A computational cognitive process model for multi-alternative multi-attribute preferential choice

by

Lena Maria Wollschläger

A thesis submitted in partial fulfillment
of the requirements for the degree of

**Doctor of Philosophy
in Psychology**

Approved Dissertation Committee

Prof. Dr. Adele Diederich
Jacobs University Bremen

Prof. Dr. Dennis Dittrich
Touro College Berlin

Prof. Dr. Angelika May
Universität Oldenburg

Dr. Keivan Mallahi Karai
Jacobs University Bremen

Date of Defense: 25. February 2019

Department of Psychology & Methods

Summary

In this thesis, a computational cognitive process model of multi-alternative multi-attribute preferential choice is proposed, revised, tested for its ability to simulate three benchmark context effects and interactions between them, and compared with earlier and more recent theories. The 2N-ary choice tree model assumes that the decision maker, given a set of N choice alternatives that are described by the same attributes, repeatedly compares pairs of attribute values and counts how often each alternative wins and loses a comparison. The number of favorable and unfavorable comparisons is stored in two separate counters per alternative and the difference of the counter states forms the preference state for the respective alternative. If the preference state for an alternative hits a negative threshold, this alternative is eliminated from the choice set and the comparison process continues without it. On the other hand, if the preference state for an alternative hits a positive threshold, this alternative is chosen and the whole process stops.

The simple choice tree model, a revised version of the 2N-ary choice tree model, introduces an additional parameter for regulating the focus on the winning or losing alternative in a comparison, which has an effect on the proportion of choices and eliminations that take place. The 2N-ary choice tree model and the simple choice tree model are both able to explain similarity, attraction, and compromise effects, three context effects that have been observed after adding a third option to a set of two choice alternatives. With its additional parameter, the simple choice tree model beyond that accounts for the positive correlation between attraction and compromise effects and the negative correlation between these two and the similarity effect, that Berkowitsch, Scheibehenne, and Rieskamp (2014) found in their recent study. The simple choice tree model is the only model that accounts for the whole range of related findings, including negative similarity, attraction, and compromise effects.

In chapter 2, the literature on similarity, attraction, and compromise effects and computational cognitive process models that explain them is reviewed. Additionally, the 2N-ary choice tree model (proposed in chapter 3) and its variant simple choice tree model (proposed in chapter 4) are interrelated with the other existing computational cognitive process models of multi-alternative multi-attribute preferential choice. In chapter 6, the importance of the dynamic aspects of such models is emphasized and it is discussed how response times could inform the modeling endeavor. Matlab code for simulating choice probabilities and choice response times with the 2N-ary choice tree model and the simple choice tree model is provided in appendix A, and results from a series of simulations based on the simple choice tree model are reported in chapter 5. In chapter 7, the impact of this thesis and possible future applications of the models proposed herein are discussed.

A computational cognitive process model for multi-alternative multi-attribute preferential choice

Lena Maria Wollschläger

To Samuel and Valeria

Acknowledgements

I would like to thank the following people for their support during the process of writing this thesis: My supervisor Prof. Dr. Adele Diederich for her continuous encouragement, trust, and patience, and for sharing her knowledge and networks. The other members of my dissertation committee, Prof. Dr. Angelika May, Prof. Dr. Dennis Dittrich, and Dr. Keivan Mallahi Karai, for fruitful discussions, challenging questions, and helpful remarks. My family and friends, especially Vasco for providing food and love to keep me going. And, last but not least, the proverbial village that helped raise the kids – Samuel and Valeria – to whom I dedicate this work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Preferential choice is something that we all engage in every day, from pressing the snooze button in the morning to drinking another glass of wine in the evening, from accepting a dinner invitation to declining a job offer, from buying a specific type of cereal to purchasing a car. From a broad perspective, such decision problems include meta-decisions about which alternatives to consider and when to choose between them. However, in the research reported in this thesis, a slightly more restricted perspective is taken, assuming that the choice alternatives, as well as relevant attributes that describe them, are readily available to the decision maker. This so-called choice from description is a relatively simple and well-defined problem, yet we are only starting to understand what drives human behavior in such situations. Open questions remain particularly for choice between three or more alternatives that are described by two or more attributes, that is, for multi-alternative multi-attribute preferential choice.

When asked to choose from several alternatives with multiple attributes, decision makers may experience preference uncertainty (March, 1978). That is, they may be uncertain about their own preferences for one attribute over the other and, what is more, for a specific combination of attribute values that describes one alternative over another combination that describes another alternative. Preference uncertainty can be overcome by constructing preferences on the spot (Bettman, 1979; Lichtenstein & Slovic, 2006, for a review). However, this leads to systematic dependencies on the task, the choice set or context, and on individual differences (cf. Payne, Bettman, & Johnson, 1992). For example, the same alternative may be chosen as best alternative and as worst alternative (task dependence), adding a third alternative to a set of two may alter the preference relation between those two options (context dependence), or a risk-averse decision maker may choose differently than a risk-seeking individual (individual differences). Here the focus is on effects of the choice set, or more precisely, on three context effects named similarity effect (Tversky, 1972b), attraction effect (Huber, Payne, & Puto, 1982), and compromise effect (Simonson, 1989). Section 2.2 contains a summary of the original findings and descriptions of some early theories that were proposed to explain them.

Theories of decision making have been traditionally categorized as either normative, prescriptive or descriptive. Normative decision theories are (economic) theories of how decisions should be made by an optimal or rational decision maker. Expected utility theory (von Neumann & Morgenstern, 1953) is a classic example from this category. Descriptive theories, on the other hand, try to describe human behavior as accurately as possible, even if it includes violations of rationality or optimality. (Cumulative) prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), for example, belongs to this category. Prescriptive theories try to bridge the gap between actual and optimal decision behavior by helping decision makers to improve their decisions. Multi-attribute utility theory (Keeney & Raiffa, 1967/1993), for example, has been used in such decision analyses.

Theories that explain similarity, attraction, and compromise effects are necessarily descriptive since the three context effects violate independence assumptions made by normative decision theories (see Tversky & Russo, 1969, for a discussion and comparison of independence assumptions). However, the theories discussed in this thesis, and particularly the model proposed in chapters 3 and 4, are not only descriptive models but beyond that computational cognitive process models. Cognitive process models are based on simple information processing mechanisms like information sampling, value comparison, or evidence accumulation, and their parameter-based mathematical formulation based on psychologically interpretable parameters allows for logically valid and precise quantitative predictions (cf. Busemeyer & Diederich, 2010). Multi-alternative decision field theory (Roe, Busemeyer, & Townsend, 2001) and the leaky competing accumulator model (Usher & McClelland, 2001, 2004) were the first computational cognitive process models of similarity, attraction, and compromise effects. In fact, they established the three context effects as a benchmark for computational cognitive process models of multi-alternative multi-attribute preferential choice and to this day serve as prototypes for such models. In section 2.3, the two theories and their accounts of the three context effects are described.

Despite their exemplary function, multi-alternative decision field theory and the leaky competing accumulator model have several shortcomings that we discuss in section 3.1. In order to overcome some of these shortcomings, the 2N-ary choice tree model is proposed in chapter 3 (and Wollschlaeger & Diederich, 2012). It is the first of six recently proposed models of multi-alternative multi-attribute preferential choice that claim to account for similarity, attraction, and compromise effects: The 2N-ary choice tree model (Wollschlaeger & Diederich, 2012), the associative accumulation model (Bhatia, 2013), the multi-attribute linear ballistic accumulator model (Trueblood, Brown, & Heathcote, 2014), the simple choice tree model (a variant of the 2N-ary choice tree model; Wollschlaeger & Diederich, 2017, and chapter 4 of this thesis), multi-attribute decision by sampling (Ronayne & Brown, 2017), and multi-alternative decision by sampling (Noguchi & Stewart, 2018). Except for multi-attribute decision by sampling, these models and their accounts of the three context effects are described in section 2.5.

The multi-attribute linear ballistic accumulator model and multi-attribute decision by sampling differ from the other four models (and from multi-alternative decision field theory and the leaky competing accumulator model) in that they do not make assumptions about the time course of information processing and preference construction. They are static models and thus not cognitive process models in a strict sense. In chapter 6, it is argued that dynamic aspects are a crucial part of multi-alternative multi-attribute preferential choice models and therefore response times should be taken into account when comparing model performance. For that, it is necessary to define optional stopping rules or decision criteria for the preference construction process, for instance in the form of thresholds for the accumulated evidence.

The 2N-ary choice tree model and its variant simple choice tree model, that are proposed in this thesis, define two such thresholds. One for eliminating alternatives from the choice set and one for choosing them. The possibility (but not necessity) to eliminate unwanted alternatives from the choice set is a unique feature of these two models and a crucial part of their account for similarity, attraction, and compromise effects. In the simple choice tree model, this feature is regulated by an additional parameter that is called focus weight. More eliminations take place for higher values of the focus weight, producing positive attraction and compromise effects but negative similarity effects. Lower values of the focus weight, on the other hand, produce positive similarity effects and negative attraction and compromise effects. A similar correlational pattern has been found in an experiment by Berkowitsch et al. (2014). But see 2.4 for an overview of recent experiments that investigate the three context effects (or variants thereof) and their interactions.

This thesis is organized as follows: In chapter 2, the literature on similarity, attraction, and compromise effects and computational cognitive process models that explain them is reviewed. Additionally, the 2N-ary choice tree model (proposed in chapter 3) and its variant simple choice tree model (proposed in chapter 4) are interrelated with the other existing computational cognitive process models of multi-alternative multi-attribute preferential choice. In chapter 6, the importance of the dynamic aspects of such models is emphasized and it is discussed how response times could inform the modeling endeavor. Matlab code for simulating choice probabilities and choice response times with the 2N-ary choice tree model and the simple choice tree model is provided in appendix A, and results from a series of simulations based on the simple choice tree model are reported in chapter 5. In chapter 7, the impact of this thesis and possible future applications of the models proposed herein are discussed.

Chapter 2

Context effects and computational cognitive process models

Lena M. Wollschlaeger, Adele Diederich

A revised version of this chapter with the title "Similarity, attraction, and compromise effects: Original findings, recent empirical observations, and computational cognitive process models." has been accepted for publication in the American Journal of Psychology on July 15, 2019.

Abstract

If a decision maker prefers A over B in one situation but B over A in another, psychologists call it a preference reversal. Preference reversals demonstrate that preferences are subject to change – contingent on the task, the context, or on individual differences. But what exactly causes preference reversals? The answer lies in the conception of preferences as constructed on the spot rather than revealed from some underlying source. However, open questions remain about the models and methods used to describe, explain and predict preference construction. We try to answer some of these questions based on the developments in the field of multi-alternative multi-attribute decision making. Similarity, attraction, and compromise effects play an important role in this endeavour, since they led to the proposition of multi-alternative decision field theory, the first computational cognitive process model of context effects. Multi-alternative decision field theory, in turn, established the three effects as a benchmark for such models and inspired several attempts to obtain them within the same experiment. To that end, 13 different variants or versions of the three effects have been introduced over the years. Besides identifying and describing those variants and versions, we show the advantages of using computational cognitive process models to explain the three effects, but also suggest to shift the focus from whole theories to the building blocks they are constructed from. Additionally, we highlight the importance of process data like response times or eye-movements for differentiating between models or mechanisms.

2.1 Introduction

Some call it part of the cognitive revolution (Payne & Bettman, 2008), some see in it a Kuhnian paradigm change (Oppenheimer & Kelso, 2015): Information processing has replaced utility theories as dominant paradigm for decision making research in Psychology. According to Kuhn (1962/2012), paradigm changes result from scientific revolutions with the following structure: Observations that cannot be explained by available theories, so-called anomalies, accumulate during periods of "normal science". The resulting crisis is resolved by "revolutionary science", which eventually leads to a new scientific paradigm.

In the case of decision science, the critical anomalies that could not be explained by utility theories were preference reversals (cf. Tversky & Thaler, 1990). For example, Lichtenstein and Slovic (1971, 1973) report that participants chose one gamble over another one, but assigned less monetary value to the chosen gamble, violating procedure invariance. Another prominent example are the framing effects reported by Tversky and Kahneman (1981) and Kahneman and Tversky (1984). Their participants chose one option over the other when consequences were described as gains, but preferred the other alternative when consequences were described as losses, violating description invariance. Preference reversals, that is, task-dependent changes in preference for option A over option B, have been reported in many different domains over the years, see Lichtenstein and Slovic (2006) for a review. Preference reversals show that decision makers do not at all behave rational in the sense of utility theories (e.g., expected utility theory, von Neumann & Morgenstern, 1953), but violate one of the most basic assumptions of these theories.

Rationality as a sine qua non for human behavior has been questioned long before the first preference reversals were observed. In his seminal paper "A behavioral model of rational choice", Simon (1955) proposes to replace "global" rationality with "bounded" or "limited" rationality that takes into account availability of information to and computational capacities of the choosing organism. He argues that the decision maker, due to limited processing capacities and in order to make a satisfactory rather than optimal choice, employs simple mechanisms instead of effortful calculations. Early theories of preference reversals (for a review, see Payne, Bettman, & Johnson, 1993), mostly within the heuristics and biases approach (Tversky & Kahneman, 1974), try to identify such simple mechanisms that may have led to the observed behavior. However, even though unifying frameworks have been developed (Payne et al., 1993; Gigerenzer, Todd, & the ABC Research Group, 1999; Shah & Oppenheimer, 2008; Gigerenzer & Gaissmaier, 2011; Hilbert, 2012), heuritics are usually domain specific and therefore remain largely disconnected from each other (Oppenheimer & Kelso, 2015). Moreover, selection of appropriate mechanisms as well as developing or learning a comprehensive set of such heuristics is actually an intractable problem (Otworowska, Blokpoel, Sweers, Wareham, & van Rooij, 2017), similar to rational choice in the sense of utility theories.

Later approaches incorporate construction of choice mechanisms on the spot, an idea originally proposed by Bettman (1979) and, slightly more general, by March (1978). Tversky, Sattath, and Slovic (1988), for example, claim that attribute weights are contingent on the task and explain preference reversals between gambles by means of the compatibility principle: Monetary elicitation methods (e.g., pricing or matching) highlight payoffs, while choice is more compatible with probabilities. Framing effects on the other hand have been associated with risk attitudes

(Tversky & Kahneman, 1986), and, in riskless choice, with loss-aversion (Tversky & Kahneman, 1991). Additionally, those theories implement context-dependent changes of reference points, that is, of baseline values for evaluating the alternatives. In their review of behavioral decision research for the years 1983 to 1991, Payne et al. (1992) conclude that construction of preferences (and/or choice mechanisms) is not only contingent on the task, but also on the context and on individual differences. In the same vein, Lichtenstein and Slovic (2006) argue that multiple theories are necessary in order to explain how preferences are constructed by different decision makers in different situations. However, they identify decision field theory (DFT, Busemeyer & Townsend, 1993; Roe et al., 2001; Busemeyer, Johnson, & Jessup, 2006) as exception to this need for multiple theories, since it is able to account for several preference reversals simultaneously. DFT's multialternative extension (MDFT, Roe et al., 2001), for example, simultaneously accounts for similarity, attraction, and compromise effects (Tversky, 1972b; Huber et al., 1982; Simonson, 1989), three so-called context effects.

