CHAPTER 1

Setting the Stage: Elements of Decision Heuristics

Introduction

When people are asked which city has more inhabitants, and they recognize one city but not the other, they usually choose the one they recognize. This simple strategy is called the recognition heuristic. German students given the alternatives San Diego and San Antonio, for instance, almost exclusively picked the correct alternative, San Diego; this was a city most of them recognized, whereas many had never heard of San Antonio (Goldstein & Gigerenzer, 2002). Explaining this behavior by focusing entirely on internal processes by, for example, postulating that San Diego was the more available option in participants' memories, does not add much beyond the fact, already known, that one city is recognized while the other is not. But *why* should one trust one's recognition memory? Goldstein and Gigerenzer analyzed the environment to look for an explanation. They showed that recognition correlates with higher values on criteria such as size, amount, or performance in many domains. The mediator they suggested are the mass media: Large cities are more often mentioned in the news, for example. Thus, recognition can be a valid cue for city size. But what will people do in situations in which recognition is not a valid cue for a certain criterion? When asked to select the larger city of a pair composed of a small neighboring city and a fictional city name, people do *not* preferentially select the recognized city, and often even do the opposite (Oppenheimer, 2003). Thus, people use recognition adaptively – as they seem to know under which circumstances recognition is valid, and under which circumstances it is not (such as when a city is known due to close geographical proximity). This example shows that analysis of the decision environment can produce powerful predictions about when a behavior will be observed and when not, under the general assumption that human cognition is adaptive.

Although ecological approaches have a long tradition in psychology (e.g., Brunswik, 1955), studying the environment in which certain cognitive tasks have to be performed is far from being found in the standard repertoire of cognitive psychologists. Maybe the spread of evolutionary thinking in psychology will increase the attention paid to the environment, but so far, cognitive theories that take the environment explicitly into account are more an exception than the rule. The existing approaches differ in their assumptions of how the environment is reflected in the human mind.

Anderson (1990), for example, views human memory as adapted to the information-retrieval requirements posed by the environment. In an impressive analysis of databases (*New York Times* headlines, word usage in speech to children, authors of e-mails), Anderson and Schooler (1991) found, for instance, that when one plots the probability that a particular item (say, a word) will occur as a function of how long it has been since it last occurred, the curve shows a strong probability increase for the most recent items, suggesting that the recency "bias" found in learning might be an adaptive response to the statistical structure of the environment. Recency of last occurrence is simply a very good predictor that an item will be encountered again at a given point in time. However, Anderson's (1990) view of how search in memory proceeds imposes the burden of optimization on the individual – search is assumed to occur in an optimal order and to stop at the optimal point. Simple heuristics that might be able to exploit the identified environmental regularities are not considered.

Shepard (2001), mainly in the domain of perception, tried to find imprints of universal principles, such as physical laws, on our brains. For example, he suggests that the three-dimensional color representation of human vision might be an adaptation to the three degrees of freedom of natural lightning on our planet: light-versus-dark variation, red-versus-green variation, and blue-versus-yellow variation. Yet areas in psychology in which deterministic laws can be applied might be rather limited, or at least they might only apply to certain processing levels. Probably not by chance, Shepard focused on color constancy and motion of objects.

Often, inferences have to be made in domains whose laws, if existent, are unknown to decision makers and which thus contain, at least subjectively, a high degree of uncertainty. Brunswik (1955) appreciated this uncertainty in how the world reveals itself to us. According to him, "the universal lawfulness of the world is of limited comfort to the perceiver or behaver not in a position to apply these laws, and he therefore must rely largely on whatever snitches of particular or semigeneralized information he may be able to assemble" (p. 209). Decision makers therefore have to rely on uncertain cues to infer a criterion. But, in contrast to the view that will be put forth in this dissertation, research in the Brunswikian tradition, exemplarily represented by multiple cue probability learning (Smedslund, 1955; for a short review, see Holzworth, 2001), has still honored the ideal of complete information representation. In these studies, people usually had to estimate the criterion value of an item based on several cues. Ideally, all available cues should be utilized, and subjective validities should accurately match ecological validities. There is no guidance on how the decision maker should search for information nor on when to stop. Not surprisingly, the prevailing benchmark model in this research area is multiple regression. From using multiple regression to also depict participants' decision behavior, it necessarily follows that simpler but potentially smart heuristics will be obscured and might be expressed in the form of inadequate beta-weights in the regression equation. Thus, whereas the uncertain relationship between cues and criterion was accepted as given, divergence on the side of the decision maker from the objectively computed ecological validities ideally should be diminished. Even Hammond (1996), one of the initiators of the field, more recently admitted that the role of multiple regression as a model for organizing information from multiple cues has been overemphasized.

A complete representation might not only be beyond the grasp of the decision maker, it might also be unnecessary. Despite bounded computational capabilities, humans possess mechanisms to cope with the complexities of life (Simon, 1990). This view allows for a more optimistic conception of our computationally bounded brains. Although the cognitive capabilities of the decision maker represent a limiting factor, there can be benefits to gain from reduction, such as speed and robustness (Hertwig & Todd, 2003). Certain regularities in the environment can be relied on without the necessity of complete representation. Cognitive heuristics are tools that exploit such environmental regularities. This chapter will present a collection of examples of this interplay between heuristics and the environment in decision making. Specifically, this chapter deals with situations in which one has to decide between two alternatives, such as which of two company shares will achieve higher returns, based on a number of cues (reasons), such as number of employees or sales figures.

Heuristics

Heuristics are conscious or unconscious cognitive strategies people rely on to solve a problem. Heuristics should be distinguished from optimization. Optimization means that one can find the best strategy for a problem, and that one can prove that there is no better one. Otherwise one cannot be sure that the strategy is the optimal one. However, most important problems are out of reach of optimization methods, because optimization is too slow, too expensive, or computationally intractable. The latter means that no mind or machine can compute the optimal strategy. Games with well-defined rules such as chess and Go are computationally intractable, and less-well structured real-world problems such as what shares to buy are out of reach of optimization, a fortiori. Thus, heuristics are a useful means to cut short an otherwise cumbersome and often interminable process.

In this dissertation, the definition of what constitutes a heuristic is based on Simon's (1955, 1956) work on models of bounded rationality and on more recent empirical studies, computer simulations, and mathematical analyses of models of heuristics (e.g., Dawes, 1979; Gigerenzer & Selten, 2001; Gigerenzer et al., 1999; Payne, Bettman, & Johnson, 1988, 1993). Following this view, a heuristic is a sequence of steps involved in a cognitive process like decision making, such as what and how much information to look for, and what to do with the pieces of information found. These steps are described in the form of explicit rules, each of which can be empirically tested. The description of the underlying cognitive processes distinguishes heuristics from as-if optimization and the so-called heuristics and biases program (e.g., Tversky & Kahneman, 1974).

As-if optimization: Unlike models of heuristics, which are process models, as-if models are not intended to describe the cognitive processes. They are outcome models, that is, their task is to predict the result of a behavior, not the behavioral process that brought about the result. Outcome and process models differ in what counts as counter-evidence for the model. A process model makes predictions about both the process of problem solving and its outcome and can be proven wrong by both kinds of data. A concrete example is provided in the next section. An as-if model, in contrast, can only be proven wrong by outcome, not by process data. As-if models often assume that outcomes can be predicted by optimization. Today, most economic theories are deliberate as-if models, assuming that people act as if they were fully rational. It is not assumed that people actually go through the mathematical process of optimization, only that their decision outcomes can be modeled as if they had done this. Similarly, psychological theories based on optimization such as Anderson's (1990, 1991) rational analysis of memory and reasoning and Nosofsky's (1991) exemplar models of classification are deliberate as-if models. Milton Friedman's (1953) defense of as-if optimization models is exemplary: Economic models are not intended to portray the process of decision making and cannot be criticized by empirically demonstrating that their assumptions are descriptively invalid. According to Friedman, even, to be important, "a hypothesis must be descriptively false in its assumptions" (p. 14). In contrast to this view, models of heuristics can be proven wrong even if they predict the actual inferences and choices of people correctly, but not their process of solving these decision problems.

Heuristics and biases: Models of heuristics also need to be distinguished from the collection of heuristics and biases that have been gathered by Daniel Kahneman, Amos Tversky, and collaborators since the early 1970s (e.g., Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1973, 1974). Unlike what is often stated (e.g., Bendor, 2003; Gilovich & Griffin, 2002), it is not just the optimistic view of human rationality that distinguishes ecologically rational heuristics from the heuristics and biases program. Researchers in this program contented themselves with labeling certain phenomena discovered in judgment and decision-making experiments without specifying the underlying process. Deviations from a norm the experimenter pre-defined as rational, such as Bayes' theorem, are considered as biases. The process that supposedly brought them about is, post hoc, merely labeled as a certain heuristic and not described. For example, the tendency of people to overestimate the degree to which others agree with them has been referred to as false consensus effect (Ross, Greene, & House, 1977), at least partially brought about through use of the availability heuristic (Ross & Anderson, 1982). The underlying, almost circular assumption is that one's own opinion is more available to oneself (for an alternative explanation of the so-called false consensus effect that does not consider it a bias but rather as consistent with application of Bayes' theorem, at least under certain circumstances, see Dawes & Mulford, 1996). The problem with vague labels such as availability or representativeness is that they can post hoc be used to explain almost everything because the underlying process is not explicitly specified, leading to vague predictions as well (e.g., Gigerenzer, 1996; Shanteau, 1989; Wallsten, 1983). The following example illustrates the negative consequences of this vagueness. The label "representativeness" has been invoked to account for both the gambler's fallacy (Tversky & Kahneman, 1974, p. 1125) and the hot-hand fallacy (Gilovich, Vallone, & Tversky, 1985, p. 295). Note that these two phenomena are exactly opposite. In the hot-hand fallacy, the intuition is that after a series of *n* equal outcomes the same outcome will occur again; in the gambler's fallacy the intuition is that after a series of *n* equal outcomes the opposite outcome will occur. Nevertheless, the notion of the representativeness heuristic is flexible enough to account for both logical possibilities (Ayton & Fischer, in press).

Demonstrating a phenomenon can inspire and guide future research, but it must not be the point where research is stopped. The goal should be rather to explain the phenomenon, to uncover the fundamental processes that bring about the outcome. Again following Simon (2001), much of scientific activity can be summarized by the paradigm: "Given the description of some natural phenomena, to find the differential equations for *processes* that will produce the phenomena" (p. 211, emphasis added). This chapter will follow a similar guideline and, given the description of the decision environment, will find decision processes that will achieve good performance without asking too much of the decision maker.

Aim

The overall goal of this chapter is to provide a framework for how we can think about the relationship between the environment and human cognitive strategies that exploit it, and for how we can study this relationship empirically.

There exists today an impressive number of models of heuristics, including satisficing (Simon, 1955), Tit for Tat (Axelrod & Hamilton, 1981), Dawes' rule (Dawes, 1979), Take The Best (Gigerenzer & Goldstein, 1996, 1999), good features (Alba & Marmorstein, 1987), Weighted Pros (Huber, 1979), the recognition heuristic (Goldstein & Gigerenzer, 1999, 2002), elimination by aspects (Tversky, 1972), categorization by elimination (Berretty, Todd, & Martignon, 1999), and QuickEst (Hertwig, Hoffrage, & Martignon, 1999). Yet there is a small body of work that analyzes the match between heuristics and the structure of environments (including experimental tasks), that is, that specifies which environmental structures can be exploited by a given heuristic. For instance, Tit for Tat is an excellent heuristic in an environment with indefinitely repeated interaction, and even more so when the possibility of preferential interaction with other Tit for Tat players exists (Axelrod & Hamilton, 1981). It is not a good heuristic for one-shot games, however. In the right environment, heuristics can be highly successful. The match between heuristic and environment is analyzed in Payne et al. (1993), Gigerenzer & Selten (2001), and Gigerenzer et al. (1999), but the analyses have not been conducted at the level of the heuristics' elements, or building blocks, but rather of complete heuristics. In line with Huber's (2000) claim for investigating partial instead of global heuristics, which are small enough to be combined in a flexible manner, this chapter narrows the focus to concentrate on the building blocks that make up the heuristics, rather than on the complete heuristics themselves. The advantage of doing so is that there are a large number of heuristics, but they are composed of relatively few building blocks. The first aim of this chapter is therefore to address the question: Which building blocks are ecologically rational in which environmental structures?

