Heuristic Decision Making

by Elke Kurz-Milcke and Gerd Gigerenzer

The study of heuristics analyzes how people make decisions when optimization is out of reach. It focuses on two questions, the first descriptive, the second normative: What are the heuristics in the adaptive toolbox? In which environments does a given heuristic succeed, and in which does it fail? Studying the adaptive toolbox involves analyzing the building blocks of heuristics and the evolved capacities they exploit. The resulting process models describe search rules, stopping rules, and decision rules. Studying the ecological rationality of heuristics reveals environmental structures in which particular heuristics are better than other strategies, including optimization. Knowledge of heuristics can help marketing researchers understand the processes by which consumers make decisions with limited time and information.
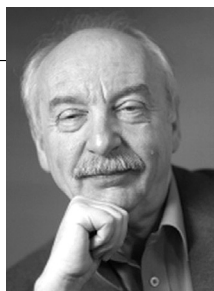
1. Introduction

Research on judgment and decision making is troubled by a conflict between how people actually make decisions and how it is thought they should make them. Con-

Elke Kurz-Milcke is a cognitive scientist working with the Institute of Mathematics at the Pedagogical University of Ludwigsburg, Reuteallee 46, 71634 Ludwigsburg, Germany, Phone: +49-7141-140-729, Fax: +49-7141-140-435, E-Mail: kurzmilcke@ ph-ludwigsburg.de

Gerd Gigerenzer is Director at the Max Planck Institute for Human Development, where he runs the Center for Adaptive Behavior & Cognition, Lentzeallee 94, 14193 Berlin, Germany, Phone: +49-30-824 06 460. Fax: +49-30-824 06 394, E-Mail: gigerenzer@ mpib-berlin.mpg.de

sider a classic review of 45 studies in which the process of decision making was investigated by means of mouse-lab, eye movement, and other process tracking techniques (*Ford et al*. 1989). The choice set varied between studies, including apartments, microwaves, and birth control methods:

"The results conclusively demonstrate that noncompensatory strategies were the dominant mode used by decision makers. Compensatory strategies were typically used only when the number of alternatives and dimensions were small or after a number of alternatives have been eliminated from consideration." (Ford et al. 1989, p. 75)

The essence of compensatory processes is that they make trade-offs between attributes, as illustrated by the weighting and adding of pros and cons in linear decision rules. Noncompensatory processes, in contrast, make no trade-offs. Among these are lexicographic heuristics such as Take The Best (see below) and elimination-by-aspect (*Tversky* 1972).

Now consider what *Keeney* and *Raiffa* (1993) have to say about lexicographic heuristics. They warn us that such a strategy

"is more widely adopted in practice than it deserves to be," "is naively simple," and "will rarely pass a test of 'reasonableness'" (pp. 77–78).

Keeney and *Raiffa* echo a broad consensus about the nature of rational choice. Virtually all theories of rational decision making, from expected utility theory to moral consequentionalism, are based on the assumption that making trade-offs is necessary for rational choice. Consequently, observations that adults rely on noncompensatory processes but children on compensatory ones are reported as strange irregularities (e.g. *Reyna/Farley* 2006). Most of the literature in psychology (e.g. *Kahneman/Slovic/Tversky* 1982), behavioral economics (e.g. *Camerer* 1995), and behavioral law and economics (*Sunstein* 2000) has resolved this conflict by blaming human information processing rather than the norms of rationality. The typical explanation is that our minds suffer from cognitive limitations, such as limited memory, that leave us little choice but to rely on simple heuristics. Why evolution would have imposed these limitations on us if they are harmful remains an open question.

This standard interpretation is emphatically not ours. In this article, we introduce marketing researchers to a body of theory and experimental research that studies the rationality of fast and frugal heuristics (*Gigerenzer/Todd/ The ABC Research Group* 1999; *Gigerenzer/Selten*

2001). Our stance is that the descriptive and normative claims are in conflict largely because of inappropriate norms, specifically norms that are based on internal criteria such as consistency rather than norms that relate to the success of heuristics in the external world. We argue instead for what we call the science of heuristics, a program that investigates how real people make decisions in an uncertain world, where optimization is often out of reach or not worth the effort.

We investigate first the empirical case for heuristics, then the normative case, and in closing deal with the methodological implications of both. For space reasons, we restrict this introduction to three classes of heuristics; for more, see *Gigerenzer* (2004a, 2006).

2. The Empirical Case for Heuristics

2.1. Lexicographic Heuristics

How do people choose between the myriad of alternatives in consumer electronics? The common assumption is that consumers weight the features, add the values for each alternative, and choose the one with the highest overall value. In this understanding of the processes underlying people's choices, a high level on one feature can compensate for low levels on others. But is this compensatory process a realistic model for how people choose from a large set of, say, SmartPhones? *Yee et al.* (in press) invited a couple of hundred respondents to complete a web-based questionnaire about SmartPhones. Respondents were introduced to a six-feature description of the products and asked to successively rank a selection of 32 such phones by clicking the picture of the phone they liked best in the set, after which their choice disappeared from the set, and continuing in this manner until all phones were ranked. Next, and importantly, the respondents were asked to rank-order two sets of four SmartPhones chosen randomly from a new set they had not seen before. This procedure allowed the authors not only to fit various choice models to the obtained data but also to test their predictive accuracy with the holdout task data. The authors tested two compensatory models and a hierarchical Bayes model (a ranked logit model), performed a linear programming estimation, and proposed a procedure for identifying the best lexicographic representation.