As Oppenheimer and Kelso (2015) point out, a Kuhnian paradigm change can only occur if a new paradigm is already available. What does this mean? First of all, Oppenheimer and Kelso (2015) interpret the term paradigm in a global sense (cf. Hacking's introductory essay to the 4th edition of *The structure of scientific revolutions*, Kuhn, 1962/2012), as something that Kuhn had called "disciplinary matrix" in the 1969 postscript to his book. They say that

> a paradigm refers to the set of practices that defines a scientific discipline at any particular period. A paradigm provides the basis for deciding (a) what phenomena to study, (b) what kinds of questions meaningfully probe for answers, (c) how these questions should be structured, (d) how an experiment is to be conducted, and (e) how the results of the investigations should be interpreted. (Oppenheimer & Kelso, 2015, pp. 278-279).

In the field of multi-alternative multi-attribute decision making research, MDFT provided the first answers to these kinds of questions. We will use similarity, attraction, and compromise effects to demonstrate how those early contributions have shaped the field and helped define the information processing paradigm, allowing for the paradigm change to take place.

In short, the three context effects or "anomalies" are changes in (relative) choice probabilities for two choice alternatives after adding a third "decoy" option to the set. Similarity, attraction, and compromise effects violate independence assumptions as, for example, the axiom of independence of irrelevant alternatives in von Neumann and Morgenstern's (1953) formulation of expected utility theory, or Luce's (1959/2012) choice axiom (see Tversky & Russo, 1969, for a discussion and comparison of independence assumptions). Despite their simplicity, the effects demonstrate that choice probabilities in multi-alternative decision making are contingent on the local context, that is, on the choice set under consideration. Due to their simplicity, on the other hand, similarity, attraction, and compromise effects have been successfully examined in numerous studies to date (e.g. Huber & Puto, 1983; Ratneshwar, Shocker, & Stewart, 1987; Wedell, 1991; Simonson & Tversky, 1992; Pettibone & Wedell, 2007; Trueblood, 2012; Trueblood, Brown, Heathcote, & Busemeyer, 2013; Berkowitsch et al., 2014; Noguchi & Stewart, 2014; Tsetsos, Chater, & Usher, 2015; Liew, Howe, & Little, 2016; Mao, 2016). However, different conditions, different experimental setups, and even different variants of the three context effects make it hard to stay on top of things. This review aims at organizing the vast evidence in literature by identifying prototypical variants of the three effects. Additionally, we will reconstruct how similarity, attraction, and compromise effects have shaped modern multi-alternative multi-attribute decision making models since multi-alternative decision field theory has been proposed to simultaneously account for the three effects. In order for this review to be self-contained, the models will be presented in some detail. The goal is to provide a common ground for this field of research rather than to thoroughly compare the different theories. Based on this common ground, we will show possible directions for future research on multi-alternative multi-attribute decision making.

This review is organized as follows: First, we summarize the original findings and early theories that explain similarity, attraction, and compromise effects separately. We then describe multi-alternative decision field theory and the leaky competing accumulator model and their simultaneous accounts of the three context effects. Afterwards we turn to more recent empirical studies, with a focus on studies that are concerned with all three effects and their interactions. We then show how four recently proposed multi-alternative multi-attribute decision making models account for the original and new findings. And finally, we propose directions for future research in this field.

2.2 Original findings and early theories

Originally, similarity, attraction, and compromise effects have been observed after adding a third alternative to a set of two clearly distinguishable options described by two attributes. Let A_1 and A_2 be two choice alternatives with two common attributes, D_1 and D_2, describing them. For example, think about two desserts described by attributes tastiness and healthiness or two outfits that differ with respect to attributes formality and comfortableness. We assume that A_1 scores high on attribute D_1 but low on attribute D_2, and vice versa for A_2. That is, D_1 is the unique strongest attribute for A_1, and D_2 is the unique strongest attribute for A_2. As for the desserts, A_1 could be a mousse au chocolat that is very tasty but not exactly healthy, and A_2 could be a fruit salad that is less tasty but healthier than the mousse au chocolat. One can think of the alternatives as placed in a two-dimensional space with dimensions D_1 and D_2, as in the left panel of figure 2.1.

Similarity effect

The *similarity effect* was named and first studied systematically by Tversky (1972b), though Debreu (1960) and Becker, DeGroot, and Marschak (1963) mentioned a similar effect before. Debreu (1960), in his review of Luce's (1959/2012) book "Individual choice behavior: A theoretical analysis", alleged the following example as violation of Luce's choice axiom (an independence assumption): A decision maker who prefers listening to a recording of a Debussy quartet (D) over listening to a recording of a Beethoven symphony (B_1) will still prefer the Debussy quartet when choosing from a ternary set including two different recordings of the same Beethoven symphony (B_1 and B_2). According to Luce's choice axiom, however, the probability for choosing Debussy should decrease with addition of the second

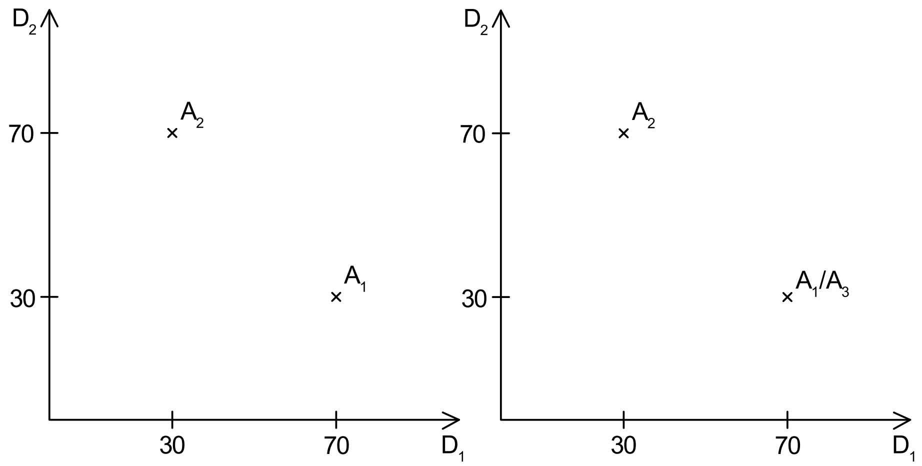

Figure 2.1: Placement of choice alternatives in the attribute space with dimensions D_1 (x-axis) and D_2 (y-axis). Left panel: Initial situation for all three context effects with alternatives A_1 and A_2. Right panel: Similarity situation with initial alternatives A_1 and A_2, and additional alternative A_3 that is similar to alternative A_1. The numbers on the axes are arbitrary.

Beethoven recording. Becker et al. (1963) tested and confirmed Debreu's (1960) assumption in a gambling situation. Their participants chose between two or three gambles of which two had dissimilar outcomes and the third gamble is similar to one of the other two. Tversky (1972b) elaborated on Debreu's (1960) example and predicted a similar effect for choice between tours of Europe and the Far East offered by two travel agencies: When indifferent between destinations and travel agencies, decision makers should be equally likely to choose each option from a binary set. Adding a third option, however, should affect only the probability for choosing the similar alternative, but not that for choosing the dissimilar alternative. For example, adding a second tour of Europe from a different travel agency to a choice set consisting of a tour of Europe and a tour of the Far East, should only affect the probability for choosing the first tour of Europe. Formally, Tversky (1972b) compared the binary choice set $\{A_1, A_2\}$ to the ternary choice set $\{A_1, A_2, A_3\}$ where A_3 is similar to one of the original alternatives. Let A_3 be similar to alternative A_1 in scoring high on attribute D_1 and low on attribute D_2 (see right panel of figure 2.1). A similarity effect (Tversky, 1972b) is observed if the probability for choosing A_1 over A_2 decreases with addition of alternative A_3 to the choice set:

$$\frac{P(A_1|A_1, A_2)}{P(A_2|A_1, A_2)} > \frac{P(A_1|A_1, A_2, A_3)}{P(A_2|A_1, A_2, A_3)}. \tag{2.1}$$

Tversky (1972b) tested for the similarity effect with three different stimulus types: Dot patterns with different sizes and densities (participants were asked to choose the pattern which contains most dots), gambles with different probabilities and outcomes (participants were asked to choose the gamble they prefer), and college applicant profiles with different intelligence and motivation scores (participants were asked to choose the most promising applicant). Participants' pooled data showed similarity effects for gambles and applicants but not for dot patterns.

Tversky (1972b) attributed this to possible differences in evaluating stimuli as unitary or composite alternatives. However, the selected attribute values (see Tversky, 1972b, p.292), the experimental setup, and the different tasks might play a role here as well. As for pooling participants' data, Tversky (1972b) noted that it would be desirable to analyze participants' data individually, but the number of collected data points did not allow for this in the reported study. Nevertheless, a similarity effect, and thus a consistent violation of Luce's (1959/2012) choice axiom or any equivalent independence assumption was observed for some of the tasks, calling for an explanation.

In order to account for the similarity effect, Tversky (1972a, 1972b) proposed the elimination by aspects (EBA) model, a multi-alternative multi-attribute choice model. According to this model, the decision maker goes through a series of aspects of the alternatives under consideration and for each aspect eliminates all alternatives that do not contain it. An aspect could be a specific manifestation of an attribute, like a price limit or a limit for the travel time, or it could be a desired characteristic of the choice alternatives, like a specific place the decision maker wants to visit or a specific ingredient she wants in her dessert. The elimination process stops as soon as a single alternative is left, which is then chosen. Choice probabilities are defined iteratively via probabilities for attending the aspects. Instead of focusing on single aspects, aspects are grouped according to the alternatives that contain them. Here, $\overline{A_1} = \{x_1, y_1, z_1, \ldots\}$ is the set of all aspects that are unique to alternative A_1, and $\overline{A_2}$ and $\overline{A_3}$ are defined accordingly for alternatives A_2 and A_3. Aspects that are shared by two alternatives are summarized as $\overline{A_1, A_2}$, $\overline{A_1, A_3}$ and $\overline{A_2, A_3}$ respectively. The group of aspects that are shared by all three alternatives, that is, $\overline{A_1, A_2, A_3}$, is ignored in the analysis since it does not lead to elimination of any alternative and thus does not bring forward the choice process. Next, Tversky (1972a, 1972b) defined a scale U on these aspect groups, assigning a positive number to each group that can be interpreted as utility or value. The probability for attending a specific group of aspects is obtained by dividing its utility by the sum of utilities of all the other groups. In the binary case, the probability for attending the unique aspects of an alternative is equal to the probability for choosing this alternative. For example, the probability for attending the unique aspects of alternative A_2, and thus for choosing alternative A_2 from the binary set $\{A_1, A_2\}$, is given by

$$P(A_2|A_1, A_2) = \frac{U(\overline{A_2})}{U(\overline{A_1}) + U(\overline{A_2})}. \tag{2.2}$$

The probabilities for choosing alternative A_1 from the set $\{A_1, A_2\}$, alternatives A_1 or A_3 from the set $\{A_1, A_3\}$, and alternatives A_2 or A_3 from the set $\{A_2, A_3\}$ are defined accordingly. The choice probabilities for the binary case are used to iteratively define ternary choice probabilities. These are obtained by summing over all the ways a choice can be achieved. For example, the dissimilar alternative A_2 (cf. figure 2.1) can be chosen in a single step by attending to an aspect that is unique to A_2 and thus leads to elimination of both A_1 and A_3. Or it can be chosen in two steps by first attending to an aspect that is contained in both A_2 and A_1 (or A_2 and A_3), leading to elimination of alternative A_1 (or A_3) and then choosing from the binary choice set $\{A_2, A_3\}$ (or $\{A_2, A_1\}$) with the probabilities defined according to equation (2.2) above. This yields

$$P(A_2|A_1,A_2,A_3) = \frac{U(\overline{A_2}) + P(A_2|A_1,A_2)\cdot U(\overline{A_1,A_2}) + P(A_2|A_2,A_3)\cdot U(\overline{A_2,A_3})}{U(\overline{A_1}) + U(\overline{A_2}) + U(\overline{A_3}) + U(\overline{A_1,A_2}) + U(\overline{A_1,A_3}) + U(\overline{A_2,A_3})}.$$

Assuming $U(\overline{A_1}) = U(\overline{A_3}) = a$, $U(\overline{A_1,A_3}) = b$, $U(\overline{A_2}) = a+b$ and $U(\overline{A_1,A_2}) = U(\overline{A_2,A_3}) = 0$ (since $\overline{A_1,A_2} = \emptyset = \overline{A_2,A_3}$), all the binary probabilities are equal to 0.5, while the ternary probabilities differ for the dissimilar alternative A_2 as compared to the similar alternatives A_1 and A_3:

$$\begin{aligned} P(A_2|A_1,A_2,A_3) = \frac{a+b}{3a+2b} > \frac{a+b(a/2a)}{3a+2b} &= P(A_1|A_1,A_2,A_3) \\ &= P(A_3|A_1,A_2,A_3). \end{aligned}$$

As a approaches 0, that is, as A_1 and A_3 share more and more of their aspects, the probability for choosing the dissimilar alternative A_2 from the ternary set approaches 0.5 while the probabilities for choosing alternative A_1 and A_3 each approach 0.25. Therefore, the EBA model predicts the similarity effect. However, it cannot account for the attraction or compromise effect.

Attraction effect

The *attraction effect* or *decoy effect* or *asymmetric dominance effect* was introduced by Huber et al. (1982) as consistent violation of the regularity principle. This principle, as presumed by the elimination-by-aspects model as well as most earlier probabilistic choice models (cf. Luce, 1977), states that additional alternatives cannot increase the choice probabilities of the original options. However, Huber et al. (1982) suggested that the relative probability for choosing alternative A_1 can be increased by adding a third alternative A_3 to the choice set that is similar to but dominated by A_1 (see figure 2.2 for possible placements of the dominated alternative A_3). For example (Huber et al., 1982), considered the choice between two six-packs of beer, of which one is cheap and of low quality (A_1) and one is expensive and of high quality (A_2). An asymmetrically dominated alternative (A_3) could be another cheap six-pack for the same price as the first one but with even lower quality. A_3 may then serve as a decoy for alternative A_1, drawing attention to it and therewith improving its evaluation and increasing its choice probability:

$$\frac{P(A_1|A_1,A_2,A_3)}{P(A_2|A_1,A_2,A_3)} > \frac{P(A_1|A_1,A_2)}{P(A_2|A_1,A_2)}. \tag{2.3}$$

Note that the same mechanism applies when a dominated decoy close to alternative A_2 is added to the choice set. This symmetric case is omitted here. Huber et al. (1982) found attraction effects in (hypothetical) choice between six-packs of beer with attributes price and quality, cars with attributes ride quality and gas mileage, restaurants with attributes driving time and food quality, lotteries with attributes chance of winning and amount to win, photographic films with attributes developing time and color fidelity, and TV sets with attributes percent distortion and reliability.