This question is addressed through theoretical analysis of which building blocks fit which structures of the decision environment.[1] From this analysis, empirically testable predictions are derived, which will be phrased in the form of concrete hypotheses. In this way, the ecological analysis will also stand against the criticism that simple heuristics are unfalsifiable because it is possible to post hoc formulate alternative building blocks (Newell, Weston & Shanks, 2003). It will be shown that concrete hypotheses can be derived a priori from such an ecological analysis and can then be tested in a very stringent way. Interestingly, such a careful ecological analysis of the fit between strategies and the experimental environment is missing in many experimental studies, among them those of Newell et al. (2003) themselves.

The second aim of this chapter is to connect the available experimental evidence with the first question: Do people exchange building blocks within a heuristic as a reaction to different or changing environmental structures? A few experimental studies have analyzed how people adapt their heuristics to the structure of environments (e.g, Bröder, 2000; Newell & Shanks, 2003; Rieskamp & Hoffrage, 1999; Rieskamp & Otto, 2004). Again, most of the work has focused on the heuristic as the unit, and not on the building blocks. This evidence is nevertheless reviewed, together with arguments for why a certain finding might support a particular hypothesis about the fit between environments and building blocks. In this way, evidence is provided only for some of the hypotheses put forward in this chapter, while others remain to be tested in future experiments.

The scope is restricted to problems that involve paired comparison, such as which of two companies' shares to buy. The focus is on inferences for which an outside criterion

[1] Under the term environment, I subsume what other authors have called *task variables* and *context variables* (Payne et al., 1993), or *characteristics of the decision problem* and *characteristics of the decision environment* (Beach & Mitchell, 1978). One reason for this is simply that I find the distinctions between these terms blurry and in some cases arbitrary. For example, according to Payne et al. (1993), correlations between cues belong to the class of context variables while number and format of cues are examples for task variables. Even though a more plausible distinction might be found, such as between characteristics of the options to choose from together with the cues to rely on on the one hand (e.g., correlation between cues), and on the other hand characteristics of the situation in which a decision is made (e.g., time pressure), certain variables cannot be unambiguously classified. For instance, is mode of presentation of cues a characteristic of the cues or of the situation? It clearly can be seen as a cue characteristic, but at the same time it can vary from situation to situation: The same information can be presented visually in form of a written list, or aurally by reading it aloud to a person. In short, environmental characteristics vary on various dimensions; dichotomy cannot always be imposed. I therefore treat the term decision environment broadly, and include characteristics of the options, the cues, as well as the decision situation.

exists to evaluate the correctness of the decision, as opposed to preferential choices. Nevertheless, whenever the ecological analysis put forward here can be connected to research in the preferential choice literature, the respective findings will be discussed as well. The analysis will be restricted to three classes of building blocks, rules for search, stopping, and making a decision – stopping and decision rules will be treated in one chapter due to their strong interrelation. The general paired comparison task is to predict which alternative, A or B, has the higher value on a criterion. The alternatives A and B are elements of a set of *N* alternatives (which can be actions, objects, events), and the prediction can be based on a set of *M* cues. In the case of binary cues, cue values "1" and "0" always stand for higher and lower criterion values, respectively (even though other labels might have been used in a particular study, such as "yes" and "no", or "+" and "–").

Building blocks of heuristics: An illustration

To illustrate the building blocks of heuristics and the way one can test process models, a study by Newell et al. (2003) serves as a paradigmatic case. The participants were presented with a series of choices between the shares of two fictional companies. In each trial, two companies were presented on a computer screen, and the participants were asked to infer which company's share would turn out to be more profitable (like in Figure 1.1).

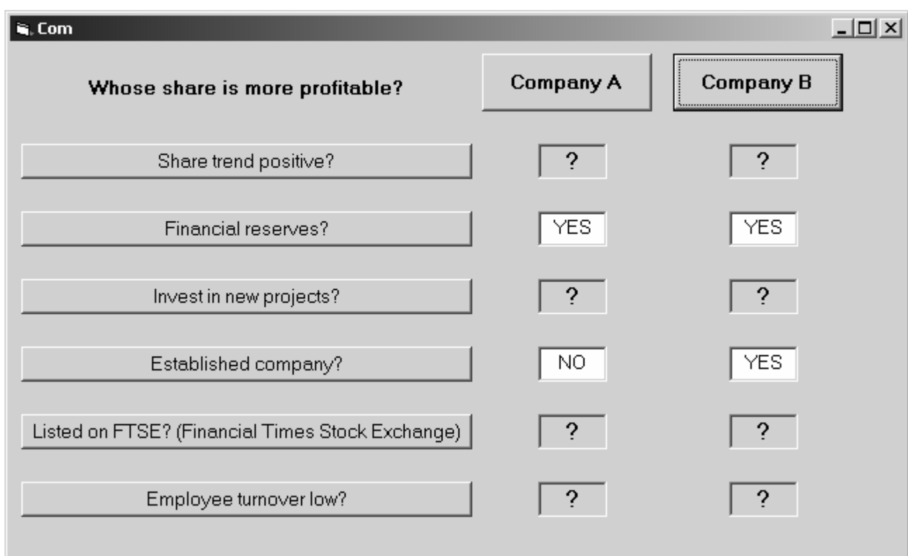

Figure 1.1: Participants in Newell et al.'s (2003) study could choose to buy up to six pieces of information in any order, as depicted above. The task was to infer which company's share is more profitable. The illustration shows a person who searched first for information concerning financial reserves but found no difference between the two companies, then continued the search and looked up whether the companies are established companies, found a difference, stopped search, and made the inference that Company B's share is more profitable.

To help find the more profitable shares, participants could acquire information concerning six cues, such as: "Does the company invest in new projects?" and "Does the company have financial reserves?" The information was given in terms of yes/no answers. In each trial, a participant could buy information about as many cues as she wanted before picking a share. The information was retrievable on a computer screen by the click of a mouse, similar to the information acquisition in the experimental program "mouse-lab" (Payne et al., 1993). The cost of information about each cue was 1p (pence). After the participant had bought as many cues as she wanted to buy, she made her choice, and feedback was given whether the answer was correct. When the answer was correct, the participant got 7 p minus the amount she had spent searching for information.

Newell et al. (2003) tested to what degree the Take The Best heuristic (Gigerenzer & Goldstein, 1996, 1999) can predict the problem solving process. Take The Best (TTB) is a process model with the following three building blocks for searching, stopping, and deciding:

(1) *Search rule:* Chose the cue with the highest validity.[2] Look up the cue values of the two objects.

(2) *Stopping rule:* If one object has a positive cue value ("1") and the other does not ("0" or unknown), then stop search and go on to Step 3. Otherwise exclude this cue and go back to Step 1. If no cues are left, guess.

(3) *Decision rule:* Predict that the object with the positive cue value ("1") has the higher criterion value.

Each of the three building blocks describes a process step, and each of them can be tested independently.

Search rule: In theory, participants can search through cues in many different ways. If they look up all six cues, there are 6! = 720 different orders. The search rule in Take The Best postulates that people will search by one of these orders, the one defined by validity v_i. To learn the validities, each participant was presented with 120 tasks like that in Figure 1.1, and outcome feedback (correct/incorrect) was given after each response. Feedback was given so that the six cues had validities v_i of .90, .85, .80, .75, .70, and .65. Which cue had which validity was counterbalanced between participants. Additionally, after 60 and 120 trials, a hint was provided about the order of cues by validity. The learning phase was followed by a test phase with 60 tasks. Figure 1.2 shows that 75% of the participants followed the search rule of Take The Best in the test phase. When in a second experiment there were only two cues, this number was 92%.

Stopping rule: The stopping rule of TTB postulates that search is stopped immediately after the first discriminating cue is found, not before and not later. Note that the stopping rule can be valid even when the search rule is violated, for instance, when people search in one of

[2] The validity v_i of a cue i is defined as $v_i = R_i / D_i$, where R_i is the number of correct predictions by cue i, and D_i is the number of pairs where the cue values of cue i differ between objects.

the 719 orders not consistent with v_i but stop after the first discriminating cue is found. Similarly, the stopping rule can be violated although the search rule is followed, for instance, when a person searches in the order of v_i but continues search after the first discriminating cue is found. Thus, the empirical result on the search rule does not constrain the stopping behavior; in principle, all of the 75% cases consistent with the search model can violate the stopping rule. Newell et al. (2003) reported that in 80% of all cases (where participants bought any information at all), participants did not buy unnecessary information. This means that in 80% of cases search did not continue beyond a single discriminating cue, and the corresponding number for the same task with two cues was 89%.

Decision rule: In theory, participants can use infinite ways to combine the information concerning six cues. This includes linear models, weighted or unweighted. If a person follows the stopping rule of Take The Best, this constrains the ways to arrive at a decision (whereas the search rules impose no constraints on the stopping and decision rules). If only one piece of discriminating information is obtained, it seems that the only reasonable decision rules left are forms of one-reason decision making, such as to decide for the alternative to which the cue points with a probability that matches the cue validity (see, e.g., Luce's choice rule, 1959). The decision rule of Take The Best is simpler, as it does not rely on probability matching: it simply goes with the alternative to which the cue points. Newell et al. (2003) report that the decision rule was followed by their participants in 89% of trials, for both six and two cues.

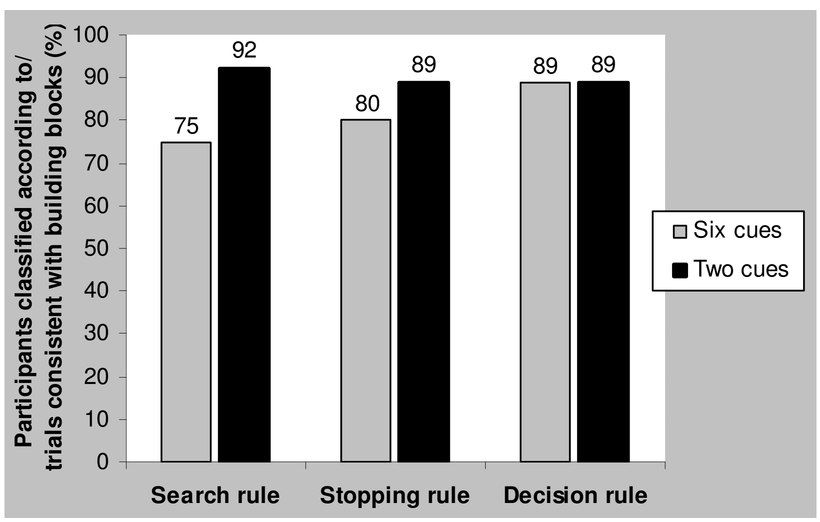

Figure 1.2: Percentage of cases in which participants' behavior was consistent with the three building blocks of TTB. The search rule column depicts percentage of *participants* classified as using TTB's search rule, the stopping and decision rule columns depict percentage of *trials* consistent with the particular building block. One experiment used six cues (as in Figure 1.1); a second one used two cues (Newell et al., 2003).

The building blocks tested in the experiment reported above may well be good models of people's behavior in the share task, but it would be inconsistent with an ecological approach to cognition to assume that people would always search by validity, always stop after having found exactly one discriminating cue, and always rely on just this one piece of information when making a decision. The remainder of this chapter will therefore be devoted to specifying the conditions under which certain building blocks are expected to be used, and, if available, reviewing empirical studies that have investigated decision making under those conditions. As the share task example has shown, there exists a strong connection between stopping and decision rules. I will therefore discuss stopping and decision rules together in one section, and work out their connection in more detail there.

Search rules

There are two visions of how the mind searches sequentially for cues: optimal search rules and heuristic search rules. Several psychological theories postulate versions of optimal search rules (e. g., Anderson, 1990, for memory retrieval; Nosofsky & Palmeri, 1997, for search for exemplars in classifications), primarily for sequential search through alternatives. For instance, Anderson (1990) assumes that memory contents are searched in an order determined by their probability of being relevant at the current moment. Like this example, most of the optimal search models make predictions about internal search only, which cannot be observed directly, so the assumption of optimized search becomes difficult to falsify.

However, the problem of finding an optimal order of cues becomes computationally intractable for large numbers of cues. For instance, for the $M = 6$ cues in the share task example, there are $M! = 720$ orderings to evaluate; if the number of cues is doubled, the number of orderings increases to 479,001,600, and if doubled once more the number of orderings is already larger than 10^{23}. The problem of finding the optimal order of cues fulfills the criterion for NP-completeness because $M! \geq 2^{M-1}$. NP-completeness denotes the complexity of an algorithm for which this condition holds, meaning that the algorithm is not to be completed in polynomial time and therefore, in practice, becomes intractable for large M (Martignon & Hoffrage, 1999). Thus, optimizing is not a likely model for how humans order cues, except for situations with a very small number of cues.