Consider the following lexicographic procedure in which the highest-valued attribute, such as "has a Microsoft operating system" – yes or no – is checked first, and only products for which the answer is positive are retained among the set of possible choices. Then a second attribute, say "flips open" is checked for the remaining SmartPhones, and again those that come out positive are retained. But what is the appropriate ordering of attributes? A decisive step in the fitting of a lexicographic procedure is to determine the best ordering of attributes, that is, the ordering that best predicts how respondents rank-order profiles of objects – here, SmartPhones. Whereas a complete enumeration of all possible orderings is still feasible for small numbers of attributes, it very quickly demands extensive computational power with a larger number. For *n* binary attributes, *n!* orderings need to be checked. Note that this computational problem is NP-hard, that is, computationally intractable (*Martignon/ Hoffrage* 2002). Yet this is only a problem for those who make the models, not for the consumers, who evidently arrive at their order of attributes on other grounds. *Yee et al.* (in press) proposed a set of algorithms for determining whether a lexicographic ordering of attributes exists that is consistent with a set of profile orderings and, in the case that none is fully consistent, for establishing the lexicographic order that best fits the respective profile order.

How well did the lexicographic model predict consumers' ranking of SmartPhones, compared to the compensatory models? The answer is, very well. Wellness of fit was equal or better in the first set of 32 phones, but most importantly, the predictive accuracy of the lexicographic heuristic in the hold-out task was better than that of the two compensatory models. For 57% of the SmartPhone respondents, the lexicographic strategy predicted at least as well as the hierarchical Bayes model did. In a study that involved choice between computers, this figure was 75%. In summary, lexicographic heuristics predicted rankings of many featured items, such as consumer electronics, better than compensatory models did.

A lexicographic heuristic generally consists of three building blocks, as illustrated by the *Take The Best* heuristic (*Gigerenzer/Goldstein* 1996):

Search rule: Look up attributes in order of validity.

Stopping rule: Stop search after the first attribute discriminates between alternatives.

Decision rule: Choose the alternative that this attribute favors.

The three building blocks specify the steps of information processing. These building blocks are fitted to each other to form an inductive device, in this case, one that decides on an alternative in a choice task. The one-good-reason stopping rule employs extremely limited search, yet the search rule adjusts for that by ordering cues according to their validities. Note that validity, which is defined as the proportion of correct choices an attribute allows among all choices, does not guarantee the "best" ordering of cues; it ignores dependencies between cues and nonetheless (or because of this) produces reasonably good and robust orders. Once an inductive device such as the Take The Best heuristic is specified, its performance can be studied in various task environments. In this way, the empirical study of ecological and bounded rationality includes both analyses of the structure of environments and of the performance of heuristics in these environments. Exposing such clearly defined inductive devices to various task environments and observing their performance relative to these environments naturally engen-

ders questions concerning their informational structures. We will return to this aspect of the research program when presenting the rational argument for heuristics.

Lexicographic heuristics can be mathematically formulated as a special case of an additive model where the weights are constrained to be noncompensatory, such as 1, 1/2, 1/4, and 1/8 for binary attributes. However, such an additive model is not psychologically equivalent to a lexicographic process. For instance, the constrained additive model postulates no order in which attributes are looked at and assumes exhaustive search for attributes. Yet a person that uses a lexicographic heuristic looks up attributes in order, and employs limited rather than exhaustive search. Although both procedures arrive at the same choice, the processes differ: The heuristic is faster and ignores information. Underlying the prevalent routine of modeling consumers' preferences with an additive model are implicit assumptions about the purpose of modeling and the form that models should take in the social sciences. Consider in this respect the practice of modeling by paramorphic representations, a variety of "as-if" models, which have been used in research on judgment and decision making to refer to the fitting of linear regression models to judgment data (see *Kurz/ Martignon* 2002 for an analysis of this research tradition). As long as other contenders could not better predict (often only better fit) the data than the linear regression model, it fulfilled its purpose as a "paramorphic representation", although it admittedly did not model judgment processes properly (*Hoffmann* 1960). Ironically, this state of affairs began to be subverted from within when this line of research begun to study unit weight models in the 1970s (e.g. *Dawes* 1979). The appearance of contenders that were nearly able to match multiple regression in terms of fit and outperform it in terms of prediction forced this standard procedure to be reconsidered. However, the unit weight model, assigning the same weight to each variable in the model and then adding these up for each alternative in the choice set, was still a strictly linear and compensatory model. Note that in this case, the term *model* refers to the outcome of a computational procedure, the fitted linear equation. An alternative to this practice are the lexicographic heuristics we have described, in which models, as in *models of ecological and bounded rationality*, signify inductive devices that specify an ordered set of steps that result in a judgment or choice.