Let m_{ij} be the attribute value for alternative A_i with respect to attribute D_j. Huber et al. (1982) differentiated between three types of decoys: (1) A_3 is called

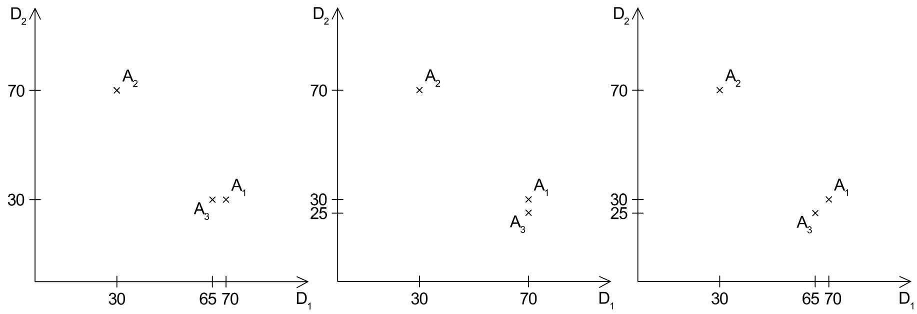

Figure 2.2: Placement of choice alternatives in the attribute space with dimensions D_1 (x-axis) and D_2 (y-axis). Left panel: Attraction situation with initial alternatives A_1 and A_2, and frequency decoy A_3 that is dominated by alternative A_1 on dimension D_1. Middle panel: Attraction situation with initial alternatives A_1 and A_2, and range decoy A_3 that is dominated by alternative A_1 on dimension D_2. Right panel: Attraction situation with initial alternatives A_1 and A_2, and range-frequency decoy A_3 that is dominated by alternative A_1 on both dimensions. The numbers on the axes are arbitrary.

frequency decoy if $m_{31} < m_{11}$ and $m_{32} = m_{12}$, that is, if A_3 is dominated by A_1 on their shared strongest dimension D_1, but equal to A_1 on their weak dimension D_2 (see left panel of figure 2.2). A_3 increases the frequency of alternatives along the first dimension. (2) A_3 is called *range decoy* if $m_{31} = m_{11}$ and $m_{32} < m_{12}$, that is, if it is equal to A_1 on D_1 and dominated by A_1 on D_2 (see middle panel of figure 2.2). A_3 increases the range of attribute values on dimension D_2. (3) A_3 is called *range-frequency decoy* if $m_{31} < m_{11}$ and $m_{32} < m_{12}$, that is, if it is dominated by A_1 on both dimensions (see right panel of figure 2.2). Huber et al. (1982) hypothesized that a frequency decoy increases the weight of dimension D_1 by either drawing attention to it or by spreading the psychological distances on that dimension. A range decoy, on the other hand, is assumed to decrease the importance of a fixed difference on dimension D_2, therewith diminishing the advantage of alternative A_2 over A_1. Both effects favor alternative A_1 and since the range-frequency decoy combines the two aspects, it is presumed to yield the highest attraction effect. However, the data reported by Huber et al. (1982) suggested that the effect is weakest for the frequency decoy and strongest for the range decoy, not showing a significant effect of more or less extreme range decoys (and thus contradicting the range effect). They concluded that the attraction effect is contingent upon the ratio of the distances between alternatives A_3 and A_1 versus A_3 and A_2.

Huber and Puto (1983) extended the investigations by including relatively inferior decoys: Alternative A_3 is called *relatively inferior decoy* if it exceeds alternative A_1 on their shared strongest attribute D_1, $m_{31} > m_{11}$, is inferior to it on their weaker attribute D_2, $m_{32} < m_{12}$, and lies below the non-concave indifference curve through alternatives A_1 and A_2. Huber and Puto (1983) suggested that the addition of an asymmetrically dominated or relatively inferior alternative to the binary choice set yields a global attraction effect, drawing preferences towards the decoy A_3 and its adjacent alternative A_1, and a local substitution effect, splitting choice probabilities between the similar alternatives A_1 and A_3. Huber and Puto

(1983) suggested that ordinal relations between the alternatives are more important than the actual magnitudes of their differences. This is consistent with the finding of Huber et al. (1982), that the extremeness of the range decoy did not influence the magnitude of the attraction effect.

Ratneshwar et al. (1987) tried to link the attraction effect in consumer choice to characteristics of the choice task, suggesting that low meaningfulness of the given information about the alternatives and a low degree of familiarity with the product category increase the attraction effect. Indeed, their studies suggested that more elaborate (that is, meaningful) information about the alternatives significantly reduces the attraction effect. The influence of familiarity had the predicted direction, but was not significant. According to the authors, this might have been due to their operationalization of familiarity and required further research. Overall, they concluded that "the interactions of person, product, and situation may be the critical factors in understanding consumer choice" (Ratneshwar et al., 1987, p. 532).

The first consistent theory for explaining the attraction effect was proposed by Simonson (1989). He called it "choice based on reasons" and suggested that consumers choose alternatives that are supported by the best reasons. In case of the attraction effect, the dominance relationship between alternatives A_1 and A_3 could serve as tie-breaker between otherwise equally preferred alternatives A_1 and A_2, in favor of the dominating option A_1. Simonson (1989) assumed the effect would be stronger when choices were anticipated to be evaluated by others or have to be justified in front of them later. The rationale behind this was that the decision maker is uncertain about the other's preferences and therefore uses objectively valid reasons like the dominance relationship to justify the choice. Indeed, Simonson (1989) found a stronger attraction effect under the high need for justification condition. Think-aloud protocols provided further evidence for his assumption and indicated that choosing alternative A_1 due to the dominance relationship required more elaboration and thus took more time than choosing alternative A_1 or A_2 due to their high value on either dimension D_1 or D_2.

Later, the attraction effect was attributed to loss aversion, the concept that choice alternatives are evaluated with respect to a reference point and the decision maker puts more weight on losses, that is, negative deviations from the reference point, than on gains, that is, positive deviations from the reference point (Tversky & Kahneman, 1991; Simonson & Tversky, 1992). While Tversky and Kahneman (1991) endowed their participants with the dominated alternative A_3 and then offered them to either keep this alternative or trade it for alternative A_1 or A_2, Simonson and Tversky (1992) compared choices from the binary set $\{A_1, A_2\}$ and the ternary set $\{A_1, A_2, A_3\}$. In both ternary choices, the dominating alternative A_1 was chosen significantly more often than the dissimilar alternative A_2, supposedly because A_1 - in contrast to A_2 - did not include a loss with respect to the reference point A_3.

Compromise effect

Based on the investigation of relatively inferior decoys by Huber and Puto (1983), the theory of reason-based choice (Simonson, 1989) predicted an additional context effect, the *compromise effect*. It may occur when a third, extreme, alternative A_3 is added to the choice set, which is neither dominated by the original alternatives A_1 and A_2 nor dominates them. For example, consider a coffee shop that sells

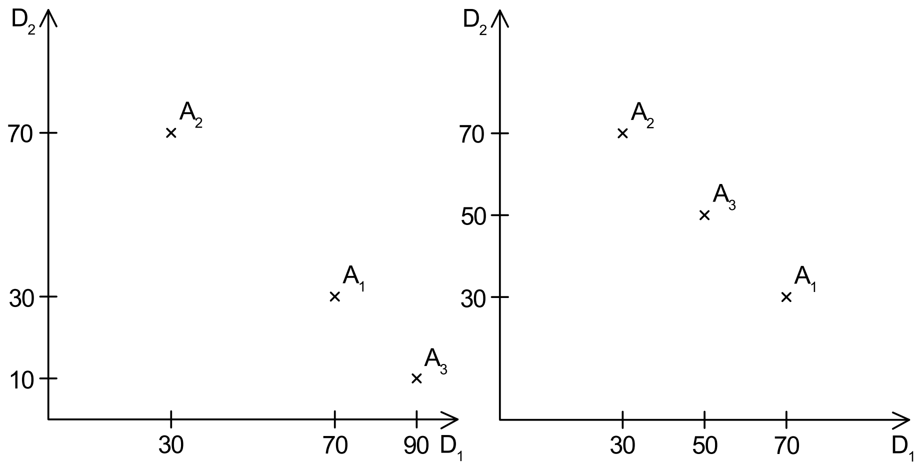

Figure 2.3: Placement of choice alternatives in the attribute space with dimensions D_1 (x-axis) and D_2 (y-axis). Left panel: Compromise situation with initial alternatives A_1 and A_2, and extreme alternative A_3 that makes A_1 a compromise between A_2 and A_3 (asymmetric compromise situation). Right panel: Compromise situation with initial alternatives A_1 and A_2, and compromise A_3 in between A_1 and A_2 (symmetric compromise situation). The numbers on the axes are arbitrary.

two sizes of coffee drinks, a small and cheap one and a tall and expensive one. Adding a third, even taller and more expensive one, should increase sales of the then medium option compared to the small option, according to the compromise effect. Formally, if A_3 is more extreme than alternative A_1 (see left panel of figure 2.3), that is, $m_{31} > m_{11}$ and $m_{32} < m_{12}$ (let m_{ij} be the attribute value for alternative A_i with respect to attribute D_j), a compromise effect is observed if A_3 increases the choice share of A_1:

$$\frac{P(A_1|A_1, A_2, A_3)}{P(A_2|A_1, A_2, A_3)} > \frac{P(A_1|A_1, A_2)}{P(A_2|A_1, A_2)}. \tag{2.4}$$

(and vice versa if the additional alternative is more extreme than alternative A_2, for example, an even smaller and cheaper coffee drink). In this article, we will refer to the compromise effect that results from adding a more extreme alternative as *asymmetric compromise effect* or *asymmetric version of the compromise effect*.

According to Simonson (1989), potential reasons for choosing the middle option from a ternary set were that it combines both attributes and is a compromise or safe option. Like dominance and relative superiority in the attraction situation, the compromise argument serves as a tie-breaker when the decision maker is indifferent between the two original choice options. However, in Simonson's (1989) experiments, the need to justify one's choice had no significant influence on the strength of the compromise effect, but the think-aloud protocols showed a similar pattern as for the attraction effect: Choosing the compromise option due to its balance between the two attributes required more elaboration and took more time than choosing one of the extreme options, e.g. A_2, due to its high value on the subjectively most important dimension, e.g. D_2 (cf. left panel of figure 2.3). Note that, the more similar the additional extreme alternative A_3 is to its adjacent alternative A_1, the more share it takes away from A_1 via the similarity effect. The theory of reason-based choice cannot explain the similarity effect and to avoid confusion of the two effects, Simonson (1989) placed the new extreme option A_3 at some distance from the original choice alternatives in his experiments.

Simonson and Tversky (1992) and Tversky and Simonson (1993) extended the concept of loss aversion to account also for the compromise effect. They assumed that, instead of focusing on a single reference point, the decision maker uses all presented alternatives as reference points for each other and evaluates the available choice options based on their disadvantages and advantages compared to the other alternatives. In the compromise setting with two extreme alternatives, A_2 and A_3, and a middle option, A_1, this leads to extremeness aversion favoring the compromise A_1, if the decision maker tries to avoid large losses (disadvantages of the extreme options compared to each other) more than to seek large gains. The compromise option features only small disadvantages (and advantages) and therefore might seem more attractive to the decision maker.

2.3 The first cognitive process models of context effects

Multi-alternative decision field theory (MDFT, Roe et al., 2001), the multi-alternative extension of decision field theory (DFT, Busemeyer & Townsend, 1993) was the first cognitive process model of context effect. DFT, unlike earlier theories of preference reversals, dissected the decision making process into microprocesses, the sequence of which was determined by psychologically interpretable parameters. More precisely, it proposed parameterized mechanisms for sampling and integrating pieces of information about the choice alternatives over time and for stopping this process and making a choice. DFT belongs to the class of *computational cognitive process models* (or *information processing models*), which have several advantages over simply verbal theories as well as complex neural models: They produce logically valid and precise quantitative predictions, are generalizable and at the same time computationally feasible (Busemeyer & Diederich, 2010). The analysis on the level of microprocesses made DFT particularly generalizable and applicable to a range of preference reversals (e.g. Roe et al., 2001; Johnson & Busemeyer, 2005; Busemeyer et al., 2006). MDFT was the first theory to consider similarity, attraction, and compromise effects simultaneously and explain them "with a common set of principles" (Roe et al., 2001, p.370), that is, by means of a specific combination of information processing mechanisms.

Together with the leaky competing accumulator model (LCA, Usher & McClelland, 2001), which was applied to the same three "anomalies" shortly afterwards (Usher & McClelland, 2004) and explained them with slightly different information processing mechanisms, MDFT set off development of a number of multi-alternative multi-attribute decision making models and established similarity, attraction, and compromise effects as benchmark for such models. Additionally, considering the three effects simultaneously inspired a series of experiments studying all three effects or variants thereof, yielding new insights about their interactions and thus about multi-alternative multi-attribute decision making in general.

Basic elements of cognitive process models of decision making

In order to describe MDFT and the LCA model, and later the more recent models, we first identify some basic elements of cognitive process models of decision making. Similar to the building blocks ("search rule", "stopping rule", and "decision rule") that Gigerenzer et al. (1999) propose for heuristics (see also Gigerenzer & Gaissmaier, 2011), or the (incomplete) taxonomy that Turner, Schley, Muller, and Tsetsos (2018) use for evaluating some of the models also discussed here, the basic elements help us, and may help other researchers in the future, to describe the models in a systematic and comparable way. The basic elements we use in the following are:

(1) Attention allocation (1.a) between attributes, and (1.b) between pairs of attribute values: Each model has to define attention probabilities for the attributes (so-called attribute weights) and – within attributes – for pairs of attribute values. **(2) Evaluation of alternatives by (2.a) selecting a focus value, and (2.b) comparing it with a reference value:** Given a pair of attribute values, each model has to assign one of the values as focus value and evaluate the corresponding alternative by comparing the focus value with a reference value (e.g., with the second value from the pair). **(3) Evidence accumulation by (3.a) setting up accumulators, (3.b) updating accumulators (including mechanisms for competition and noise), and (3.c) defining stopping rules:** Each model has to specify how many accumulators gather evidence or preference (for each alternative), how these accumulators are updated over time, and when and why updating stops and a decision is made.

Note that all the theories discussed here model so-called choice from description, that is, they assume that all relevant information about the choice alternatives is readily available for the decision maker to sample and evaluate. Let n_a be the number of alternatives under consideration, $\{A_i\}_{i=1,\ldots,n_a}$, and n_d the number of attributes, $\{D_j\}_{j=1,\ldots,n_d}$, that characterize them. The decision maker is provided with one attribute value per alternative per attribute, that is, $n_a \times n_d$ attribute values in total. Let $m_{ij}, i \in \{1,\ldots,n_a\}, j \in \{1,\ldots,n_d\}$ be the attribute value for alternative A_i with respect to attribute D_j.

Multi-alternative decision field theory

Before explaining how MDFT (Roe et al., 2001) accounts for the three context effects, we shortly describe its core mechanisms by means of the basic elements:

(1) Attention allocation: MDFT assumes that attention over time switches stochastically between attributes. This is reflected in the model by momentary attribute weights $w_j(t), j \in \{1,\ldots,n_d\}$, with $P(w_j(t) = 1, w_l(t) = 0 \forall l \neq j) = \omega_j$ for all $t \geq 0$, and $0 \leq \sum_{j=1}^{n_d} \omega_j \leq 1$. ω_j can be interpreted as attention probability for attribute D_j. Within attributes, however, MDFT assumes that all possible pairs of attribute values are attended simultaneously.