The second vision employs heuristic rules for search. Ordering cues by validities is such a method. It does not generally result in the optimal order, which can be defined as the one that gives the highest accuracy for a set of N alternatives. One reason is that v_i ignores dependencies between cues. Computer simulations with nine cues (for inferring which of two cities has a larger population) showed that despite its simplicity, v_i led to an order that was better than 98% of all possible orders (Martignon & Hoffrage, 1999). Furthermore, heuristic search rules tend to be more robust than the optimal order when information is scarce and uncertain. This result can be obtained when one tries to predict new data, rather than fit data already obtained. Martignon and Hoffrage (1999, p. 134) split the set of cities into two halves and calculated for one half the optimal order as well as the order by v_i. As mentioned above, search by validity resulted in inferences that were better than most orders, but the optimal order did better, by definition, in data fitting. These two orders were then tested in the second half of the cities, which is known as cross-validation. The result was that search by validity now led to higher accuracy than the optimal order: it was more robust. The point is that the optimal order in one sample is not necessarily the optimal order in a new sample, and in this precise sense, a robust heuristic search rule can be said to do "better than optimal" in prediction, although never in fitting.

Three general forms of search rules will be distinguished: *random search*, *search in a fixed order* (as in TTB) – whereby different ordering criteria, to be introduced below, might be used – and *recency search* – whereby the search order is continuously updated based on which cue most recently led to a (correct) decision.[3]

The general postulate put forward is that heuristic rules for search depend on the structure of the environment, more precisely the structure as it is known to a person. The term "environment" is used as shorthand for the structure of both natural environments, such as the actual situation of a person faced with the problem of where to invest spare money, and experimental tasks, such as the situation of a participant in Newell et al.'s (2003) study given paired comparisons between hypothetical company shares that vary on six cues.

In the following, different types of environments are examined: *Unfamiliar environments*, *stable versus changing environments*, and *environments with outcome feedback versus without outcome feedback*. Another environmental variable to be explored in the context of stable environments with feedback is *costs for acquiring information*. An environment is stable from the point of view of a problem solver when its characteristics (the properties of the cues and the alternatives, such as the cue validities) do not change during the time of problem solving. The experimental share task is an example of a stable environment. Outcome feedback refers to whether the problem solver receives information concerning the correctness of her decisions. The share task provides both stability and feedback. Many real environments are located in between these two pairs of poles, and graded distinctions can be made. For the purpose of exposition, however, the two dimensions are treated as binary. The summary of the theory about how search rules depend

[3] I will not talk about a basic distinction between two kinds of search principles often made in the preferential choice literature, that is, the distinction between attribute-based and alternative-based search (e.g., Bettman, Johnson, Luce, and Payne, 1993; Payne et al., 1988; Russo & Dosher, 1983). In the inference tasks reported here, search usually proceeds from cue to cue, that is, participants can choose one cue after the other and are informed about the values *both* alternatives have on that cue. But one could also imagine that cue values are requested separately for alternatives. This would leave open the possibility that first all the cue values of one alternative are checked before examining the second alternative. Attribute-based, or cue-based search, proceeding within one cue dimension across alternatives, is often called dimensional, whereas alternative-based search is often called holistic. Attribute-based search is the equivalent to the cue-to-cue search pattern I exclusively look at here when talking about search rules for inferences.

on environmental structure is shown in Table 1.1 and will be explained in the following sections. First, search for information in an environment about which a person is ignorant will be considered.

Table 1.1: Search rules dependent on environmental structures.

Random search	Ordered search	Recency search
Unfamiliar environments	Stable environments; ordering criterion will depend on: - availability of outcome feedback - information costs	Changing environments; criterion whose recency will be tracked depends on: - availability of outcome feedback

Unfamiliar environments

If a person is faced with a problem and does not know which cues are more relevant than others, she has to deal with an "unfamiliar environment." What the person still might know are the cue directions. For instance, in the first trials of the share experiment (Figure 1.1), participants only knew that "share trend positive" is good and "financial reserves" are good but not which of the two is more important. In this state of ignorance, the search rule is:

> <u>Random search:</u> Choose a cue randomly and look up the cue values of the two objects.

The term "random" is used in a broad sense, including quasi-random, unsystematic choices of cues, not necessarily the use of a random device. When knowledge about cues is subsequently acquired, random search is likely to be replaced by one of the search rules in the following sections.

The Minimalist heuristic employs random search (otherwise the heuristic proceeds like Take The Best). Studies using 20 different prediction tasks, such as predicting which of two female celebrities people find more attractive, showed that despite its simplicity, accuracy was substantially above chance, approaching that of a multiple regression model by 8 percentage points in data fitting and 3 percentage points in prediction (Czerlinski et al., 1999).

Random search is also part of heuristics to solve problems beyond paired comparisons. For instance, satisficing searches through alternatives in a quasi-random fashion; that is, the alternatives are somehow encountered, but the satisficer is not assumed to impose an order on the sequence in which she encounters the alternatives (Simon, 1955). Sequential mate search models often implicitly assume the same quasi-random order in which potential spouses are encountered (Miller & Todd, 1998). In these tasks, what is unknown are not the cues but the alternatives, whereas in paired comparison tasks the alternatives are given.

Stable environments without outcome feedback

An environment in which characteristics of cues (such as validity and discrimination rate, see below) and alternatives do not change much is a stable environment. For instance, the experimental environment in the share experiment is an artificially stable environment, because the experimenters did not vary the characteristics during the 180 trials. In the share task, however, participants received feedback after each decision; this section is about situations without outcome feedback. These include environments in which feedback comes late, is too expensive to obtain, or is ambiguous. For instance, physicians often are not sure whether they have looked at the right cues to decide which of two treatments would be best for a patient, because patients may not return when the treatment is effective, or it may take years to determine that the treatment was effective (Gigerenzer, 2002). How would search proceed in the absence of learning through feedback?

In the absence of feedback, cues can be ordered according to how often they discriminate between the alternatives. For instance, if almost all companies have a positive share trend, this cue does not discriminate often and is in this sense not very informative. For binary cues, cue values of [1;0] and [0;1] discriminate between alternatives, whereas cue values of [0;0] and [1;1] do not. The discrimination rate d_i of a cue i is

$$d_i = \frac{D_i}{P}$$

where D_i is the number of pairs where the cue values of cue i differ between objects, and P is the total number of pairs. Search by d_i amounts to the following search rule:

> Search by discrimination: Choose the cue with the highest discrimination rate. Look up the cue values of the two objects.

For a paired comparison task with binary cues and a class of N alternatives (such as companies), the discrimination rate can also be computed from the set of alternatives without performing all possible paired comparisons:

$$d_i = \frac{2x_i y_i}{1 - \frac{1}{N}}$$

where x_i and y_i are the relative frequencies of the cue values.

From this formula it is easy to see that for large N, the discrimination rate is approximately $2x_iy_i$, and that the discrimination rate of a cue is highest when its cue values (0 and 1) occur equally often. Thus, search by discrimination can be executed without actually calculating d_i. A simpler method for learning and estimating the order of discrimination rates is to observe whether the cue values match, that is, are equally frequent. Let X_i and Y_i denote the absolute frequencies of the two cue values (0 and 1) of cue i in a given environment, with $N = X_i + Y_i$. Then, the values of cue i match if $|X_i - Y_i| = 0$, or match best if $|X_i - Y_i| = \min$. This allows for a simpler implementation:

> <u>Search by discrimination':</u> Choose the cue with the best matched cue values. Look up the cue values of the two objects.

Search by best matched cue values, that is, by $|X_i - Y_i| = \min$, leads to the same order as d_i.[4] Thus, simple tallies of the frequency of occurrence of the two cue values can replace the computation of d_i. Now, the first hypothesis can be formulated:

> *Hypothesis A (no outcome feedback):* In stable environments without outcome feedback, random search in the initial trials will turn into search by discrimination in subsequent trials.

So far, no experimental tests of Hypothesis A are available, as all studies on the use of simple heuristics for paired comparisons provided outcome feedback or informed participants directly about cue validities. A broad class of environments without feedback about the correctness of one's decision is studied in the literature on preferences, as opposed to inferences, where, by definition, an objectively correct answer does not exist. This literature, however, has only rarely studied how attributes of the choice alternatives are searched for. Mainly it has taken the order a participant used and attributed it to some unspecified subjective preference, boiling down to the finding that search order follows subjective attribute weights (e.g., Aschenbrenner, Albert, & Schmalhofer, 1984; Saad, 1999). There are some studies demonstrating a positive effect of the *range* of cue values, which increases discriminability when dealing with continuous cue values, on the subjective weight attached to that cue (Beattie & Baron, 1991, Experiment 6; Fischer, 1995; Meyer & Eagle, 1982). This, however, can at best be seen as indirect evidence for the influence of cue range and variance on search order, as search was not directly measured in these studies but other weight-elicitation methods were used.

---

[4] The proof is as follows. The order of d_i is the same as that of x_iy_i, because they relate to one another by a linear function $2/(1-1/N)$. The order of x_iy_i is the same as that of X_iY_i , because they are identical except for a constant N^2. Finally, the order of X_iY_i is identical to that of $\min|X_i - Y_i|$, because both functions peak at $X_i = Y_i$ and fall monotonically and symmetrically at both sides.

To summarize, in environments with no or insufficient feedback, search can nevertheless do better than chance by going through cues sequentially according to their potential to discriminate. The benefit of search by discrimination is that it is more likely to encounter an informative cue early, and that, as a consequence, search can be stopped, which in turn allows one to make faster and more frugal decisions.

Stable environments with outcome feedback

In situations with feedback, search can be guided by the predictive power of a cue, not only by its discriminating power. In a stable environment, the predictive power of a cue is constant and can be defined by its validity v_i:

$$v_i = \frac{R_i}{D_i}$$

where R_i is the number of correct predictions by cue *i*, and D_i is the number of pairs where the cue values of cue *i* differ between objects.

Cue validity tells us, for all cases where the cue discriminates, the probability that the decision will be correct. If $v_i = .50$, the predictive power of cue *i* is at chance level; values greater than .50 measure the proportion of correct predictions a cue produces above chance.

Learning to order cues

Stable environments with feedback allow for three types of learning: evolutionary, social, and individual. Evolutionary learning is slowest. For instance, a female guppy comes already equipped with a search rule for mates (Dugatkin, 2000). When she has to decide between two potential mates, the most important cue seems to be the extent of orange color. If one male has considerable more than the other, this cue is sufficient to stop search and decide in favor of him. Otherwise, the second cue seems to be mate copying, that is, if she has seen one of the two potential candidates mating earlier with another female, she favors him. A female guppy does not have to learn individually what cues to look for, and in which order. In the case of evolutionary learning, the assumption is that a cue (such as orange coloring) predicts the fitness of the potential mate, and thereby that of the offspring, at least in past environments. In humans, social learning is generally the fastest way to learn cue orders. For instance, medical students are instructed about diagnostic cues for certain diseases by their teachers; they do not generally learn this through individual experience. Both evolutionary and social learning can provide orderings of cues, and the two examples from animal biology and medical diagnosis illustrate that it is not always easy in the real world to determine the degree to which an individual's search order corresponds to the actual validities, or other criteria.

Individual learning is generally slower than social learning, but faster than evolutionary learning. In environments where an individual has to learn the predictive power of cues from feedback, she does not know the v_i values but has to estimate these from samples. Gigerenzer and Goldstein (1999) proposed that people estimate validities by the proportion

$$\hat{v}_i = \frac{\textit{number of correct decisions made by cue}_i}{\textit{number of decisions made by cue}_i}$$

in the sample of paired comparisons encountered so far.

Although this approach is adequate when sample sizes are large and of similar magnitude for all cues, Lee and Cummins (2004) pointed out correctly that it is problematic in environments where sample sizes for some of the cues are very small. For instance, when one cue has received 100 correct decisions out of 100 cases, and a second cue has 1 correct out of 1 case, both cues would have an estimated validity of 1. To correct for this problem, Lee, Chandrasena, and Navarro (2002) suggested a Bayesian approach to estimating validities, assuming uniform priors, which results in the following estimate:[5]

$$\hat{v}_i = \frac{\textit{number of correct decisions made by cue}_i + 1}{\textit{number of decisions made by cue}_i + 2}$$

This estimate is based on the same reasoning and has the same structure as Laplace's rule of succession (see Gigerenzer & Murray, 1987). It helps to avoid the consequences of unequal sample sizes. In the example just mentioned, the estimated validities would no longer be the same, but .99 and .67, respectively. As cues make more and more decisions, the two estimates converge.

Paying attention to validity only and ignoring discrimination rate leads to the following search rule:

> <u>Search by validity:</u> Choose the cue with the highest validity v_i. Look up the cue values of the two objects.