2.2. Recognition Heuristic

Consumers are often able to give sophisticated descriptions of differences in taste among various brands of peanut butter, beer, or red wine. Do these matter when consumers choose products? In an experiment, participants had a choice between three jars of peanut butter (*Hoyer/Brown* 1990). In a pre-test, one brand had been rated as higher quality, and participants could identify the higher-quality product 59 % of the time in a blind test (substantially higher than chance, which was 33 %).

With another group of participants, the researchers put labels on the jars. One was a well-known national brand that had been advertised heavily and which all participants recognized; the other two were brands they had never heard of before. The critical situation was when the experimenters put the higher-quality peanut butter into the jars with the unrecognized labels, and the participants were asked to taste and choose. Would the same percentage of participants still opt for the best-tasting peanut butter? No. This time 73 % chose the low-quality product with the recognized brand label, and only 20 % the high-quality product. Name recognition was more influential than taste perception. In a second tasting test, the researchers put exactly the same peanut butter into three jars, again two with unrecognized labels and one with a recognized brand name. The result was nearly identical. In this case, 75 % of the participants chose the jar with the recognized brand, even though its content was the same as that in the two other jars. Marking one brand with a higher price than the other two had minimal effect. Altogether, taste and price mattered little compared with the influence of brand name recognition.

For the simplest case of choosing between two alternatives, the underlying process can be described by the *recognition heuristic*:

If you recognize one alternative but not the other, then choose the recognized alternative.

If there is a large set of alternatives, however, brand name recognition can only determine the consideration set, that is, the consumers who follow this principle consider only alternatives they recognize and reject others. Here, the total set of alternatives is first reduced to the consideration set, and a further heuristic principle is needed to pick one alternative. One illustration is a heuristic an American professor of business uses for buying a stereo set:

When you buy a stereo, choose a brand you recognize and the second-least expensive model.

The professor's rationale is that if he has heard of a company, it is likely because its products are good. His justification for the additional step is that the quality of stereo technology has reached a level at which he is no longer able to hear the difference, so it does not matter which stereo he purchases – except for the cheapest and potentially less reliable model that companies manufacture for the low-price market. This rule saves time, and likely protects him from being taken in.

A large experimental literature documents that people tend to rely on the recognition heuristic in situations when it is ecologically valid (see below), from choosing a college based on name recognition to predictions of the outcomes of tennis matches in Wimbledon (*Goldstein/ Gigerenzer* 2002). The recognition heuristic takes advantage of the evolved capacity of recognition memory – face, voice, or name recognition – whereas the lexicographic heuristics exploit recall memory.

2.3. 1/*N* Heuristic

How to invest your money in *N* assets? In 1990, Harry Markowitz received a Nobel Prize in Economics for his theoretical work on optimal asset allocation. He addressed a vital investment problem that everyone faces in some form or other, from saving for one's retirement to earning money on the stock market. Markowitz showed that there is an optimal portfolio that maximizes the return and minimizes the risk. Nevertheless, for his own retirement investments, he relied on a simple heuristic, the *1/N rule*:

Allocate your money equally to each of N funds.

There is considerable empirical evidence for this heuristic: about 50% of people studied rely on it, and most consider only about 3 or 4 funds to invest in. Researchers in behavioral finance have criticized this behavior as too simple. Note that 1/*N* is not a financial investment heuristic; its range is much broader. When children divide the treats they collected at Halloween, the same heuristic is at work. In the ultimatum game, a majority of adults offer a 50/50 split, sometimes slightly corrected in favor of the proposer. In other experimental games, the strategy to divide equally is known as the equity heuristic. Another incarnation of the heuristic is LaPlace's principle of ignorance, where the prior probabilities are determined in exactly the same way.

For all three of these heuristics – lexicographic, recognition, and 1/*N* – evidence exists that people rely on them when making consumer decisions. Although they are different, they share some important features. Each heuristic allows for fast decisions without wasting any time and is frugal, ignoring part of the information. All three can be amazingly successful and accurate, as we will see in Section 3. These three heuristics are instances of a larger family of heuristics that have been studied (*Gigerenzer* 2004a). For the role of social heuristics and emotions, see *Gigerenzer/Selten* (2001).

2.4. Methodological Implications

The three heuristics illustrate the general empirical claim that people often base their decisions on heuristics, especially in situations when it is ecologically rational to do so (see below). The empirical case for heuristics has methodological implications for marketing research. We consider here three important ones (*Gigerenzer/Todd/The ABC Research Group* 1999).

Beware of the routine use of conjoint analysis. If a majority of consumers rely on lexicographic heuristics, or the recognition heuristic, then the routine use of conjoint analysis will lead to misleading results. Conjoint analysis assumes that people use a linear trade-off rule and determines the best-fitting weights (unlike its parent, conjoint measurement theory, which tests whether the necessary and/or sufficient axioms are fulfilled). One of the authors of the SmartPhone study, John Hauser, had relied on conjoint analysis for most of his academic career until he began actually testing complex compensatory strategies against lexicographic ones. He found little evidence for the former, but strong evidence for lexicographic heuristics, consistent with the review by *Ford et al.* (1989) cited in the introduction. Routine fitting of linear models, from conjoint analysis to multiple regression, does not detect lexicographic and other heuristic strategies, even if consumers rely on them.