(2) Evaluation of alternatives: In fact, MDFT assumes that each of the n_a attribute values on the momentarily attended dimension D_j is contrasted with the average of the remaining $n_a - 1$ attribute values on the same dimension, that is, all attribute values serve as focus value and reference value simultaneously. The evaluation results in so called valences

$$v_{ij} = m_{ij} - \frac{\sum_{k \neq i} m_{kj}}{n_a - 1}.$$

Note that the valences sum up to 0 for each attribute $D_j, j \in \{1, \ldots, n_d\}$:

$$\sum_{i=1}^{n_a} v_{ij} = 0.$$

(3) Evidence accumulation: Over time, the valences are gathered in n_a accumulators, $\{S_i\}_{i=1,\ldots,n_a}$, that is, separately for each alternative. The initial accumulator states $S_i(0), i \in \{1, \ldots, n_a\}$, are set to 0. For each alternative $A_i, i \in \{1, \ldots, n_a\}$, the accumulator state $S_i(t)$ at time t depends on the current accumulator state $S_i(t-1)$, the current accumulator states $S_k(t-1), k \neq i$ for the other alternatives, and the momentary valence

$$v_i(t) = \sum_{j=1}^{n_d} (\omega_j(t) \cdot v_{ij}) + \xi_i(t)$$

for alternative A_i at time t, where $\xi_i(t)$ is a normally distributed error or noise term. Overall, accumulators are updated according to the following equation:

$$S_i(t) = \delta \cdot S_i(t-1) + \sum_{k \neq i} b_{ik} S_k(t-1) + v_i(t),$$

with decay parameter $0 \leq \delta \leq 1$, and distance-dependent inhibition factor $b_{ik} \leq 0$. The negative interconnections between preference states, reflected by the inhibition factors $b_{ik}, i, k \in \{1, \ldots, n_a\}$, are stronger for similar alternatives (located close to each other in the attribute space) than for dissimilar ones (located far from each other in the attribute space). This yields lateral inhibition, a local competitive influence between alternatives. Evidence accumulation stops either at a *fixed* (and usually externally imposed) *stopping time* T, in which case the alternative with the highest preference state is chosen. Or accumulation stops as soon as the preference state for one of the alternatives exceeds a positive threshold θ^+. In this case, the stopping time is set by the decision maker and called *optional stopping time* (Busemeyer & Diederich, 2002). Roe et al. (2001, Appendix B) provide mathematical formulas for the choice probabilities for fixed stopping times. Only this version of the model has been applied to data so far (e.g., Berkowitsch et al., 2014).

The following explanations that MDFT provides for the similarity, attraction, and compromise effects are also based on fixed stopping times, that is, no decision threshold is required and the choice probability for an alternative A_i at time T is equal to the probability that $S_i(T) > S_k(T)$ for all $k \neq i$ (tied alternatives are chosen with equal probability).

According to MDFT, the **similarity effect** is caused by a positive correlation between the valences of the similar alternatives A_1 and A_3 (cf. right panel of figure 2.1) and a negative correlation between the valences of those two and the dissimilar alternative A_2. Whenever the decision maker focusses on dimension D_1, the preference states of alternatives A_1 and A_3 increase and the preference state of alternative A_2 decreases, and vice versa for dimension D_2. The positively correlated alternatives A_1 and A_3 split their choice probabilities, leading to an advantage for the dissimilar, negatively correlated option A_2. High variance of the normally distributed error terms $\xi_i, i \in \{1, \ldots, n_a\}$ strongly diminishes the effect since it covers the correlation. Lateral inhibition, too, diminishes the similarity effect but is necessary for explaining the other two effects.

As for the **attraction effect**, lateral inhibition in MDFT promotes the dominating option A_1 (cf. figure 2.2) by negating the negative preference for the dominated adjacent alternative A_3. There is no such effect for the distant alternative A_2, which leads to an overall advantage of alternative A_1.

Instead of explaining the asymmetric version of the **compromise effect** described above (cf. left panel of figure 2.3), Roe et al. (2001) introduced a *symmetric compromise effect*, where a compromise option A_3 is added in between the two original choice alternatives A_1 and A_2 (see right panel of figure 2.3). MDFT explains the symmetric version of the compromise effect by assuming lateral inhibition between the compromise and each of the extreme alternatives, but not between the two extreme options. Since the mean valence input is zero for all three alternatives, lateral inhibition operates only on momentary fluctuations of valences. By that, the preference for choosing the compromise is negatively correlated with the preferences for choosing either of the extreme options and thus the differences between the compromise option and each of the extremes are positively correlated. Like in the similarity situation, the positively correlated options (that is, the extremes A_1 and A_2) split their choice probabilities, leading to an advantage for the negatively correlated compromise alternative A_3. Asymmetric versions of the compromise effect, as observed by Simonson (1989) and described above, can be explained within MDFT only by assuming different psychological distances between the compromise and each of the extreme alternatives, leading to differing strengths of lateral inhibition.

Overall, MDFT predicts the similarity effect for relatively low levels of lateral inhibition and a low variance in the normally distributed error terms. On the other hand, the compromise effect is predicted for higher levels of lateral inhibition and higher variance in the error terms. For the attraction affect, the standard deviation of the error term plays a minor role, but high lateral inhibition promotes the effect. On the individual level, the similarity and the compromise effect should thus be negatively correlated and the attraction effect should be positively correlated with the compromise effect and negatively correlated with the similarity effect. There are, however, levels of lateral inhibition and variance in the error terms where all three effects occur. The individual differences predicted by MDFT were later indeed observed by Berkowitsch et al. (2014), but see below for a detailed description of the findings.

Leaky Competing Accumulator Model

The LCA model (Usher & McClelland, 2001, 2004) has been shown to account for the three context effects shortly after MDFT. Again, we describe the core mechanisms of the model by means of the basic elements before explaining how it accounts for the three context effects:

(1) Attention allocation: Like MDFT, the LCA model assumes that attention over time switches stochastically between attributes. The attribute weights $\omega_j, j \in \{1, \ldots, n_d\}$, with $0 \leq \sum_{j=1}^{n_d} \omega_j \leq 1$, determine the momentary attribute weights $w_j(t)$, via the probability equation $P(w_j(t) = 1, w_l(t) = 0 \forall l \neq j) = \omega_j$ for all $t \geq 0$. All possible pairs of attribute values are attended simultaneously, like in MDFT.

(2) Evaluation of alternatives: In contrast to MDFT, however, the LCA model assumes that alternatives are evaluated asymmetrically, based on loss aversion. Let $i \neq k$ and $m_{ij} < m_{kj}$. Then the negative difference $(m_{ij} - m_{kj})$ for evaluating alternative A_i and the positive difference $(m_{kj} - m_{ij})$ for evaluating alternative A_k are weighted differently, according to the loss-averse value function

$$v(x) = \begin{cases} log(1+x), & \text{if } x \geq 0 \\ -log(1-x)(log(1-x))^2, & \text{if } x < 0. \end{cases} \tag{2.5}$$

With respect to the basic elements, this means that the smaller attribute value in a pair is more likely to be selected as focus value than the greater attribute value.

(3) Evidence accumulation: Like MDFT, the LCA model defines n_a leaky competing choice units $S_i, i \in \{1, \ldots, n_a\}$, with $S_i(0) = 0$, for the n_a choice alternatives. For each alternative $A_i, i \in \{1, \ldots, n_a\}$, the state $S_i(t)$ of the choice unit or accumulator at time t depends on the current state $S_i(t-1)$, the current accumulator states $S_k(t-1), k \neq i$ for the other alternatives, and the momentary input

$$I_i(t) = I_0 + \sum_{j=1}^{n_d} \left(w_j(t) \cdot \sum_{k \neq i} v(m_{ij} - m_{kj}) \right)$$

for alternative A_i at time t, where $I_0 > 0$ is a positive constant that prevents negative input values due to loss aversion. I_0 can be interpreted as promoting the available alternatives into the choice set (Usher & McClelland, 2004). Overall, accumulators are updated according to the following iterative equation:

$$S_i(t) = \delta \cdot S_i(t-1) + (1-\delta) \left(I_i(t) - \beta \sum_{k \neq i} S_k(t-1) - \xi_i(t) \right),$$

with decay parameter $0 \leq \delta \leq 1$, global inhibition factor $0 \geq \beta < 1$, and normally distributed noise term $\xi_i(t)$. The LCA model implements inhibition via a global factor β, reducing the states of the choice units proportionally to the overall activation. The leaky competing choice units are restricted to be non-negative, preventing the promotion of alternatives via negated lateral inhibition. The updating process stops at a fixed time T and the alternative with the highest preference state at that time is chosen.

The LCA model offers the following explanations for the three context effects. The **similarity effect** is explained by a subtle interplay between the loss-averse value function (equation 2.5), stochastic attention switching between attributes, inhibition and leaky integration of activations. The loss-averse value function (together with the positive constant I_0) promotes choice options that have less distant competitors. The two similar choice options A_1 and A_3 both have one distant competitor, A_2, while the dissimilar alternative A_2 has two distant competitors, A_1 and A_3 (cf. figure 2.1). Thus, the loss-averse value function leads to a disadvantage for alternative A_2. However, due to stochastic attention switching between attributes, the activations for A_1 and A_3 covary and the two alternatives share their choices. This leads to an advantage for the dissimilar option A_2. This advantage is further promoted by inhibition, which, though implemented globally, leads to a higher degree of competition between the similar alternatives A_1 and A_3 due to their covarying activations. Finally, the leaky integration of activations plays a crucial role in explaining the similarity effect since it allows activations to recover from time intervals where the similar alternatives A_1 and A_3 dominate the choice set due to the loss-averse value function. Overall, activations change between patterns favoring the similar options (a negative similarity effect) and patterns favoring the dissimilar option (a positive similarity effect).

The explanation for the **attraction effect** is based on the loss-averse value function. The dominating alternative A_1, which has only one distant competitor, A_2, is favored over the dissimilar alternative A_2, which has two distant competitors, A_1 and A_3 (cf. figure 2.2). Stochastic attention switching and inhibition play a minor role for the attraction affect. Since alternative A_3 is dominated by alternative A_1, activation for A_3 remains low throughout the process, preventing these mechanisms from altering the choice probabilities. Leaky integration, however, slightly diminishes the magnitude of the attraction effect.

The loss-averse value function also accounts for the (symmetric and asymmetric) **compromise effect**. The compromise alternative, A_3 in the symmetric version of the effect and A_1 in the asymmetric version (cf. figure 2.3), has two medium distant competitors, the two extreme alternatives. It is thus favored over the extreme options, which have each one medium distant competitor, the compromise option, and one distant competitor, the other extreme alternative. Leaky integration reduces the magnitude of the compromise effect. On the other hand, a medium level of inhibition can increase the magnitude of the compromise effect. The LCA model does not make any predictions about individual differences in context effect patterns. However, explanations for the attraction and compromise effect rely on the loss-averse value function, while the similarity effect is diminished by loss aversion. Furthermore, the level of inhibition might play a role in explaining such differences since it has different effects on the occurrence and magnitude of the three effects.

2.4 Recent empirical observations

Until recently, MDFT and the LCA model have not been tested on data. One reason for this was that, as Tsetsos, Usher, and Chater (2010) pointed out, "the three effects have not all been obtained in a single experiment" (p.1287). Since explaining similarity, attraction, and compromise effects simultaneously has become a benchmark for multi-alternative multi-attribute decision making models, several attempts have been made to obtain the three effects in a single experiment or at least within a single experimental paradigm (e.g., Trueblood, 2012; Trueblood et al., 2013; Berkowitsch et al., 2014; Noguchi & Stewart, 2014; Trueblood, Brown, & Heathcote, 2015; Liew et al., 2016; Turner et al., 2018). However, only the studies reported by Berkowitsch et al. (2014), Liew et al. (2016, Experiment 1), and Turner et al. (2018, Study 2) considered the "original" effects as describe above, where relative choice shares are compared between choice sets with two and three alternatives. Trueblood (2012), Trueblood et al. (2013), Noguchi and Stewart (2014), Trueblood et al. (2015), and Liew et al. (2016, Experiment 2) instead studied "ternary" variants of the three effects, based on the comparison of choice shares in situations with three choice alternatives each. Before summarizing the experimental results, we describe those ternary variants of the three context effects.

Ternary variants of the three context effects

The first ternary variant of the **compromise effect** was already proposed by Simonson (1989), who tested the effect "by moving choice set position" (p.165),

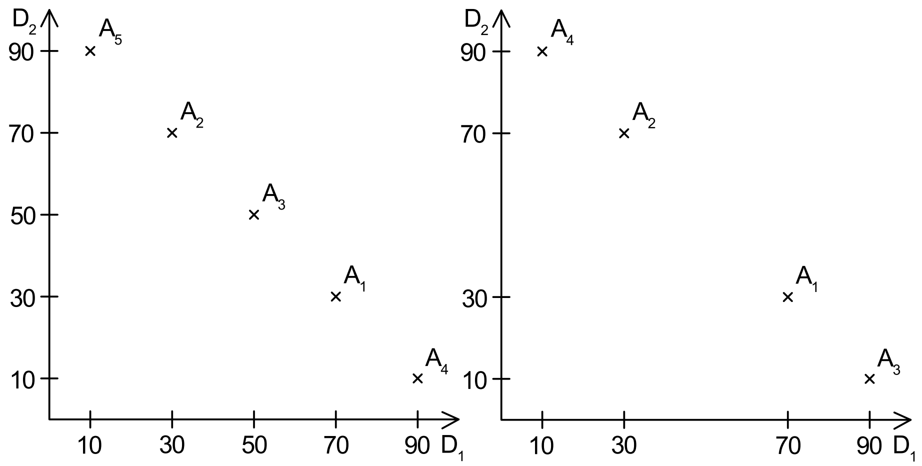

Figure 2.4: Placement of choice alternatives in the attribute space with dimensions D_1 (x-axis) and D_2 (y-axis). Left panel: Situation for testing the compromise effect by moving choice set position (ternary symmetric compromise effect). The three ternary choice sets are $\{A_4, A_1, A_3\}$, $\{A_1, A_3, A_2\}$, and $\{A_3, A_2, A_5\}$. Right panel: Alternative ternary variant of the compromise effect (ternary asymmetric compromise effect). The two choice sets are $\{A_3, A_1, A_2\}$ and $\{A_1, A_2, A_4\}$. The numbers on the axes are arbitrary.

constructing three ternary choice sets from five choice alternatives (ternary symmetric compromise effect, see left panel of figure 2.4): Starting with the two original choice alternatives A_1 and A_2 (with A_1 scoring high on attribute D_1 but low on attribute D_2, and vice versa for alternative A_2, see above), Simonson (1989) added a third alternative A_3 with medium values on both dimensions in between them. Furthermore, he introduced two extreme alternatives, A_4 and A_5, with A_4 being more extreme than A_1 (i.e., $m_{41} > m_{11}$ and $m_{42} < m_{12}$) and A_5 being more extreme than A_2 (i.e., $m_{52} > m_{22}$ and $m_{51} < m_{21}$). The three ternary choice sets are $\{A_4, A_1, A_3\}$, $\{A_1, A_3, A_2\}$, and $\{A_3, A_2, A_5\}$. Alternative A_1 is a compromise option in the first set but an extreme option in the second set, alternative A_2 is a compromise option in the third set but an extreme option in the second set, and alternative A_3 is a compromise option in the second set but an extreme option in the first and third set. Thus, a ternary symmetric compromise effect is observed if $P(A_1|A_4, A_1, A_3) > P(A_1|A_1, A_3, A_2)$ or $P(A_2|A_3, A_2, A_5) > P(A_2|A_1, A_3, A_2)$ or $P(A_3|A_1, A_3, A_2) > P(A_3|A_4, A_1, A_3)$ or $P(A_3|A_1, A_3, A_2) > P(A_3|A_3, A_2, A_5)$. Trueblood (2012) and Liew et al. (2016, Experiment 2) considered this variant of the compromise effect in their experiments, see table 2.1.