This search rule is a heuristic rule; it does not try to find the optimal order for a given set of cues and alternatives. For instance, it does not try to compute conditional validities or partial correlations to account for dependencies between cues. As noted before, when the number of

---

[5] Lee and Cummins (2004) also present the Bayesian estimate as a solution to the problem that validity does not reflect discrimination rates, arguing that the uncorrected estimate "does not take into account how often the cue discriminates when its validity is calculated" (p. 345). However, validity is defined so that it does not take account of discrimination rates, thus interpreting the Bayesian estimate this way makes it no longer an estimate of validity. In the next section, search rules that attend to both validity and discrimination will be introduced.

cues is large, such an attempt would be fruitless, because the optimal order cannot be computed by any mind or machine. Furthermore, even when the number of cues is small enough so that an optimal order can be determined in a given sample, in noisy environments the optimal order may not generalize well to new samples because of overfitting (Martignon & Hoffrage, 1999).

> *Hypothesis B (outcome feedback):* In stable environments that provide outcome feedback, random search in the initial trials will turn into search by validity in subsequent trials.

It will be shown below that empirical evidence from several studies supports search by validity in stable environments that provide outcome feedback. Before going into the details, alternatives to search by validity that also depend on outcome feedback will be considered. The key feature of search by validity is that it ignores the discriminating power of cues. Other search rules track discrimination rate in addition to validity.

Combining validity and discrimination

How can a heuristic for search pay attention to both validity and discrimination? One could postulate that cues are ordered by $v_i d_i$, but that would involve estimating the validities and the discrimination rates and multiplying them, which poses a problem for a process model (although not for an as-if model). Let us call the balanced feeling between validity and discrimination the "usefulness" $u_i = v_i d_i$ of cue i. In what follows, it will become clear that a mind can order cues according to usefulness without computing either v_i or d_i.

Note that the denominator (D_i) of v_i changes from cue to cue if the discrimination rate changes, because D_i is the number of discriminating pairs. Through multiplication of discrimination rate and validity, one gets rid of D_i, and now the denominator is P_i, the number of all pairs in which a cue has been looked up:

$$u_i = \frac{R_i}{P_i}$$

The "cognitive" trick is that usefulness, unlike validity, does not need to normalize the number of correct inferences against the number of discriminating pairs and thus keeps information about the discriminating power of a cue, and at the same time simplifies the computation. Even more, if all cues are looked up or presented equally often and thus do not differ in their presentation base rates, as in the learning phase of the share task, the denominator P_i is the same for all cues, P. Then, the order of u_i can be computed in an even simpler way, namely, by ordering cues by the number of correct discriminations R_i:

$$\text{order}(u_i) = \text{order}(R_i)$$

Usefulness thus measures the product of validity times discrimination rate, but without having to mentally multiply the two rates. It does so by a simple count of the number of correct answers due to cue *i* among all answers. Paying attention to both validity and discrimination is described in the following search rule:

> Search by usefulness: Choose the cue with the highest usefulness. Look up the cue values of the two objects.

Earlier work has discussed a related rule that was called search by success (Martignon & Hoffrage, 1999). In a simulation, the accuracy of search by success was only about 2 percentage points lower than that of search by validity. Like usefulness, success traces both the validity and the discriminating power of cues, but unlike usefulness, it complicates the computation by including correct guesses. The success s_i of a cue *i* amounts to its usefulness plus the proportion of correct decisions expected from guessing:

$$s_i = \frac{R_i + 0.5(P_i - D_i)}{P_i} = u_i + \frac{1 - d_i}{2}$$

where $P_i - D_i$ is the number of pairs in which a cue *i* does not discriminate.

Despite their similarity, usefulness and success do not always lead to the same rank order of cues. Figure 1.3 shows how usefulness and success combine validity and discrimination rate. The darker curves show cues with equal usefulness; the lighter curves show cues with equal success. Curves for success are steeper, that is, a constant loss in validity has to be compensated by a higher increase in discrimination. For instance, a cue with a validity of 1.00 and a discrimination rate of .45 has the same usefulness as one with a validity of .90 and a discrimination rate of .50, but the first has a higher success than the second.

Also the order of s_i can be computed in a simpler way than indicated by the formula above. If the learning phase consists of presenting only one cue at a time, and one has to make inferences only with this cue, then the learner can easily observe the total number of correct inferences (including correct guesses). In this case, again the following holds:

order (s_i) = order (R)

with R denoting the total number of correct decisions resulting from cases in which the cue discriminates plus cases of guessing.

> Search by success: Choose the cue with the highest success. Look up the cue values of the two objects.

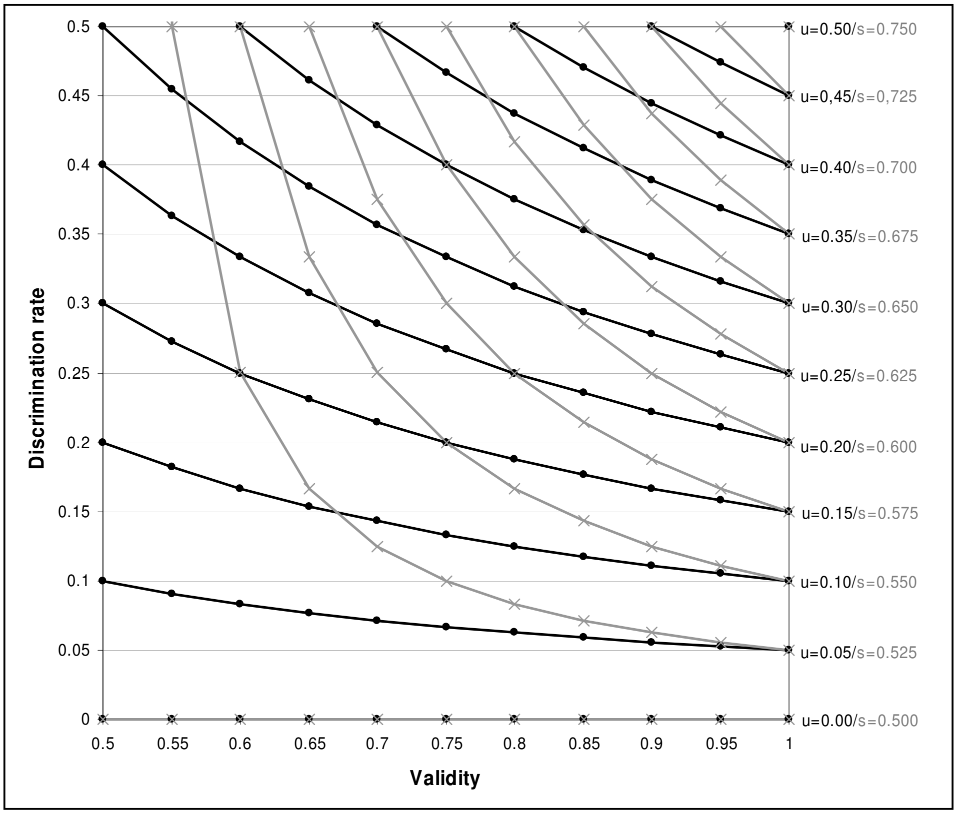

Figure 1.3: Isoquants for different values of usefulness (black lines) and success (grey lines) are shown as function of validity and discrimination rate. All points on an isoquant have the same usefulness, or success, respectively. The success isoquants are steeper than the usefulness isoquants, which implies that a fixed decrease in validity (if one moves from right to left) needs to be compensated for by a higher increase in discrimination rate for success compared to usefulness.

The ease of computation of the two measures that combine validity and discrimination rate thus depends on the learning phase:

> *Hypothesis C (learning phase):* In stable environments that provide outcome feedback, the search rule is cued by the structure of the learning phase. If in the learning phase all cues are displayed simultaneously, then one can observe the number of correct inferences by each cue relative to *P*, that is, the reference class is equal for all cues. This simultaneous learning favors search by usefulness in a subsequent test phase in which cues are searched sequentially. If the learning phase consists of presenting only one cue at each time, and one has to make an inference only with this cue, then the learner can observe the total number of correct inferences including guessing relative to *P*. This single-cue

learning should elicit search by success in a subsequent multiple-cue inference task.[6]

Newell, Rakow, Weston, and Shanks (2004) used simultaneous presentation of cues in the learning phase and report that search by success predicts the data better than search by validity. But this empirical data has not tested success against usefulness as a competing hypothesis, and a closer look at the data they provide reveals that there is just as much evidence for search by usefulness as there is for success. This experiment is reported in more detail below.

Information costs

The payoff function in an experiment can be of different kinds. In most experiments, participants' payoff depends on one of three kinds of information, or some combination: whether the decision was correct, how fast it was, and how frugal it was (how many cues were looked up). These three dimensions are accuracy, speed, and frugality. For instance, the payoff function in the share experiment was a combination of accuracy and frugality, whereas time did not play a role. In this section, the argument is put forward that in stable environments, the search rule in a heuristic adapts to the payoff function. The concept of relative information costs is defined as follows:

<u>Definition:</u> Relative information cost I is the cost of a piece of information (i.e., the values of the two objects on a cue) relative to the gain of a correct answer:

$$I = \frac{cost}{gain}$$

Note that cost and gain are defined in explicit monetary costs. They need to be distinguished from implicit "cognitive costs," which are difficult to measure and are not considered here (for approaches to measuring cognitive effort by counting elementary information-processing units, EIPs, see, e.g., Johnson & Payne, 1985; Newell & Simon, 1972). To invoke any motivation for search (as opposed to mere guessing) in the first place, an environment must have the feature $0 \leq I \leq 0.50$ in the case of a paired comparison. Otherwise, guessing would achieve higher net payoff.

If $I = 0$, that is, when information costs nothing, then there is little incentive (besides time spent) to search for highly discriminating cues. In memory-based search, $I = 0$, and thus, there is little incentive to sacrifice validity for discrimination. This leads to the following hypothesis:

---

[6] Note that there is no shortcut for the computation of validities that circumvents the required division by the number of discrimianting cases per cue. This might make validity difficult to learn. This issue will be dealt with in more detail in Chapter 3.

> *Hypothesis D* (*internal memory*): In stable environments that provide outcome feedback, search by validity is more frequent when search occurs in memory rather than in external storages. The difference will increase with the search costs in external storages.

There seems to be only indirect evidence so far. Bröder and Schiffer (2003) compared memory-based inferences with "inferences from givens," that is, with information that is fully displayed on a computer screen. Participants learned attribute patterns – consisting of four cues – of suspects in a murder case, and their task was to infer which of the two suspects was more likely to have committed the murder. To solve the task, they had to search for information in their memories. In two conditions that differed in memory load, the authors classified 72% of their participants in the memory load condition and 56% in the no load condition as Take The Best users (and recall that search by validity is one of its building blocks). A second experiment was devised to rule out the possibility that the effect might be due to the particular material that had been used. One group of participants received all the necessary information on a computer screen while solving the task; that is, they did not search for the information in memory. In this screen condition, only 20% of participants could be classified as Take The Best users, compared to 44% in a memory condition, consistent with Hypothesis D. Therefore, the effect cannot be attributed to the specific material the authors used. Unfortunately, the authors tested on the level of heuristics, not on the level of search rules, which makes the evidence indirect.

Search in external storages such as libraries and the Internet often has direct or indirect costs, such as money, time, and opportunity costs. For instance, in the share experiment, each cue costs money, and cues that rarely discriminate tend to lead to losses. This leads to the following more general hypothesis in environments where knowledge about validity and discrimination is available:

> *Hypothesis E (external search costs):* The higher I ($0 \leq I \leq 0.50$), the more search rules are guided by discrimination rate relative to validity. For low I, the observed frequency of the three search rules v, s, and u, should be $v > s > u$, that is, validity should be observed most often. For high I, the order is $u > s > v$, and for intermediate values, s should be most frequent.

Several experimental studies have used information costs in stable environments with feedback. The proportion of cases in which people followed the rule "search by validity" varied between studies. As mentioned before, Newell et al. (2003) reported that 75% and 92% of participants followed search by validity in two experiments, with six cues and I = 1/7, and two cues and I = 1/5, respectively. Newell and Shanks (2003) tested search by validity in three experiments with a similar mouse lab task and four cues. In the first experiment, 5 out of 16 participants (31%) followed the search rule (2 participants followed the search rule when costs were relatively low with I = 1/10, and 3 followed it when I = 1/5), but the learning phase was too short to provide sufficient information to learn the cue validities. When this problem was solved in the second and third experiment, which both used relatively high information costs of I =1/5, the percentage of participants who were reported to follow search by validity increased to 83% (10 out of 12), and 92% (22 out of 24).[7] When Bröder (2000) introduced information costs, 40% of participants in the low cost condition (I = 1/100) and 65% in the high cost condition (I = 1/10) were classified as TTB users, but he did not test the search rule independently; classifications were solely based on decision outcomes.

