Klaus Lucas · Peter Roosen
Editors

Emergence, Analysis and Evolution of Structures

Concepts and Strategies Across Disciplines

Katsikopoulos, K. V., & Gigerenzer, G. (2010). Decision structures on the basis of bounded rationality. In K. Lucas & P. Roosen (Eds.), *Emergence, analysis and optimization of structures: Concepts and strategies across disciplines* (pp. 214–226). Berlin: Springer Verlag.

Springer

3.12 Decision Structures on the Basis of Bounded Rationality

Konstantinos V. Katsikopoulos and Gerd Gigerenzer

To the great relief of passengers all around the world airplanes take off, fly, and land every day. Mathematicians, physicists, mechanical and aeronautical engineers have created a large store of know-how on the technical aspects of aviation. The management of the aviation industry, however, remains challenging. Consider, for example, making decisions under uncertainty. We do not really know how to route passengers to jointly optimize time, safety, and fuel efficiency. And, even if we focused on just one objective, say time, the routing problem may be computationally intractable.

Like in aviation, many decision problems in large-scale technical systems resist formulation in optimization terms for various reasons. Typical difficulties are multiple criteria, information that is not easily available and uncertainties of various kinds. Even when they can be formulated as optimization problems, computing an optimal solution often turns out to be intractable. What to do? The approach typically taken in the natural sciences and engineering is to optimize a simplification of the original problem.

This often works well but it can be difficult to know if the simplifications hold and what is the loss when they are violated. More broadly, the optimization approach – with some exceptions, e.g., Taguchi methods – is not tailored to handle issues like uncertainty, robustness, and flexibility that are increasingly recognized as fundamental in the management of engineering systems [Pap04]. Finally, the interface between optimization methods and practitioners often lacks transparency, usability, and acceptance [KOCZ93].

Overall, engineers are trained in the rigorous theory of optimization but all too often, when they graduate, they seem to find themselves using heuristics that worked in the past. At the Max Planck Institute for Human Development in Berlin, our team of life, natural, and social scientists, as well as historians, philosophers, mathematicians, and engineers, has, for more than ten years, been studying heuristics for decision making under uncertainty. Our research program can be viewed as the study of bounded rationality, a popular concept in the social sciences [Sim55]. One of our interests is in modeling the heuristic cognitive processes laypeople and some practitioners – medical doctors and mock jurors – use. This chapter samples some answers to the normative question of how well the heuristics perform. The research has used decision-making tasks that do not directly relate to technical concerns but we also speculate how it can be applied to the management of engineering systems. We start with some general comments on the heuristic view of bounded rationality.

3.12.1 Bounded Rationality: Fast and Frugal Heuristics

Bounded rationality is often interpreted as optimization under constraints, where the constraints are due to impoverished cognitive ability or incomplete information [Con96]. This interpretation is contrary to Simon's [Sim55] who emphasized satisficing, such as picking any outcome that exceeds a pre-determined aspiration level (as opposed to picking only an optimal outcome). Furthermore, this interpretation does not allow much progress as it reverts the study of decision making back to the study of logic, probability, and calculus, while excluding psychology.

Our approach to bounded rationality takes an ecological rather than logical view. It does not study optimal, internally consistent decisions but decisions that surpass aspiration levels with regard to external criteria like speed, accuracy, robustness, and transparency. This fits well with engineering where the focus is not so much on internal consistency but on external performance. Gigerenzer, Todd, and colleagues [GTtARG99, TGtArgip19] model the decisions with a breed of simple rules of thumb, called *fast and frugal heuristics*, which use a minimum of time, information, and computational resources.

The heuristics can be understood from a Darwinian perspective. First, because evolution does not follow a grand plan, there is a patchwork of heuristics, tailored to particular problems. This gives flexibility to the bounded rationality approach. Second, just as evolution produces adaptations that are bound to a particular ecological niche, the heuristics are not rational or irrational, per se, but only relative to an environment. Note that the study of the interaction between the decision-maker and the environment – emphasized in the ecological approach to psychology [Bru55, Sim56] – is missing in the optimization-under-constraints approach.