Trueblood et al. (2013), Noguchi and Stewart (2014), and Trueblood et al. (2015) studied yet another variant of the **compromise effect**, comparing choice probabilities in two ternary choice sets based on the original asymmetric version of the effect (ternary asymmetric compromise effect, see right panel of figure 2.4 and table 2.1): Starting with the original choice alternatives A_1 and A_2, Trueblood et al. (2013) introduced two extreme alternatives A_3 and A_4 such that A_1 became a compromise between A_2 and A_3, and A_2 became a compromise between A_1 and A_4.

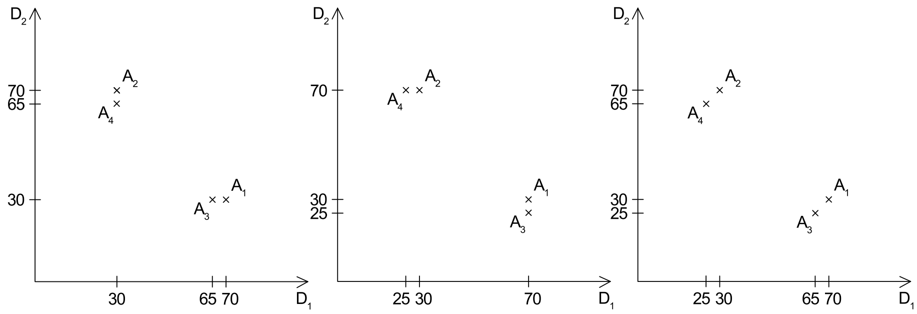

Figure 2.5: Placement of choice alternatives for testing the ternary variants of the attraction effect in the attribute space with dimensions D_1 (x-axis) and D_2 (y-axis). The two choice sets are $\{A_1, A_2, A_3\}$ and $\{A_1, A_2, A_4\}$. Left panel: A_3 and A_4 are range decoys (ternary range attraction effect). Middle panel: A_3 and A_4 are frequency decoys (ternary frequency attraction effect). Right panel: A_3 and A_4 are range-frequency decoys (ternary range-frequency attraction effect). The numbers on the axes are arbitrary.

The two ternary choice sets are $\{A_3, A_1, A_2\}$ and $\{A_1, A_2, A_4\}$ and a ternary asymmetric compromise effect is observed if $P(A_1|A_3, A_1, A_2) > P(A_1|A_1, A_2, A_4)$ or $P(A_2|A_1, A_2, A_4) > P(A_3|A_3, A_1, A_2)$.

For testing the **attraction effect**, Trueblood (2012), Trueblood et al. (2013), Noguchi and Stewart (2014), Trueblood et al. (2015), and Liew et al. (2016, Experiment 2) constructed two ternary choice sets from four choice alternatives, following Wedell (1991, see figure 2.5 and table 2.1): Again starting with the original alternatives A_1 and A_2, they introduced two asymmetrically dominated decoys, A_3 and A_4, with A_3 being similar to but dominated by A_1, and A_4 being similar to but dominated by A_2. Further differentiation is possible with regard to the dimensions on which the decoys are dominated, that is, whether they are frequency decoys, range decoys or range-frequency decoys (see above). The two ternary choice sets are $\{A_1, A_2, A_3\}$ and $\{A_1, A_2, A_4\}$ and a ternary variant of the attraction effect is observed if $P(A_1|A_1, A_2, A_3) > P(A_1|A_1, A_2, A_4)$ or $P(A_2|A_1, A_2, A_4) > P(A_2|A_1, A_2, A_3)$.

For testing the **similarity effect**, Trueblood (2012), and Liew et al. (2016, Experiment 2) constructed four ternary choice sets from seven alternatives (ternary symmetric similarity effect, see left panel of figure 2.6 and table 2.1): Starting with the original alternatives A_1 and A_2, and the symmetric compromise A_3, they introduced four alternatives that are each similar to one of the other alternatives: A_4 is similar to but slightly more extreme than A_1, A_5 is similar to but slightly more extreme than A_2, and A_6 and A_7 are each similar to the compromise A_3, though lying on different sides of A_3. Here, A_6 lies between A_3 and A_2, and A_7 lies between A_3 and A_1 (but both closer to A_3 than to A_1 or A_2). The four ternary choice sets are $\{A_1, A_3, A_6\}$, $\{A_3, A_1, A_4\}$, $\{A_2, A_3, A_7\}$, and $\{A_3, A_2, A_5\}$. A (variant of the) similarity effect is observed if $P(A_1|A_3, A_1, A_4) < P(A_1|A_1, A_3, A_6)$ or $P(A_3|A_1, A_3, A_6) < P(A_3|A_3, A_1, A_4)$ or $P(A_2|A_3, A_2, A_5) < P(A_2|A_2, A_3, A_7)$ or $P(A_3|A_2, A_3, A_7) < P(A_3|A_3, A_2, A_5)$.

Trueblood et al. (2013), Noguchi and Stewart (2014), and Trueblood et al. (2015) constructed only two ternary choice sets from four choice alternatives

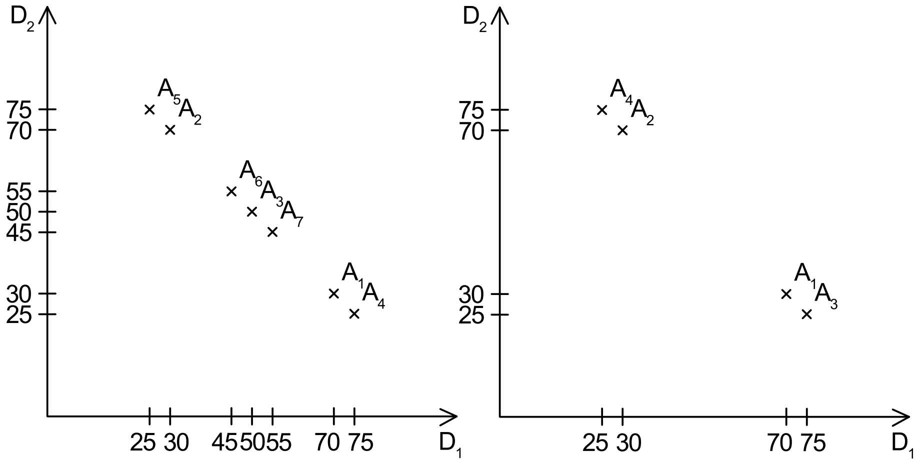

Figure 2.6: Placement of choice alternatives for testing two different ternary variants of the similarity effect in the attribute space with dimensions D_1 (x-axis) and D_2 (y-axis). Left panel: The four choice sets are $\{A_1, A_3, A_6\}$ and $\{A_3, A_1, A_4\}$, and $\{A_2, A_3, A_7\}$ and $\{A_3, A_2, A_5\}$ (ternary symmetric similarity effect). Right panel: The two choice sets are $\{A_1, A_2, A_3\}$ and $\{A_1, A_2, A_4\}$ (ternary asymmetric similarity effect). The numbers on the axes are arbitrary.

for testing another variant of the **similarity effect** (ternary asymmetric similarity effect, see right panel of figure 2.6 and table 2.1): Starting with the original alternatives A_1 and A_2, the authors introduced two additional alternatives A_3 and A_4, with A_3 similar to A_1 and A_4 similar to A_2. The two choice sets are $\{A_1, A_2, A_3\}$ and $\{A_1, A_2, A_4\}$ and a (variant of the) similarity effect is observed if $P(A_1|A_1, A_2, A_3) < P(A_1|A_1, A_2, A_4)$ or $P(A_2|A_1, A_2, A_4) < P(A_2|A_1, A_2, A_3)$. Please refer to table 2.1 for an overview of the 13 variants and versions of the three context effects and their usage in the seven studies described in the next section.

Empirical observations

Trueblood (2012) observed the three context effects in three separate experiments within the same inference paradigm. Participants had to "infer which suspect out of a set of three is most likely to have committed a crime, on the basis of two separate eyewitness testimonies" (p.963). The strengths of the eyewitness testimonies were given as ratings on a 0-100 scale, so the task was similar to choosing a college applicant based on intelligence and motivation scores as in Tversky's (1972b) original study of the similarity effect. Trueblood (2012) reported that in the ternary symmetric similarity experiment all four probability inequalities are significant across participants. In the ternary attraction experiment, both probability inequalities were significant across participants for each decoy type, that is for range decoys and frequency decoys as well as range-frequency decoys. In the ternary symmetric compromise experiment, three out of four probability inequalities were significant across participants. Individual choice probabilities were presented only in scatter plots. However, those plots suggest that a considerable number of participants

	similarity			attraction						compromise			
	binary	ternary symmetric	ternary asymmetric	binary range	binary frequency	binary range-frequency	ternary range	ternary frequency	ternary range-frequency	binary asymmetric	binary symmetric	ternary symmetric	ternary asymmetric
(Tversky, 1972b)	+												
(Huber et al., 1982)				+	+	+							
(Simonson, 1989)										+		+	
(Trueblood, 2012)		+					+	+	+			+	
(Trueblood et al., 2013)			+				+	+	+				+
(Berkowitsch et al., 2014)	+					+				+			
(Noguchi & Stewart, 2014)			+						+				+
(Trueblood et al., 2015)			+				+	+	+				+
(Liew et al., 2016), Exp 1		+					+	+	+			+	
(Liew et al., 2016), Exp 2	+					+				+			
(Turner et al., 2018)	+					+					+		

Table 2.1: Variants and versions of similarity, attraction, and compromise effects used in seven studies that tried to obtain the three effects in a single experiment or within the same experimental paradigm. The original studies by Tversky (1972b), Huber et al. (1982), and Simonson (1989) are included for reference.

actually showed negative context effects, and reverse probability inequalities hold for those participants.

Trueblood et al. (2013) demonstrated the three context effects in perceptual choice. Participants had to choose out of three rectangles with different widths and heights (the two attributes) the one with the supposedly largest area. A similar task had been used by Tversky and Russo (1969). Trueblood et al. (2013) tested similarity, attraction, and compromise effects in three separate experiments. In the ternary asymmetric similarity experiment, both probability inequalities were significant across participants and for 69% of individual participants. In the ternary attraction experiment, all probability inequalities were significant across decoy types and across participants, though only the range decoy and the range-frequency decoy produced a significant attraction effect across participants. Individually, 69% of participants showed a significant attraction effect with the range decoy, 61% showed a significant attraction effect with the range-frequency decoy, and 59% showed a significant attraction effect with the frequency decoy. In the ternary asymmetric compromise experiment, both probability inequalities held, though they were not significant across participants. Nevertheless, 66% of participants showed the effect.

Berkowitsch et al. (2014) tested the original (binary vs. ternary set) versions of the similarity, attraction, and compromise effects in a within-subject consumer choice paradigm. Several days before the main experiment, participants completed a matching task, in which they "repeatedly filled in missing attribute values (e.g., price) so that two products (e.g., a heavier and a lighter racing bike) became equally attractive" (p.1336). During the main experiment, participants had to choose between triplets of products that consisted of the two matched options and a third option according to the similarity, attraction, and asymmetric compromise situations described above. Relative choice shares of the "target" and "competitor" options in the ternary sets were compared to 0.5, the assumed choice probability for the previously matched options in a (not tested) binary choice situation. Across participants, Berkowitsch et al. (2014) observed a strong binary range-frequency attraction effect, a reliable binary asymmetric compromise effect and a weak binary similarity effect. For 19% of participants, all three effects reached significance. However, on the individual level, the authors reported a positive correlation between the relative choice shares of the target options in the attraction and compromise conditions as well as a negative correlation between the relative choice shares in the attraction and similarity conditions, and in the compromise and similarity conditions. Thus, participants who showed the similarity effect were less likely to also show the attraction and compromise effect, and vice versa. This confirms the individual differences predicted by multi-alternative decision field theory (Roe et al., 2001) and described above.

Noguchi and Stewart (2014) studied ternary variants of the three context effects in a within-subject consumer choice paradigm while recording eye-movements. Similar to the study by Berkowitsch et al. (2014), a strong (ternary range-frequency) attraction effect, a medium (ternary asymmetric) compromise effect and a weak (ternary asymmetric) similarity effect were observed across participants. Noguchi and Stewart (2014) divided the participants into two groups according to their performance in 10 "catch trials" which included one option that dominated the other two on both dimensions. All three context effects were significant for "engaged" participants who did not make any mistakes in the catch trials. For the remaining participants, only the attraction and the compromise effect reached significance. Regarding eye-movements, the authors focused on transitions between pieces of information. They found that more transitions were made within attributes and between alternatives, and concluded that "alternatives are repeatedly compared in pairs on single dimensions" (p.44) during multi-alternative multi-attribute choice.

Trueblood et al. (2015) replicated the three perceptual choice experiments from Trueblood et al. (2013) in a single experiment, exploring also individual differences in context effect patterns. Like Berkowitsch et al. (2014), they tested if the relative choice shares for the "target" options deviated from 0.5, supposedly comparing binary with ternary choices but without actually assessing binary choice probabilities (nor matching rectangle sizes to being perceived as equally large). Overall, a strong (ternary asymmetric) similarity effect, a medium (ternary) attraction effect and a weak (ternary asymmetric) compromise effect were observed. Even though this pattern deviates from the study by Berkowitsch et al. (2014), the remaining results are quite similar: Trueblood et al. (2015) observed that only a relatively small proportion of participants (23.6%) showed all three context effects and that attraction and compromise effects were positively correlated with each other and negatively correlated with the similarity effect.

Liew et al. (2016) criticized previous studies for analyzing grouped data (Trueblood, 2012; Trueblood et al., 2013) or relative choice shares (Berkowitsch et al., 2014; Trueblood et al., 2015). The former, because averaging over participants does not take into account individual differences between participants, and the latter, because relative choice shares ignore decoy choice proportions which may indicate dimensional biases. Instead, they proposed to cluster data according to choice frequencies before analyzing them. They replicated the three inference experiments from Trueblood (2012) and the consumer choice experiment from Berkowitsch et al. (2014). Four clusters were revealed for the ternary symmetric similarity effect inference experiment, with three of them showing a positive similarity effect and one of them showing no similarity effect. Two clusters were revealed for the ternary attraction effect inference experiment, with the bigger cluster showing no attraction effect but a strong dimensional bias towards dimension D_1, and the smaller cluster showing a weak attraction effect. Six clusters were revealed for the ternary symmetric compromise effect inference experiment, with the largest two clusters showing a strong negative compromise effect (that is, a preference for one of the extreme alternatives), and the third largest cluster showing a strong positive compromise effect. The remaining clusters all showed a weak negative compromise effect. For the consumer choice experiment, three clusters were revealed. Participants in the first cluster showed a positive binary range-frequency attraction effect, a positive binary asymmetric compromise effect, and a negative binary similarity effect. Participants in the second cluster showed a positive binary similarity effect, no binary range-frequency attraction effect and a negative binary asymmetric compromise effect together with a dimensional bias towards dimension D_2. Participants in the third cluster showed a positive binary similarity effect, no binary range-frequency attraction effect and a negative binary asymmetric compromise effect together with a dimensional bias towards dimension D_1. Overall, the differences between clusters are remarkable, and some of them even show reversed effects as compared to the averaged data.