Beware of the difference between "small world" and "real world" judgments. As noted by *Ford et al.* (1989), compensatory processes are more often observed when participants are put into what we call a "small world" with only a few alternatives and attributes. That is, reducing the realistic choice between a large number of brands that vary on many attributes to only a few in an experiment may change the decision process. Consequently, caution is warranted in concluding that the findings in the restricted "small world" generalize to the unconstrained setting.

Beware of the difference between memory-based and menu-based judgments. Heuristics emphasize the process of search, and it has been shown that search in memory differs from search outside memory (*Hertwig et al.* 2004). Search outside of memory occurs for example in libraries and on the Internet. Lexicographic heuristics such as Take The Best are more common in tasks that involve search in memory (*Bröder/Schiffer* 2003). In an experiment, it thus matters whether all the relevant information is displayed in front of the participant, or whether the person is asked for judgments requiring search for information in memory. If search is not allowed in an experiment, the results suggest that people more often make trade-offs.

3. The Normative Case for Heuristics

In Section 2, we argued that people rely on heuristics. But why don't people optimize and weigh and add? As mentioned before, the traditional answer is cognitive limitations. For example, "employing simplifying heuristics is a rational approach to decision making only *because* of our cognitive limitations." (*Korobkin* 2003, pp. 1292–1293). We disagree. There are important reasons for using heuristics that do not relate to cognitive limitations but instead reside in the environment. The study of the ecological rationality of a heuristic answers the question of what environmental structures a heuristic can exploit. This normative program looks at both so-called optimization methods and heuristics and ascertains in which environments one strategy works better than another. It teaches us the conditions in which a simple heuristic can be more accurate than optimization, and vice versa. Internal definitions of rationality such as consistency are not the single yardstick for determining the rationality or irrationality of heuristics. Their rationality is ecological, not logical.

3.1. Computational Intractability

Loosely speaking, a problem is called computationally intractable if there is no machine or mind that can find the optimal (best) solution in a reasonably short time (say, a millennia or the time since the Big Bang). The class of intractable problems includes well-defined games such as chess, computer games such as Tetris, the traveling salesman problem, and all ill-defined problems such as finding the best mate or business partner. It also includes all problems with more than one goal and vaguely defined goals such as happiness (*Gigerenzer* 2004b). Most problems in AI are computationally intractable (*Reddy* 1988).

If a problem is computationally intractable, it makes little sense to assert that people rely on Bayes' rule or some other optimization method in order to solve it. Moreover, the related claim that people behave *as if* they optimized is equally unrealistic. If neither mind nor machine can find the best solution, heuristic methods may well be the only way to cope with this class of problems. Computational intractability is one important reason why lexicographic decisions are not necessarily second-best, as *Keeney* and *Raiffa* (1993) assert. The unconditional philosophy of optimization, we argue, belongs to theology rather than to rational decision theory.

Consider consumer choice of cell phones on the Internet, as in Hauser's study. A complete decision tree is computationally intractable with 2^n exits, that is, it increases exponentially with the number of attributes. This number holds for binary attributes, otherwise it is higher. By contrast, heuristics such as Take The Best and fast and frugal trees ("pruned" decision trees) are computationally tractable because they only search for very few attributes. Note that Take The Best establishes a good-enough order, not an optimal one; the problem of ordering attributes in an optimal way is NP-hard (*Schmitt/ Martignon* 2006). Although the optimal order can clearly be found with small *n*, performance lacks in robustness. That is, with a new sample or population, the determined order ceases to be the optimal one, and can be worse than the simple order used by Take The Best (*Martignon/Hoffrage* 1999).

3.2. Robustness

Unlike chess, the asset allocation problem is computationally tractable and allows for optimization. Yet it illustrates a second reason why heuristics might be preferred to optimization: the possible severe consequences of estimation errors. Consider the 1/*N* heuristic again.
How much better is optimizing than 1/*N*? A recent study compared twelve optimal asset allocation policies with the 1/*N* rule in seven allocation problems, such as allocating one's money to ten American industry portfolios. The optimal rules included Markowitz's mean-variance policy as well as Bayesian and non-Bayesian models. Despite their complexity, none of the optimal rules could

outperform the 1/*N* rule on various financial measures (*DeMiguel/Garlappi/Uppal* 2006).

How can a heuristic strategy be better than an optimizing one? At issue is not computational intractability but robustness. The optimization models did better in data fitting (adjusting their parameters to the data of the past ten years) than the simple heuristic, but worse in predicting the future. Thus, they overfitted the past data. The 1/*N* heuristic, in contrast, does not estimate any parameter and therefore cannot overfit.

The important point is still to come. 1/*N* is not always better than optimization. But when is it? The study of the *ecological rationality* of a heuristic gives us the answer. Three relevant environmental features for the performance of simple heuristics are:

(i)   the predictive uncertainty of the problem,
(ii) the number *N* of assets, and
(iii) the size of the learning sample.