Taken together, there seems to be evidence supporting the use of search by validity as a process model. This evidence seems to fly in the face of Hypothesis E, as it was obtained despite varying information costs, and search by validity seems rather to increase with increasing information costs than decrease. Although none of these experiments tested other search rules besides search by validity, one should see search by validity decrease with higher information costs if Hypothesis E is correct. A closer look into the experimental design solves this apparent contradiction. All experiments used systematically constructed (rather than natural) environments, and all seem to have used constant discrimination rates.[8] When the discrimination rates of cues are constant, search by validity is the same as search by usefulness or success. Thus, the experiments could not distinguish between the three search rules, nor were they designed to do so. The fact that these experiments report a tendency, by direct or indirect evidence, that search by validity increases when I increases is fully consistent with concluding that search by, say, usefulness increases because the order by validity and usefulness coincided. The general finding that ordered search increases can easily be explained: The more expensive information is, the more selective should a decision maker be in what information to look up first.

This discussion shows that it is important to distinguish between experiments that embody the structure of a natural environment and those that use an artificially created structure. In natural environments, discrimination rates vary. There seem to be only two experiments in which the structure of a natural environment was embedded in the experimental task (Läge, Hausmann, Christen, & Daub, 2004; Lee & Cummins, 2004). Unfortunately, in the experiment by Lee and Cummins (2004) no search was involved as inferences had to be made from givens, that is, information about all cues was automatically and simultaneously provided to participants. In Läge et al.'s (2004) study, participants had to

---

[7] The validities in of the four cues were tightly spaced, .80, .75,.70, and .69. There was no way to reliably learn the difference between that last two validities. Consequently, inversions of the last two cues are ignored here.

[8] Bröder's (2000) environment consisted of all 120 paired comparisons of the 16 different cue patterns that can be formed by 4 cues. To be able to construct 16 different cue patterns from 4 cues, all cues have to have maximum discrimination rates, otherwise the number of different cue patterns would be < 16. Similarly, Newell and Shanks' (2003) environment consisted of all 120 pairs of 16 distinct cue patterns formed by 4 cues. Newell et al. (2003, Experiment 1) used 6 cues resulting in 64 distinct cue patterns – again only possible when every cue had the same and maximum discrimination rate.

search for cues. The authors used the German cities data set from Gigerenzer and Goldstein (1996) but replaced the names of the cities with fictitious names of Chinese cities, and relabeled cues accordingly. The task was still to select the larger of two cities. Participants were informed about the cues' validities and discrimination rates prior to each decision through display of the respective numerical values. These measures thus did not have to be learned in a separate phase of the experiment. Relative costs for cues were 1/10. The authors looked at many different search rules. On the majority of trials (66%), search followed criteria that combine validity and discrimination rate (e.g., usefulness, success, the additive combination of validity and discrimination rate, etc.).

In the Newell et al. (2004) study, briefly mentioned above in connection with Hypothesis C, discrimination rates of all four cues also vary, and all four ordering criteria discussed here (discrimination rate, validity, success, and usefulness) lead to different orders. Although the authors did not look at the usefulness search order in particular, it can easily be computed from the validities and discrimination rates they provide. Unfortunately – probably because the authors were not interested in looking at usefulness – success order and usefulness order differ only in the two lowest-ranking cues. In two experiments that were identical in the learning and decision phases, the order of cues according to the number of trials in which they were purchased follows success in Experiment 1, and usefulness in Experiment 2. The same holds for the ratings of the subjects. Relative information costs were 1/6. A theoretical reason favors usefulness over success, as the design of the learning phase speaks against counting correct guesses. As mentioned, all cues were presented simultaneously in the learning phase. Thus, the computation of the success of each cue would have required considering for all non-discriminating instances the expected accuracy from guessing, even though participants mostly did not have to guess, as at least one of the four cues was likely to discriminate.[9]

In sum, the support for search by validity comes from studies in which discrimination rates did not vary, and therefore v, s, and u all predict the same search order. The few studies that used data for which the predictions of the different criteria vary support the hypothesis that both discrimination rate and validity of a cue influence its position in the search order. This evidence was obtained when search proceeded in the external environment and information had direct costs. Search by usefulness and success might, depending on the learning phase, have the additional advantage of being relatively easy to learn. For search in memory, search by validity is predicted – whether the so-far supporting evidence for this hypothesis also holds when different search criteria are tested directly and against each other remains to be seen.

---

[9] Note that also sequential cue search and application of the stopping rule of Take The Best would not provide the data necessary for s, as no guess is made if a cue does not discriminate; search continues for another cue, unless the entire set of cues is exhausted.

Changing environments

Search rules have generally been modeled assuming a stable environment. That is, the validities, or discrimination rates, of the cues are assumed to be stable across time and alternatives, which can be seen from the fact that the model parameters are estimated from the total set of alternatives, such as the total number of correct discriminations by a cue, R_i. Also experimental tests of search rules have usually been performed by putting participants in stable environments, as in, for instance, Newell et al. (2003).

The term "changing environment" describes environments where the characteristics of cues (such as validities and discrimination rates) change systematically over time or across sets of alternatives. This change can occur to different degrees and on different time scales. For instance, in human history preference for high sugar- and high fat-content food provided an adaptive advantage because food was scarce, and thus high-energy food should be preferred. However, in modern societies with abundant around-the-clock supply of food of all kinds and at the same time a lack of physical demands, high sugar and fat content no longer represent healthy nutrition – the cues' direction might even have reversed (Eaton, Konner & Shostak, 1988). Similarly, discrimination rates of cues can change over time or between situations, such as when one predicts first the profitability of shares for insurance companies that vary much in their employee turnover rates, and then for a group of consulting companies that *all* have high turnover.

The cognitive challenge is how to guide search when a person does not know in what direction an environment will change, or how much, or whether it will change at all. There is empirical evidence in the problem-solving literature – going back to Duncker's (1935) experiments – that people tend to try the solution that worked the last time they were faced with a new problem. This habit has been referred to as "Einstellung" (mental "set"; Luchins, 1942). Similar evidence comes from decision-making experiments and game theory, where a "win-stay, lose-shift" heuristic has been demonstrated to work very well in the Prisoner's Dilemma game (Nowak & Sigmund, 1993). In the context of information search in paired comparisons, relying on what worked the last time has two meanings: in environments without feedback, it means using the cue that allowed stopping search the last time, through discrimination; in environments with feedback, it means using the cue that stopped search the last time *and* made a correct decision.

In environments without outcome feedback, the following search rule can track changing structures:

> Einstellung search (by discrimination): Choose the cue that stopped search on the most recent problem. Look up the cue values of the two objects.

To explain that logic of this search rule, consider the task in Figure 1.1 but assume that there is no feedback. In the first trial, a participant may have tried "financial reserves?" but both companies had financial reserves. Then she looked up which company was an established company, and the answer was Company A but not B. Because the companies differ, this cue is informative. Einstellung search assumes that she enters “established company?” into a memory, at this point a string with length one:

<established company?>

The search rule predicts that, in the next task, she will try “established company?” first. If this cue discriminates again, the memory will not be changed. If it does not discriminate, but “share trend positive?” does, the memory will be enlarged to two items, with the most recent one first, resulting in

<share trend positive?; established company?>

In the next task, she will try “share trend positive?” first, and if it discriminates, the memory does not change. Otherwise, she tries “established company?”, and if it discriminates, the order of the two will be reversed. If neither of the two makes a difference, a third cue will be looked up, and if it discriminates, it will be ranked first in a memory that is now of length 3, and so on. The maximum memory span this search rule needs is M–1 (because if M–1 cues failed to discriminate, there is only one cue left to look up and its position does not need to be remembered). In the example in Figure 1.1, $M = 6$, so the maximum memory span needed is 5. The rule does not need to compute frequencies or probabilities, nor pay attention to reference classes.

Different from search by discrimination, Einstellung search has a recency bias. The focus on most recent information is the factor that enables the rule to adapt quickly to changing environments. The important difference between Einstellung search and all other search rules discussed so far is that it needs no learning phase to get reliable estimates for ordering cues. Still, if cues are ranked according to d_i, the top cue has by definition the highest probability for being the one that stopped search in the most recent problem, the second cue has the second-highest probability, and so on. When it is uncertain whether the environment is stable, Einstellung search has substantive advantages over d_i because it has a short memory and quickly can adapt to a changing environment.

Einstellung search can be generalized to situations in which outcome feedback is provided:

> <u>Einstellung search (by correctness):</u> Choose the cue that allowed for a correct inference on the most recent problem. Look up the cue values of the two objects.

The logic is identical to Einstellung search by discrimination, except that Einstellung search by discrimination is guided by whether a cue allowed for a correct answer and not by whether a cue merely discriminated. Consider for example the task in Figure 1.1 with feedback, as in the original experiment. If the feedback says that Company B is the correct answer, the cue "established company?" is entered into a task-specific memory and placed at the top of a string of maximum length M–1. In the next trial, it will be tried first, and the quick updating of the order in the string is the same as for Einstellung search.

Just as Einstellung search by discrimination is related to discrimination rate, Einstellung search by correctness is related to usefulness. Usefulness can also be expressed as the unconditional probability that a particular cue will lead to a correct decision. Therefore, if cues are ranked according to u_i, the top cue has by definition the highest probability for being the one that led to a correct decision in the most recent problem, the second cue has the second-highest probability, and so on. The important difference is again that Einstellung search by correctness does not need a learning phase and can pick up changes in the environment quickly, from one case to the next. This leads to the following hypothesis for changing environments:

> *Hypothesis F (changing environments):* The more a decision environment changes over time in statistical properties such as discrimination rates and validities of cues, the more frequently search rules with small memory are used relative to others.

If the discrimination rates or validities change over time in the share experiment, Einstellung search could quickly react to change because it has a short memory. Search by discrimination or usefulness would adapt slowly to change because the updating of discrimination rates and usefulness keeps earlier experience in memory. Unfortunately, such an experimental test using changing environments has not yet been conducted.

It is clear that search by discrimination or usefulness (or any other criterion computed across the whole set of alternatives) would not do well in changing environments, but how would Einstellung search do in stable environments? Would it, conversely, always do badly in stable environments? Not necessarily.

Note first that every stable environment looks to a novice like a changing environment in the first steps of learning. Thus Einstellung search should have an advantage when information about an environment, stable or unstable, is scarce. A comparison between Einstellung search by discrimination (which only looks at discrimination and ignores outcome feedback) and search by validity in an environment with feedback, assuming different levels of knowledge, provides evidence for this prediction. Recall that whether an inference is correct only matters for search by validity. This seems to be a quite unequal competition. In predicting which of two cities has more inhabitants, search by validity (as embedded in Take The Best) outperformed Einstellung search by a wide margin when the learning set included 100% of the cities (data fitting), and by a similar, only slightly smaller margin when the learning set included 50% of the cities and the strategies were then applied to the other 50% (cross-validation). A multiple regression model outperformed Einstellung search by a similar margin. However, when the learning phase had only limited information, that is, up to 20% of the cities, with half of their cue values missing, then Einstellung search outperformed regression, and with even more limited knowledge of only 10% of the cities also search by validity (Gigerenzer & Goldstein, 1999). This is a quite striking demonstration of the power of simplicity when uncertainty is high. The power disappears the more information becomes available.

But the instability of early learning is not the only factor that can make Einstellung search successful also in stable environments. The recency bias contained in Einstellung search makes the rule highly sensitive to the order of presentation of paired comparisons, enabling it to track information contained in that order, such as when the paired comparisons are presented in a systematic rather than random order. By ordering the alternatives in a systematic way, a stable environment is transformed into a changing one. Even Einstellung search by discrimination can lead to astonishing performance under these circumstances, without paying attention to correctness, as the following surprising result demonstrates.

In early simulations with the German city size data for the book "Simple heuristics that make us smart" (Gigerenzer et al., 1999), an at first inexplicable result was obtained: The one-reason decision making heuristic Take The Last, differing from Take The Best in its search rule, which is Einstellung search by discrimination, outperformed other strategies by some 10 percentage points (Gigerenzer, 2000, pp. 189-190). How could this happen? A closer look at the simulations revealed that the secret behind Take The Last's success lay in systematic testing. Instead of randomly creating pairs of cities, as done in later simulations, the simulations started with the largest city and compared it with the remaining smaller cities in descending order. Then the second largest city was compared to all smaller ones, and so forth. Larger cities tend to have positive values on many cues but are more likely to differ from smaller cities on the more valid cues. Whereas less valid cues have a relatively scattered distribution of positive and negative values across the ordered list of objects, highly valid cues by definition are very likely to have positive values for large, and negative values for small cities. Thus, assuming equal discrimination rates, highly valid cues more often discriminate the large cities in the set from the small cities than do less valid cues. If now the order of paired comparisons is constructed such that cities are always compared to increasingly smaller ones, Einstellung search helps to find cues on which larger cities have positive values whereas smaller cities have negative values, and these happen to be at the same time very valid in this kind of testing scenario.