Finally, and importantly, heuristics exploit core psychological capacities, such as the ability to track objects. This is exactly what allows the heuristics to be simple, yet successful. For example, consider a pilot who spots another plane approaching, and fears a collision. How can she avoid it? A simple heuristic that works is to look at a scratch in her windshield and observe whether the other plane moves relative to that scratch. If it does not, dive away quickly.

In short, in our view, bounded rationality deals with simple and transparent heuristics that require minimum input, do not strive to find a best and general solution, but nevertheless are accurate and robust. What do these heuristics look like?

3.12.2 The Recognition Heuristic

Imagine you are a contestant in a TV game show and face the $1,000,000 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. Some two thirds of undergraduates at the University of Chicago did [GG02]. 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 that we tested achieve? Despite a considerable lack of knowledge, nearly all of the Germans answered the question correctly. 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 following fast and frugal heuristic can be used.

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 of faces, voices, and, as here, of names. 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 that person as a key determinant of its use, Journal of Experimental Psychology: Learning, Memory, and Cognition [CM87].

Intuitively, the recognition heuristic is successful when ignorance is systematic rather than random, that is, when recognition is 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 F.A. Cup third-round soccer matches, such as Manchester United versus Shrewsberry Town. Ayton and Önkal [AO97] 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% versus 66% correct). Their predictions were consistent with the recognition heuristic in 627 out of 662 cases (95%). More generally, a number of experimental studies have found that if the accuracy of the recognition heuristic, α, is substantial (i.e., exceeds, say, 0.7), people use the heuristics in about 90% of all cases [Nay01].

The recognition heuristic implies several counterintuitive phenomena of human decision making that cannot be deduced from any other theory we are aware of. For instance, recognition information tends to dominate further knowledge, in rats as well as in people, even if there is conflicting evidence [GG02, PH06]. Here we concentrate on the counterintuitive finding that less information can increase accuracy.

3.12.3 The Less-is-More Effect

Assume that a person recognizes n out of N objects. The probability of being able to use the heuristic equals the probability of recognizing exactly one object in a sample of two, or

$$r(n) = \frac{2n(N-n)}{N(N-1)}. \tag{3.20}$$

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

$$k(n) = \frac{n(n-1)}{N(N-1)}. \tag{3.21}$$

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

$$g(n) = \frac{(N-n)(N-n-1)}{N(N-1)}. \tag{3.22}$$

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

$$f(n) = r(n)\alpha + k(n)\beta + g(n)\left(\frac{1}{2}\right). \tag{3.23}$$

Definition 1. *The* less-is-more effect *occurs when there exist* n_1 *and* n_2 *so that* $n_1 < n_2$ *but* $f(n_1) > f(n_2)$ *with* $n_1, n_2 \in 0, 1, ..., N$.

Definition 2. *The* prevalence, p, *of the less-is-more effect is the proportion of pairs* (n_1, n_2) *with* $n_1 < n_2$ *for which* $f(n_1) > f(n_2)$ *the less-is-more effect occurs.*

The prevalence p of the less-is-more effect varies between zero – no effect – for increasing $f(n)$, and unity – there is always an effect – for strictly decreasing $f(n)$. The prevalence of the less-is-more effect depends on the person's α and β. For $\alpha = 0.8$ and $\beta = 0.6$, simple enumeration yields $p = 1/3$. More generally, the following holds [RK04].

Result 1 The less-is-more effect occurs (i.e., $p \neq 0$) if and only if $\alpha > \beta$. The effect becomes more prevalent (i.e., p increases) as α increases or β decreases. The assumption is that α and β are independent of n.

At 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 is condition formalized by $\alpha > \beta$.