Typically, the larger the uncertainty and the number of assets and the smaller the learning sample, the greater the advantage of the 1/*N* heuristic. Since the uncertainty of funds is large and cannot be changed, we focus on the learning sample, which was ten years of data. When would the optimization models begin to outperform the heuristic? The authors report that with 25 and 50 assets to allocate one's wealth, the optimization policies would need a window of 250 and 500 years, respectively, to eventually outperform the 1/*N* rule.

3.3. Less Is (Sometimes) More

The recognition heuristic illustrates another feature of heuristic-friendly environments: situations in which lack of name recognition is informative. In an environment where a firm first increases product quality, which in turn increases name recognition, reliance on mere brand names is a better-than-chance strategy. Yet one needs a beneficial degree of ignorance to use this heuristic: If one has heard of all products, than it cannot be applied. Ignorance is beneficial if the unfamiliar brands or alternatives tend to be lower in quality than the familiar ones. One way to measure this is the recognition validity α:

$$\alpha = R / (R + W), \tag{1}$$

where *R* and *W* are the numbers of correct and incorrect judgments, respectively, for all pair-wise comparisons to which the heuristic can be applied (i.e., when one has heard of one but not the other alternative). The recognition validity applies to situations in which a measurable outcome criterion exists. For instance, the collective recognition of semi-ignorant amateur players led to better predictions of the Wimbledon Gentlemen's matches in 2003 than the ATP Rankings and the seeding of the Wimbledon experts did (*Serwe/Frings* 2006). Note that this measure can be determined for the individual person as well, and thus reflect his or her particular state of ignorance in a domain. For instance, the recognition

validity of laypeople and amateur tennis players in predicting the outcomes of the Wimbledon tennis matches is typically around .70, and the validity for judging the population of foreign cities is around .80. That is, in those cases where a person has not even heard of one of the two alternatives (players, cities, and so on), he or she nevertheless has a 70 % to 80 % chance of getting the prediction right. That is often more than a highly knowledgeable person can hope to achieve. More specifically, given that the number n of alternatives (e.g., players) a person has heard of out of the total number of alternatives N (e.g., the 128 players in Wimbledon) is known, the proportion c of correct judgments made by this person when relying on the recognition heuristic can be computed in the following way:

$$c = 2\left(\frac{n}{N}\right)\left(\frac{N-n}{N-1}\right)\alpha + \left(\frac{N-n}{N}\right)\left(\frac{N-n-1}{N-1}\right)\frac{1}{2} + \left(\frac{n}{N}\right)\left(\frac{n-1}{N-1}\right)\beta,$$

where $(n \leq N)$. (2)

The three summands on the right side of the equation represent three possible states of knowledge: with the first, only one of the alternatives is recognized; with the second, none of the two alternatives is recognized; and with the third, both are recognized. Given the first state, the recognition heuristic can be applied, and the proportion is thus multiplied by α. Given the second state, the person has to guess, and the probability of success is 1/2. Given the third state, where both objects are recognized, the heuristic cannot be applied, and the judgment is based on whatever knowledge a person has, which is measured by the validity of knowledge β:

$$\beta = R_k / (R_k + W_k),$$ (3)

where R_k and W_k are the numbers of correct and incorrect judgments, respectively, for all cases where both alternatives are recognized. This analysis provides us with an answer to the question of when relying on the recognition heuristic is a reasonable strategy. It is reasonable if $\alpha > .5$, that is, one's lack of recognition carries information. Most interestingly, the recognition heuristic can lead to a less-is-more effect if

$$\alpha > \beta.$$ (4)

The conditions under which this counter-intuitive phenomenon occurs and the respective experiments are reported in *Goldstein* and *Gigerenzer* (2002). For a simple illustration consider the following study. American students were asked which city has a larger population, San Diego or San Antonio. About 63% gave the correct answer, San Diego (*Goldstein/Gigerenzer* 1999). Next, German students were asked the same question. They knew little about San Diego, and many had never heard of San Antonio. What proportion of Germans found the right answer? The result was 100%. How can it be that people with less knowledge are more accurate than those who know considerably more? The answer is that the Germans relied on the recognition heuristic, whereas the Americans could not. They had heard of both cities, and knew too much.

The fact that consumers rely on the recognition heuristic can be exploited by non-informative advertisement. Recall that if a firm invests in product quality, and as a consequence, word of mouth or the media increase the firm's brand name recognition, the recognition heuristic is a reasonable guide to shopping. Of course, this process can be shortcut with extensive advertisement where firms invest huge amounts of money to buy a place in consumers' recognition memory or increase the fluency with which their brands' name and image is processed. Companies such as Benetton, for instance, do not even provide any information about their product but are only concerned with increasing their brand-name recognition. If more than one alternative is recognized, degrees of processing fluency can make the difference (for the *fluency heuristic*, see *Schooler/Hertwig* 2005). Generally, for a firm to take advantage of consumers' recognition-based heuristic processing, it is important that consumers hear and see its brand name but also that they are prevented from hearing and seeing the competitors' names.

