

Chapter 2

Fundamentals on Decision-Making Behavior

This introductory chapter describes the fundamentals for later analysis, modeling and discussion of choice tasks and behavior. Figure 2.1 depicts the basic elements of the choice process which are relevant for the present work. On the left hand side, we see the general problem the decision makers are faced with: the choice task. Generally speaking, a choice task defines the problem of choosing the preferred out of a discrete and finite set of alternatives. The decision makers' preferences determine what the preferred alternative is. Thus, in Sect. 2.1, we define both choice tasks and preferences.

In recent decades, a large stream of research has shown empirically that choice behavior is contingent on the characteristics of both choice tasks and the decision makers (Einhorn and Hogarth 1981; Gigerenzer and Selten 2001; Payne et al. 1993). Following this body of literature, we assume that decision makers have a number of decision strategies in their mind and apply them contingent upon such factors such as the complexity of the choice task, which describes how difficult a choice task is for a decision maker (Payne et al. 1993). Hence, in Sect. 2.2, we provide a review of the different decision strategies which have been described in the literature so far.

In the subsequent Sect. 2.3, we then allude to different methods how researchers can observe and measure the decision process and determine the decision strategies which best explain the observed behavior and the final choice.

Finally, we provide a detailed literature review on different measures of complexity and the influence of complexity on decision-making behavior in Sect. 2.4.

2.1 Choice Tasks and Preferences

Choice tasks belong to the class of preferential decision problems. Preferential decision problems are characterized by the information elements which are given to the decision maker and by the goal statement of the decision (Payne et al. 1993). Concerning the information elements, preferential decision problems consist of three different components: (1) a number of alternatives, (2) events or contingencies

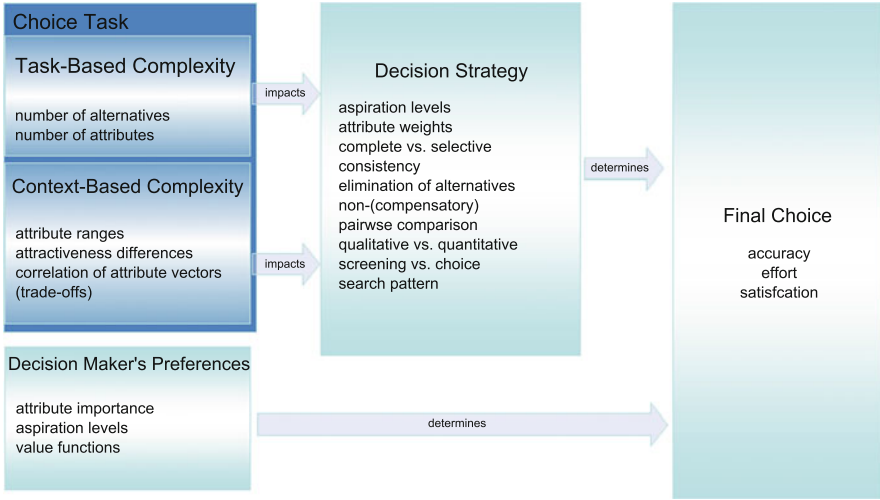


Fig. 2.1 The choice process with its central elements discussed in this chapter

that relate actions to outcomes and the probabilities for these events, and (3) values associated with outcomes (Payne et al. 1993, p. 20). Some information elements might be unknown. For instance, some alternatives might not be given but have to be generated by some optimization algorithm (Gettys et al. 1987; Keller and Ho 1988). The goal statement describes the outcome of the decision. Either the decision makers (1) select the most attractive alternative, or they (2) rank order the alternatives according to their attractiveness, or (3) they give a numerical value representing the attractiveness of every separate alternative.

In the present work, we focus on preferential decision problems in the context of purchase decisions. Following others, we refer to them as *choice tasks* (Bettman and Zins 1979). We define choice tasks as preferential decision problems with a limited number of alternatives where all information elements are known and there are no uncertainties. Thus, we do not deal with probabilities of events but assume that the outcome of alternatives is known. Furthermore, the goal of a choice task is to select the most attractive alternative, e.g., the product the consumer likes to purchase.

More formally, a choice task is a multi-alternative multi-attribute problem, which consists of n alternatives $alt_j, j = 1, \dots, n, n \geq 2$ which are described by a_{ij} attribute levels, one for each of the m attributes, $attr_i, i = 1, \dots, m$ (Harte and Koele 2001; Keeney and Raiffa 1993). Attribute levels are concrete occurrences of the attributes, where each attribute can have a different number of possible occurrences $|A_i|$. As an example, imagine a set of different cell phones (alternatives) which are described across different attributes, such as price, brand, and battery runtime. Each cell phone is specified by the three attribute levels it takes for each of the three attributes, such as €100, Samsung, and 48 h battery runtime for cell phone A or €150, Nokia, and 60 h battery runtime for cell phone B. For example, if

Table 2.1 Example of a product-comparison matrix with two alternatives and three attributes

Attribute	Cell phone A	Cell phone B
Price	100	150
Brand	Samsung	Nokia
Battery runtime	48 h	60 h

there are 10 different brands available, $|A_{brand}|$ would be equal to 10, as the attribute brand can take 10 possible levels. We assume that each alternative is described by the same attributes.

Typically, in the context of purchase decisions, choice tasks are displayed in form of *product-comparison matrices* (Häubl and Trifts 2000). A product-comparison matrix displays each alternative in a column. An example of such a product-comparison matrix is given in Table 2.1.

The space of attribute levels differs with respect to scales and ranges, which makes it hard to compare and trade-off alternatives against each other; e.g., while the price is measured as a ratio level in a currency, the color is described by words and induces only a nominal scale.¹ We assume that decision makers have preferences over attribute levels. These preferences can have a simple ordinal form “yellow is preferred over red”. Sometimes, it is further assumed that decision makers assign so-called attribute values to attribute levels (Eisenfuhr and Weber 2002). These attribute values reflect the degree of attractiveness the decision maker assigns to the attribute level. Each decision maker hence has m value functions, v_i , that assign attribute values to all available attribute levels, $v_i(a_{ij})$. Typical examples of value functions are displayed in Fig. 2.2. In this figure, v_{price} is a decreasing function, as the highest attribute value is assigned to the lowest price. However, value functions do not have to be monotonic, as is often the case for the attribute *size*. The figure shows an example, where the decision maker prefers a medium sized product to small or large ones. Finally, the right figures shows an example of the attribute color, which has nominal attribute levels.

2.2 Decision Strategies

“The aim of research on multi-attribute evaluation processes is to describe the process taking place between the presentation of the information and the final evaluation” (Harte and Koele 2001, p. 30). Many researchers have formalized common operations which describe people’s decision-making behavior and inferred decision strategies which describe the process of acquiring, evaluating, and comparing information elements of choice tasks (Beach 1990; Hogarth 1987; Payne et al.

¹Nominal scales use mere labels; ordinal scales indicate a relative position (e.g., graduations); interval scales indicate the magnitude of differences without a fixed zero point (degree Celsius), and ratio scales indicate the magnitude of differences and fix a zero point (e.g., size, price).

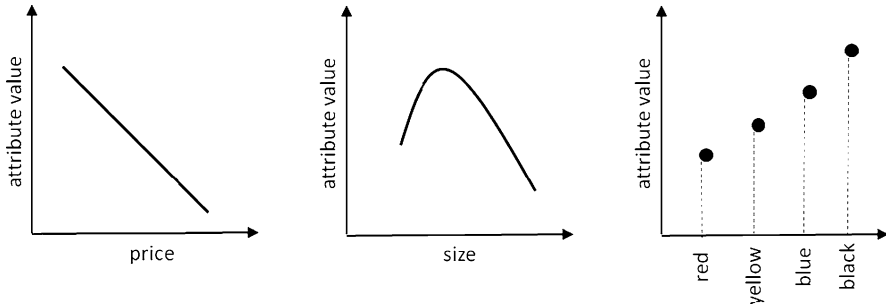


Fig. 2.2 Three examples of value functions

1993; Russo and Doshier 1983; Tversky 1969, 1972). Generally speaking, a decision strategy is defined as “a sequence of operations used to transform an initial stage of knowledge into a final goal state of knowledge in which the decision maker feels that the decision problem is solved” (Payne et al. 1992, p. 109). In our context, a decision strategy describes the process which decision makers follow when choosing an alternative in a choice task.

This section reviews decision strategies in detail, as they are the state-of-the-art for describing consumer purchase behavior. Yet, to gain a deep understanding of the common elements of and differences between decision strategies, we start by pointing out decision strategy characteristics which we retrieved from another literature review (e.g., Payne et al. 1993; Riedl et al. 2008).

2.2.1 Characteristics

2.2.1.1 Compensatory Versus Non-Compensatory

Decision strategies consist of compensatory and non-compensatory strategies. In compensatory strategies, a low value of an attribute of a product can be compensated by a high value on a different attribute. A typical statement of a consumer applying a compensatory strategy would be “I would prefer a cell with integrated navigation system, but not if I have to pay too much extra for that”. Thus, compensatory strategies allow for trade-offs among attributes. In contrast, non-compensatory strategies do not allow attribute values on one attribute to compensate for low values on another attribute. An example for a non-compensatory strategy is “I choose the cheapest product.”