Finally, Turner et al. (2018) tested "binary vs. ternary set" variants of the similarity, attraction, and symmetric compromise effects within Trueblood et al.'s (2013) perceptual choice paradigm. The data Turner et al. (2018) reported is averaged across participants and dimensions, that is, individual differences or dimensional biases are not taken into account. Their operationalization yielded a positive binary range-frequency attraction effect, a negative binary similarity effect (which they attributed to their placement of the similarity decoy in-between the original alternatives), and no binary symmetric compromise effect. As for the compromise effect, Turner et al. (2018) argued that the ternary variants of the compromise effect found by Trueblood et al. (2013, 2015) could actually be artifacts. However, another difference between the two studies is that Trueblood et al. (2013, 2015) used an asymmetric version of the compromise effect, while Turner et al. (2018) used a symmetric version. Like Berkowitsch et al. (2014) and Trueblood et al. (2015), Turner et al. (2018) found a negative correlation between the attraction effect and the similarity effect. Furthermore, they reported a weak negative correlation between the attraction effect and the compromise effect and no correlation between the compromise effect and the similarity effect, unlike in previous studies. Again, this might be due to Turner et al.'s (2018) use of the symmetric version of the compromise effect.

2.5 Recent cognitive process models of context effects

In parallel to those empirical observations of the three context effect, since 2012, theoretical advancements in explaining them have been made as well. In fact, three cognitive process models (and at least two static models) have been proposed that are able to explain all three effects, with at least partly different information processing mechanisms than both MDFT and the LCA model. The three cognitive process models are the 2N-ary choice tree model (2NCT, Wollschlaeger & Diederich, 2012) and its variant simple choice tree model (SCT, Wollschlaeger & Diederich, 2017), the associative accumulation model (AAM, Bhatia, 2013), and multi-alternative decision by sampling (MDbS, Noguchi & Stewart, 2018). The multi-attribute linear ballistic accumulator model (MLBA, Trueblood et al., 2014) and multi-attribute decision by sampling (MADS, Ronayne & Brown, 2017) also account for the three context effects but are static models und thus not cognitive process models in a strict sense. In the last section of this review, we want to show possible directions for future research on multi-alternative multi-attribute decision making based on the insights from recent empirical and theoretical advancements. Therefore, and in order for this review to be self-contained, we will now describe the 2NCT/SCT models, AAM, the MLBA model, and MDbS by means of the basic elements identified above and show how they account for the three context effects. Again, let n_a be the number of alternatives under consideration, $\{A_i\}_{i=1,\ldots,n_a}$, n_d the number of attributes that characterize them, $\{D_j\}_{j=1,\ldots,n_d}$, and $m_{ij}, i \in \{1,\ldots,n_a\}, j \in \{1,\ldots,n_d\}$ the attribute value for alternative A_i on dimension D_j.

(Simple) 2N-ary choice tree model

Since the 2NCT model (Wollschlaeger & Diederich, 2012) and the SCT model (Wollschlaeger & Diederich, 2017) overlap in large part, we describe them together and highlight differences where applicable.

(1) Attention allocation: The 2NCT/SCT models assume that attention over time switches stochastically between attributes and, within attributes, between pairs of attribute values. This is reflected in the model on the one hand by attribute weights ω_j, $j \in \{1,\ldots,n_a\}$, with $0 \leq \sum_{j=1}^{n_d} \omega_j \leq 1$, like in MDFT and the LCA model, and on the other hand by comparison weights $p_{\{ij,kj\}}$, $i \neq k \in \{1,\ldots,n_a\}$:

$$p_{\{ij,kj\}} = p_{ij,kj} + p_{kj,ij} = \frac{d_{ij,kj} + d_{kj,ij}}{\sum_{l \neq m \in \{1,\ldots,n_a\}} (d_{lj,mj} + d_{mj,lj})},$$

with $d_{ij,kj} = |m_{ij} - m_{kj}|$. The comparison weights serve as attention probabilities for pairs of attribute values within attributes in the 2NCT/SCT models. Note that the comparison weights depend on the relative difference of attribute values, such that pairs that differ more get more attention.

(2) Evaluation of alternatives: In order to evaluate alternatives in the 2NC-T/SCT models, one attribute value within a pair is selected as focus value and the other automatically becomes the reference value. Only the alternative associated with the focus value is evaluated, by an ordinal comparison of the two values. In the original version of the 2NCT model, both attribute values in a pair are selected as focus value equally often. However, it is possible to apply LCA's asymmetric, loss-averse value function to the differences $(m_{ij} - m_{kj})$, which leads to a focus shift towards the smaller value in each pair (cf. Wollschlaeger & Diederich, 2012). In the SCT model (Wollschlaeger & Diederich, 2017), selection of the focus value depends on a so-called focus weight $\lambda, 0 \leq \lambda \leq 1$. Let $i \neq k$ and $m_{ij} < m_{kj}$. Then m_{ij}, the smaller one of the two values, is selected as focus value with probability λ, and m_{kj}, the larger one, is selected as focus value with probability $1 - \lambda$. By setting $\lambda > 0.5$, that is, by shifting the focus towards the smaller value of each pair, the SCT model is able to mimic loss-aversion as implemented in the LCA model.

(3) Evidence accumulation: The 2NCT/SCT models, in contrast to MDFT and the LCA model, define two accumulators or counters per alternative, that is,

$2 \times n_a$ counters in total. The positive counter S_i^+ accumulates evidence in favor of choosing alternative $A_i, i \in \{1, \dots, n_a\}$, the negative counter S_i^- accumulates evidence against choosing it. The initial counter states are set to zero, $S_i^{\pm}(0) = 0$, and for each comparison between two attribute values the state of one counter is increased by one. For example, let $i \neq k$ and $m_{ij} < m_{kj}$. If m_{ij} is the focus value and m_{kj} is the reference value at time t, the negative counter for alternative A_i is updated, $S_i^-(t) = S_i^-(t-1) + 1$, and if m_{kj} is the focus value and m_{ij} is the reference value, the positive counter for alternative A_k is updated, $S_k^+(t) = S_k^+(t-1) + 1$. For each counter, the updating probability $p_i^{\pm}, i \in \{1, \dots, n_a\}$ at each time t is composed of attribute weights, comparison weights, and the focus weight λ:

$$p_i^{\pm} = \sum_{j=0}^{n_d} p_{ij}^{\pm} \cdot \omega_j,$$

with

$$p_{ij}^- = \sum_{k:(m_{ij} < m_{kj})} \lambda \cdot p_{ij,kj}, \quad p_{ij}^+ = \sum_{k:(m_{ij} > m_{kj})} (1-\lambda) \cdot p_{ij,kj},$$

for $j \in \{1, \dots, n_d\}$, $p_{i0}^{\pm} = \frac{1}{2 \cdot n_a}$, and $\omega_0 = 1 - \sum_{j=1}^{n_d} \omega_j$. Note that $\sum_{i=1}^{n_a} (p_i^+ + p_i^-) = 1$. On average, only one counter per time step is updated, leading to competition between the choice alternatives. Furthermore, randomness is introduced to the counter updating process via $p_{i0}^{\pm}$, given that $\omega_0 > 0$. Wollschlaeger and Diederich (2012) additionally implement leakage and inhibition into the 2NCT model, in order to mimic MDFT and the LCA model more closely. However, both mechanisms are omitted later, thus we do not describe them here. The counter updating process in the 2NCT/SCT models stops when enough evidence has been accumulated to make the required choice. For each alternative, the difference of the states of the positive and the negative counter at time t, $Pref_i(t) = S_i^+(t) - S_i^-(t)$, the so-called momentary preference state, is constantly compared to two thresholds, a positive threshold θ^+, and a negative threshold $\theta^- = -\theta^+$. If the momentary preference state for alternative A_i hits the positive threshold, the process stops and A_i is chosen. If, on the other hand, the momentary preference state for alternative A_k hits the negative threshold, A_k is eliminated from the choice set and the process continues with the remaining alternatives until one of them is chosen or until all but one of them have been eliminated.

Three interacting mechanisms produce similarity, attraction, and compromise effects in the 2NCT/SCT models: (1) Attention allocation between pairs of attribute values based on normalized differences, (2) the possibility to eliminate unwanted alternatives from the choice set, and (3) weighting of attributes based on salience.

The **similarity effect** is produced as follows: The first mechanism leads to a higher impact of dissimilar alternatives on the updating probabilities and thus faster evidence accumulation for alternatives with more distant competitors. In the similarity setting, evidence for the dissimilar alternative A_2 (cf. figure 2.1) with two distant competitors A_1 and A_3 is accumulated fastest. In the 2NCT model without loss-aversion and in the SCT model with medium to low $\lambda \leq 0.5$, this leads to increased choice probabilities for alternative A_2 and thus a similarity effect. On the other hand, in the 2NCT model with loss aversion and in the SCT model with $\lambda > 0.5$, this leads to a negative similarity effect, since the dissimilar alternative A_2 is eliminated from the choice set with a relatively high probability. Both effects, the negative and the positive similarity effect, can be either strengthened or reversed by means of the attribute weights.

The **attraction effect** is produced as follows: In the attraction setting, like in the similarity setting, evidence for the dissimilar alternative A_2 (cf. figure 2.2) is accumulated fastest. In the 2NCT model with loss-aversion and in the SCT model with $\lambda > 0.5$, this leads to faster elimination of alternative A_2, followed by elimination of the dominated alternative A_3, choice of the dominating alternative A_1, and thus an attraction effect. On the other hand, in the 2NCT model without loss-aversion and in the SCT model with $\lambda \leq 0.5$, the increased choice probability for the dissimilar alternative A_2 leads to a negative attraction effect. Again, the attribute weights are able to strengthen or reverse both the negative and the positive attraction effect.

The **compromise effect** is produced as follows: In the compromise setting, evidence for the two extreme alternatives (cf. figure 2.3) is accumulated fastest since they have both one distant competitor (the other extreme alternative) and one medium distant competitor (the compromise alternative), while the compromise alternative has two medium distant competitors (the two extreme alternatives). Note that the differences between the alternatives are normalized before being serving as attention probabilities and therefore only relative distances play a role here. The following explanations thus apply to symmetric and asymmetric versions of the compromise effect. In the 2NCT model with loss-aversion and in the SCT model with $\lambda > 0.5$, faster evidence accumulation for the extreme alternatives leads to their elimination, leaving the decision maker with the compromise option. On the other hand, in the 2NCT model without loss-aversion and in the SCT model with $\lambda \leq 0.5$, faster evidence accumulation for the extreme alternatives increases their choice probability, leading to a negative compromise effect. The interplay between attribute weighting and the other mechanisms in the 2NCT/SCT models is quite complex in the compromise setting. In summary it can be said that choice of an extreme alternative becomes more likely with unbalanced attribute weights.

Associative accumulation model

(1) Attention allocation: Like MDFT and the LCA and 2NCT/SCT models, the AAM assumes that attention switches stochastically between attributes over time. However, the attribute weights that drive attention allocation in the AAM depend on accessibility of the attributes, defined as a weighted sum of attribute values:

$$a_j = \alpha_0 + \sum_{i=1}^{n_a} \omega_i \cdot m_{ij},$$

with salience weights ω_i for the alternatives and constant α_0 that moderates the strength of the associative bias. Additional parameters ("exchange rates") $\alpha_j, j \in \{1, \ldots, n_a\}$ might be necessary to obtain comparability of attributes. However, all α_j have been set to one in present applications of the AAM. Accessibility is higher for attributes with larger attribute values and attributes that are present in more alternatives or more salient alternatives (that is, alternatives with larger salience weights, e.g., reference points). In order to obtain attribute weights, the accessibility values are normalized to sum up to one:

$$w_j = \frac{a_j}{\sum_{l \in \{1,\ldots,n_d\}} a_l}.$$

Within attributes, the AAM assumes that all pairs of attribute values are attended simultaneously, like MDFT and the LCA model.

(2) Evaluation of alternatives: More precisely, the AAM assumes that all alternatives are evaluated simultaneously but separately within attributes, that is, no comparison of attribute values takes place. Instead, a value function $v(x)$ with

$$v(x) = \begin{cases} x^{1/2} & \text{for “good” attributes, and} \\ -x^{1/2} & \text{for “bad” attributes,} \end{cases}$$

that is nonnegative and increasing for positive attributes and nonpositive and decreasing for negative attributes, is applied to the attribute values. Again, note that additional parameters (“exchange rates”) $\alpha_j, j \in \{1, \ldots, n_a\}$ might be necessary to obtain comparability of attributes.

(3) Evidence accumulation: Like MDFT and the LCA model, the AAM defines n_a choice units $S_i, i \in \{1, \ldots, n_a\}$, with $S_i(0) = 0$. Over time, the accumulators are updated according to the following equation:

$$S_i(t) = \delta \cdot S_i(t-1) + v(m_{ij}) + \xi_i(t),$$

with decay parameter $\delta > 0$, and normally distributed noise term $\xi_i(t)$. Evidence accumulation stops either at a fixed time T, in which case the alternative with the highest preference state is chosen, or as soon as the preference state for one of the alternatives exceeds a positive threshold θ^+.

The AAM explains similarity, attraction, and compromise effects by means of two interacting mechanisms: (1) similarity based covariance of preferences, and (2) choice-set dependent changes of the associative bias.

Like MDFT and the LCA model, the AAM explains the **similarity effect** by means of covarying preferences for similar alternatives. Since all alternatives are evaluated simultaneously with respect to the momentarily attended attribute, the counter states or preference states for the similar alternatives, A_1 and A_3 (cf. figure 2.1), increase and decrease together. When it comes to making a choice and the preference states of the similar alternatives are both high, either of them is equally likely to be chosen. This results in a disadvantage of the similar alternatives A_1 and A_3 compared to the dissimilar alternative A_2. On the other hand, adding a third alternative to a choice set of two, increases the attribute weight of the strongest attribute of the new alternative, biasing choice towards alternatives that score high on this attribute as well. In the similarity situation, this may lead to an advantage of the two similar alternatives A_1 and A_3, reversing the similarity effect.

As for the **attraction effect**, the added dominated alternative A_3 (cf. figure 2.2) biases attention towards the attribute that is high for both the dominated alternative A_3 and the dominating alternative A_1, leading to an advantage of the dominating alternative A_1 over the dissimilar alternative A_2. However, since the dominated and dominating alternative are also similar to each other, covarying preferences are able to cover or even reverse the attraction effect.