As an example, Goldstein and Gigerenzer discuss three Parisian sisters who have to compare all pairs of cities from the most populous $N = 100$ German cities [GG02]. All sisters have $\alpha = 0.8$ and $\beta = 0.6$, but they vary on the number of recognized objects: The youngest sister has $n = 0$, the middle sister has $n = 50$, and the eldest sister has $n = 100$. When $\alpha > \beta$ the less-is-more effect is predicted: for the middle sister, $f(50) = 0.68$, while for the eldest sister $f(100) = 0.60$. Accuracy for $\alpha = 0.8$ and $\beta = 0.6$, based on Eqs. 3.19, 3.20, 3.21, 3.22, interpolated for all n, is graphed in Fig. 3.81 (solid curve; the dashed curve will be explained below).

A less-is-more effect can emerge in at least three different situations. First, it can occur between domains, that is, when the same group of people achieves 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 most populous German cities (such as Cologne versus Frankfurt), they scored a median 71.0% (mean 71.1%) correct on their own cities but slightly higher on the less familiar German cities, with a median of 73.0% correct (mean 71.4%). This effect was

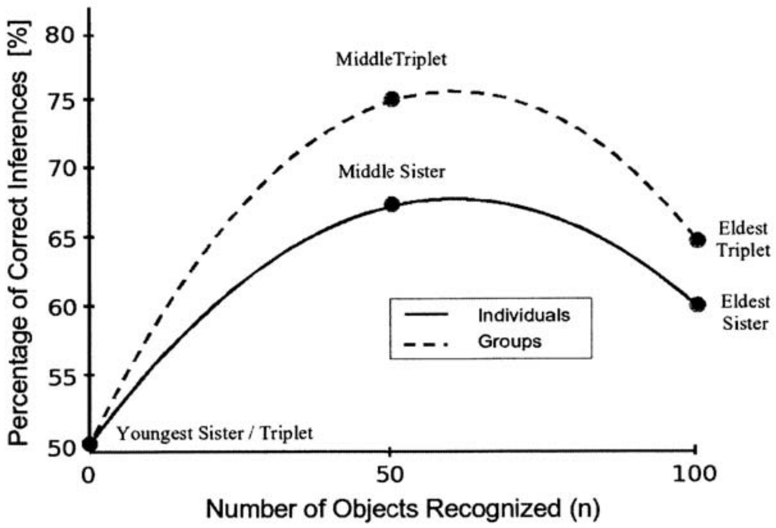

Fig. 3.81 Predicted accuracy as a function of number of recognized objects for $\alpha = 0.8$ and $\beta = 0.6$, for individuals (*solid curve*) and three-member groups that use the majority rule (*dashed curve*).

obtained despite a handicap: Many Americans already knew the three largest U. S. cities in order, and did not have to make any inferences [GG02]. 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 [Hof95]. Second, a less-is-more effect can occur during knowledge acquisition, that is, when an individual's performance curve first increases but then decreases again. Finally, the effect can occur between two groups of people, when a more knowledgeable group makes fewer correct 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 Detroit or Milwaukee is more populous [GG02]. Furthermore, Reimer and Katsikopoulos [RK04] ran a study where 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: Milan or Modena?" were displayed. The task of the group was to find the correct answer through discussion, and they were free to use whatever means. The correct solution is difficult to prove by an individual group member; thus one might expect that the majority determines the group decision [GH97].

The accuracy, $G(n)$, of a group using the majority rule is calculated as follows. Assume first that the group is *homogeneous* (i.e., all members have equal α, β, and n) and *independent* (i.e., the recognition and inference processes of members are independent given the values of the criterion on the objects). Let $F(i)$ be the probability of exactly i members being accurate and the group, using the majority rule, being correct. Finally let the group have m members, $c(m, i)$ be the number of ways in which i objects can be sampled out of m objects without replacement, and majority $(m) = (m+1)/2$ if m is odd, and $= (m/2)$ if m is even. Then, the following holds, where $i \geq$ majority (m):

$$F(i) = c(m,i)f(n)^i(1-f(n))^{m-i}(1/2), \quad i = \text{majority } (m) \quad \& \; m \text{ is even}$$
$$= c(m,i)f(n)^i(1-f(n))^{m-i}, \quad \text{otherwise.} \tag{3.24}$$

$$G(n) = \Sigma_{i=\text{majority}(m),\ldots,m} F(i). \tag{3.25}$$

The application of (3.24) and (3.25) for $\alpha = 0.8$, $\beta = 0.6$, and $m = 3$ is illustrated in Fig. 3.81 (dashed curve). A less-is-more effect is again predicted and p = 1/3. More generally, the following holds [RK04].