2.2.1.2 Attribute Weights

Attribute weights, w_i , reflect how important each attribute is. Hence, they express how much each attribute influences a decision. For a decision maker with a high

weight for price, for instance, the product price is a very important attribute. While some strategies involve an explicit weighting of attributes and thus assume a ratio scale on their values, others require an ordinal scale, assume equal importance of all attributes, or assume no order at all.

2.2.1.3 Attribute-Wise Versus Alternative-Wise Processing

A strategy can induce an attribute-wise or an alternative-wise comparison. Attribute-wise behavior describes a decision maker who picks one attribute, compares its attribute levels across all alternatives, and then moves to the next attribute. Comparing all products first regarding their color and then regarding their price is an example of attribute-wise processing. Alternative-wise strategies, by contrast, assume the decision maker to sequentially evaluate alternatives regarding all or at least several of their attribute levels.

2.2.1.4 Pairwise Comparison

Some strategies compare alternatives sequentially in pairs. Usually they eliminate the weaker alternative of that pair and keep the stronger alternative for forming a new pair.

2.2.1.5 Aspiration Levels

In some strategies, attribute levels are compared to aspiration levels. Aspiration levels (or cutoff-values) can be interpreted as thresholds or acceptable levels. Formally, for each attribute level, a_{ij} , we can assign whether the aspiration level is violated, $asp(a_{ij}) = 1$, or not $asp(a_{ij}) = 0$. Dependent on the strategy, an alternative is either immediately excluded from further consideration once an aspiration level is not met or this attribute level is marked to be negative. An example of an aspiration level would be a maximum price we are at most willing to pay for a product.

2.2.1.6 Consistency Across Attributes/Alternatives

A consistent decision strategy evaluates the same number of attribute levels per attribute/alternative. For an inconsistent strategy, the number of attribute levels which is considered per attribute/alternative varies.

2.2.1.7 Complete Versus Selective

A complete decision strategy considers all attribute levels. Therefore, each complete decision strategy is by definition consistent. A selective strategy leaves out certain

attribute levels from consideration. This can be done in a consistent way (leaving out the same amount of attribute levels per alternative/attribute), or in an inconsistent way.

2.2.1.8 Elimination of Alternatives

Some decision strategies eliminate alternatives to narrow down the choice task until only the preferred alternative is left.

2.2.1.9 Qualitative Versus Quantitative Reasoning

Decision strategies can be distinguished by the way they evaluate attribute levels. Quantitative strategies require counting, adding, or multiplying, while qualitative ones evaluate by simply comparing attribute levels or values with one another.

2.2.1.10 Screening Versus Choice

This characteristic describes a tendency rather than a strict characteristic. Decision strategies differ in their goal to either narrow down a larger set of alternatives quickly in the screening process, or to choose one out of a small set of alternatives. The result of the screening phase is the consideration set which consists of only a few alternatives which the decision makers prefer most. Some of the strategies are more capable of building a consideration set than leading to a final choice. Although very similar at first glance, this characteristic is different to *elimination of alternatives*. While all screening decision strategies eliminate alternatives, choice decision strategies can either further eliminate alternatives or just choose an alternative out of the whole consideration set without explicit elimination of inferior ones.

2.2.2 Types

In the following paragraphs, we describe different strategies in detail. By means of an analysis of the literature, we have identified fifteen decision strategies. Therefore, our summary extends current overviews (Payne et al. 1993; Riedl et al. 2008). Moreover, in contrast to current overviews, we aim at a more formal description.

Table A.1 in the appendix lists all strategies, the main literature sources, as well as alternative name conventions of the strategies. In addition, Fig. A.1 gives an overview of the different strategies and their characteristics.

1. **EQW (Equal Weight Heuristic):** Decision makers select the alternative with highest utility, ($\max_{j=1,\dots,n} [u(alt_j) = \sum_{i=1}^m v_i(a_{ij})]$). Each attribute is assumed

to be of equal importance to decision makers. Thus, the utility of an alternative is defined as the sum of all attribute values.

2. **WADD (Weighted Additive Rule):** The normative rule in the decision making literature. It assumes that a decision maker computes a utility for each alternative. WADD defines the utility function as the sum of the weighted attribute values ($\max_{j=1,\dots,n} [u(alt_j) = \sum_{i=1}^m w_i v_i(a_{ij})]$).
3. **ADD(Additive Difference Strategy):** When using this strategy, a decision maker iteratively performs pairwise comparisons of alternatives until only one candidate is left. The decision maker computes the utility difference between two alternatives as $diff(alt_k, alt_l) = \sum_{i=1}^m w_i [v_i(a_{ik}) - v_i(a_{il})]$. If $diff(alt_k, alt_l) > 0$, alt_l is eliminated and alt_k is compared to the next alternative. If $diff(alt_k, alt_l) < 0$, alt_k is eliminated.
4. **MCD (Majority of Confirming Dimensions Heuristic):** Analogous to ADD, a decision maker compares alternatives pairwise. In contrast to ADD, however, decision makers do not assign utility values, rather they decide whether they prefer a_{ik} over a_{il} . The difference of two alternatives is computed as: $diff(alt_k, alt_l) = \sum_{i=1}^m D(a_{ik}, a_{il})$, where $D(a_{ik}, a_{il}) = 1$ if $v_i(a_{ik}) > v_i(a_{il})$, $D(a_{ik}, a_{il}) = -1$ if $v_i(a_{ik}) < v_i(a_{il})$, and 0 otherwise.
5. **FRQ (Frequency of Good and/or Bad Features Heuristic):** The decision maker distinguishes between good, neutral, and bad attribute values. For this purpose, $frq(a_{ij}) = 1$ if a_{ij} has a desired attribute level, $frq(a_{ij}) = 0$ if the attribute level of a_{ij} has no influence on the decision maker's choice decision, and $frq(a_{ij}) = -1$ if the attribute level of a_{ij} is not attractive for the decision maker. The literature describes three variants of FRQ. A decision maker can choose the alternative (a) with the highest number of good attribute levels, (b) with the lowest number of bad attribute levels, or (c) a decision maker could consider both good and bad attribute levels ($\max_{j=1,\dots,n} [\sum_{i=1}^m frq(a_{ij})]$).
6. **COM (Compatibility Test):** An alternative alt_j is eliminated if its attribute values violate the corresponding aspiration levels more than k times, where k is specified by the decision maker: $\sum_{i=1}^m asp(a_{ij}) > k$.
7. **CONJ (Conjunctive Strategy):** A decision maker removes an alternative alt_j if at least one of its attribute values violates an aspiration level: $\sum_{i=1}^m asp(a_{ij}) \geq 1$.
8. **SAT (Satisficing Heuristic):** Decision makers sequentially consider alternatives in the order in which they occur in the choice task. They select the first alternative where all attribute levels are above the corresponding aspiration levels ($\forall i asp(a_{ij}) = 0$). In contrast to CONJ, which assumes the decision makers to evaluate every alternative, SAT stops as soon as one alternative which meets all aspiration levels is identified.
9. **SAT+ (Satisficing-Plus Strategy):** A variant of SAT. Decision makers consider only m^* attributes. They select the alternative where all m^* attribute levels are above the corresponding aspiration levels.
10. **DIS (Disjunctive Strategy):** The decision makers remove an alternative alt_j if all of its attribute values violate an aspiration level:

$$\sum_{i=1}^m asp(a_{ij}) = m.$$

11. **DOM (Dominance Strategy):** A decision maker chooses the alternative that dominates all other alternatives. An alternative alt_k dominates alt_l if all attribute levels are at least as good and at least one attribute level is better : $\forall i \ v_i(a_{ik}) \geq v_i(a_{il}) \wedge \exists i \ v_i(a_{ik}) > v_i(a_{il})$. If no alternative dominates all other alternatives, the decision maker chooses no alternative.
12. **MAJ (Simple Majority Decision Rule):** Same as DOM – but the decision maker always chooses an alternative. The decision maker chooses the alternative whose attribute levels are best on the highest number of attributes.
13. **EBA (Elimination by Aspect Strategy):** Decision makers sort the attributes $attr_i$ according to their weight w_i . Starting with the attribute with the highest weight, they iteratively remove alternatives alt_j if the value of the i th attribute does not meet the aspiration level ($asp(a_{ij}) = 1$). The strategy stops if there is only one alternative left or all attributes are considered. In the original version, EBA considers attributes not deterministically but probabilistically according to attribute weights (Tversky 1972).
14. **LEX (Lexicographic Heuristic):** Decision makers consider the attribute $attr_h$ with the highest attribute weight and select the alternative alt_j whose attribute value has the highest value: $\max_{j=1,\dots,n} v_h(a_{hj})$. If this returns more than one alternative, they iteratively compare the remaining alternative across the next most important attribute until there is only one alternative left.
15. **LED (Minimum Difference Lexicographic Rule):** This rule selects an alternative analogous to LEX, but two attribute values a_{ij} and a_{ik} are considered to be equal if $v_i(a_{ij}) - v_i(a_{ik}) < \Delta_i$, where Δ_i is the threshold above which a decision maker notices a difference between two attribute values (Luce 1956; Tversky 1969).

2.3 Measuring Decision-Making Behavior

We aim at explaining and understanding observed decision-making behavior and hence follow a descriptive approach. There are basically two descriptive research paradigms for studying choice behavior. The outcome-based approach fits the mathematical models of the relation between attribute values and evaluations of alternatives (Harte and Koele 2001; Rieskamp and Hoffrage 1999). Structural modeling and comparative model fitting are the two representatives of this paradigm. The second research paradigm is process-oriented and focuses on the sequence of information acquisition and evaluation steps. Representatives of process tracing techniques are verbal protocols, Mouselab, and eye tracking which record the attribute levels observed, the length of observation time as well as the sequence in which they are looked at (Einhorn and Hogarth 1981).

