how this non-compensatory lens model can also capture the processes postulated by fast and frugal heuristics.

Before we turn to a second class of cue-based heuristics for paired comparisons, we note that lexicographic heuristics can also be applied to other tasks such as risky choices. For example, the priority heuristic (Brandstaetter, Gigerenzer, & Hertwig, 2006) specifies how people choose between two lotteries such as, for example, receiving 40 dollars for sure and receiving either 100 or 0 dollars, where each outcome obtains with a probability of

50%. Like take-the-best, the priority heuristic goes through cues one at a time (where cues are now the values and probabilities of the possible outcomes in each lottery) and decides based on the first cue that discriminates between the options. Brandstaetter et al. (2006) found that the priority heuristic could predict peoples' choices overall better than other popular choice models such as cumulative prospect theory (Tversky & Kahneman, 1992). Katsikopoulos and Gigerenzer (2008) analytically showed that the heuristic implies, rather than just being consistent with, a number of major empirical phenomena in risky choice, such as common consequence effects and the fourfold pattern of risk attitude reversals. In other words, phenomena such as the Allais paradox are implied by a simple heuristic that does not transform values and probabilities in complex nonlinear fashion, as prospect theory and other modifications of expected utility theory need to do. Ford et al. (1989) reviewed 45 studies of peoples' choices and reported that non-compensatory models, similar to the priority heuristic, were the models used most often.

Tallying Heuristics

Recall that lexicographic heuristics are a simplification of linear models where cues are weighed but not added. *Tallying* heuristics are another simplification of linear models where cues are added but not weighed. That is, object *A* is inferred to have a higher criterion value than object *B* if the sum of cue values on *A* is higher than the sum of cue values on *B*, or if $\sum_i c_i(A) > \sum_i c_i(B)$ (if the sums are equal, one object is picked randomly).

In the psychology of judgment and decision making, early work on tallying was done by Dawes and Corrigan (1974) and Einhorn and Hogarth (1975). There are also other applications of tallying, for example, in finance, where it refers to investors allocating an equal amount of wealth to each asset in their portfolio. This heuristic is also known as 1/N, where N is the number of assets or alternatives. Behavioral finance researchers such as Benartzi and Thaler (2001) have considered tallying to be suboptimal, calling it naive diversification. Interestingly, it was found more than 20 years ago that naive diversification yields as good risk-adjusted returns as optimization models; and, in a more recent study, DeMiguel, Garlappi, and Uppal (2007) did a comprehensive simulation study and corroborated that tallying (1/N) is as effective as complex finance optimization models. The next section shows that lexicographic and tallying heuristics

enjoy high predictive accuracy in the paired comparison task.

The Ecological Rationality of Take-the-Best and Tallying

Dawes and Corrigan (1974) showed that tallying can have higher predictive accuracy than linear regression. Einhorn and Hogarth (1975) provided statistical reasons for this such as the absence of sampling error in the estimation of weights. Czerlinski, Gigerenzer, and Goldstein (1999) replicated this finding in 20 real-world paired comparison tasks, such as predicting which Chicago public high school has the higher dropout rate based on 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 were predicting people's attractiveness judgments, homelessness rates, adolescents' obesity at age 18, etc.) They tested linear regression, tallying, take-the-best, and a simple Bayesian network called naive Bayes (see Martignon & Hoffrage, 2002).

Czerlinski et al. (1999) split each data set in half, and the parameters of each model (e.g., cue weights and cue validities) were estimated. The part of the data set used to estimate the parameters of the models is the *training set*. The parameter estimates were used to predict, for each model, the paired comparisons on the other half of the data set, called the *test set*. This process was repeated 1,000 times to average out random variation. In the test set, the predictive accuracy of the models was as follows: regression 68%, tallying 69%, take-the-best with dichotomized cues 71%, naive Bayes 73%, regression with un-dichotomized cues 76%, and take-the-best with un-dichotomized cues 76%.

One interpretation of these results is that using *less* complex computations and sophisticated information can lead to *more* predictive accuracy. That is, simplicity can, under some conditions, lead to robustness (yet, as the performance of tallying shows, there are limits to the benefits of simplicity).