Result 2 Assume a homonegenous and independent group, using the majority rule. Then, (i) less-is-more effect is predicted if and only if $\alpha > \beta$ and (ii) the prevalence of the effect equals the prevalence for one member. The assumption is that α and β are independent of n.

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%), the group voted for B [RK04]. This rose to 76% if two members used recognition.

Group members letting 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. Broadly consistent with Result 2, Reimer and Katsikopoulos [RK04] observed that when two groups had the same average α and β (that were such that $\alpha > \beta$), 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% of the cities and those in a second group 80%, but the first group got 83% answers correct in a series of over 100 questions, whereas the second only 75%. 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.

3.12.4 Cue-Based Heuristics

When recognition is not valid, or people recognize all objects, heuristics can involve search for reasons or, in psychological jargon, *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 [Dar69, pp. 232–233]. 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 [Wal50] optimizing models of sequential analysis, several psychological theories postulated versions of sequential search and stopping rules [BT93]. 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 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 cue-based heuristics – that exploit some core capacities. Social heuristics exploit the capacity of humans for social learning and imitation (imitation needs not result in learning), which is unmatched among the animal species. For instance, the following heuristic generates social facilitation [Lal01]:

Do-what-the-majority-does 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-does 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 [BR85, GGH$^+$01].

Darwin, however, seems to have based his decision on cues. We will describe two classes of heuristics that search for cues. Unlike optimizing models, they do not weight and add cues. One class of heuristics dispenses with adding, and searches cues in order (a simple form of weighing). These are called *lexicographic heuristics.* Another class dispenses with weighting and simply adds, or *tallies,*

3.12.5 Lexicographic and Tallying Heuristics

We again consider the comparison task in which both objects are recognized. The decision is made on the basis of n binary cues $c_1, c_2, ..., c_n$, $n \geq 2$ (for any object a cue c_i equals 1 or 0). Since the 18th century, a popular decision rule for paired comparisons is the linear rule [KMM02] (2003). In this rule, cue c_i has a weight w_i; weights can be estimated by, say, minimizing the sum of squared differences between the predictions of the rule and the observations. For an object A with cue values $c_i(\text{A})$, the score $\Sigma_i c_i(\text{A}) w_i$ is computed and the object with the higher score is picked. If the scores are equal, an object is picked randomly. *Tallying* is a linear rule where weights are equal to unity, an old idea in psychological measurement [Gul50].

Take The Best is a heuristic in the lexicographic tradition. First, cues are ordered by decreasing validity, where the validity v_i of cue c_i is the conditional probability that the cue points to the larger object ($c_i = 1$ on the larger object and $c_i = 0$ on the other object) given that the cue discriminates between the objects [$c_i(\text{A}) \neq c_i(\text{B})$]. (Without loss of generality it can be assumed that $1 \geq v_i \geq 1/2$). After cues are ordered, the decision maker inspects the first cue. If this cue points to one of the objects then this object is taken to be larger. 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.

One-cue decision making has been observed in high-stake decisions. British magistrates tend to make bail decisions on the basis of one good reason only [Dha03, DA01], and so do British general practitioners when they prescribe lipid-lowering drugs [DH01]. Many parents rely on one cue to decide on which doctor to drive to in the night when their child becomes seriously ill [Sco02].

Both *take the best* and *tallying* are naïve in the sense that they do not take cue dependencies into account. While at first glance they might appear simplistic, simulation studies have shown that naïve heuristics compare remarkably well to statistical benchmarks. Three decades ago, Dawes and Corrigan (1974) convincingly argued that tallying can have greater predictive accuracy than linear regression. Einhorn and Hogarth [EH75] provided statistical reasons for this, including the absence of sampling error in the estimation of weights. Czerlinski, Gigerenzer, and Goldstein [CGG99] replicated this finding in twenty real-world datasets, emphasizing the concepts of *overfitting* and *robustness*. To define these, we distinguish between a learning sample on which a model estimated its parameters and the test sample on which the model is tested. Both samples are randomly drawn from the same population.