2.3.1 Outcome-Based Approach

2.3.1.1 Structural Modeling

Structural modeling assumes that decision makers maximize their utility. The structural modeling approach then predominantly uses regression analysis to find the parameters which achieve the best fit given the assigned values or the final choice and a predefined utility model (Brehmer 1994). Usually, the assumed utility model is the weighted additive rule (WADD, Sect. 2.2.2), which is fitted to the data by linear regression. WADD is a very robust model and currently provides the best fit compared to competing utility models (Dawes and Corrigan 1974; Dawes 1979; Harte and Koele 2001).

Since the full range of decision strategies is not covered and only parameters of utility maximizing models are estimated, researchers try to approximate non-compensatory decision making from the parameters (Swait and Adamowicz 2001). If, for instance, one attribute has a much higher weight than others, one could infer that this attribute predominantly determines the decision, speaking in favor of a non-compensatory strategy.

The basic assumption of structural modeling approaches is that the overall utility value $u(alt_j)$ of an alternative j can be decomposed into two additively separable parts, (1) a deterministic component $d(alt_j)$, and (2) a stochastic component – the error term ϵ_j – representing, for instance, unobserved attributes affecting choice: $u(alt_j) = d(alt_j) + \epsilon_j$. The deterministic component $d(alt_j)$ is assumed to have an additively separable linear form. Given m attributes with each $|A_i|$ possible attribute levels it can be written as

$$d(alt_j) = \sum_{i=1}^m \sum_{k=1}^{|A_i|} \beta_{ik} X_{jik} \quad (2.1)$$

where X_{jik} is a binary variable indicating whether the alternative j contains the occurrence k of attribute i . β_{ik} is the part-worth utility of occurrence k of $attr_i$. The part-worth utility reflects with which value each occurrence contributes to the utility of an alternative. In case of WADD, which assumes the weighted linear additive utility function, for instance, if the occurrence k is included in alt_j , $X_{jik} = 1$, $\beta_{ik} = w_i v_i(a_{ij})$. In case of EQW, it would be $\beta_{ik} = v_i(a_{ij})$. The part-worth utilities for each respondent can be estimated via a Multinomial Logit model or a Hierarchical Bayes approach (Lenk et al. 1996).

The range of part-worths for an attribute i can be used to evaluate the importance of this attribute (Verlegh et al. 2002). If, for instance, the part-worths for price have a high range, this indicates that prices might be very important to the individual. While, if part-worths of, for instance, color, differ only slightly for the different colors, this might indicate that the individual does not care much about color. Hence, we can compute attribute weights from part-worths and $\sum_{i=1}^m \sum_{k=1}^{|A_i|} \beta_{ik} X_{jik} = \sum_{i=1}^m \sum_{j=1}^n w_i v_i(a_{ij})$.

Although WADD explains large portions of the final decisions, many researchers have examined that WADD is remote from describing the actual decision process. First, WADD assumes no interaction among attribute values. This means that attribute values contribute to the utility of an alternative independently from other attribute values of the same or other alternatives. Yet people report that their value functions depend on what other alternatives are available (Ford et al. 1989; Payne 1976). Second, weighting and summing up the attribute values of all alternatives does not describe the actual process decision makers follow (Maule and Svenson 1993). This loose link between the assumed utility model and the observed behavior is the main problem of structural modeling approaches. Bröder and Schiffer (2003a) argue that structural modeling approaches can only reveal information about the process “in case there is a clear formal link between theoretical descriptions of these processes and the expectations about decision outcomes” (p. 195). They point out that using regression to estimate, for instance, the WADD strategy is lacking this formal link. Consequently, WADD is less a model to describe human decision making but rather a kind of a general-purpose analytical tool to make general conclusions. For further discussion on structural modeling and a description of the analysis of variance as another method for fitting mathematical models, see Harte and Koele (2001).

2.3.1.2 Comparative Model Fitting

Bröder and Schiffer (2003a) recently introduced the term *comparative model fitting* for an approach which gains increasing interest. In comparative model fitting, several models compete for the best explanatory power for a given data set (Dhami and Ayton 2001; Garcia-Retamero et al. 2007; Gilbride and Allenby 2004; Hoffrage et al. 2000). Parameters of the competing models are defined a priori including appropriate error models, and maximum likelihood methods are used to compare final choices with the predictions of the models (Glöckner 2009). Thus, in contrast to structural modeling approaches, different decision strategies and not just a utility-maximizing strategy are assumed. Furthermore, the final choice is sufficient for comparing the explanatory power of competing models.

Bröder and Schiffer (2003a), for instance, use a Bayesian method to assess whether a simplified variant of LEX, EQW, or WADD provide the best likelihood function for explaining observed decision-making behavior. If for an observed decision, the likelihood of one strategy exceeds the likelihood of the other two, it is classified accordingly. Furthermore, the authors assume that respondents make errors when applying one or the other strategy and they incorporate a uniform distribution for modeling this error.

Recently, also other researchers have developed approaches for identifying non-compensatory strategies, such as the conjunctive and disjunctive strategies and elimination by aspects (Gilbride and Allenby 2006; Hauser et al. 2010; Kohli and Jedidi 2007).

According to Cutting (2000) the problems of comparative model fitting are the specification of measures a priori and an equally balanced number of free parameters per model to ensure that each model provides the same degree of flexibility. Furthermore, comparative model fitting still has the same problem a structural modeling approach has: if strategies make the same predictions, then we cannot distinguish between different strategies (Glöckner 2009; Rieskamp and Hoffrage 1999). Unfortunately, it is not unusual that strategies explain the same choices (Batley and Daley 2006; Bröder 2000; Bröder and Schiffer 2003b; Lee and Cummins 2004; Louviere and Meyer 2007; Rieskamp and Hoffrage 1999, 2008). Thus, the authors propose to use additional measures, in particular decision times and confidence judgment in order to differentiate between strategies.

2.3.2 Process Tracing

In contrast to outcome-based approaches, where parameters of decision strategies are estimated such that they best explain the final choice, in process tracing, patterns of information acquisition are observed. Thus, process tracing methods should be able to detect which information is acquired when and for how long. From these patterns, many of the characteristics which have been described in Sect. 2.2.1 can be identified. Process tracing techniques can, for instance, distinguish whether the decision maker acquires information selectively and consistently. These characteristics, in turn, suggest that certain strategies were applied by the decision maker (Payne 1976). Since some characteristics cannot yet be identified with process tracing techniques, often only subgroups of strategies but not the exact strategy can be determined. When we observe a complete, consistent and alternative-wise information acquisition process, for instance, WADD, EQW, FRQ might have been used (see Table A.1 for a listing of decision strategies and their characteristics). In the following section, we describe the three most prominent techniques for process tracing: verbal protocols, Mouselab and eye tracking.

2.3.2.1 Verbal Protocols

Verbal protocols can be recorded simultaneously or retrospectively, depending on whether the decision makers describe their behavior during the search process or thereafter. Retrospective processing is criticized for a potential lack of validity, as the decision maker might describe the decision-making behavior based on their own psychological theory rather than an accurate description of the own cognitive steps during the search (Fidler 1983; Svenson 1989). The more common approach is therefore to ask decision makers to think aloud simultaneously to the decision process (Ball et al. 1998; Bettman and Park 1980; Biggs et al. 1985; Nisbett and Wilson 1977; Selart et al. 1998; Todd and Benbasat 2000).

An analysis of the validity of the simultaneous approach of thinking aloud showed that although it tends to slow down the process, it does not seem to change the sequence of thoughts during the process (Ericsson and Simon 1980, 1993). Furthermore, as long as the respondent only reports contents of the short-term memory which are simple to verbalize, the protocol reflects the decision making sufficiently. However, Russo et al. (1989) criticize both simultaneous and retrospective verbal protocols. They argue that simultaneous protocols might interfere with the decision task and found that in retrospective protocols respondents did forget or not report truly what had actually happened.

Besides a potential lack of validity, several other disadvantages have caused a more and more rare usage of verbal protocols. First, transcribing and coding of statements is time-consuming (Reisen et al. 2008). Now, much more sophisticated and automatized techniques are available due to the advancement in information technology in recent years (see Mouselab and eye tracking). Furthermore, coding is prone to biases since sometimes it is unclear which model to fit to the verbal statement. Third, verbal protocols are usually incomplete and do not fully cover the thinking process (Somerén et al. 1994). Someren et al. (1994) remark, “occasionally protocols contain ‘holes’ of which it is almost certain that an intermediate thought occurred here” (p. 33).

2.3.2.2 Mouselab and Information Display Boards

Mouselab is currently the predominant method used in research and is an improved, automatized version of an older method, the so-called Information Display Boards (Bettman et al. 1990; Payne 1976; Payne et al. 1993; Reisen et al. 2008). When using Information Display Boards, decision makers have to manually pull cards out of envelopes to retrieve the attribute level information and put them back in the appropriate envelope afterwards. Thus, at most one attribute level is uncovered at any one time. Besides a higher effort for the decision makers to acquire information on the attribute levels, data observation in Information Display Boards is tedious, as data is recorded just by observing decision makers’ behavior and there is no automated way of data collection.