The study of the conditions under which a model, such as take-the-best, achieves high predictive accuracy is the study of the *ecological rationality* of the model. Recent work has distinguished the mathematical conditions under which tallying and take-the-best are accurate (Baucells, Carrasco, & Hogarth, 2008; Hogarth & Karelaia, 2007; Katsikopoulos & Fasolo, 2006; Katsikopoulos & Martignon, 2006; Martignon & Hoffrage, 2002; for a review see Katsikopoulos, 2011). For example, when cue validities are equal, tallying achieves

maximum accuracy among all possible inference models; when some cues have much higher validities than others, take-the-best achieves maximum accuracy (Katsikopoulos & Martignon, 2006). Using an insight from machine learning, Gigerenzer and Brighton (2009) conjectured that the heuristics are more robust than neural networks, classification and regression trees, and other complex algorithms because the heuristics have smaller variance in their predictions.

We end our tour of fast and frugal heuristics by considering a family of heuristics developed for assigning objects to categories.

Fast and Frugal Trees

A middle-aged man is taken to the hospital with complaints of intense chest pain. The doctors have to decide quickly whether he is at a low risk of having ischemic heart disease and needs just a regular nursing bed, or he is at a high risk and should be rushed to the emergency room. This is a decision where resources—such as time, information, and computation—are limited, there is pressure to be accurate, and the stakes are high. The fast and frugal heuristics research program has provided some answers as to how professionals and laypeople make, or should make, accurate categorizations with limited resources, by using simple trees.

We first introduce some elements from the general theory of trees for categorization. In a categorization task, the decision maker has to assign objects to one of several mutually exclusive categories based on the values of the objects on cues. In the example above, the objects are the patients, there are two categories—a low or high risk of ischemic heart disease—and the cues are the available pieces of medical information such as readings from an electrocardiogram.

A categorization tree can be graphically represented by the root node, on the tree's first level, and subsequent levels with one cue processed at each level (see Figure 33.2). There are two types of nodes. First, a node may specify a question about the value of the object on a cue; the answer then leads to another node at the next level, and the process continues in this way. The root node is of that type. For nodes of the other type there is an *exit*; the object is categorized and the process stops. In sum, starting from the root node and answering a sequence of questions, an exit is reached and the object is categorized. For trees to be easy for people to understand and apply, they should not have too many levels or cues. For example, Figure 33.2 shows

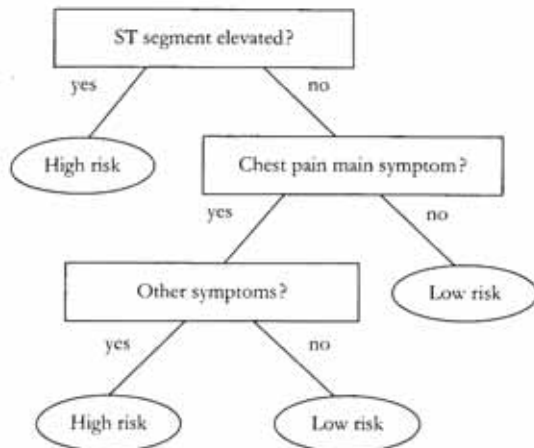


Figure 33.2 A fast and frugal tree for categorizing patients as having a high or low risk of ischemic heart disease (for more details, see Green & Mehr, 1997).

such a tree for the ischemic heart disease problem (Green & Mehr, 1997).

A categorization tree is *fast and frugal* if and only if it has at least one exit at each level (Martignon, Katsikopoulos, & Woike, 2008). According to this definition, the tree in Figure 33.2 is fast and frugal. If a second question were asked for all patients with an elevated ST segment, the tree would not have been fast and frugal.

Fast and frugal trees are “minimal” in the sense of using the fewest number of question nodes and still involving all available cues, one at a time. Fast and frugal trees specify a number of cognitive processes. They specify how information is searched for, how the search is stopped, and how a decision is taken based on the obtained information. For example, a

physician using the tree in Figure 33.2 first looks up the ST segment, then the chest pain, and finally other symptoms. There are a number of simple ways of ordering cues, but we will not describe them here; see Martignon et al. (2008).

Fast and frugal trees have been applied to a number of fields such as educational training (Wottawa & Hossiep, 1987), engineering (Katsikopoulos & Fasolo, 2006), law (Dhami, 2003), and medicine (Green & Mehr, 1997; Fischer et al., 2002), in order to describe what people do or to prescribe what people should be doing.