Einstellung search by correctness has not been tested yet in simulations. However, due to its focus on correctness, it will probably achieve slightly higher accuracy and slightly lower frugality than Einstellung search by discrimination and is thus to be preferred in most situations in which outcome feedback is available.

To summarize, in stable environments without feedback, the hypothesis was put forward that search follows discrimination. In stable environments with feedback, search will in contrast follow validity, usefulness, or success. Three hypotheses emerge: (1) Search in memory will favor validity; (2) the format of learning – simultaneous or single cue – will favor usefulness and success, respectively; and (3) low relative information costs will elicit search by validity, intermediate costs search by success, and high costs search by usefulness. Although there is some evidence in support of the first two hypotheses, the evidence is flawed by the fact that different search orders were confounded in the design of the respective experiments, both in Börder and Schiffer (2003) with regard to the first hypothesis, and, to some extent, also in Newell et al. (2004) with regard to the second hypothesis. Future research has to test different criteria against each other, also taking into account compatibility with the kind of learning phase that is used. This requirement also holds for testing the third hypothesis – so far, again due to confounding of different search orders (e.g., Bröder, 2000; Newell & Shanks, 2003; Newell et al., 2003), the evidence for increased use of search by validity under high information costs conditions has to be treated as preliminary and needs to be tested more rigorously. Unstable environments require that instead of using cue orders that are fixed by a criterion computed across a data sample, such as validity, more weight is given to recent information. Search by recency of either the last discrimination (if no feedback is available) or the most recent correct decision made by a cue fulfills this condition and is thus hypothesized to be observed when people have to make decisions in environments whose properties change over time.

Stopping and decision rules

A search rule gives search a direction, in memory or in external storages. It does not contain a criterion when to stop search. Search could continue in principle infinitely – in memory, libraries, the Internet, or other storages. Search without a stopping rule is called exhaustive search. It is possible in certain situations where the number of cues is constrained, such as in experiments with abstract stimuli that limit the number of cues to those presented, such as in the share task (Newell et al. 2003). In this paradigmatic example (see Figure 1.1), one can at most search for six stimuli. Intentionally, this makes search in long-term memory inapplicable. If the problem involves real rather than hypothetical companies, such as predicting whether Intel or IBM shares will be more profitable, exhaustive search – if it ends at all – could consume considerable time.

The question of when to stop has generated two kinds of answers, just like the question of what search order to use. One answer is to model stopping within an optimizing framework, and classical work comes from statistics, namely Wald's (1947) sequential analysis. Models for sequential search with optimal stopping rules have been proposed in economics (e.g., Stigler, 1961) and psychology (e.g., Anderson, 1990; Busemeyer & Rapoport, 1988). These models are often called optimization under constraints. The key idea is that search is stopped when the costs of further search exceed its benefits. For instance, in one of Anderson's (1990) models, search in memory stops when costs of retrieving the next record in terms of retrieval time exceed the benefit of retrieving it.

Heuristic search, in contrast to exhaustive search, needs stopping rules, and, in contrast to optimized stopping, does not compute the optimal stopping point. Stopping search at a reasonable point is an ability that every intelligent system needs, otherwise it may be slow, caught in irrelevant details, or incapable of making a decision. Thus, it is surprising that there are only a few theories that formulate the conditions of stopping, and few experiments that test when people stop looking for further information. Stopping rules can be distinguished according to how many cues they let pass before they stop. Unlike in the section on search rules, this part will not deal primarily with environmental characteristics that a priori exclude certain rules, or at least unambiguously favor certain rules over others. Rather, environmental features that vary in degree are discussed, and it is this degree on which it depends to what extent certain stopping rules are favored over others. Therefore, the different stopping rules will be introduced before addressing the question of under which circumstances they are ecologically rational.

Stopping by number of discriminating (informative) cues

One stopping rule that has already been introduced is Take The Best's. It stops when the first cue is found that discriminates between options, that is, shows a positive value for one option, and a negative value for the other. It thus stops when the minimum requirement for making an informed decision is fulfilled.

<u>One-reason stopping:</u> Stop search after the first informative cue is found.

One-reason stopping is a special case of taking a tally of m_d discriminating cues ($1 \leq m_d \leq M$).

<u>Stop by tallying:</u> Stop search after m_d informative cues are found.

Consider search by validity and stopping after the second cue that discriminates, that is, $m_d = 2$. Looking up a second cue is less frugal, but can one expect more accurate inferences in return? At first glance, it seems unreasonable to stop after the second discriminating cue. There are two possible results. If the second cue points to the same alternative as the first, the decision will be the same and accuracy will not improve. If the second cue points to the opposite alternative, then again any reasonable decision rule will decide with the first cue, because it has the higher validity. Thus, in either case, one cue is as good as two. If there are costs, then one cue is better than two. Without further considerations, this result could lead to the hypothesis that in paired comparisons, search stops after the first or third (or $m_d > 3$) discriminating cue, but not after the second discriminating one.

But there is a problem. Imagine a person who uses a tally rule with three discriminating cues. The first two cues point into the same direction, thus she decides to stop because of the same logic that was just used to prove that two cues are not better than one cue. If two discriminating cues point into the same direction, it does not make sense to look for a third – one should do so only if the second cue points in a different direction from the first. Thus, the general assumption of stopping rules by discrimination seems to have a deep problem. The idea of tallying by discrimination is therefore given up, and instead I turn to stop by tallying what can be called positive evidence for one alternative.

> <u>Stop by tallying':</u> Stop search after m_d informative cues are found that point to one alternative.

In the case of stopping after the first discriminating cue ($m_d = 1$), the two rules do not differ. One discriminating cue also means that there is one piece of positive evidence for one of the alternatives. The simplest version that makes a difference would be

> <u>Take Two:</u> Stop as soon as there are two cues that point to one alternative.

Stopping by a fixed number of cues

All the stopping rules considered so far stop depending on whether a cue discriminates (i.e., is informative). They can violate guidelines of the kind: always check the following three (or four, five) cues before you make a decision. Social guidelines often are in the form of a fixed number of cues, rather than a fixed number of discriminating cues. For instance, when interviewing two applicants for a job in a bank, the interviewer may go through the protocol of getting information about college grades, IQ, the field in which the applicant majored, professional experience, and foreign language abilities, as prescribed by the bank's guidelines, and record all of this information. For her decision about to whom to offer the job, however, the interviewer may use only two pieces of information. The social protocol of the bank prescribes not stopping until the list of cues is gone through, and the interviewer decides not to violate it. Similarly, doctors often feel compelled to take a fixed number of measurements from their patients when there is the suspicion of a disease, but this amounts to defensive information search. If the patient dies, the doctor can defend herself by proving that she did all the tests, although she may not use many of the test results for making a treatment recommendation. Stopping after a fixed number of cues amounts to ($1 \leq m_f \leq M$):

> Fixed-number stopping rule: Stop search after the values of m_f cues have been looked up (whether the cues discriminate or not).[10]

The simplest case is:

> M_f=1 rule: Stop search after the values of one cue have been looked up (whether this cue discriminates or not).

This frugal stopping rule seems to be ecologically rational in environments with real-valued cues rather than with binary cues, because real values avoid a high rate of decisions by guessing.

Dependencies between stopping and decision rules

Many psychological experiments study cognition in tasks that, explicitly or implicitly, exclude search for cues. Often, tasks are used in which all cues are already displayed in front of the participant in a questionnaire-like form. In addition, to eliminate search for additional cues in memory, alternatives are chosen about which the participant has no specific knowledge, or which are hypothetical (such as "Company A" and "B" in the share task), so that the person cannot retrieve further cues and cue values from memory. The reason for elimination of search is often to avoid the possibility that cues other than those displayed could influence the final judgment. For instance, many theories of judgment focus on the question of how a person integrates *given* information. Consistent with this type of question, many models of human judgment do not model search, and as a consequence also do not include theories about stopping search, but are instead models of the decision outcome, such as multiple regression models of vicarious functioning predominant in the field of multiple cue probability learning (see Cooksey, 1996; Holzworth, 2001), and other weighted additive models, such as the large array of expected utility theories (see Fishburn, 2001). By modeling and testing these various decision rules, the implicit (but rarely explicit) assumption is that the process of information acquisition does not constrain decision rules. However, this independence does not always hold.

As laid out above, any stopping or decision rule can be applied regardless of which search rule has been used. Both general types of stopping rules can be combined with each of the seven search rules defined earlier. In other words, search does not constrain stopping. However, one-reason stopping constrains the range of possible decision rules, and in that

[10] There is also the possibility of hybrid stopping rules that combine stopping by number of discriminating cues and stopping by a fixed number of cues. For example, a decision maker might, when information is costly, aspire to use one-reason decision making, but at the same time set herself a limit as to how much information she will maximally purchase, such as "Stop when a discriminating cue is found, but only look for a maximum of m_f cues. If no discriminating cue is found until that point, guess." For simplicity, such rules are ignored in the analysis put forward in this chapter.

regard poses a theoretical challenge to many cognitive models. One-reason stopping excludes all decision rules that weigh and add the cue values of several cues. When search is stopped after just one discriminating cue, any reasonable decision maker will also rely on that cue. Whenever the hypothesis is put forward that one-reason stopping will be used, I therefore also predict that people will decide in favor of the alternative to which the one discriminating cue points.[11] Because of the strong dependency between one-reason stopping and decision making, both building blocks are treated together in this section.

In the remainder of this chapter, I will lay out how the performance of different stopping rules depends on the structure of the decision environment and thus derive testable predictions about people's behavior. The statistical structures of decision environments to be analyzed in more detail are the "*compensatoriness*" *of cues weights*, and *correlations between cues* (i.e., the degree of *information redundancy* in an environment). Then, the impact of *information costs* will again be explored. Finally, the influence of the *format of presentation of cue information* is considered. How stopping rules depend on environmental structures is summarized in Table 1.2.

Table 1.2: Stopping rules dependent on environmental structures

Stopping by tallying ($m_d > 1$)	One-reason stopping ($m_d = 1$)	Fixed stopping ($m_f < M$)
- compensatory environments - low-redundancy environments - low information costs	- non-compensatory environments - high-redundancy environments - high information cost (direct or indirect) - search in memory - propositional cue format	- (very) high information costs

Compensatory versus non-compensatory environments

Consider an environment with M binary cues $c_1, \ldots, c_M$, and a linear strategy, such as multiple regression, that orders the cues by their beta-weights $w_1, \ldots, w_M$. An environment is called non-compensatory if every weight w_i is larger than the sum of the weights $w_{i+1} + w_{i+2} + \ldots + w_M$ (Martignon & Hoffrage, 1999). An example is the set of weights 1, 1/2, 1/4, 1/8, 1/16. Figure 1.4 shows an example of a compensatory and a non-compensatory environment.

---

[11] Chapter 2 of this dissertation will substantiate this claim and provide empirical evidence for the effect of stopping on the decision rule that is used.

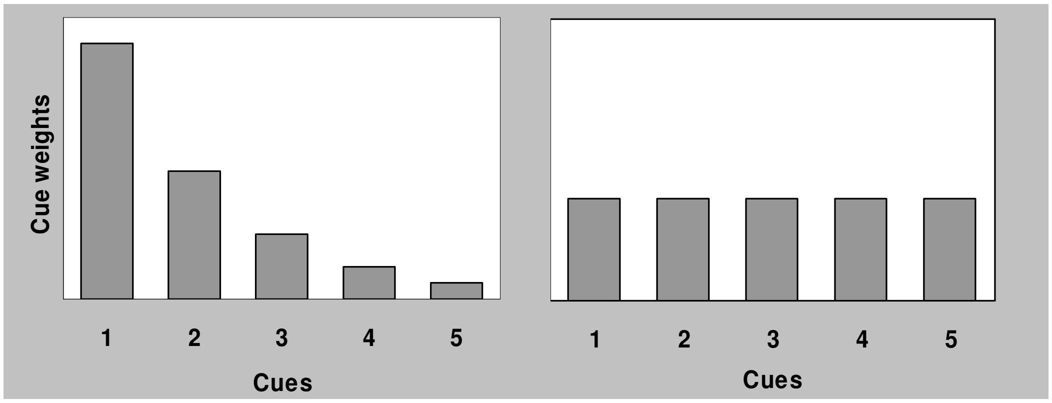

Figure 1.4: A problem structure with non-compensatory weights (left), and one with compensatory weights (right).