Mouselab is a software that keeps track of information acquisition, response time, and choices by recording mouse movements in a product-comparison matrix on a computer screen (Bettman et al. 1990). At the beginning of the choice task, all attribute levels are hidden behind boxes. Only by clicking or moving the cursor on each box can the respondent retrieve the attribute level. In the original version, the box is hidden again once the cursor moves away (Payne et al. 1988).

Mouselab covers the decision process more completely than verbal protocols but because of the complete coverage, we no longer can distinguish the pure information acquisition stage from the cognitive evaluation stage (Svenson 1979). In a newer version of Mouselab, the so-called MouseTrace, the authors try to overcome these deficits (Jasper and Shapiro 2002). MouseTrace incorporates the notion of decision stages by allowing the software to record different process measures for different

time periods. However, as the number of stages as well as the number of mouse clicks, movements, or choices per stage must be defined in advance, a certain structure of the process is assumed which might not reflect the actual way of decision making at all.

Besides the problem of ignoring the difference between pure information acquisition and the evaluation and comparison of alternatives which neither Mouselab nor eye tracking can resolve currently, a main disadvantage of Mouselab is the potential lack of external validity as it might bias the decision process. First, moving the mouse for turning cards causes additional effort and might therefore keep the decision makers from considering certain attribute levels. Second, a matrix of hidden information does not reflect the situation consumers are faced with in online webstores and is therefore not appropriate for analyzing online purchase behavior. For an overview on 45 studies which either used Information Display Boards or verbal protocols, see Ford et al. (1989).

2.3.2.3 Eye Tracking

Due to the increased speed of computer processors and computer vision techniques as well as reduced costs for the required hardware equipment, eye tracking has become an alternative way to keep record of the decision process (Duchowski 2007; Lohse and Johnson 1996; Reisen et al. 2008). In the field of human computer interaction, Nielsen (1993) was one of the first to provide the idea that eye trackers could be used to infer the users' intentions by reading their eye movements. Eye tracking is the process of measuring eye gazes or eye movements with eye tracker systems which differ with respect to the different video techniques they apply (Duchowski 2007). In desktop-based human computer interaction, for instance, the currently applied systems are head-mounted, stationary eye trackers (Pfeiffer 2010). They consist of two units. The user wears a camera unit, while at the PC a computer-vision system is installed.

In contrast to Mouselab or its variants, with eye trackers it is unnecessary to hide information since the eye tracker system is able to precisely record fixations on attribute levels or other pieces of information (Lohse and Johnson 1996; Reisen et al. 2008). It keeps track of not only the exact position and sequence of the fixations, but also of the length of fixations. There is a pre-specified time threshold for eye fixations which must be met in order to interpret the fixation as actual information acquisition.

The lack of external validity of Mouselab is addressed by eye tracking as it allows to capture information acquisition on the level of the decision maker's visual system and creates a situation which is closer to a natural purchasing process than other techniques do (Russo 1978; Russo and Doshier 1983; Russo and Leclerc 1994). Others even argue that eye tracking does not affect information acquisition costs at all if compared to natural purchasing decisions (Lohse and Johnson 1996). Svenson (1979) found that if the amount of information is too high (i.e., the number of alternatives and attributes is large), process tracing is imprecise. However, in our

days, eye tracking has become a precise method for studying user behavior and interface design. Already in 1996, Lohse and Johnson pointed out:

It is also important to mention that the discussion of eye tracking equipment in the literature [...] is based on equipment that is over 20 years old. Just as computers have rapidly increased in power and performance, eye tracking equipment has also improved significantly. Old eye equipment had limited precision and accuracy for detecting small regions on a display. Current eye tracking systems are able to detect regions as small as 1.5 square centimeters. [...] Thus, it seems time to reevaluate the importance of eye tracking equipment as a tool for process tracing studies (p. 42).

Now, in 2010 eye tracking is able to detect regions as small as 0.38 cm^2 .² Typically, current systems operate in a spatial volume of 30 cm^3 around the screen and sample with rates around 600 Hz.

Lohse and Johnson (1996) were the first to address the question whether Mouselab and eye tracking techniques influence the decision process. In their study, when using eye tracking, respondents needed less time for the decision, looked more often at attribute levels already looked at before (reacquisition), but the proportion of cells accessed at least once (breadth of search) was lower than when using Mouselab. Furthermore, when using eye tracking, respondents had a more inconsistent search and they looked more attribute-wise. Reisen et al. (2008) did a comparable study and confirmed that when using eye tracking, respondents needed less time, did more reacquisitions, accessed more cells in total and had a more inconsistent search. However, they contradict some other findings by Lohse and Johnson (1996). First, the percentage of attribute levels which were considered at all was not significantly different from Mouselab. Second, they did not find any difference in the the sequence in which information was acquired. Under both conditions, respondents compared alternatives across their attributes. Because of the contradicting results, the two research groups draw different conclusions. While Lohse and Johnson (1996) concluded that the process tracing method does influence the decision process, Reisen et al. (2008) concluded the opposite.

Negative effects stemming from Mouselab were found by Dieckmann et al. (2009): They argue that decisions can be less effective when measured with Mouselab since moving the mouse causes additional cognitive effort. Hence, less effort is spent on the information acquisition itself. A consequence might be that different decision strategies are used when Mouselab instead of eye tracking is chosen for process measurement.

In a recent study, we compared Mouselab and eye tracking to shed more light on the discussion. We did not only analyze whether the process tracing method influences the decision-making process but also whether it influences the final choice (Meißner et al. 2010). In line with Lohse and Johnson (1996) and Reisen et al. (2008), respondents needed more time in the Mouselab condition, but in contradiction to both studies, they acquired the same amount of information as in the

²The SMI EyeLink II eyetracker system used in the present work, for instance, has a gaze position accuracy of 0.5° and respondents posit themselves in 40–60 cm distance from the screen ($\tan(0.5^\circ) \times 40 \text{ cm} = 0.35 \text{ cm}$). A 0.35 cm radius forms a circle of 0.38 cm^2 .

eye tracking condition. We confirmed all other results by Lohse and Johnson (1996) in that when using eye tracking the respondents had a more inconsistent and a more attribute-wise search. The difference for the search pattern was highly significant, and in total in Mouselab respondents searched more alternative-wise than with eye tracking. Yet, in the eye tracking conditions, respondents switched from a more attribute-wise search to a more and more alternative-wise search, the more tasks they had already completed. In summary, we agree with Lohse and Johnson (1996) that the process tracing method influences the process, in particular since both the consistency of search and the search pattern differed in the Mouselab vs. the eye tracking condition. However, when comparing final choices, we found no significant difference with respect to the estimated preference model. We estimated the linear additive model (WADD) with a structural modeling approach based on the choices respondents had made and found no differences in the estimated utility functions. Thus, although the process tracing method influences the decision process, the final choices were similar.

In summary, we argue along the lines of Lohse and Johnson (1996), who claim that the relevance of the difference in the decision-making process depends on the nature of the research question. Because of the differences, a pure descriptive research has its limits since observed information acquisition patterns are dependent on the used measurement instrument and, thus, generalizations should be made with caution. However, as soon as the effect of independent variables on the decision process is tested, the process measuring technique only affects the result if there is an interaction between manipulation and measuring technique. If there is no such interaction, the process measuring technique may only increase or decrease the main effect and therefore should not influence the nature of the research conclusion. Nevertheless, such interaction effects can hardly be anticipated in advance. We therefore suggest that as long as resources for eye tracking are available, eye tracking should always be the preferred method as it is able to record the decision process more precisely (Norman and Schulte-Mecklenbeck 2010; Seth et al. 2008).

2.3.2.4 Outcome-Based Versus Process Tracing

In this section, we have discussed the differences between two methods for determining peoples' decision-making behavior: outcome-based and process tracing methods. While the outcome-based approach estimates parameters of decision strategies based on observed final choices, the process tracing approaches, Mouselab and eyetracking, observe the sequence and the amount in which people retrieve information. Thus, they measure information acquisition. Verbal protocols, in contrast, also may provide information about what information the participants actually process.

The question remains whether, in general, outcome-based or process tracing methods should be applied when analyzing decision-making behavior. Since outcome-based and process tracing approaches focus on different aspects of decision making, we argue that they can complement one another well. There are

only few studies which have used such a multi-method approach (Biggs et al. 1985; Lohse and Johnson 1996; Rieskamp and Hoffrage 1999, 2008). Among them Covey and Lovie (1998) conclude, “(...) a skilfully combined approach can help to draw up a more complete picture which could not be so easily achieved by either method on their own” (p. 33).

Yet, because of cost and time restrictions experimental researchers might not have the resources (time, money, technical possibilities, etc.) to use both approaches. In general, we share the belief that novel steps in the study on individual decision-making behavior should first start with the observation of final choices (a widespread procedure in decision theory and economics, see e.g., Bröder and Schiffer 2003a; Caplin and Dean 2010). One reason is that one has to take into account that process tracing techniques are only able to record acquired information, which does not necessarily mean that all this information is also processed and evaluated (Payne et al. 1978). Second, the insights almost always relate to general categories such as overall alternative-wise or attribute-wise information acquisition, rather than to the level of the particular strategy used. Thus, it is hard to determine the exact decision strategy which has been applied just based on process data (Bröder and Schiffer 2003a). Nevertheless, as several researchers have shown, when applying the outcome-based approach, competing models might lead to the same choices, although they rely on different decision-making processes (Batley and Daley 2006; Bröder 2000; Bröder and Schiffer 2003b; Lee and Cummins 2004; Louviere and Meyer 2007). In these cases, process tracing methods are obligatory to learn about how decision makers actually behave.³ Adomavicius and Tuzhilin (2005) point out that in these cases, one should refrain from only using outcome-based approaches which serve as “as if” models because they are only judged by the degree to which they predict behavior, and not by the degree to which they actually explain real decision-making behavior.