Fast and frugal trees differ from many other categorization models in three respects: First, the trees are deterministic, that is, they predict that an object belongs to a category with probability 1 or 0. The motivation for this is that under conditions of limited time, information, and computation, people might not attempt or be able to compute probabilities. For example, Green and Mehr (1997) found that medical doctors refused to use a logistic regression model that computed the probability of a patient having ischemic heart disease. Second, fast and frugal trees can make only qualitative predictions about response times. For example, according to the tree in Figure 33.2, a patient with an elevated ST segment would be categorized faster than a patient without an elevated ST segment. Finally, cues are not combined and are considered one at a time. This feature of fast and frugal trees makes it easier for them to generalize well to new patients.

The Robustness of Fast and Frugal Trees

Figure 33.3 compares the accuracy of two types of fast and frugal trees (called Zig and Max; for details,

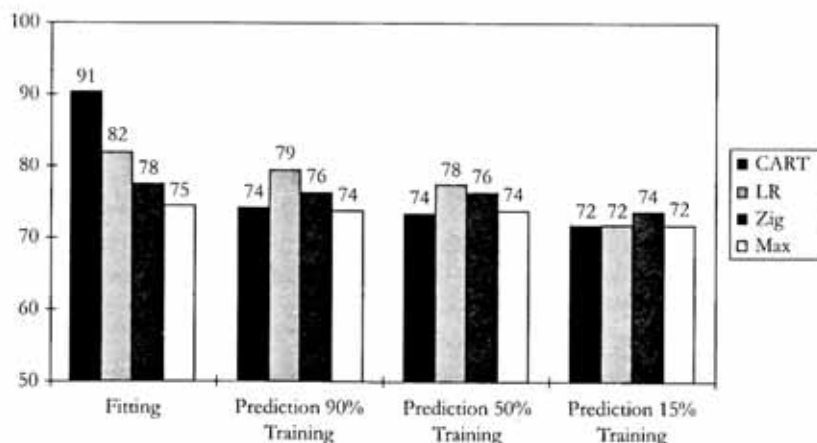


Figure 33.3 Average performance of four categorization models (classification and regression trees, or CART; logistic regression, or LR; and two types of fast and frugal trees called Zig and Max; see Martignon et al., 2008, for details), across 11 medical problems, in fitting and three cases of prediction (training set was 90%, 50%, or 15% of the whole dataset).

see Martignon et al., 2008) to two benchmarks from statistics and artificial intelligence, logistic regression (LR) and the classification and regression trees of Breiman et al. (CART; 1984). We used 11 medical categorization problems from the UC Irvine Machine Learning Repository. The accuracy of the four models was assessed in four cases each. In the first case, the parameters of the models were estimated based on all data, that is, the models were *fitted* to the whole dataset. In the other three cases, the parameters of the models were estimated based on a subset of the data (90%, 50%, or 15% of the whole dataset), and the same parameters were used to assess how well the models *predicted* the remaining data. Because typically decision makers do not have access to all data, the three prediction cases are more relevant tests of the models.

As is often the case, the more complex models (LR and CART) did much better in fitting. But a good categorization model needs to predict unknown cases, not (just) explain the past by hindsight. In prediction, the simple Zig trees matched or came close to the accuracy of CART and logistic regression, while Max lagged a few percentage points behind. Zig even outperformed CART and LR when there was little available information (15% of the whole dataset).

This completes our survey of fast and frugal heuristics. In sum, we argued that fast and frugal heuristics (i) require realistic amounts of resources such as time, information, and computation; (ii) have high predictive accuracy; (iii) describe the decisions people make and the processes that lead to the decisions; and (iv) are easy to understand, apply, and explain. It seems then that these heuristics are good candidates on which to base cognitive interfaces for decision making. How exactly to do this is still an open research question, and below we speculate on it.

Future Directions

We are not aware of comprehensive research programs that have focused on building interfaces that support heuristic decision making by laypeople and professionals. This is in contrast with the volume of written publications and software packages implementing utility-based decision analysis. By utility-based decision analysis we mean practical procedures that aim at getting people to make decisions in line with subjective-expected-utility or multi-attribute-utility theory (Edwards, Miles, & von Winterfeldt, 2007). It is beyond our scope to evaluate the success of this approach, even though we have earlier alluded to the work of Klein and his

colleagues that challenges the acceptance and results of utility-based decision analysis; for more on this point with respect to the design of engineering systems, see Clausning and Katsikopoulos (2008). Instead we consider some differences between the principles of utility-based decision analysis and of the alternative, which could be called heuristic-based decision analysis.