In a non-compensatory environment, no weighted linear rule can outperform the faster and more frugal Take The Best when both use the same order of cues (Martignon & Hoffrage, 1999). Thus, in a non-compensatory environment, a one-reason stopping rule (part of Take The Best) is ecologically rational. In contrast, in a compensatory environment, a tallying stopping rule allows for a tallying decision rule, that is, a unit-weight linear rule, which can also perform at the level of multiple or logistic regression (e.g., Dawes & Corrigan, 1974; for application of tallying rules to classification, see Forster, Martignon, Vitouch, & Gigerenzer, 2002).

Leaving the area of inferences under uncertainty for a moment, one can find non-compensatory structure in human-made rule systems that are supposed to be easy to apply and should allow fast decisions. Traffic rules provide a good example: In Germany, the right-of-way is regulated in a non-compensatory way. When the traffic lights are red, you have to stop even if there is a right-of-way sign at your side of the crossing and the other driver comes from the left-hand side. A red light cannot be compensated for by the presence of two contradicting cues of minor "impact." And even three pieces of contradicting evidence (green light, right-of-way sign, you come from the right) cannot overrule the stopping gesture of a police officer regulating the traffic at a crossing. These rules allow us to make fast and predictable decisions when driving, which is crucial for maneuvering safely among many other road users. The error-prone careful weighing and summing of pieces of evidence for or against alternative courses of action would have disastrous consequences here. This example also shows that individuals are not only able to adjust their decision behavior to certain situational demands but can also change their environments so that they fulfill conditions that allow people to make fast decisions through one-reason stopping and deciding. But let us now turn back to the former case and see for inferences under uncertainty whether people adapt their stopping and decision behavior to the (non-) compensatoriness of the environment.

Knowledge about an environment's structure, for example, from learning through feedback, should lead to the use of stopping rules that are adapted to the structure of the environment:

> *Hypothesis G (non-compensatory information):* One-reason stopping rules are used more frequently in a non-compensatory environment than in a compensatory environment.

There are two experiments that provide a test of Hypothesis G. In an experiment by Rieskamp and Otto (2004, Study 1), participants took the role of bank consultants who had to evaluate which of two companies applying for a loan was more creditworthy. For each company, six cue values, such as qualification of employees and profitability, were provided. One group of participants encountered a non-compensatory decision environment, meaning that in about 90% of the cases, the outcome feedback they received was determined by dominance of the first discriminating cue rather than integration of several cues. For the second group, feedback was determined in a compensatory way, meaning that in about 90% of the cases, the more creditworthy company was determined by a weighted additive rule, which for each alternative calculates the sum of cue values multiplied by the corresponding cue validities and selects the alternative with the highest score. Did people adapt their heuristics to the structure of the environments? This was indeed the case: In the non-compensatory environment, the choices consistent with Take The Best increased from 28% to 71%, whereas in the compensatory environment they decreased to 12%. People learned—without instruction—that in different environments different heuristics are successful.

In an experiment by Bröder (2003, Experiment 2), a share task was used. Half of the participants were assigned to an environment whose cue weights were almost non-compensatory (i.e., the cue weights of the payoff function were 47, 25, 17, and 10), and the other half to a compensatory environment (the corresponding weights were 32, 26, 22, and 20). In the non-compensatory environment, 77% of the participants were classified as using Take The Best, but only 15% in the compensatory environment, consistent with Hypothesis G. The corresponding numbers for the use of a compensatory strategy were 15% and 60%, which is also consistent with Hypothesis G, because stopping by tallying (or by a fixed number of cues > 1) is a necessary condition for any compensatory strategy.

Non-compensatory environments are characterized by a dispersion of cue weights across a wide range, that is, a high variability of cue weights. While it is presumably difficult for a decision maker to evaluate whether a certain environment mathematically fulfills the condition of being non-compensatory, she might very well have had the experience that in some decision environments, certain cues are much better than others and much more often lead to correct decisions than others, whereas in other environments, cues differ less in their predictive power. Also Bröder's (2003) "non-compensatory" environment was not strictly speaking non-compensatory, but only close to it. Nevertheless, a high proportion of participants predominantly used Take The Best. Therefore, a weaker variant of Hypothesis G can be formulated:

> *Hypothesis H (variability of cue validities):* The larger the variability of cue validities, the more frequently one-reason stopping rules are used.

Bröder (2000, experiment 2) compared a high dispersion environment (with cue validities ranging from .54 to .96) and a low dispersion environment (from .66 to .90) in terms of prevalence of different decision strategies. He found only a small (non-significant) difference in the number of participants who could be classified as one-reason decision makers: Under the low dispersion condition, 20% of the participants were classified as TTB users; under the high dispersion condition, this percentage rose to 35%. But there might be an obvious reason for these low percentages. Search and stopping was prevented by the experimental setup: All were cues presented simultaneously, which might make ignoring information difficult. I will talk about the effect of preventing search in more detail below.

The issue of variability of cue structure is not necessarily restricted to inferences. The concept of a non-compensatory attribute structure can also help to explain decisions in the domain of preferential choice, or, more broadly speaking, decisions where an objective outside criterion on which to evaluate accuracy is missing. Because of the missing outside criterion, validities cannot be computed, but people assign different importance weights to attributes of options to choose among, and search through attributes accordingly (e.g., Aschenbrenner et al., 1984; Saad, 1999). Strong preferences for certain attributes could create a situation comparable to a non-compensatory cue structure. This would mean, for example, that positive values on important attributes cannot be compensated for by negative values on other attributes about which one cares less. Related to this, it has, for example, been shown that strong preferences immunize people to preference reversals due to framing outcomes in terms of losses or gains (Wang, 1996).

Similarly, in the domain of moral preferences, which often involve difficult decisions or even dilemmas, holding moral values, or protected values, can facilitate decision making (e.g., Baron & Spranca, 1997). These values are so named to indicate that they are protected from otherwise difficult trade-offs (Fiske & Tetlock, 1997). In the domain of moral decisions the problem of incommensurability becomes especially immanent. How would one translate, say, the value of hundreds of animal and plant species becoming extinct through slashing and burning in the Amazon rain forest, and the benefits expected from breeding cows on the destroyed area, into a common currency (let alone dilemmas that might require assigning monetary value to human life)? Many people refrain from trading human life, love of one's family, and so forth for money or material goods, characterized to the point by the slogan "no blood for oil." Some values are regarded as priceless and *not* to be *compensated* by money, and attempts to do so are viewed as immoral. Again, immunity to framing manipulations has been demonstrated for people holding protected values in the area of environmental protection (Tanner & Medin, in press). Let us now have a closer look at the issue of trade-offs.

Correlations between cues

Inter-attribute correlations have received attention especially in the area of preferential choice (e.g., Bettman et al., 1993; Fasolo, McClelland & Lange, 2004; Gilliland & Schmitt, 1993; Johnson, Meyer & Ghose, 1989), where negative correlations are a frequently occurring characteristic. One example is the negative correlation between price and quality of a product inherent in many consumer choices. In experiments on preferential choice, people have often had to choose between gambles that differ, for example, in the amount to be won and the probability of winning, both very often negatively correlated. When correlations between attributes are negative, this creates conflict for the decision maker: One will rarely find a pair of alternatives to choose from where all attributes point to one alternative. Rather, in negatively correlated environments, different attributes tend to point to different alternatives. Therefore, there are trade-offs, and environments can be considered as "unfriendly" in that regard.

In contrast, negative inter-cue correlations are rare in inference problems. Mostly, cues correlate positively with each other due to the constraint that they all have to correlate positively with the criterion. (Chapter 2 of this dissertation will explore this connection in more detail.) These on average positive correlations often create high information redundancy. High information redundancy in an environment means that looking up more information often does not reveal new information. Thus in terms of accuracy, stopping after one discriminating cue is found should not fall far behind looking up more information and integrating it. One does not need to trade accuracy for frugality in an environment with positively correlated cues. Again, although a decision maker might be unable to estimate correctly the numerical value of the correlation between cues, she might very well note that certain cues quite often make the same prediction under conditions of high redundancy. Similarly, she can note that certain cues often contradict each other and make opposite predictions under conditions of low redundancy. This leads to the following hypothesis:

> *Hypothesis I (information redundancy):* The higher the average correlation between cues, the more frequently one-reason stopping rules are used.

The second chapter of this dissertation is dedicated to the question of how well different heuristics perform in high- versus low-redundancy environments, and whether people respond adaptively to this environmental structure. Support for Hypothesis I has indeed been found and is reported there (Figure 2.2 on p. 65 can be consulted to gain an impression).

But can Hypothesis I be turned around, predicting that in low-redundancy environments, people should search for information more extensively and apply, for example, stopping by tallying? Although one-reason stopping and deciding has been shown to be less accurate in unfriendly environments with no or negative inter-cue correlations (Shanteau & Thomas, 2000)[12], there is a psychological benefit to be gained from limiting information search. In low-redundancy environments, even more so when cues are negatively correlated, there often exists conflict between cues. Here, one-reason stopping can serve as a strategy that avoids conflict. For the preference domain, it has been suggested that considering many pieces of information makes explicit trade-offs that can be uncomfortable to face (Hogarth, 1980). Individual differences might exist in whether one prefers facing the conflict to achieve higher performance although it makes a decision harder, or avoiding the conflict through one-reason stopping and settling for a slightly lower performance. This is another question that will be addressed empirically in the second chapter of this dissertation.

Information costs

If search is in external sources, the question of costs has a clear meaning and can affect the stopping rule. As defined above, relative information costs are the costs of a piece of information relative to the gain from a correct decision. When information is costly, stopping early is a very straightforward way to save money.

> *Hypothesis J (external search costs):* The higher the relative information cost I ($0 \leq I \leq 0.50$), the more frequently one-reason stopping rules are used.

In Experiment 1 of Newell et al. (2003), relative information costs were 1/7, with six cues, and in Experiment 2, I was 1/5, with two cues. As mentioned above, in all cases where participants searched for information (rather than guessing), the proportion of cases in which subjects did not search beyond a first discriminating cue was .80 and .89, respectively. This result is consistent with Hypothesis J.

For $I = 1/10$, Läge et al. (2004) report similarly high proportions for the use of a one-reason stopping rule. In 76% of decisions, search was stopped after the first discriminating cue was found. Like the Newell et al. (2003) study, this study provides indirect evidence only as I is compared across experiments, which differ in several respects. To test the hypothesis directly, relative costs must be varied *within* one experiment.

---

[12] However, even setting aside the fact that negative inter-cue correlations are rare in inference problems, I will show in simulations (reported in detail in Chapter 2) that when cue validities are widely dispersed, TTB hardly suffers from low redundancy. Only when there is both low redundancy and little variability of cue validities does TTB's accuracy drop. Again, a comparison with the preference domain seems to suggest itself, where, as reported above, strong preferences or moral convictions help to solve otherwise difficult trade-offs.

There are three studies that varied information costs in such a way. Bröder (2000, Experiment 3) showed that when the relative information costs I increased from 1/100 to 1/10, the proportion of people classified as using Take The Best increased from 40% to 65%. However, classifications were based on decisions only. Bröder did not look at process measures and thus does not report how many people precisely followed Take The Best's stopping rule, that is, one-reason stopping.

Newell and Shanks (2003, Experiment 1) did look at stopping separately. Their analysis of stopping rules shows very clear results: In the low-cost condition (I = 1/10), only in 36% of the non-guessing trials was adherence to a one-reason stopping rule observed. When I was increased to 1/5, this proportion rose to 85%. This result strongly supports Hypothesis J.

The experiment that will be reported in more detail in the third chapter of this dissertation also revealed a higher proportion of one-reason stopping under high (I = 3/20) compared to low (I = 1/20) relative costs. Five cues were available. When costs were low, participants stopped buying information after one discriminating cue was found in 51% of all cases. When costs were high, this percentage was 70% (see also Figure 3.6 on p. 108). In turn, however, this result implies that in the low-cost condition, the proportion of search that continued after one or more discriminating cues had already been found was relatively high. Those cases might be examples of stopping by tallying, a stopping rule that is easier to afford when costs are low. In the introduction to this section, arguments were put forward for stopping by tallying positive evidence for one alternative rather than stopping by tallying discriminations per se. The simplest case that allows for compensatory decision making, while still ensuring moderate frugality, is Take Two, that is, stopping when two discriminating cues are found that point in the same direction (but not after two cues that point in different directions). To test the alternative tallying rule, I only looked at the subset of responses in which search continued until at least a second discriminating cue was found. In half of these cases, the second discriminating cue pointed in the same direction as the first and only 14% of participants continued searching. In the other half of these cases, the second discriminating cue pointed in the other direction. Now, participants continued to search in 83% of the cases. Figure 1.5 illustrates this finding. This represents strong evidence for the tallying rule with m_d = 2, which is adaptive in a compensatory environment when costs are relatively low. Also the experiments reported in Chapter 2 will provide evidence for this sensitivity to positive evidence.