For studying consumer purchase decisions, predicting market shares is the most important goal. Therefore, outcome-based based methods seem to be most appropriate. On the other hand, for studying human decision-making behavior and understanding cognitive processes from a psychological perspective, process tracing methods seem to be most appropriate, as they reveal more details of the decision process.

2.4 Complexity of Choice Tasks

2.4.1 *Task-Based Versus Context-Based Complexity*

Faced with a choice task, previous research has found that besides personal characteristics (e.g., cognitive ability and prior knowledge) and the social context (e.g., accountability and group membership), certain problem characteristics

³We will present a novel approach, which addresses this problem, in Chap. 4.

influence decision-making behavior (Ford et al. 1989; Payne et al. 1993). Problem characteristics influence how difficult the choice task itself is for the decision maker and how much effort the decision maker needs for the decision. Similar to Payne et al. (1993), among the problem characteristics, we distinguish between *task-based complexity* and *context-based complexity* effects. We define task effects to be general aspects of the decision task which are independent of attribute levels and from the decision maker's preferences while context effects depend on particular attribute levels and often depend on decision makers' preferences.

In this section, we provide a literature review of empirical studies on the influence of task-based complexity and context-based complexity on decision-making behavior. Since there are many different task and context effects and their influence has been tested on many different aspects of decision-making behavior, we have structured this section into four parts. In the first part, we explain different task and context effects. In Sect. 2.4.2, we describe the variables which are used to measure the effect on decision-making behavior. Thus, from a methodological perspective, in the first section, we discuss the independent variables and in the second part, the dependent variables of the experimental studies which follow in the third and the fourth part. The third part summarizes studies on task-based complexity and the fourth part studies on context-based complexity. In addition to that, we provide a detailed summary of the reviewed literature in the appendix in Table . This table resumes the experimental design, process tracing methods, independent and dependent variables, and results.

2.4.1.1 Task-Based Complexity

As defined above, task-based complexity describes general aspects of choice tasks such as the number of alternatives and attributes (amount of information), time pressure, information display (e.g., sequential vs. simultaneous presentation of alternatives), and the response mode (e.g., selecting one alternative vs. stating a rank order of alternatives). There is a huge body of literature on the influences of these different task effects on decision-making behavior (Bettman et al. 1991; Conlon et al. 2001; Ford et al. 1989; Payne et al. 1992). Hence, it is beyond the scope of the present work to give a complete review of all different kinds of task-based complexity. Therefore, we need to focus on those aspects which are relevant for the overall goal of the present work: the design of a DSS for webstores which supports consumers' decision-making behavior. Because of the anonymity in the Internet, operators of online webstores have limited knowledge about some aspects of the choice task environment of the online shoppers. For instance, they do not know whether the online shopper is under time pressure. Other aspects are pre-specified by the setting of an online webstore (e.g., is the response mode usually the selection of one product rather than a rank order of products). Hence, from our viewpoint, the most interesting factor is the amount of information displayed, that is, the number of alternatives and attributes. For details on other effects which are not addressed in the present work, the reader is referred to Ford et al. (1989) and Payne et al. (1993).

As the number of alternatives and attributes increases, people use simplifying decision strategies which neglect information rather than more effortful compensatory strategies (Bettman et al. 1991; Conlon et al. 2001; Ford et al. 1989; Payne et al. 1992). An increasing number of alternatives and attributes, for example, makes a decision more difficult since the decision maker has to consider more information for making a choice (Biggs et al. 1985; Payne et al. 1993; Timmermans 1993) and experiences information overload.

2.4.1.2 Context-Based Complexity

Concerning task effects, we focus on the number of attributes and alternatives. But what information is actually shown about the products? Context effects deal with that question. They concern the particular attribute levels, such as the price of each product (e.g., low, medium, high price) and their relationship to one another. Moreover, they might also require knowledge on the decision maker's preferences if they take into account not only attribute levels but also attribute values or weights (see Sect. 2.1).

Payne et al. (1993) provide an overview of different context factors: *Framing effects*, *the quality of the alternative set*, and *similarity*. Among them, *framing effects* and *the quality of the alternative set* are less relevant in the context of online purchase decisions. The quality of the alternative set plays a role in risky decisions, for instance decisions between gambles, and are thus irrelevant for the context of purchase decisions in non-risky environments. Framing effects usually refer to wording effects, such as whether choice tasks are formulated as gains or losses. In one famous study, Tversky and Kahneman (1981) asked subjects to choose between two alternative programs to combat an outbreak of an Asian disease which is expected to kill 600 people. In the one condition, alternatives are formulated in positive wording, for instance, "the program will result in 200 people being saved", in the other condition the wording was negative "the program will result in 400 dead people". Even though both alternatives resulted in the same number of deaths, subjects decided for different alternatives, depending on the wording. Although, framing effects are very relevant and an interesting research area, from our viewpoint, they are less relevant for online purchase decisions. We will examine choice tasks where alternatives (products) are described by the technical details where the wording plays a minor role and there is little flexibility to frame the details in either negative or positive wording.

Hence, we will now focus on the remaining factor, the *similarity of alternatives*. Consider two situations: one in which the attribute values of prices, designs, and brands are similar, and another in which several attributes vary sharply. The general observation is that the more similar alternatives are, the harder the choice task is (higher complexity). One reason might be that the more similar the values, the harder it is for the decision maker to eliminate one of the alternatives based on only few attribute-wise comparisons (White and Hoffrage 2009).

According to Payne et al. (1993), similarity of alternatives can be measured in different ways, namely *dominance structures*, *trade-offs* reflecting the degree of conflicting alternatives, *attribute ranges*, and *attractiveness differences*. It is rather unlikely that there is a dominant alternative in the context of purchase decisions, as a lower price usually compensates low values on other attributes and vice versa (Fasolo et al. 2003). It is therefore of little interest to investigate the influence of a dominant alternative for the present work. Thus, we will consider only *trade-offs*, *attribute ranges* and *attractiveness differences*. For details on dominance structures, the framing effect and the quality of the alternative set, the reader is referred to Klein and Yadav (1989), Payne et al. (1993), Garbarino and Edell (1997), and Swait and Adamowicz (2001).

A *trade-off* occurs when there are two different attributes, $attr_l$, $attr_k$, and two different alternatives alt_p , alt_q , where alt_p is better for $attr_l$ and worse for $attr_k$ than alt_q : $v_l(a_{lp}) > v_l(a_{lq})$ and $v_k(a_{kp}) < v_k(a_{kq})$. Each trade-off requires the decision maker to balance one against another attribute level and hence increases complexity (Payne et al. 1993). Trade-offs are measured by computing the average pairwise correlation of attribute vectors (Fasolo et al. 2009; Luce et al. 1997), where the attribute vector of $attr_l$ consists of attribute levels and is defined as: $attr_l = (a_{l1}, a_{l2}, \dots, a_{ln})$. When attribute vectors are correlated positively, one alternative beats the other alternatives on most attributes. Table 2.2 shows two examples with three alternatives each, $n = 3$, and three attributes, $m = 3$. Instead of attribute levels, we display their attribute values, $v_i(a_{ij}) \in [0, 1]$ for a fictive decision maker, so the larger $v_i(a_{ij})$ the better is a_{ij} for the decision maker. The price is best (0.8) for alternative 1 and gets worse for alternative 2 (0.3) and for alternative 3 (0.2) (see Table 2.2a). The two attribute vectors price and size are correlated positively ($CORR((0.8, 0.3, 0.2), (0.6, 0.5, 0.4)) = 0.93$). The same holds for the other pairs of attribute vectors, price and weight, as well as size and weight. The average correlation of all three pairs of attribute vectors is 0.9. Thus, the choice task has a high positive correlation and few trade-offs (in this case, there are even no trade-offs). In Table 2.2b price and size have a negative correlation ($CORR((0.8, 0.3, 0.2), (0.4, 0.7, 0.9)) = -0.97$). Price and weight also have a negative correlation of -0.36 , while weight and size have a slightly positive correlation of 0.11. On average this choice task is correlated negatively (-0.4). This negative correlation indicates strong trade-offs.

Table 2.2 Example of positive and negative correlation of attribute vectors

Attribute	alt_1	alt_2	alt_3
(a) Positive correlation of attribute vectors			
Price	0.8	0.3	0.2
Size	0.6	0.5	0.4
Weight	0.7	0.6	0.3
(b) Negative correlation of attribute vectors			
Price	0.8	0.3	0.2
Size	0.4	0.7	0.9
Weight	0.3	0.5	0.3

In consumer decision making, DM often have to trade off attribute levels against each other because they generally want low prices but high quality (Fasolo et al. 2005). Often, decision makers avoid dealing with trade-offs and use simplifying, non-compensatory strategies (see LEX, EBA, etc. in Sect. 2.2), “to escape from the unpleasant state of conflict induced by the decision problem itself” (Shephard 1964, p. 277).