Practitioners of utility-based decision analysis initially suggested that decision makers should accept the quantitative calculations that underlie utility theories, but lately have concentrated more on the acceptance of the qualitative aspects of decision problems as these are modeled in utility theories (Keeney, 2004). In the case of risky choices, these aspects are that decision problems consist of choosing among a number of options, each of which has a number of possible outcomes, values, and probabilities. Furthermore, there are two basic broad principles that specify how an option should be chosen: the *more-is-better* maxim and the making of *trade-offs*. The *more-is-better* maxim asserts that the decision maker should use all available information. Second, low values on an attribute of an option (e.g., low probability of winning a prize) can and should be compensated by high values of another attribute of that option (e.g., high value of the prize). It is interesting that these principles are almost taken as axioms, with no evidence offered for their effectiveness and with whatever justification being offered having to do with the theoretical possibilities of losses for violators of *more-is-better* or *trade-off* making. Also, such losses have been discussed in thought experiments but have not been observed in practice (for details on this point, see Clausning & Katsikopoulos, 2008).

Most of the heuristics presented here, such as the recognition heuristic and take-the-best, do not follow the *more-is-better* maxim or implement trade-offs. Still, the heuristics achieve accuracy in a very relevant sense, that is, in predicting the correct answer in decisions not yet made. It would be a big step for cognitive engineers to dare to build and test interfaces that do not assume or promote classic principles such as the *more-is-better* maxim or making trade-offs. This step, however, may likely be just the beginning of a whole new set of challenges like the following: (1) How can cognitive engineers convince users accustomed to utility-based aids that more is not always better or that it is not always necessary to make trade-offs? (2) How can they facilitate a user's transition from utility-based aids to heuristic-based aids? (3) How can they build a

hybrid system that adaptively switches between utility-based and heuristic-based aids?

References

- Baucells, M., Carrasco, J. A., & Hogarth, R. M. (2008). Cumulative dominance and heuristic performance in binary multi-attribute choice. *Operations Research*, *56*(5), 1289–1304.
- Benartzi, S., & Thaler, R. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review*, *91*, 79–98.
- Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. Oxford, England: Oxford University Press.
- Brandstatter, Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: A process model of risky choice. *Psychological Review*, *113*, 409–432.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Chapman and Hall.
- Broeder, A., & Newell, B. R. (2008). Challenging some common beliefs: Empirical research within the adaptive toolbox metaphor. *Judgment and Decision Making*, *3*, 205–214.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.
- Clausing, D. P., & Katsikopoulos, K. V. (2008). Rationality in systems engineering: Beyond calculation or political action. *Systems Engineering*, *11*(4), 309–328.
- Corey, G. A., & Merenstein, J. H. (1987). Applying the acute ischemic heart disease predictive instrument. *Journal of Family Practice*, *25*, 127–133.
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97–118). New York, NY: Oxford University Press.
- Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *81*(2), 95–106.
- DeMiguel, V., Garlappi, L., & Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, *22*, 1915–1953.
- Dhmi, M. K. (2003). Psychological models of professional decision-making. *Psychological Science*, *14*, 175–180.
- Edwards, W., Miles, R. F., Jr., & von Winterfeldt, D. (Eds.). (2007). *Advances in decision analysis: From foundations to applications*. New York, NY: Cambridge University Press.
- Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, *73*, 221–230.
- Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Decision Processes*, *13*, 171–192.
- Fischer, J. E., Steiner, F., Zucol, F., Berger, C., Martignon, L., Bossart, W., ... Nadal, D. (2002). Using simple heuristics to target macrolide prescription in children with community-acquired pneumonia. *Archives of Pediatric and Adolescent Medicine*, *156*(10), 1005–1008.
- Ford, J. K., Schmitt, N., Schechtman, S. L., Hults, B. H., & Doherty, M. L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes*, *43*, 75–117.
- Gigerenzer, G. (2004). Fast and frugal heuristics: The tools of bounded rationality. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 62–88). Oxford, England: Blackwell.
- Gigerenzer, G. (2008). *Rationality for mortals*. New York, NY: Oxford University Press.
- Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*, 107–143.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669.
- Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.
- Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.). (2011). *Heuristics: The foundations of adaptive behavior*. New York, NY: Oxford University Press.
- Gigerenzer, G., & Kurz, E. M. (2001). Vicarious functioning reconsidered: A fast and frugal lens model. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 342–348). Oxford, England: Oxford University Press.
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75–90.
- Green, L., & Mehr, D. R. (1997). What alters physicians' decisions to admit to the coronary care unit? *Journal of Family Practice*, *45*(3), 219–226.
- Hammond, K. R., & Stewart, T. (2001). *The essential Brunswik: Beginnings, explications, applications*. Oxford, England: Oxford University Press.
- Hillier, F. S., & Lieberman, G. J. (2001). *Introduction to operations research*. McGraw-Hill.
- Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, *114*(3), 733–758.
- Johnson, E. J., & Goldstein, D. G. (2003). Do defaults save lives? *Science*, *302*, 1338–1339.
- Katsikopoulos, K. V. (2009). The conceptual connection between lens models and fast and frugal heuristics: A process approach. *Theory and Psychology*, *19*(5), 688–697.
- Katsikopoulos, K. V. (2011). Psychological heuristics for making inferences: Definition, performance, and the emerging theory and practice. *Decision Analysis*, *8*(1), 10–29.
- Katsikopoulos, K. V., & Fasolo, B. (2006). New tools for decision analysts. *IEEE Transactions on Systems, Man, and Cybernetics: Systems and Humans*, *36*(5), 960–967.
- Katsikopoulos, K. V., & Gigerenzer, G. (2008). One-reason decision-making: Modeling violations of expected utility theory. *Journal of Risk and Uncertainty*, *37*(1), 35–56.
- Katsikopoulos, K. V., & Martignon, L. (2006). Naive heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology*, *50*, 488–494.
- Keeney, R. L. (2004). Making better decision makers. *Decision Analysis*, *1*(4), 193–204.
- Kirlik, A. (Ed.). (2006). *Adaptive perspectives on human-technology interaction: Methods and models for cognitive engineering and human-computer interaction*. New York, NY: Oxford University Press.
- Klein, G., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision-making on the fireground. *Proceedings of the Human Factors and Ergonomics Society 30th Annual Meeting*, *1*, 576–580.
- Klein, G., Orasanu, J., Calderwood, R., & Zsombok, C. E. (1993). *Decision making in action: Models and methods*. Cambridge, MA: MIT Press.