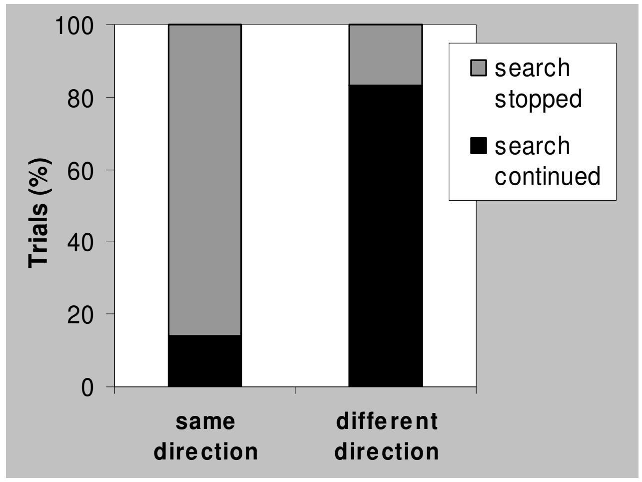

Figure 1.5: After two discriminating cues that point in the same direction have been found, search was continued in only a few cases (left). When, in contrast, the second discriminating cue pointed in a different direction from the first, search was continued in most cases (see Chapter 3 for the details of the experiment).

What Environments Support Fixed-Stopping Rules?

In environments where the relative search costs are high, stopping by discrimination may be counterproductive. Consider the following example. The payoff for a correct inference is $10, the cost of getting information about one cue, $4. The first cue did not discriminate. According to any stopping rule based on discrimination, one should now look up the second cue. But after the second cue is bought, the maximal win would be $2, whereas stopping search and guessing promises an expected value of $3. This leads to the following hypothesis:

> *Hypothesis K (one cue only):* The higher the relative information cost *I*, the more frequently the $m_f = 1$ rule is used.

Recall that in the second experiment reported by Newell et al. (2003), *I* was increased from 1/7 per cue in the first experiment (with six cues) to 1/5 per cue (two cues). Given these relatively high costs, 29% of their participants relied on the $M = 1$ rule and simply guessed whenever the first cue was not informative, and these participants did so consistently on *all* trials. The authors did not report the percentage of trials—computed across all participants—that was consistent with this one-trial stopping rule, but as 29% of participants used it on *all* trials, this percentage must be higher or at least equally high. Unfortunately the proportion of fixed-trial stopping in the first experiment is unknown due to incomplete response recording by the experimental program, so it is unclear whether the proportion increased from the first to the second experiment, as predicted by Hypothesis K. The data therefore only provide the indirect evidence that there indeed is a considerable proportion of participants who adopt one-trial stopping when information costs are high.

In the study by Läge et al. (2004), with a relatively low I = 1/10 (four-cue environment), in 5% of the decisions search was stopped after one or more cues without having found a discriminating cue. Unlike in Newell et al.'s Experiment 2, *none* of the participants used the M = 1 rule in all trials, or as the predominant stopping rule, which is consistent with Hypothesis K. Again, however, the comparison is across experiments. In the high-cost condition of the experiment reported later in Chapter 3, the tendency to stop search before a discriminating cue has been found does not substantially differ between high- and low-cost conditions (10% when I = 3/20, compared to 7% when I = 1/20).

Time pressure

Information search does not necessarily need to involve direct monetary costs to favor frugality. Time pressure, too, favors fast and frugal decision making. It limits the search time for information and should therefore increase the use of a stopping rule that ends search quickly.

> *Hypothesis L (indirect search costs):* The shorter the time available for making a decision, the more frequently one-reason stopping rules are used.

What does experimental data have to say about the effects of time pressure on human decision behavior? Rieskamp and Hoffrage (1999) presented participants with a choice between different companies from which they could buy shares. Participants either worked under low time pressure (50 sec for each choice), or high time pressure (20 sec). The authors tested many different strategies. Classification of participants according to their decisions revealed that under high time pressure, participants who used a generalized form of TTB represented the largest group (46%). Under low time pressure, participants who used a weighted additive strategy constituted the largest group (42%).

Sequential search without direct information costs

Is it possible to go even further, and postulate that just having to make the effort to search actively for cues will increase the use of one-reason deciding making? According to Gigerenzer and Todd (1999, p. 23), having to search for information, in either the external or internal environment, is a minimum requirement for being able to test models of ecological rationality, as they rely on search as a central component.

> *Hypothesis M (search excluded):* One-reason decision making is more frequent when cues have to be searched for than when search and stopping is prevented by the design of the experiment.

This seems to be true at least for internal environments. In the study by Bröder and Schiffer (2003) quoted above in the section on search rules, a large percentage of participants (on average 64%) could be classified as using TTB when cue values had to be retrieved from memory. As second experiment with the same stimulus material directly compared search in memory versus inferences from givens. Indeed, a higher percentage of participants predominately used TTB in the condition where information had to be searched for in memory (44% vs. 20%). Of course, when search proceeds in memory, process measures that indicate search order and stopping point cannot be recorded directly, which is why these numbers are based on decision outcomes.

However, more direct evidence for the use of different stopping rules could be gathered through reaction time differences. If people stop searching after having found one discriminating cue, one would expect reaction time differences between different kinds of decision pairs, for example, between one item [cue profile of alternative A: 10000; alternative B: 00000] and another [alternative A: 00001; alternative B: 00000].

With search in external environments, the situation looks different. Experiments by Bröder (2000) show that when information search was excluded and inferences had to be made from givens, only 28% of participants were classified as using TTB (Experiment 2), but when they had to search for information at some cost (I = 1/10 and 1/100), as in the share task, this number increased to 53% of participants (averaged across both cost conditions; Experiment 3). This result emerged from a comparison between experiments, and sequential search and information costs were confounded. When Bröder (2000, Experiment 4) directly tested the effect of successive versus simultaneous presentation of cues, he found that successive presentation alone (without direct information costs) did *not* increase the proportion of TTB users in his sample – it was 15% in both conditions. Only when cues had to be revealed successively *and* some monetary investment had to be made to get the information (I = 1/10) did the percentage of TTB users rise, to 65%.

An explanation for the divergent findings for search in memory versus search in the external environment without direct information costs might be that search in memory bears the additional load that the information of a discriminating cue has to be stored in working memory if search continues. This poses an additional cognitive challenge beyond the mere integration of information, which, in a sense, might also be considered as costs that have to be paid for information search. Yet there might be factors that can mitigate the demands of compensatory decision making even when information has to be retrieved from memory, as the effects of different presentation formats suggest.

Presentation format

Bröder and Schiffer (2003) have found a high proportion decisions consistent with TTB when information had to be retrieved from memory. In their first experiments, cited above, the information that had to be memorized before the decision-making task was in the form of verbal lists. This format might have favored the use of one-reason decision making. The combination of sequential search and one-reason stopping might be more likely when the information is coded in propositional rather than pictorial form. Propositional forms include verbal information transmitted aurally, or written text transmitted visually. Pictorial representations, such as photos and figures, may lead to more holistic encoding that makes it difficult to retrieve separate pieces of information sequentially, and thus also difficult to ignore some information.

> *Hypothesis N (information format):* One-reason decision making is more frequent when information is presented in propositional form than when information is pictorially presented.

Experimental evidence for Hypothesis N comes again from the studies by Bröder and Schiffer (2003). As reported earlier, participants had to make memory-based inferences about which of two suspects in a murder case was more likely to have committed the murder. In an additional study (Experiment 4), the authors tested whether the format of presentation of cue information influences the decision strategy that is used.

Participants either had to learn verbally presented cue profiles in the form of written lists or cue profiles presented in the form of pictures, that is, pictures of the suspects with their respective attributes. In the group that had memorized verbal cue profiles, participants whose decisions were predominately in line with TTB represented the largest proportion (64%, excluding participants who predominantly guessed). In contrast, the decisions of participants who had memorized pictorial information were best predicted by a compensatory strategy (weighted additive or equal weight rule): 69% of participants could be classified accordingly (again excluding guessers, and averaged across two pictorial conditions whose details do not matter for the current purpose).

These results are in line with the hypothesis that verbal information is more likely to be retrieved sequentially, thus suggesting one-reason stopping. Pictorial information might give rise to a more holistic representation that, when retrieved, provides information about several attributes simultaneously, thus suggesting information integration in a very similar way to simultaneously available cues in the external environment.

Discussion

Newell et al. (2003) questioned the falsifiability of the idea of an adaptive toolbox. From their finding that not all people obey Take The Best on all decisions, the authors conclude that "if the only way to explain these violations within the fast-and-frugal framework is through the post hoc invention of new heuristics or building blocks, then the framework begins to appear dangerously unfalsifiable" (p. 94). What the authors forgot to take into account, however, is that careful analysis of the environment allows very precise predictions both about decision outcomes and processes, making fast and frugal heuristics, in contrast to continuous parameterization, falsifiable with regard to both. It is worth noting that Newell et al. (2003) indeed did not undertake such a thorough environmental analysis to derive their hypotheses, although the idea of ecologically rational building blocks rests on the careful analysis of the environment at its very core.

Careful analysis of the environment does not imply an exhaustive cost–benefit analysis as in optimization under constraints. Ecologically rational building blocks are not based on a complete representation of the environment, nor derived from it. There is reduction to salient, relatively easy to perceive, and at the same time valid and robust features. The environmental features examined here are chosen to be very salient to the decision maker. Moreover, environmental features themselves often limit which building blocks could possibly be applied, thus simplifying the selection on the part of the decision maker. Additionally, stopping seems to drive deciding, thus constraining the range of decision rules available for selection.

With regard to search rules, for example, the absence of outcome feedback a priori rules out several search orders, thus limiting the focus of the decision maker on cases in which a cue discriminates. Also, it was suggested that the design of the learning phase strongly favors certain search rules over others. These predictions, along with the predictions about search in changing environments, remain to be tested thoroughly. Other environmental features might leave room for more building blocks to select from. But at least for information costs, taking discriminative power of cues into account seems an obvious way of responding to this feature. Indeed, there is accumulating evidence for adaptivity of search in response to (direct and indirect) information cost. People seem to be more selective in which cues they look up first when information is expensive, although the question of the exact ordering criterion – validity, or rather a criterion that combines both validity and discrimination rate, such as usefulness – remains open due to confounds in the respective experiments (Bröder, 2000; Bröder & Schiffer, 2003; Newell & Shanks, 2003; Newell et al., 2003; and even, to a lesser extent, Newell et al., 2004).

Similarly, on stopping and decision rules, there by now exists a range of studies confirming high proportions of one-reason stopping and deciding when one has to pay for information (Bröder, 2000, Experiments 3 & 4; Newell & Shanks, 2003; Newell et al., 2003), or when there are indirect costs of information search, resulting from time pressure

(Rieskamp & Hoffrage, 1999) or from the necessity to retrieve information from memory (Bröder & Schiffer, 2003). For other statistical properties of the decision environment, adaptive stopping has yet to be demonstrated. For example, adaptivity to the variability of cue validities would need to be addressed in a study that refrains from presenting all information simultaneously (as in Bröder, 2000, Experiment 2). The second chapter of this dissertation will be devoted to the property of information redundancy, an environmental feature thus far ignored in experimental studies on simple inference heuristics.

Whereas all these questions are about when a certain predefined building block will be applied, one might also ask how easily the rules can be constructed in the first place. Also here, the selection of rules that were considered in this chapter was motivated by simplicity.

For the suggested stopping rules, computational complexity seems to be less of an issue. Searching until one informative cue is found, and if you can afford it or you have experienced that it improves accuracy, looking for a second piece of evidence, does not require complex calculations.

Computationally complex rule-construction processes are more of an open question for the proposed search rules. But as especially the examples of search by usefulness and success show, there exist shortcuts that avoid complex computation and thus make these processes more plausible. The criteria of usefulness and success can be derived from simple tallies, and there are findings that people are very good at keeping track of the frequency of occurrence, or even do so automatically (Hasher & Zacks, 1984). Yet still, of course, how simple the construction of some of the rules really is remains a question for further experimental investigation. Especially search by validity, which normalizes the number of correct decisions against the number of discriminations and thus requires two tallies as well as computation of a division, might run into problems in terms of psychological plausibility. The third chapter of this dissertation will therefore take one step back and, for the building block of search, propose simple rules for deriving orders, to be tested in simulations and experiments.