The second context-specific effect refers to *attribute ranges*, which can be measured by the number of different attribute levels displayed per attribute. For instance, the number of different brands or the number of different colors that are available for the products. If alternatives all have similar attribute levels, then this might reduce the effort as there are few distinct attributes that will have to be considered (Shugan 1980). Instead of counting the number of different attribute levels, a more advanced approach is to measure the standard deviations of the attribute values (Böckenholt et al. 1991; Iglesias-Parro et al. 2002). Most researchers have argued that decision makers find more similar alternatives more difficult to solve and this is the reason why they spend more time on the decision (Biggs et al. 1985; Böckenholt et al. 1991).

Finally, the *attractiveness differences* describes the differences in utilities of the offered alternatives and thus relies on some utility maximizing strategy, such as WADD. If one alternative has a much higher utility than others, in other words the attractiveness difference is large, then it should be easy for decision makers to find this utility maximizing alternative, and thus complexity is low (Böckenholt et al. 1991; Swait and Adamowicz 2001).

In contrast to other works, we like to separate attractiveness differences clearly from attribute ranges (see Payne et al. 1993; Böckenholt et al. 1991 for a less clear separation). While attribute ranges measure to what extent the single attributes differ from one another, for instance measured by computing the variance of attribute values on each attribute, the attractiveness differences refer to the difference of utilities of alternatives. The following examples display two choice tasks with high attribute ranges (see Table 2.3). In Table 2.3a, the alternatives hardly differ with respect to their total utility (0.425 vs. 0.475 vs. 0.425), while in Table 2.3b the alternatives differ a lot in respect to their total utility (0.8 vs. 0.375 vs. 0.225). However, both have the same average variance of attribute levels (attribute range) of 0.11.

Swait and Adamowicz (2001) introduced a measure of complexity which captures both the (1) complexity resulting from the quantity of information presented in a choice task and (2) the attractiveness difference: the entropy $H(ct)$. The entropy of a choice task, ct , is defined as

$$H(ct) = - \sum_{j=1}^n \pi(alt_j) \log \pi(alt_j), \quad (2.2)$$

where $\pi(alt_j)$ is the probability that alternative j is chosen. The choice probability $\pi(alt_j) = \frac{u(alt_j)}{\sum_{j \in CT} u(alt_j)}$, of an alternative j results from the utility value $u(alt_j)$ of that alternative in relation to the utility values of all alternatives in the set of

Table 2.3 Two examples with both high attribute ranges but different attractiveness differences

Attribute	alt_1	alt_2	alt_3	Variance
(a) High range and low utility difference				
Price	0.8	0.5	0.2	0.09
Size	0.1	0.5	0.9	0.16
Weight	0.2	0.8	0.2	0.12
Brand	0.6	0.1	0.4	0.06
Total utility	0.425	0.475	0.425	0.11
(b) High range and high utility difference				
Price	0.8	0.3	0.2	0.10
Size	0.6	0.3	0.2	0.04
Weight	1	0.5	0.1	0.20
Brand	0.8	0.4	0.4	0.05
Total utility	0.8	0.375	0.225	0.11

choice tasks, CT . The utility values are estimated by means of a MNL model and Hierarchical Bayes for each individual.

If there is an alternative in a choice task with choice probability one and the others therefore have probabilities of zero, the entropy, $H(ct)$, reaches its minimum value of zero. On the contrary, if the choice probability of all alternatives in a choice task is equal, the entropy will reach its maximum value of $\log n$, where n is the number of alternatives. Thus, the entropy depends not only on the similarity of the utilities of alternatives but also on the number of alternatives in the choice task. The more alternatives are in the choice task, the higher the entropy can be. For a given number of alternatives, n , the entropy is always greater or equal than for any other set with less alternatives n' : $2 \leq n' \leq n$.

The authors argue that the number of attributes and their correlation only effects entropy indirectly, since they influence the choice probabilities, $\pi(alt_j)$. From our viewpoint, this effect should be included more directly into the model as the indirect influence stays vague.

In the following paragraphs, we will review studies on the influence of the number of alternatives and attributes on decision making as well as the influence of the context-based measurements described above. Before we summarize the empirical evidence for the influence of complexity on decision-making behavior, we want to describe how decision-making behavior is typically measured.

2.4.2 Variables for Describing Decision-Making Behavior

In Sect. 2.3, we discussed how decision-making behavior can be analyzed with outcome-based and process tracing approaches. In general, research on outcome-based approaches focuses more on the development of better models and methods

and less on the influence of complexity (for an exception see Conlon et al. 2001; DeShazo and Fermo 2002; Swait and Adamowicz 2001). Thus, the following literature review summarizes mainly process tracing studies. As we will see in the following paragraphs, some of the variables typically measured correspond to characteristics of decision-making behavior as described in 2.2.1. Thus, Table A.1 can be used to determine which decision strategies fit best to the observed behavior.

1. *Depth of search* measures the absolute number of attribute levels which were assessed, including repeated processing of the same attribute levels (Huber 1980).
2. *Breadth of search* measures the number of considered attribute levels in relation to the amount of available information (Biggs et al. 1985). Thus, it counts whether an attribute level was assessed at least once, excluding repeated processing of the same attribute levels. A respondent who considers each attribute level at least once has a breadth of search of 1, and thus would use a decision strategy which is complete (see characteristic *complete* in Sect. 2.2.1).
3. A related measure describes the *consistency* of search (see characteristic *consistent* in Sect. 2.2.1), which is determined by the number of different attribute levels considered per alternative and per attribute or the time they were considered (Biggs et al. 1985; Timmermans 1993). In case the number of considered attribute levels per attributes is the same for each attribute, the search is consistent with respect to attributes. In case the number of considered attribute levels per alternatives is the same for each alternative, the search is consistent with respect to alternatives. This measure helps us to determine whether a compensatory or non-compensatory strategy was used. A consistent search with a breadth of search of one, for instance, militates in favor of a compensatory strategy (see Table A.1, the only consistent and complete but non-compensatory strategy is DOM).
4. The *search pattern* describes the sequence in which information is acquired. When decision makers tend to consider first several attributes of the same alternative before proceeding to the next alternative, the search pattern is alternative-wise. When, in contrast, they compare alternatives across attributes, we speak of attribute-wise search patterns. The most common form of operationalizing search patterns is the search index (SI) (Payne 1976). The SI sets the number of attribute-wise transitions (r_{attr}) and alternative-wise transitions (r_{alt}) in relation. It varies from -1 to $+1$, with -1 indicating completely attribute-wise search and $+1$ indicating completely alternative-wise search:

$$SI = \frac{r_{alt} - r_{attr}}{r_{alt} + r_{attr}}. \quad (2.3)$$

Alternative-wise and attribute-wise transitions can be measured with Mouselab and eye tracking methods. In case verbal protocols are used, researchers count statements reflecting comparisons among alternatives vs. comparisons among attributes (Timmermans 1993). Other researchers using verbal protocols,

approximate the search pattern by analyzing how often decision makers mention the elimination of alternatives (Klein and Yadav 1989). Furthermore, when respondents spend more time on alternative-wise (attribute-wise) processing, this speaks in favor of alternative-wise (attribute-wise) search patterns (Bettman et al. 1993).

5. Another prevalent measure is *effort*. This can be measured by the total time taken for the choice. Sometimes, in addition to using time for measuring effort, researchers also measure the respondents' self-reported effort (Iglesias-Parro et al. 2002; Luce et al. 1997).
6. Besides effort, decision *accuracy* is an interesting factor, but few studies are able to measure it. That is because accuracy is usually defined as the relation between the utility value of the chosen alternative and the utility value of the alternative selected when using the normative WADD strategy. Thus, highest accuracy is achieved when the WADD strategy is applied. Since only very few studies determine the utility function, they are not able to measure accuracy. Some studies use self-reported accuracy measures instead (i.e., Klein and Yadav 1989) or they present subjects inference tasks where the optimal choice is independent from people's preferences (Rieskamp and Hoffrage 2008).

In the following two sections, we describe what effects task-based and context-based complexity have on these variables which describe decision-making behavior.

2.4.3 *Influence of Task-Based Complexity*

The influence of the number of alternatives and attributes has been the subject of a number of studies. In a meta-study by Ford et al. (1989), for example, 20 out of 33 studies examine that influence. These results as well as a newer study by Timmermans (1993) report that an increasing amount of information leads to decreasing depth of search and decreasing search time. This indicates the limits of decision makers' cognitive capabilities which are not sufficient for processing all available information. Hence, decision makers apply selective and non-compensatory strategies.

Timmermans (1993) finds that an increasing amount of information leads to an increased depth of search. However, for an increasing number of attributes this only happens until decision makers exceed their cognitive capabilities. When this point is reached, increasing the number of attributes does no longer lead to an increased depth of search. Furthermore, with an increasing amount of information, the breadth of search decreases. The reason might be that in such cases, decision makers apply selective and non-compensatory strategies which, in general, do not consider attributes with low importance and hence the more attributes there are, the more information is just neglected.

Studies on search patterns lead to similar results as the studies on the depth and breadth of search. When the amount of information is low (complexity is

low), complete and compensatory strategies are used. An increasing amount of information increases the usage of non-compensatory strategies (Billings and Marcus 1983; Johnson and Meyer 1984; Olshavsky 1979; Staelin and Payne 1976). Biggs et al. (1985) and Shields (1980) measure the number of observed attributes per alternative and conclude that a relationship between the search pattern and the amount of information exists. With increasing amount of information, the variance of considered attributes per alternative increases, thus more inconsistent strategies are applied. The authors further argue that this result indicates an attribute-wise search and – according to Billings and Marcus (1983) and Crow et al. (1980) – might stand for strategies which are based on elimination such as EBA and CONJ. They further observe that at the end of the search process, people tend to use more alternative-wise search.