- Laland, K. N. (2001). Imitation, social learning, and preparedness as mechanisms of bounded rationality. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.
- Magee, C., & Frey, D. D. (2006). Experimentation and its role in engineering design: Linking a student design exercise with new results from cognitive psychology. *International Journal of Engineering Education*, 22(3), 478–488.
- Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, 52, 29–71.
- Martignon, L., Katsikopoulos, K. V., & Woike, J. K. (2008). Categorization with limited resources: A family of simple heuristics. *Journal of Mathematical Psychology*, 52(6), 352–361.
- Pachur, T., Broeder, A., & Marewski, J. N. (2008). The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? *Journal of Behavioral Decision Making*, 21, 183–210.
- Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 983–1002.
- Pearson, S. D., Goldman, L., Garcia, T. B., Cook, E. F., & Lee, T. H. (1994). Physician response to a prediction rule for the triage of emergency department patients with chest pain. *Journal of General Internal Medicine*, 9, 241–247.
- Pichert, D., & Katsikopoulos, K. V. (2008). Green defaults: Information presentation and pro-environmental behavior. *Journal of Environmental Psychology*, 28, 63–73.
- Reimer, T., & Katsikopoulos, K. V. (2004). The use of recognition in group decision-making. *Cognitive Science*, 28, 1009–1029.
- Rothrock, L., & Kirlik, A. (2003). Inferring rule-based strategies in dynamic judgment tasks: Toward a noncompensatory formulation of the lens model. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 33(1), 58–72.
- Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, 23, 727–741.
- Todd, P. M., Gigerenzer, G., & ABC Research Group. (2012). *Ecological rationality: Intelligence in the world*. New York, NY: Oxford University Press.
- Tversky, A., & Kahneman, D. (1974). Heuristics and biases: Judgment under uncertainty. *Science*, 185, 1124–1130.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297–323.
- Wottawa, H., & Hossiep, R. (1987). *Grundlagen Psychologischer Diagnostik: Eine Einführung* [Foundations of Psychological Diagnostics: An Introduction]. Hogrefe Verlag.