As for the (asymmetric) **compromise effect**, the added extreme alternative A_3 (cf. left panel of figure 2.3) biases attention towards its strongest attribute, D_1. This leads to an advantage of the compromise option A_1 over the other extreme option A_2. Note that the AAM does not predict that the compromise option is chosen more often than the newly added extreme option, it just predicts a change in relative choice shares of the two original choice alternatives. With this, it is theoretically able to also explain the symmetric compromise effect (cf. right panel of figure 2.3). Explaining the symmetric compromise effect is further facilitated by the concavity of the value function $v(x)$. However, if the newly added extreme option A_3 is positioned close to the compromise option A_1 in the attribute space, the compromise effect transitions into a similarity effect.

Multi-attribute linear ballistic accumulator model

In the MLBA model, comparison and evaluation of alternatives takes place in a static and deterministic preprocessing stage prior to the proposed (linear) evidence accumulation. Therefore, the MLBA model is not a cognitive process model of decision making in a strict sense. We discuss it here regardless, since Trueblood et al. (2014) place it in this category, and it has been repeatedly mentioned in the same breath with the other models discussed here.

(1) Attention allocation: The MLBA model does not make any assumptions about attention allocation.

(2) Evaluation of alternatives: Choice alternatives are evaluated based on their attribute values in a static and deterministic preprocessing stage. In a first step, parameterized curvature is added to the attribute space, transforming indifference lines into indifference curves via

$$\left(\frac{x}{a}\right)^{\gamma} + \left(\frac{y}{b}\right)^{\gamma} = 1.$$

For $\gamma > 1$, the curves are convex and intermediate options are preferred to extreme options, for $\gamma < 1$ they are concave and extreme options are preferred to intermediate options. Note that the labels "intermediate" and "extreme" here refer to attribute dispersion, that is, to the attribute space independent of the current choice set. Let $m'_{ij}, i \in \{1, \ldots, n_a\}, j \in \{1, \ldots, n_d\}$, be the transformed attribute values. In a second step, the difference for each pair of transformed attribute values within attributes is calculated, $d_{ij,kj} = (m'_{ij} - m'_{kj}), i \neq k \in \{1, \ldots, n_a\}$, and multiplied with a distance- and direction-dependent weight

$$w_{ij,kj} = \begin{cases} exp(-\lambda_1 |m'_{ij} - m'_{kj}|), & \text{for } (m'_{ij} - m'_{kj}) \geq 0, \text{ and} \\ exp(-\lambda_2 |m'_{ij} - m'_{kj}|), & \text{for } (m'_{ij} - m'_{kj}) < 0. \end{cases}$$

And lastly, the weighted differences are added up per alternative, yielding the input for the linear evidence accumulation:

$$m_i = I_0 + \sum_{k \neq i} \sum_{j=1}^{n_d} w_{ij,kj} \cdot d_{ij,kj}, \tag{2.6}$$

with constant $I_0 > 0$ to prevent negative values.

(3) Evidence accumulation: The MLBA model, like MDFT, the LCA model and the AAM, defines n_a choice units $S_i, i \in \{1, \ldots, n_a\}$. In contrast to the other models though, the inital accumulator states are drawn from a uniform distribution with support $[0, A]$: $S_i(0) \in \mathcal{U}(0, A)$. Starting from the initial accumulator states, evidence accumulation is linear and deterministic, with slopes drawn from normal distributions with means m_i (defined in equation 2.6) and standard deviation σ. Evidence accumulation stops as soon as the first counter hits a positive threshold θ^+. Note that, in most applications of the MLBA model so far, only choices but no response times were analyzed, and thus the parameters A, σ, and θ^+ have been fixed to $A = 1, \theta^+ = 2$, and $\sigma = 1$.

The three context effects are so to say hard coded into the MLBA model by means of three mechanisms in the preprocessing stage.

The **similarity effect** emerges from the different weights, λ_1 and λ_2, for evidence in favor of and against choosing each alternative. Compared to the additional alternative A_3 (cf. figure 2.1), the dissimilar alternative A_2 has a large advantage on dimension D_2 and a large disadvantage on dimension D_1, and the similar alternative A_1 has a small advantage on one dimension and a small disadvantage on the other dimension. If advantages loom larger than disadvantages in the MLBA model, that is, if $\lambda_1 > \lambda_2$, the dissimilar alternative A_2 benefits more than the similar alternative A_1 from the addition of alternative A_3.

The **attraction effect** emerges from the distance dependence of the weights $w_{ij,kj}$. Since the additional dominated alternative A_3 (cf. figure 2.2) is closer to the dominating alternative A_1 than to the dissimilar alternative A_2, the comparison between A_3 and A_1 receives more weight than the comparison between A_3 and A_2. Therefore, alternative A_1 benefits more than alternative A_2 from the addition of alternative A_3. Note that the two mechanisms producing similarity and attraction effects in the MLBA model are opposed to each other and each could potentially cover the other effect (but see Tsetsos et al., 2015; Trueblood et al., 2015, for a discussion).

The **compromise effect** emerges from the parameterized curvature that is added to the attribute space. As mentioned above, intermediate alternatives are preferred to extreme alternatives for $\gamma > 1$. Note that this mechanism mainly produces symmetric compromise effects (cf. figure 2.3). However, the distance dependent weights support both symmetric and asymmetric compromise effects, since comparisons involving the compromise receive more weight than comparisons between the two extremes.

Multi-alternative decision by sampling

The original theory of decision by sampling (DbS, Stewart, Chater, & Brown, 2006) assumes that preferences are constructed by means of binary, ordinal comparisons of the choice alternative's attribute values with reference values sampled from long-term memory. Recently, two extensions of DbS have been proposed that both account for the three context effects, but with different mechanisms: Multi-attribute decision by sampling (MADS, Ronayne & Brown, 2017), and multi-alternative decision by sampling (MDbS, Noguchi & Stewart, 2018). While MADS assumes that the available choice alternatives in a trial affect the distribution from which reference values are sampled, MDbS implements a similarity-based mechanism for attention allocation between available choice alternatives. However, since MADS is a static model, we only discuss MDbS here.

(1) Attention allocation: Based on an eye tracking study of the three context effects (Noguchi & Stewart, 2014), MDbS assumes that attention in multi-alternative multi-attribute decision making switches between attribute values, and attention probability is higher for attribute values that are similar to other available attribute values. Let m_{ij} and m_{kj} be two attribute values with $i \neq k \in$

$\{1, \ldots, n_a\}$, and $j \in \{1, \ldots, n_d\}$. MDbS defines the similarity of m_{ij} to m_{kj} as

$$s_{ij,kj} = \exp\left(-\lambda \frac{|m_{ij} - m_{kj}|}{|m_{kj}|}\right),$$

with similarity parameter λ. Note that, in general, $s_{ij,kj} \neq s_{kj,ij}$. Let

$$s_{ij} = \sum_{k \neq i \in \{1,\ldots,n_a\}} s_{ij,kj}$$

be the summed similarity of attribute value m_{ij} to the other attribute values on the same dimension. Then the probability for focusing on attribute value m_{ij} is obtained by dividing its summed similarity value with the summed similarity values of all attribute values across attributes:

$$p_{ij} = \frac{s_{ij}}{\sum_{l \in \{1,\ldots,n_a\}} \sum_{m \in \{1,\ldots,n_d\}} s_{lm}}.$$

The probability p_j for attending attribute $D_j, j \in \{1, \ldots, n_d\}$ can be easily derived from this definition, though it is not explicitly used in MDbS:

$$p_j = \sum_{i \in \{1,\ldots,n_a\}} p_{ij}.$$

(2) Evaluation of alternatives: MDbS assumes that alternatives are evaluated in pairwise comparisons of attribute values within attributes. Given an alternative $A_i, i \in \{1, \ldots, n_a\}$, that is supposed to be evaluated with respect to attribute $D_j, j \in \{1, \ldots, n_d\}$, every other attribute value $m_{kj}, k \neq i \in \{1, \ldots, n_a\}$, on the same dimension is equally likely to serve as reference value for the focus value m_{ij}. The probability that m_{ij} is favored over m_{kj} is defined by means of a logistic function:

$$P(m_{ij} \text{ favored over } m_{kj}) = \begin{cases} \frac{1}{1+\exp\left(\delta_1 \cdot \left(\frac{|m_{ij}-m_{kj}|}{|m_{kj}|} - \delta_0\right)\right)}, & \text{for } m_{ij} > m_{kj} \\ 0 & \text{for } m_{ij} \leq m_{kj}, \end{cases}$$

with sensitivity parameters δ_0 and δ_1. This function allows for "soft" comparisons: For $\delta_0 = 50$ and $\delta_1 = 0.1$, for example, a difference of 10% is detected with probability 0.5 and a difference of 20% is detected with probability > 0.99 (Noguchi & Stewart, 2018). Like in the original DbS, comparisons in MDbS are ordinal, that is, it only counts whether an attribute value is favored over an other one but not how much the two values differ. Overall, the probability that an attribute value m_{ij} wins a comparison is the average over the winning probabilities for all possible pairwise comparisons with respect to the same attribute:

$$P(m_{ij} \text{ wins a comparison}) = \frac{\sum_{k \neq i \in \{1,\ldots,n_a\}} P(m_{ij} \text{ favored over } m_{kj})}{n_a - 1}.$$

(3) Evidence accumulation: In MDbS, the evidence for each alternative is the number of won comparisons of the associated attribute values. It is gathered in one accumulator or counter per alternative, $S_i, i \in \{1, \ldots, n_a\}$, with initial counter state $S_i(0) = 0$. The probability that the state of counter S_i is increased is

$$p_i = \sum_{j \in \{1,\ldots,n_d\}} (p_{ij} \cdot P(m_{ij} \text{ wins a comparison})).$$

Like the 2NCT/SCT models, MDbS does not explicitly implement competition between accumulators, however, competition naturally emerges from the definition of attention probabilities for the attribute values as normalized summed similarity. Furthermore, there is no noise in MDbS's evidence accumulation. MDbS implements a relative stopping rule, such that a decision is made when the maximum of counter states exceeds the average of all counter states by θ. For demonstration purposes, Noguchi and Stewart (2018) set $\theta = 0.1$. In this case, a decision is made as soon as one unit of evidence is accumulated, making the model basically static. For more general cases, the authors provide a closed form solution for choice probabilities, based on a tutorial by Diederich and Busemeyer (2003).

MDbS explains similarity, attraction, and compromise effects by means of three interacting mechanisms: (1) "Soft" comparisons that ignore similar alternatives, (2) changes in attention probabilities due to additional alternatives in the choice set, and (3) changes in probabilities for winning comparisons due to additional alternatives in the choice set.

As for the **similarity effect**, MDbS assumes that the dissimilar alternative A_2 (cf. figure 2.1) profits more from addition of A_3 than the similar alternative A_1, since A_1 and A_3 are too similar for passing the sensitivity threshold of the "soft" comparison mechanism. This explanation lasts even though A_1 and A_3 overall receive more attention than alternative A_2, due to their similarity. However, the strength of the similarity effect is moderated by the similarity parameter λ. The higher λ, the weaker the similarity effect, and even negative similarity effects may occur.

As for the **attraction effect**, MDbS assumes that the dominating alternative A_1 (cf. figure 2.2) profits more from addition of the dominated alternative A_3 than the dissimilar alternative A_2, since A_1 and A_3 receive more attention due to their similarity, and the probability for winning a comparison increases significantly for the dominating alternative A_2. Note that this explanation is valid only for sufficiently different alternatives A_1 and A_3, because otherwise the "soft" comparison mechanism undermines the effect. On the other hand, the comparison mechanism allows MDbS to account for differences between range, frequency, and range-frequency decoys. Like for the similarity effect, the similarity parameter λ moderates the strength of the attraction effect, but in reversed direction: The higher λ, the stronger the attraction effect. A negative attraction effect is very unlikely to occur in MDbS, except for very similar alternatives A_1 and A_3.

As for the **compromise effect**, MDbS assumes that the addition of an extreme alternative A_3 (cf. left panel of figure 2.3) increases the attention/evaluation probability for the compromise option A_1 since it has the highest similarity value with two medium distant competitors compared to the extreme alternatives A_2 and A_3 with one distant and one medium competitior each. Since similarity is relative in MDbS, this explanation is valid for both symmetric and asymmetric compromise effects. Again, the strength of the compromise effect is moderated by the similarity parameter λ, in the same direction as for the attraction effect: The higher λ, the stronger the compromise effect. Note that the similarity parameter allows MDbS to account for correlations between the three context effects, as, for example, observed by Berkowitsch et al. (2014). However, MDbS is not able to explain negative compromise effects as reported, for example, by Liew et al. (2016).

2.6 Summary and outlook

Following Oppenheimer and Kelso (2015), who identified a Kuhninan paradigm change in the field of decision making research, we described in detail how three "anomalies" in multi-alternative multi-attribute decision making drove the change from utility theories to information processing theories in this sub-field. The three so-called context effects, similarity, attraction, and compromise effects, demonstrate that preferences are contingent on the choice set under consideration and thus violate independence assumptions of rational choice theories and early probabilistic choice models. The original findings reported by Tversky (1972b), Huber et al. (1982), and Simonson (1989) have first been explained simultaneously by multi-alternative decision field theory (MDFT, Roe et al., 2001), the multi-alternative extension of decision field theory (DFT, Busemeyer & Townsend, 1993), and shortly afterwards by the leaky competing accumulator model (LCA, Usher & McClelland, 2001, 2004). These models are so-called computational cognitive process models that dissect the decision making process into parameterized microprocesses, describing the construction of preferences over time. Together, MDFT and the LCA model set off development of a number of multi-alternative multi-attribute decision making models and established similarity, attraction, and compromise effects as benchmark for such models. That is, they defined what phenomena multi-alternative multi-attribute decision making models should account for, and how such models should be constructed (namely from parameterized microprocesses). In other words, they (partly) defined the information processing paradigm for multi-alternative multi-attribute decision making research. However, until recently, MDFT and the LCA model have not been tested on data, partly due to a lack of simultaneous observations of similarity, attraction, and compromise effects in a single experiment or even within one experimental paradigm.

Since 2012, several such experimental paradigms have been developed, most of them using ternary variants of the three context effects in order to increase effect sizes. It turned out that, on an individual level, participants who show the similarity effect ususally show negative attraction and compromise effects, and participants who show attraction and compromise effects usually show a negative similarity effect (e.g. Berkowitsch et al., 2014). This explains the difficulty in obtaining all three effects simultaneously and at the same time calls for models that are able to explain similarity, attraction, and compromise effects as well as the opposite or negative effects and the interactions between them.

Also since 2012, several new computational cognitive process models that account for the three context effects have been proposed. Each of these models has some unique features, but they are all constructed from parameterized microprocesses, like MDFT and the LCA model, and some of them draw additional inspiration from the early theories of preference reversals presented in section 2 above.

The unique feature of the 2N-ary choice tree model (Wollschlaeger & Diederich, 2012) and its variant simple choice tree model (Wollschlaeger & Diederich, 2017) is the possibility to eliminate unwanted alternatives from the choice set. While this feature is reminiscent of the theory of elimination by aspects (Tversky, 1972a, 1972b), and allows the 2NCT/SCT models to mimick EBA for appropriate choices of reference points (and, in SCT, focus value), it is actually an implementation of (relative) loss avoidance or loss aversion (cf. Tversky & Kahneman, 1991). In the SCT model, the elimination mechanism is parameterized by means of the focus weight, allowing the decision maker to either avoid losses or to approach gains. There, avoiding losses is associated with attraction and compromise effects (and negative similarity effects), while approaching gains is associated with similarity effects (and negative attraction and compromise effects).