Definition 3. *A model M overfits the learning sample if an alternative model M' exists such that M has a smaller error than M' in the learning sample but a larger error in the test sample. In this case, M' is called the more robust model.*

Figure 3.82 shows the accuracy of three heuristics compared to linear regression, averaged across 20 real-world problems [CGG99], e.g., to predict 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

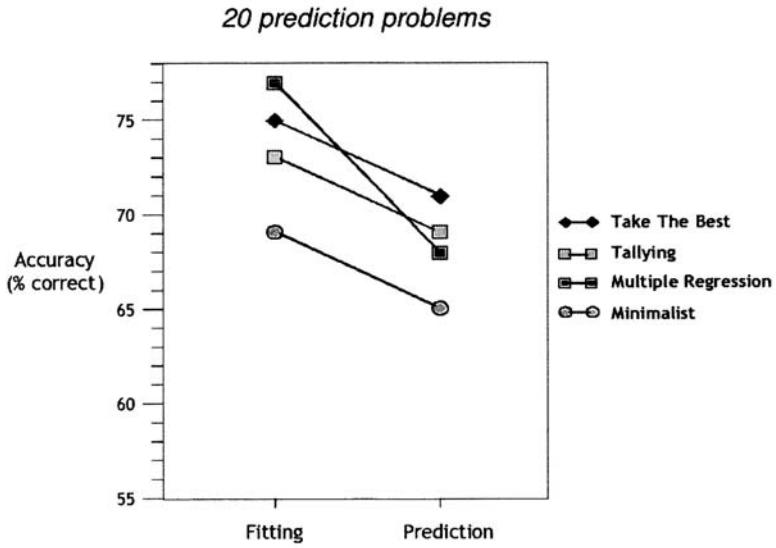

Fig. 3.82 Robustness of three heuristics and linear regression, averaged across 20 real-world problems.

of the students on various standardized tests. (Other problems were to predict people's attractiveness judgments, homelessness rates, adolescents' obesity at age 18, etc.) The three heuristics were *take the best*, *minimalist* (which is a lexicographic heuristic that searches cues in random order), and *tallying*. *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, which amounted to an average of 7.7 cues. How accurate are the heuristics?

Linear regression had the best fit. However, the true test of a method concerns its predictive accuracy, which was tested by cross-validation, that is, the four methods learned their parameters on half of the data (learning sample), and were tested on the other half (test sample). Figure 3.82 shows that *regression* over-fitted the data relative to both *take the best* and *tallying*. 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). By extracting too much information from the data, one will get a better fit but 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 dispenses with both weighting and adding, extracts too little information from the data.

In general, the predictive accuracy of a model increases with its fit, and decreases with its number of adjustable parameters, and the difference between fit and predictive accuracy grows smaller with larger number of data points [Aka73, FS94]. 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 cue that allows one to make a decision and ignore the rest of the cues, 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 [HT04]. Yet experts base their judgments on surprisingly few pieces of information [Sha92], and professional handball players make better decisions when they have less time [JR03]. People can form reliable impressions of strangers from video clips lasting half a minute [AR93], shoppers buy more when there are fewer varieties [IL00], and zero-intelligence traders make as much profit as intelligent people do in experimental markets [GS93]. 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 [SWM$^+$02]. Less can be more.

Beyond computer simulations, mathematical analyses have also been used to investigate the accuracy of heuristics [MH02, HK05, KM00]. In the case where cues are conditionally independent (i.e., independent given the values of the criterion on the objects), the optimality of *take the best* (that searches cues in the order $c_1, c_2, \ldots, c_n$) and *tallying* can be characterized as follows [KM00].