The ecological analysis of the recognition heuristic specifies one condition under which less is more. It is an example of conditions under which (i) ignoring available information is beneficial, as in lexicographic heuristics, (ii) less time is beneficial, as in studies with expert golfers who play better when their time is restricted (*Beilock et al.* 2004), and (iii) social change is enabled when the major actors are partially ignorant (*Gigerenzer* in press, chap.11).

3.4. Methodological Implications

Test the predictive accuracy of strategies, not fitting. Model testing in the social sciences is often performed by fitting parameters to given data (*Roberts/Pashler* 2000). The R^2 of fitting typically looks more impressive than that of prediction. Yet we want theories that predict in foresight, not in hindsight. For instance, the mean-variance models of optimal asset allocation performed better than $1/N$ in hindsight, that is, in fitting its parameters to past data, but less in foresight, that is, for predicting future investment performance. By data fitting only, one would have arrived at the wrong conclusion that the optimization models do better. Similarly, Take The Best is less accurate in fitting than multiple regression, but on average, more accurate in cross-validation, that is, prediction (*Czerlinski/Gigerenzer/Goldstein* 1999). A complex strategy that is more accurate than a simple heuristic in fitting but less accurate in prediction is said to overfit. The importance of testing the predictive power of various strategies holds equally for the normative question (e.g., in which environments is $1/N$ rational?) and the descriptive question (e.g., do people follow a heuristic?).

Apart from cross-validation, also known as out-of-sample prediction, there are other ways to measure the predictive accuracy of strategies. One is out-of-population prediction, as when a clinical testing procedure validated in one hospital is applied to the population in a different one. A third kind of prediction has been rarely

investigated: measuring the robustness of strategies when environments change unexpectedly, such as when animals encounter a new predator species, or when a product is introduced into a new segment of the market. It seems that in order to protect against the consequences of surprises, behavior has to be suboptimal relative to the known world (*Bookstaber/Langsam* 1985).

Investigate the ecological rationality of heuristics. Rather than automatically assuming that heuristics are second-best solutions, it is necessary to empirically study the structure of environments in which a heuristic works and fails. To do this, one has to specify a currency such as predictive accuracy or frugality, and then use analysis or computer simulation in order to compare the performance of various heuristics. In addition, models of heuristics need to be precisely specified. Mere verbal labels such as availability and representativeness are insufficient; they are too vague to allow for a study of their ecological rationality (*Gigerenzer* 1996). This article has presented some examples; for more, see *Goldstein et al.* (2001) and *Hogarth* and *Karelaia* (2005, 2006).

4. Four Visions of Rationality

Rationality, in spite of its image to the contrary, comes in distinct flavors. These can be classified according to a few ideals or themata. Three key ideals to which classical theories of rationality aspire are optimality, universality, and omniscience. Optimality means that a theory of rationality is about the very best action or strategy, not just a good-enough one. The ideal of omniscience assumes by default that the decision maker has complete information about all relevant aspects of the task. Universality is the ideal that there is one and only one calculus or rationality. The origins of these three ideals are related to the seminal developments in mathematics beginning with the introduction of the differential and integral calculus and the taming of uncertainty and chance in the formulation of probability theory throughout the 17th and 18th century. Here is a short sketch.

The ideal of optimality was made feasible by developments in mathematics, such as computing the answer to the question of what happens "at the limit." Optimization, such as finding the maximum or minimum of a function, became both feasible and desirable. Leibniz, who discovered differential calculus, envisioned the Universal Characteristic, a universal calculus that could provide the answers to all our problems. The remarkable achievements in formal mathematics transported ideals of omniscience, such as using our intellect for driving out all uncertainties from the world. The French astronomer and physicist Pierre-Simon Laplace, who made seminal contributions to probability theory, promoted the fiction of an ideal being – later known as Laplace's superintelligence or demon – who knows everything about the past and present, and can calculate the future. Note that the demon calculates the future, which corresponds to

the task of prediction rather than data fitting. Why so much of social science has been enticed into interpreting the task as one of fitting the past is itself an interesting question.

These themata have variably shaped the four main conceptions of rationality that underlie present-day understandings of cognition and decision making (*Gigerenzer* 2006):

(1) unbounded rationality
(2) optimization under constraints
(3) cognitive illusions
(4) ecological rationality

Unbounded rationality assumes optimality, universality, and omniscience. In this view, cognitive agents behave as if they were able to find the optimal strategy, one that maximizes some criterion. Maximization of expected value or expected utility are among such commonly found criteria. Unbounded rationality is assumed in many a theory, from economics to optimal foraging theories in biology: firms, individuals, and animals know all relevant behavioral options and the benefits, costs and probabilities of their consequences.