Timmermans (1993) reports similar results of a study in which she uses verbal protocols. In choice tasks with low complexity, respondents show more alternative-wise search. With an increasing amount of information, however, they compare alternatives more attribute-wise and eliminated more alternatives. She concludes that this indicates a two-step decision process.

Finally, studies reveal that respondents spend more time, the more complex the task is (Olshavsky 1979; Onken et al. 1985; Payne and Braunstein 1978).

In general, the influence of the number of alternatives on the depth and breadth of search, the search pattern, and the time is quite clear, while the influence of the number of attributes is still not completely understood. Furthermore, conclusions on the applied strategy are sometimes retrieved solely from search patterns, for instance researchers often conclude a compensatory strategy from alternative-wise processing (Conlon et al. 2001; Crow et al. 1980; Lee and Lee 2004; Olshavsky 1979; Payne et al. 1993). However, as can be seen in Table A.1, not all decision strategies with alternative-wise processing are compensatory (i.e., CONJ, DIS, SAT), nor are all strategies with attribute-wise processing non-compensatory (i.e., ADD, MAJ, and MCD). Hence, conclusions on the actual decision strategy applied based on only one or two process tracing measures should be drawn with caution.

2.4.4 Influence of Context-Based Complexity

Whilst the influence of the number of attributes and alternatives has come under a lot of academic scrutiny, the influence of context-based complexity is barely understood. We address this lack of research and start by providing a detailed overview of the context-based measures that influence the different dependent variables. As pointed out earlier, we focus only on three measures relevant to our later studies, namely *attribute ranges*, *trade-offs*, and *attractiveness differences*.

2.4.4.1 Attribute Ranges and Depth/Breadth/Consistency of Search

Several studies show that the smaller the attribute ranges, and thus the more similar the alternatives are, the higher is the depth and breadth of search. In Payne et al. (1988), subjects have to choose one out of two gambles. They operationalize the attribute range by the variance of probabilities of the two alternative gambles (low vs. high). If the variation of probabilities is high, the alternatives have low similarity. They show that subjects process less information (less depth of search) in case of high variance. Two follow-up studies support this result and also show that the complementary effect is true: high similarity, thus low variance, leads to an increased depth of search. Results are even stronger if many trade-offs are prevalent (Bettman et al. 1993; Payne et al. 1993).

In the first of their two studies, Böckenholt et al. (1991) also manipulate the attribute range. They show respondents two alternatives with attribute levels ranging from -6 to $+6$. In the small range condition, the two alternatives differ between 0 to 3 levels per attribute, while in the high range condition, they differ either 4 or 5 or 6 attribute levels. Their results show that respondents increase their breadth of search in case of low attribute ranges. Although the authors call this manipulation *attractiveness difference*, this is rather a manipulation of attribute ranges since they do not calculate the overall attractiveness of alternatives.

Biggs et al. (1985) draw similar conclusions by also manipulating the range of attributes. Instead of computing the variance, they design choice tasks where attribute levels differ at least one and at most three levels on a scale with a total of eleven levels. They combine this with the manipulation of the task complexity (operationalized through the number of attributes and alternatives). The authors observe that in case of constant task complexity, a high similarity causes an increase of both breadth and depth of search.

The same studies show that the lower the variance of attribute levels, the more consistent is the search. Subjects spend an evenly distributed amount of time on both attributes and alternatives (Bettman et al. 1993; Payne et al. 1988). Furthermore, subjects consider a constant amount of attribute levels per attribute in case of low variance and an inconsistent amount in case of high variance (Biggs et al. 1985).

In summary, in cases of high similarity, decision makers tend to search for more information, both in terms of breadth and depth of search, and they acquire information more consistently. Most strategies which are consistent and complete strategies are compensatory, such as ADD, EQW, FRQ, MAJ, MCD, WADD (see Table A.1). The only exception is DOM, which is consistent and complete but non-compensatory.

2.4.4.2 Attribute Ranges and Search Pattern

Studies on search pattern and attribute ranges are less common and results are less clear. Payne et al. (1988) and Bettman et al. (1993) find a positive SI, i.e. alternative-wise processing, in cases of low variance of attribute levels and a negative one in cases of high variance.

In contrast, Iglesias-Parro et al. (2002) do not find any influence of attribute ranges on the search patterns with Mouselab. They only find a negative SI in both conditions of low and high attribute range, indicating an overall attribute-wise search. In their study, they pre-define a utility function as well as specific weights. The subjects were then asked to behave as though this utility function were their own. In choice tasks with two alternatives, they then vary the variance and the mean of differences for the two attribute levels of each attribute, as well as the correlation of attribute vectors. In three studies, they examine the influence of these three context-based measures of complexity in different combinations.

Biggs et al. (1985) also partly considered search patterns with Information Display Boards and verbal protocols yet they did not explicitly report their results in the paper since they use the pattern only indirectly assigning one out of four strategies to the observed choice behavior. They report some evidence that subjects might use compensatory strategies in cases of increasing similarity, but use non-compensatory strategies in cases of increasing amount of information. However, their results must be interpreted with caution as they observed only seven subjects.

In general, people have a tendency to use more alternative-wise search patterns and to apply more compensatory strategies in cases of low variance. However, due to the small number of respondents in Biggs et al. (1985) the latter result needs further evidence.

2.4.4.3 Attribute Ranges and Effort and Accuracy

Payne et al. (1988) observe that decision makers need less time and thus spend less effort in case of high variances than in case of low variances. Moreover, Bettman et al. (1993) report high interaction effects between the attribute ranges and another context effect, the trade-offs measured in terms of the correlation of attribute vectors (see Sect. 2.4.1.2). In case of low variance and negative correlation (many trade-offs), subjects need the most amount of time for the decision. Furthermore, the relative accuracy as well as the number of utility-maximizing decisions decreases in case of low variance and negative correlation. Thus, in decisions where attribute ranges are small and there are a lot of trade-offs among the alternatives, decisions seem to be very difficult, because decision makers spend more effort but achieve low accuracy.

In their first study, Iglesias-Parro et al. (2002) do not control for trade-offs. They observe an increased amount of time and thus more effort with increasing attribute range when measured with the mean difference of attribute values for each attribute. However, it is hard to interpret these results without controlling for the amount of trade-offs inherent in the decision, as high variances can occur in scenarios with a dominant alternative (easy task) as well as in scenarios with a lot of trade-offs (difficult task). Consequently, in their following two studies, they also manipulate the amount of trade-offs but do not find significant interaction effects. They find again that with increasing attribute range, the time increases – an observation which is not in agreement with others (Bettman et al. 1993; Payne et al. 1988).

In summary, the results concerning the influence of attribute ranges on effort and accuracy are contradictory. We think that this is because attribute ranges alone cannot describe the difficulty of a choice task sufficiently. As explained before, low variance can occur in cases with low complexity where there is one dominant alternative which is only slightly better than others. But low variance can also occur in situations with high complexity where we have many trade-offs and no dominant alternative.

2.4.4.4 Trade-Offs and Depth/Breadth/Consistency of Search

In the previous paragraph, it was suggested that trade-offs were an important factor influencing the decision-making process. Trade-offs are usually measured by computing the average pairwise correlation between attribute vectors. In case of negative pairwise correlation between attribute vectors, there are trade-offs in the choice task. Imagine a choice task of two alternatives, where alt_1 dominates alt_2 . In this case, the attribute vectors correlate positively as attribute values always decrease, going from alternative 1 to alternative 2. If, on the other hand, there are trade-offs and no dominant alternative, some attribute values increase and some decrease, because across some attributes alt_1 beats alt_2 and across others it is the other way around. Thus, the average correlation between all pairwise attribute vectors would be low.

In case decision makers adopt compensatory strategies, and hence actually consider all trade-offs, negative correlations make a choice more complex (Fasolo et al. 2009; Payne et al. 1993). Positive correlations, though, simplify choices, since simpler, non-compensatory strategies can be used in order to early eliminate “weaker” alternatives early on without losing too much accuracy (Bettman et al. 1993; Luce et al. 1997).

Correlation structures, and hence trade-offs, influence the depth of search and the consistency. Bettman et al. (1993) observe an increased depth of search as well as a more consistent search with negative correlations. Luce et al. (1997) also observe an increased depth of search. In one of their studies, 41 students had to choose among five jobs which were described along four attributes on a 7-point scale ranging from *best* to *worst*. In the group with few trade-offs, the average correlation between all pairs of attribute vectors was 0.14, in the group with high trade-off, it was -0.31 (the authors call the conditions “high” vs. “low conflict”). In a pre-study, the authors also acquire subjects’ individual attribute weights. Thus, in contrast to the attribute values which were prescribed by the Likert scale on the subjects, the attribute weights were directly rated by each subject. The authors take the attribute weights to also manipulate in-between subjects whether the trade-offs involve more important attributes (high trade-off difficulty) or less important attributes (low trade-off difficulty). They find a main effect for the first variable, the attribute correlation. In case of negative correlation structures, subjects’ depth of search increases. For the second independent variable, the trade-off difficulty, no main effect was found.

However, there was a significant interaction effect. In case of negative correlations on important attributes, the depth of search was highest.