Result 3 For conditionally independent cues, *Take The Best* is optimal if and only if $o_i > \Pi_{k>i}(o_k)$, where $o_i = v_i/(1 - v_i)$.

Result 4 For conditionally independent cues, *tallying* is optimal if and only if $v_i = v$.

3.12.6 Other Tasks

The paired comparison task is related to other tasks such as deciding whether an object is larger than a certain threshold (classification) or judging how large the object is (estimation). Fast and frugal heuristics have been studied for these tasks as well and have been again found to perform well compared to standard benchmarks [HR08]. For example, Katsikopoulos, Woike, and Brighton [Pro] found that simple, fast and frugal trees can make more robust classifications than discriminant analysis and trees used in artificial intelligence do [BFOS84]. Research on other decision tasks is reviewed by Gigerenzer [Gig04].

3.12.7 Fast and Frugal Heuristics in Technology Development?

According to Moses [Pap04], one of the main goals of the study of technological systems is to deal with changes that occur during the life cycle of these systems. Change can be dealt with actively by building flexibility into the system, that is, allowing the system to perform a number of functions. Change can also be dealt with passively by building a robust system, that is, a system that does not lose much of its performance when conditions vary. Both flexibility and robustness are necessary when there is uncertainty in the system. One might even say that uncertainty represents a chance for improvement in that it motivates flexibility and robustness (perhaps this is what Moses means when he talks about "viewing uncertainty as an opportunity" [Pap04]). The final challenge is to combine these properties with transparency and usability so that the system will be accepted by its users.

Our research program was not designed to study technological systems. Our results about the normative success of heuristics were obtained in decision tasks such as comparing dropout rates in highschools. We did not study how engineering students and practitioners compare, say, two product designs. Of course, at a certain level of abstraction, these are very similar tasks, but we do not want to downplay the potential influence of context. Thus, we see our results as making a methodological suggestion about a new program of research. We believe that a fast-and-frugal-heuristics approach to making decisions in engineering systems may be helpful.

Fast and frugal heuristics tend to be robust. There are more results than those presented here, to this effect. For example, Brighton [Bri06] has pitted the heuristics against powerful machine learning methods (such as Quinlan's ID3 method) using the minimum description length as a criterion of robustness. He found that, in many cases, heuristics compressed the data more than the machine learning methods.

With their focus on external outcomes, heuristics implement more practical intelligence than do mathematical methods that target full internal rigor. Furthermore, heuristics are less ambitious than methods that try to work all the time; heuristics are problem-specific and information-structure-specific. A given heuristic may be applied successfully only to those comparisons, estimations, classifications, or choices with certain statistical properties (i.e., flat or very skewed distribution of cue validities). Taken together, however, heuristics cover a wide spectrum of decision tasks. The set of heuristics has been called the *adaptive toolbox* [GS01].

The adaptive toolbox is a flexible system for decision making: To build the heuristics in it, one combines different building blocks (rules for searching for information, e.g., by validity, with rules for deciding based on the available information, e.g., use only one cue). The building blocks themselves are based on core psychological capacities (e.g., recognition). The toolbox allows the introduction of new heuristics by *(i)* combining existing building blocks in new ways or by *(ii)* creating new building blocks based on newly discovered capacities.

Finally, fast and frugal heuristics are transparent: They are easy to understand and apply (and, hence, are more acceptable). There are two reasons for this. First, heuristics are expressed as clear and simple sequential algorithms (e.g., *Take The Best*). Second, the way they represent the information they use is consistent with people's cognitive representations (e.g., in *Take The Best* validities can be cast in terms of frequencies, not conditional probabilities; see also [HLHG00]). Perhaps for these reasons, some practitioners, such as medical doctors, advocate the substitution of classical decision analysis with fast and frugal heuristics [EEER01, Nay01, KF00]. Just as some successful methods for engineering decision making – Pugh's concept selection [patICoED81] – heuristics can be used to generate, rather than to suggest or impose, new possibilities. We want to encourage academics and practitioners to explore the potential of fast and frugal heuristics in engineering.