Optimization under constraints drops the ideal of omniscience, and introduces search for information, including search costs. Internal constraints such as limited memory and external constraints such as information costs imply that unbounded rationality is out of reach for humble humans, as opposed to demons. Optimization under constraints builds (some of) these limits into the theories, but retains the ideal of optimality. Such theories can in fact impose a great deal of new requirements upon the knowledge decision makers need to have about the costs and benefits of attaining certain options. For instance, finding the optimal stopping point in information search requires additional knowledge to determine the point where the benefit of further search is neutralized by its costs. Thus, the intention of attaining more realism by dropping omniscience for information search while retaining optimality is easily frustrated by new and unrealistic demands on computational resources incurred from the ideal of optimal cost-benefit analyses.

The study of cognitive illusions, also known as the heuristics and biases program (*Kahneman/Slovic/Tversky* 1982) differs from the previous two approaches in that it does not assume that humans are intrinsically rational. It has produced a long list of biases and influenced many fields, among them social psychology and behavioral decision making. It also helped to create new fields such as *behavioral economics* and *behavioral law and economics*. Although it appears diametrically opposed to rationality, with and without constraints, this only holds for the descriptive conclusions that people suffer from cognitive illusions. The cognitive illusions program does not criticize the norms of logic or optimization in the two previous programs. With this image of rationality, deviations in human judgment and reasoning are interpreted as fallacies.

Ecological rationality, in contrast, replaces all three heavenly ideals. The study of heuristic decisions replaces optimization. This is not to say that there is no room for optimization: When "small worlds" are studied, for instance, optimization can be a possibility. The ideal of universality is replaced by that of modularity. Heuristics come in the plural, and the adaptive toolbox is a substitute for Leibniz's dream of a universal calculus. Finally, the ideal of omniscience is replaced by the study of decision making under limited information and time. The search and stopping rules, however, are not based on optimality but on robust rules as defined above. Thus, ecological rationality is based on satisficing rather than optimizing, modularity rather than universality, and limited search rather than omniscience. The less-is-more phenomena document that limited information can actually beneficial, as can cognitive limitations (*Hertwig/ Todd* 2003). Most importantly, this program brings new themata into the foreground, such as computational tractability and robustness. These reflect the complexity and the uncertainty of the world in which humans, as opposed to demons, live.

5. The Science of Heuristics

The study of fast and frugal heuristics asks two key questions: What heuristics are in the adaptive toolbox, and what are their building blocks? In which environments does a given heuristic work, and in which would another heuristic be better? The fact that people rely on a multitude of heuristics is well-documented in the empirical literature (e.g. *Bröder* 2003; *Bröder/Schiffer* 2003), as is the fact that these heuristics are often used in an adaptive way (e.g. *Payne/Bettman/Johnson* 1993). These results remind us that consumers do not always weigh and add when they make decisions, and that they may vary their decision processes from situation to situation, according to the perceived ecological rationality. The principle of robust decisions reminds us that less is sometimes more. Instead of trying to optimally integrate everything, good decisions in the real world need to know what information to ignore, and heuristic rules of search and stopping search provide models of this intuitive skill.

References

Beilock, S.L./Bertenthal, B.I./McCoy, A.M./Carr, T.H. (2004): Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills, in: Psychonomic Bulletin & Review, Vol. 11, pp. 373–379.

Bookstaber, R./Langsam, J. (1985): On the optimality of coarse behavior rules, in: Journal of Theoretical Biology, Vol. 116, pp. 161–193.

Bröder, A. (2003): Decision making with the "adaptive toolbox": Influence of environmental structure, intelligence, and working memory load, in: Journal of Experimental Psychology, Vol. 29, pp. 611–625.

Bröder, A./Schiffer, S. (2003): Take The Best versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format, in: Journal of Experimental Psychology: General, Vol. 132, pp. 277–293.

Camerer, C. (1995): Individual decision making, in: Kagel, J.H./ Roth, A.E. (Eds.): The handbook of experimental economics, Princeton, NJ, pp. 587–703.

Czerlinski, J./Gigerenzer, G./Goldstein, D.G. (1999): How good are simple heuristics? in: Gigerenzer, G./Todd, P.M./The ABC Research Group: Simple heuristics that make us smart, New York, pp. 97–118.

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

DeMiguel, V./Garlappi, L./Uppal, R. (2006): 1/N, Unpublished manuscript, EFA 2006 Zurich Meetings, available at SSRN: http://ssrn.com/abstract=911512.

Ford, J.K./Schmitt, N./Schechtman, S.L./Hults, B.H./Doherty, M.L. (1989): Process tracing methods: Contributions, problems, and neglected research questions, in: Organizational Behavior and Decision Processes, Vol. 43, pp. 75–117.

Gigerenzer, G. (1996): On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996), in: Psychological Review, Vol. 103, pp. 592–596.

Gigerenzer, G. (2004a): Fast and frugal heuristics: The tools of bounded rationality, in: Koehler, D.J./Harvey, N. (Eds.): Blackwell handbook of judgment and decision making, Oxford, UK, pp. 62–88.

Gigerenzer, G. (2004b): Striking a blow for sanity in theories of rationality, in: Augier, M./March, J.G. (Eds.): Models of a man: Essays in honor of Herbert A. Simon, Cambridge, MA, pp. 389–409.