To sum up, the more trade-offs among the alternatives, the more information is considered and the more consistent is the search. Thus, more trade-offs might result in the application of more compensatory strategies. Recall that literature on attribute ranges and depth/breadth/consistency of search and search patterns has come to the conclusion that the more similar alternatives are, the more compensatory strategies are used. Depending on how we define similarity in the context of trade-offs, these two results do contradict each other. Payne et al. (1993) have said that “In general, if the alternatives are similar, the attributes will be positively correlated; if they are very dissimilar, the attributes are negatively correlated” (p. 59). If we follow this interpretation, the results contradict. The more similar alternatives are with respect to their attribute range, the more compensatory is the search. The more similar alternatives are with respect to trade-offs, the less compensatory search is. From our viewpoint, the statement by Payne et al. (1993) neglects an important fact which we have already brought up in the discussion on attribute ranges and effort/accuracy. Alternatives with positive correlations can either have low attribute range or high attribute range. In the first situation, where we have positive correlation and low attribute ranges, alternatives only differ slightly. Hence, we have high similarity among alternatives. In contrast, in case of positive correlations and high attribute ranges one alternative, or at least a few, stand out clearly from the others and might even dominate them. In this case, we would speak of low similarity among alternatives. Thus, a choice task cannot be interpreted to be similar or non-similar alongside trade-offs alone. We always have to take into account other measures, such as attribute ranges and differences in attractiveness, in order to speak of similar or dissimilar alternatives. From the studies, we can therefore only conclude that the more trade-offs, the more likely is it that more compensatory strategies are used, but we should be careful to interpret these results in terms of similarity.

2.4.4.5 Trade-Offs and Search Pattern

Both Iglesias-Parro et al. (2002) and Luce et al. (1997) find no effect of trade-offs on the search pattern. Luce et al. (1997) find significant influence in the conditions with negative correlations where the trade-offs are along the most important attributes. In these cases, subjects process information attribute-wise. In a further analysis, the authors show tentative evidence that subjects start off with attribute-wise processing and then switch to alternative-wise processing. Furthermore, they support the view that subjects search longer attribute-wise in case of negative correlations on the most important attributes.

In contrast to Luce et al. (1997) and Iglesias-Parro et al. (2002), Bettman et al. (1993) and Fasolo et al. (2003) measure a positive search index in the context of negative correlations and thus an alternative-wise search. Fasolo et al. (2003) for instance, let subjects choose four times between five different digital cameras described by eight attributes and record their process with a web-based process

tracing technique, while Bettman et al. (1993) let subjects choose between gambles. Because of the contradictory results, more research needs to be done in order to draw clear conclusions.

2.4.4.6 Trade-Offs and Effort and Accuracy

All studies observe that in case of negative correlations, the time increases significantly. Thus, in case of trade-offs, the choice task seems to be more complex and increases effort (Bettman et al. 1993; Fasolo et al. 2009; Iglesias-Parro et al. 2001; Luce et al. 1997). Some also show that accuracy decreases with increasing trade-offs (Bettman et al. 1993).

2.4.4.7 Attractiveness Differences and Depth/Breadth/Consistency of Search

Böckenholt et al. (1991) manipulate attractiveness difference and overall attractiveness of alternatives. The overall attractiveness reflects whether both alternatives are very attractive (high utility) or rather unattractive (low utility) for the subjects. The attractiveness differences were manipulated by creating choice tasks which presented an overall difference in the first two attributes of either one level at most, or one level at least (large difference). Each subject has to make 30 choices among two alternatives. They find that subjects consider a higher percentage of attribute levels (breadth of search) if the attractiveness differences are low and if the overall attractiveness are low. However, no interaction effect between attractiveness differences and attractiveness of alternatives is found.

Swait and Adamowicz (2001) use a structural modeling approach to measure the influence of complexity on decision-making behavior. They hypothesize that the degree to which participants tend to be utility maximizers depends on the entropy in a choice task (see (2.2)). They compare the estimated attribute weights in choice tasks with low vs. high entropy. They find that subjects tend to focus on only a few attributes when entropy increases, because the estimated attribute weights from the model have a high variance. Thus, in case of high entropy, one can interpret the results in terms of an inconsistent search and a focus on only a few attributes.

2.4.5 Discussion

In this section, we have summarized how task-based and context-based complexity influences decision-making behavior. Specifically, we have focused on the amount of information, attribute ranges, trade-offs, and attractiveness differences. Furthermore, we have discussed that attribute ranges and attractiveness difference define how similar alternatives are, while trade-offs alone are not able to measure similarity.

The results which relate to the number of attributes and alternatives show that people use simpler strategies in case of an increasing amount of information. That is, they apply more non-compensatory strategies. This result is more distinct for the influence of alternatives than for the influence of attributes.

While there are plenty of studies on the amount of information, relatively few studies examine the influence of similarity and trade-offs, although the existing studies give evidence for their strong influence on decision-making behavior. Since we have not found any detailed overview of these studies in the literature, we gave a detailed overview of all studies and summarized them in Table in the appendix.

A large number of trade-offs as well as a small range of attribute values lead both to an increase in information acquisition (breadth and depth of search) and more consistent search. In combination, these two observations indicate the usage of more compensatory strategies in case of similar alternatives with many trade-offs.

Usually, compensatory strategies cost more effort than non-compensatory strategies. Thus, in case of low attribute ranges and many trade-offs, we would expect an increase in effort. However, there are contradictory results for the influence of the attribute range on effort, while the result for trade-offs is more clear. In case of many trade-offs, more time is spent on the choice task.

When we measure the influence of attribute ranges and trade-offs with search patterns, the results are unclear. The few studies on attribute ranges and search patterns find no influence at all (Iglesias-Parro et al. 2002) or lack a profound subject pool (Biggs et al. 1985). Furthermore, studies contradict in respect to the influence of trade-offs. One study finds some evidence of a more attribute-wise search in case of trade-offs on the most important attributes (Luce et al. 1997), others a more alternative-wise search (Bettman et al. 1993; Fasolo et al. 2003).

Consequently, literature suggests that trade-offs and attribute ranges might influence the usage of compensatory vs. non-compensatory strategies but not the search pattern. Because compensatory and non-compensatory strategies can both induce an attribute and alternative-wise search, these results do not necessarily contradict each other.

Only few studies exist on the influence of attractiveness difference. Böckenholt et al. (1991) find that more information is considered if the difference is low. Swait and Adamowicz (2001) find the contrary effect. If entropy is high, and thus the attractiveness difference is low, subjects focus on only a few attributes. However, as their entropy measure takes into account both the number of alternatives and attractiveness difference, the two effects might overlap.

Swait and Adamowicz (2001) and Biggs et al. (1985) are two of the few studies which consider both task- and context based complexity. From our viewpoint, more research should not only focus on different measures of context-based complexity and their interactions, but more research should also follow the notion of analyzing both task- and context-based effects.

Most critical to all of the mentioned studies is that respondent's utility functions are not measured on the individual level. Already Böckenholt et al. (1991) outlined that one has to be aware that participants may value attributes and alternatives differently. Consequently, the similarity of alternatives and the amount of trade-offs

depends on individual preferences and therefore is difficult to manipulate as an independent variable.

Hence, a large quantity of research has either neglected the existence of individual preference structures or used very restrictive assumptions for the part-worth utilities of attribute values (i.e., the attribute value which is ascribed to a certain attribute level is set to the same value for all respondents). Attribute levels are, for instance, displayed as integer values from 1 [very poor] to 10 [very good], so instead of naming an exact price, such as “coffee machine A is worth €189.99 and machine B €129.99”, they record “coffee machine A is very poor in terms of price [value 1] and B is good [value 8] (Biggs et al. 1985). It thus appears that unless the value functions are measured accurately, the context-based measurement cannot be precisely determined.

2.5 Conclusions

In this chapter, we provided an overview of the fundamentals which are relevant to the present work. We defined the most important terms, such as choice task, decision strategy, preference as well as task- and context-based complexity. We have seen that a variety of different decision strategies have been described in the literature so far, covering all kinds of different decision-making behaviors. Moreover, we stressed the fact that decision-making behavior is contingent upon factors such as the characteristics of the choice task or of the decision maker. Hence, in order to build DSS that assist people in their decision processes, we need to understand these contingencies.

To study such contingencies it is necessary to empirically observe and analyze decision-making behaviors by using methods such as structural modeling, comparative model fitting and process tracing. Furthermore, we reviewed the literature pertaining to the influence of complexity on decision-making behavior. A main result of this literature review is the lack of research on different context-based complexities. Specifically, the effect of context-based complexity on search patterns is unclear. Moreover, the relationship between the two similarity measures, attribute ranges and attractiveness differences, and trade-offs needs to be scrutinized further. Another gap in current research is that many studies use rough estimates for measuring parameters and recording the observed data. First, they use process tracing methods such as Mouselab that might influence the decision process. Second, they do not measure decision makers’ preferences in detail, but use some simple proxy and aggregated utility functions (for an exception see Klein and Yadav 1989).

We postulate that a more thorough measurement of both preferences and the decision process, by, for instance, using advanced preference measurement methods from marketing and by recording the decision process with eye tracking will improve validity.



<http://www.springer.com/978-3-7908-2768-2>

Interactive Decision Aids in E-Commerce

Pfeiffer, J.

2012, XXII, 250 p., Hardcover

ISBN: 978-3-7908-2768-2

A product of Physica-Verlag Heidelberg