Gigerenzer, G. (2006): Bounded and rational, in: Stainton, R.J. (Ed.): Contemporary debates in cognitive science, Oxford, UK, pp. 115–133.

Gigerenzer, G. (in press): Gut feelings: The intelligence of the unconscious, New York.

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

Gigerenzer, G./Selten, R. (Eds.) (2001): Bounded rationality: The adaptive toolbox, Cambridge, MA.

Gigerenzer, G./Todd, P.M./The ABC Research Group (1999): Simple heuristics that make us smart, New York.

Goldstein, D.G./Gigerenzer, G. (1999): The recognition heuristic: How ignorance makes us smart, in: Gigerenzer, G./Todd, P.M. /The ABC Research Group (Eds.): Simple heuristics that make us smart, New York, pp. 37–58.

Goldstein, D.G./Gigerenzer, G. (2002): Models of ecological rationality: The recognition heuristic, in: Psychological Review, Vol. 109, pp. 75–90.

Goldstein, D.G./Gigerenzer, G./Hogarth, R.M./Kacelnik, A./ Kareev, Y./Klein, G./Martignon, L./Payne, J.W./Schlag, K. (2001): Why and when do simple heuristics work? in: Gigerenzer, G./Selten, R. (Eds.): Bounded rationality: The adaptive toolbox, Cambridge, MA: MIT Press, pp. 173–190.

Hertwig, R./Barron, G./Weber, E.U./Erev, I. (2004): Decision from experience and the effect of rare events, in: Psychological Science, Vol. 15, pp. 534–539.

Hertwig, R./Todd, P.M. (2003): More is not always better: The benefits of cognitive limits, in: Hardman, D./Macchi, L. (Eds.): The psychology of reasoning and decision making: A handbook, Chichester, UK, pp. 213–231.

Hogarth, R.M./Karelaia, N. (2005): Simple models for multi-attribute choice with many alternatives: When it does and does not pay to face tradeoffs with binary attributes, in: Management Science, Vol. 51, pp. 1860–1872.

Hogarth, R.M./Karelaia, N. (2006): "Take-the-best" and other simple strategies: Why and when they work "well" with binary cues, in: Theory and Decision, Vol. 61, pp. 205–249.

Hoyer, W.D./Brown, S.P. (1990): Effects of brand awareness on choice for a common, repeat-purchase product, in: Journal of Consumer Research, Vol. 17, pp. 141–148.

Kahneman, D./Slovic, P./Tversky, A. (Eds.) (1982): Judgment under uncertainty: Heuristics and biases, Cambridge, UK.

Katsikopoulos, K./Martignon, L. (2006): Naive heuristics for paired comparisons: Some results on their relative accuracy, in: Journal of Mathematical Psychology, Vol. 50, pp. 488–494.

Keeney, R.L./Raiffa, H. (1993): Decisions with multiple objectives, Cambridge, UK.

Korobkin, R. (2003): Bounded rationality, standard form contracts, and unconscionability, in: University of Chicago Law Review, Vol. 70, pp. 1203–1295.

McGraw, A.P./Tetlock, P.E./Kristel, O.V. (2003): The limits of fungibility: Relational schemata and the value of things, in: Journal of Consumer Research, Vol. 30, pp. 219–229.

Martignon, L./Hoffrage, U. (1999): Why does one-reason decision making work? A case study in ecological rationality, in: Gigerenzer, G./Todd, P.M./The ABC Research Group (Eds.): Simple heuristics that make us smart, New York, pp. 119–140.

Martignon, L./Hoffrage, U. (2002): Fast, frugal and fit: Lexicographic heuristics for paired comparison, in: Theory and Decision, Vol. 52, pp. 29–71.

Payne, J.W./Bettman, J.R./Johnson, E.J. (1993): The adaptive decision maker, Cambridge, UK.

Reddy, R. (1988): Foundations and grand challenges of Artificial Intelligence: AAAI Presidential Address, in: AI Magazine, Vol. 9, pp. 9–21.

Reyna, V.F./Farley, F. (2006): Risk and rationality in adolescent decision making: Implications for theory, practice, and public policy, in: Psychological Science in the Public Interest, Vol. 7, pp.1–44.

Roberts, S./Pashler, H. (2000): How persuasive is a good fit? A comment on theory testing, in: Psychological Review, Vol. 107, pp. 358–367.

Schmitt, M./Martignon, L. (2006): On the complexity of learning lexicographic strategies, in: Journal of Machine Learning Research, Vol. 7, p. 55–83.

Schooler, L.J./Hertwig, R. (2005): How forgetting aids heuristic inference, in: Psychological Review, Vol. 112, pp. 610–628.

Serwe, S./Frings, C. (2006): Who will win Wimbledon? The recognition heuristic in predicting sports events, in: Journal of Behavioral Decision Making, Vol. 19, pp. 321–322.

Sunstein, C.R. (Ed.) (2000): Behavioral law and economics, Cambridge, UK.

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

Yee, M./Hauser, J./Orlin, J./Dahan, E. (in press): Greedoid-based non-compensatory two-stage consideration-then-choice inference, in: Marketing Science.