The unique feature of the associative accumulation model (AAM, Bhatia, 2013) is an associative process between choice alternatives and attributes. Attributes that are present in many choice alternatives or dimensions on which alternatives score relatively high receive higher weights, leading to changes in preferences when adding new alternatives to the choice set. Mechanisms based on changing attribute weights have been discussed before, for example by Huber et al. (1982), Tversky et al. (1988), and Wedell (1991), but Bhatia (2013) was the first to implement such a mechanism in a cognitive process model. The AAM is able to explain positive and negative similarity and attraction effects and interactions between them, since the mechanisms that produce the two effects are opposed to each other. However, the model is not able to produce negative compromise effects and does not explain the interactions between the compromise effect and the other two context effects.

The multi-attribute linear ballistic accumulator model (MLBA, Trueblood et al., 2014) neglects dynamic aspects altogether and the three context effects are hard-wired into the model. While this allows for mathematical tractability, the declared unique feature of the MLBA model, the resulting response times are actually meaningless. The mechanisms that explain similarity and attraction effects in the MLBA model are opposed to each other, that is, the model is in principle able to produce the negative correlation between the two effects. The mechanism that produces the compromise effect is able to produce negative compromise effects as well, but there is no interaction with the other two effects.

The unique feature of multi-alternative decision by sampling (MDbS, Noguchi & Stewart, 2018) is a similarity-driven attention allocation mechanism which favors alternatives that are similar to other available choice options. On the other hand, a "soft" comparison mechanism prevents similar alternatives from boosting each other. These two mechanisms allow MDbS to account for the interactions between similarity, attraction, and compromise effects, but negative attraction effects are very unlikely to occur.

Outlook

We said in the introduction, that Lichtenstein and Slovic (2006) identified decision field theory as an exception to the need for multiple theories to explain how preferences are constructed by different decision makers in different situation. Yet, the different theories described in this review have been compared with each other mostly on exactly the same level as early theories of preference reversals, that is, on the level of whole theories.

Only recently, Turner et al. (2018) made a first attempt to compare theories on the level of the microprocesses they are based on. They identified the information processing mechanisms used by multi-alternative decision field theory, the leaky competing accumulator model, the associative accumulation model and the multi-attribute linear ballistic accumulator model and compared 432 different combinations of those microprocesses in a so-called switchboard analysis. The switchboard analysis was based on data from Turner et al.'s (2018) study 2, a perceptual choice experiment of similarity, attraction, and symmetric compromise effects. We applaude their ideas and efforts, but also see several shortcomings that need to be addressed in the future. First of all, the experiment was restricted to the symmetric version of the effect (cf. figure 2.3). This restriction favors mechanisms tailored to explain this version of the effect, as implemented in AAM and the MLBA model, and consequently also favors those models. Future studies should include both, symmetric and asymmetric, versions of the compromise effect. A similar criticism pertains to the attraction effect, since Turner et al.'s (2018) experiment features only range-frequency decoys, ignoring possible differences between those and range decoys or frequency decoys (cf. Huber et al., 1982, and figures 2.2 and 2.5). Future studies should include all three kinds of dominated decoys.

More generally, the capability to explain additional effects that have been observed in multi-alternative and/or multi-attribute decision making should be explored for the models discusses here, and particulary for their mechanisms. Bhatia (2013), for example, proposes to consider the alignability effect (Markman & Medin, 1995), the less is more effect (Simonson, Carmon, & O'Curry, 1994), the endowment effect (Thaler, 1980), the status quo bias (Samuelson & Zeckhauser, 1988), and two effects reported by Tversky and Kahneman (1991): The improvements versus tradeoffs effect, and the advantages and disadvantages effect. Additionally, Bhatia (2013) discusses phantom decoys, that is, unavailable alternatives (cf. Pratkanis & Farquhar, 1992), as a way to study preference oderings for the whole (ternary) choice set. Highhouse (1996), for example, uses phantom decoys to study preference orderings for job candidates in an attraction situation, Pettibone and Wedell (2000, 2007) use phantom decoys with consumer products in attraction and compromise situations, Usher, Elhalal, and McClelland (2008) use them to compare decision field theory and the leaky competing accumulator model in a compromise situation, and Trueblood and Pettibone (2015) use phantom decoys in a perceptual study of attraction and compromise effects.

Another way to study preference orderings for ternary choice sets is so-called best-worst scaling, where participants are asked to state the most preferred and the least preferred alternative in a choice set (cf. Louviere, Flynn, & Marley, 2015). Hawkins et al. (2013) and Hawkins et al. (2014) explore the capability of the multi-attribute linear ballistic accumulator model to account for this kind of data. Some of the problems they raise could be solved by implementing an elimination threshold as proposed by Wollschlaeger and Diederich (2012, 2017) for selecting worst alternatives. An elimination mechanisms could also help with explaining choice deferral, that is, delaying or avoiding a decision. So far, choice deferral has been implemented into decision field theory as additional choice alternative (Busemeyer et al., 2006) and into the associative accumulation model as time limit to the deliberation process (Bhatia & Mullett, 2016).

Time limits touch on another important topic (maybe *the* most important topic) for cognitive process models of decision making: The temporal development of preferences. Early preference reversal experiments, including the original context effects studies (Tversky, 1972b; Huber et al., 1982; Simonson, 1989), relied on paper-and-pencil measurements, that is, participants had to indicate on a piece of paper which alternative they preferred. Accordingly, early theories of context effects only predicted choice probabilities. Computational cognitive process models like decision field theory, on the other hand, are theoretically able to predict also choice response times. Practically, however, this ability has been mosty neglected in the multi-alternative, multi-attribute decision making literature, as also in the recent comparative study by Turner et al. (2018). One reason for this is the lack of mathematical tractability in most of these models. While Turner et al. (2018)

argue that model fit and model complexity are more important than mathematical tractability, it remains unclear whether their likelihood-free estimation method is able to deal with response times. Theories like the (simple) 2N-ary choice tree model (Wollschlaeger & Diederich, 2012, 2017) and multi-alternative decision by sampling (Noguchi & Stewart, 2018), on the other hand, predict response times based on the discrete methods promoted by Diederich and Busemeyer (2003), a quite promising approach. With computer-based measurements as standard in modern psychological studies, response times can easily be recorded and should be used to differentiate between models and information processing mechanisms.

Response times can be interpreted as a kind of process data. Other kinds include think-aloud protocols, mouse-tracing, and eye-tracking (Schulte-Mecklenbeck, Kuehberger, & Ranyard, 2011). While think-aloud protocols and mouse-tracing might interfere with information processing, eye-tracking is relatively unobtrusive and could provide additional insights about attention allocation and comparative evaluation of choice alternatives, two of the basic elements identified above (see Ashby, Johnson, Krajbich, & Wedel, 2016, for a recent review of eye-movement research in judgment and decision making). As Cohen, Kang, and Leise (2017) point out, cognitive process models of decision making could also be extended to explicitly account for attention allocation and eye movements. Such an approach has been adopted, for example, by Johnson and Busemeyer (2016). Further inspiration can be found in models of eye movements in reading, which have been developed and applied for several decades (Trukenbrod & Engbert, 2014; Clifton Jr et al., 2016).

The main topic of this review have been context effects, that is, changes in choice probabilities contingent on choice set composition. However, it would be interesting to explore for the cognitive process models discussed here, whether they are able to account for other kinds of preference reversals as well. As we said in the introduction, preference construction is contingent on the task, the context, and on individual differences (Payne et al., 1992). Of course there are also interdependencies between the different causes of preference reversals. Chang and Liu (2008) and Cataldo and Cohen (2018), for example, observe reversed compromise and similarity effects, respectively, for certain task and presentation formats. A recent research topic in frontiers in psychology (see Houpt, Yang, & Townsend, 2016, for the editorial) deals with modeling individual differences in perceptual decision making. There is a lack of similar studies for multi-alternative multi-attribute (preferential) decision making.

Chapter 3

The 2N-ary choice tree model

Lena M. Wollschlaeger, Adele Diederich

This chapter has been published in Frontiers of Psychology under the title "The 2N-ary choice tree model of N-alternative preferential choice" on June 20, 2012.

Abstract

The $2N$-ary choice tree model accounts for response times and choice probabilities in multi-alternative preferential choice. It implements pairwise comparison of alternatives on weighted attributes into an information sampling process which, in turn, results in a preference process. The model provides expected choice probabilities and response time distributions in closed form for optional and fixed stopping times. The theoretical background of the $2N$-ary choice tree model is explained in detail with focus on the transition probabilities that take into account constituents of human preferences such as expectations, emotions or socially influenced attention. Then it is shown how the model accounts for several context-effects observed in human preferential choice like similarity, attraction and compromise effects and how long it takes, on average, for the decision. The model is extended to deal with more than three choice alternatives. A short discussion on how the $2N$-ary choice tree model differs from the Multi-alternative Decision Field Theory and the Leaky Competing Accumulator model is provided.

3.1 Introduction

Life is full of decisions: Be it the selection of clothing in the morning or of menu for lunch, the question which car to buy or if taking cold medication is necessary. This type of decisions is called preferential choice and has been subject of numerous investigations within the field of decision theory (Koehler & Harvey, 2007, for a review). Several effects have been observed when the decision maker has more than two choice options (multi-alternative preferential choice). Hick's Law (Hick, 1952; Hyman, 1953), originally defined in the context of stimulus detection paradigms, postulates a dependency of deliberation time on the number of alternatives. In particular, it states that a linear increase of the number of equally attractive alternatives to choose from leads to a logarithmic increase of the time that passes until the decision is made. Furthermore, a decision maker who is indifferent between two choice alternatives from a given choice set may change the preference for one or the other alternative when the choice set is enlarged, i.e., the local context may affect the decision and generate preference reversals. Similarity effects (Tversky, 1972a), attraction effects (Huber et al., 1982) and compromise effects (Simonson, 1989), for instance, depend on a third alternative that is added to a choice set of two equally attractive but dissimilar alternatives. If the third alternative is very similar to one of the others, the two similar alternatives share their choice frequency and are both chosen less often than the dissimilar one (similarity effect). If the third alternative is similar to one of the others but slightly inferior, it promotes the similar one and increases its choice frequency compared to the dissimilar one (attraction effect). If the third alternative is a compromise between the other two, the decision maker will prefer the compromise to the other alternatives (compromise effect). Besides those preference reversals that emerge from local context, there might also be influence from background context (Tversky & Simonson, 1993) like a reference point outside of the choice set which – together with the loss-aversion principle (Kahneman & Tversky, 1979) – affects evaluation of the given alternatives.

One challenge for (cognitive) modelers is to think of a model which predicts decision making behavior for multi-alternative preferential choice tasks in general but also accounts for all the aforementioned effects. Another challenge is to formulate the model such that (expected) response times and choice probabilities can be calculated and the model parameters conveniently estimated from the observed choice times and choice frequencies.

Decision Field Theory (DFT, Busemeyer & Townsend, 1992, 1993) and its multi-attribute extension (Diederich, 1997) predict choice response times and choice probabilities for binary choice tasks. Both approaches provide closed form solutions to calculate these entities. Since then, several attempts have been made to extend this kind of models to multi-alternative preferential choice tasks: Multi-alternative Decision Field Theory (Roe et al., 2001) and the Leaky Competing Accumulator (LCA) model (Usher & McClelland, 2001, 2004) predict choice probabilities for three alternative choice tasks but cannot account for optional choice times, i.e., the time the decision maker needs to come to a decision. Both approaches, however, do account for fixed stopping times, i.e., for an externally determined time limit. Furthermore, multi-alternative DFT and the LCA model both account for the similarity, attraction and compromise effects using computer simulations to predict the patterns. To do so, Roe et al. (2001) interpret DFT as a connectionist network and implement distance-dependent inhibition between the alternatives. Usher and McClelland (2001, 2004) add insights from perceptual choice and neuropsychology to the multi-alternative DFT and propose for their LCA model direct implementation of loss-aversion by means of an asymmetric value function and global inhibition instead of distance-dependent inhibition.

Our $2N$-ary choice tree model builds on the previous approaches and tries to overcome some of their problems. It is a general model for choice probabilities and response times in choice between N alternatives with D attributes. As such, it provides a way to calculate expected response times, response time distributions and choice probabilities in closed form by determining the time course of an information sampling process via a random walk on a specific tree. It is able to account for similarity, attraction and compromise effects which have been most challenging for previous models. In contrast to previous approaches, the 2N-ary choice tree model accounts for these effects without additional mechanisms like inhibition or loss-aversion and is thus more parsimonious. However, it is possible to implement these mechanisms if the situation requires it.

First, we describe the structure of the $2N$-ary choice tree and the implementation of the random walk on it in general, including a discussion of initial values and stopping rules. Then we define expected choice probabilities and reaction times and state that these exist and can be calculated in finite time. The proof of this statement is given later in the paper. It is not essential for understanding the theory; we provide it rather as completing the theoretical derivations. Next, we show how to derive transition probabilities from the given alternatives in a specific choice set and therewith define the random walk for that set. A psychological interpretation of their constituents is given afterwards. Finally, we demonstrate the predictive power of our model by showing several simulations for choice situations producing the similarity, attraction and compromise effect and calculate expected hitting times and choice probabilities. We conclude with a comparison of the $2N$-ary choice tree with its closest competitors, the multi-alternative DFT and the LCA model.

3.2 The $2N$-ary choice tree model

Making an informed decision usually implies sampling of information about the alternatives under consideration. In Psychology, information sampling processes (e.g. Townsend & Ashby, 1983; Luce, 1986, for review; LaBerge, 1962; Laming, 1968; Link & Heath, 1975; Townsend & Ashby, 1983; Luce, 1986; Ratcliff & Smith, 2004) have a long tradition and proven to be an adequate tool for detailed interpretation of decision making processes, mostly in perception as they provide insight about accuracy and time course of these processes. Poisson counter models (e.g. Pike, 1966; Townsend & Ashby, 1983; LaBerge, 1994; Diederich, 1995; Van Zandt, Colonius, & Proctor, 2000; Smith & Van Zandt, 2000) are a special class of information sampling models that assume constant amounts of information being sampled at Poisson distributed points in time. (Multi-alternative) DFT (Busemeyer & Townsend, 1993; Roe et al., 2001) and the LCA model (Usher & McClelland, 2004) make use of information sampling principles in modeling preferential choice under uncertainty. Both models assume one counter per alternative and all of these counters are updated once per fixed time interval until one of them reaches a threshold. The amounts to update the counters depend on comparison of the alternatives and on already sampled information. In our $2N$-ary choice tree model, only one counter per fixed time interval is updated with a fixed amount, but the probability for each counter to be updated depends on comparison of the alternatives and on already sampled information. With regard to its constituents it is thus based on the same principles as both DFT and the LCA model. As only one counter is updated per iteration, the next time for a specific counter to be updated depends on the given probabilities. Hence the technical component of the $2N$-ary choice tree model resembles a counter model.

$2N$-ary choice trees

In contrast to the aforementioned models, the $2N$-ary choice tree model assigns two counters to each of N alternatives in a given choice set. One of them samples positive information, i.e. information in favor of the respective alternative, the other one samples negative information, i.e. information against it. Their difference describes the actual preference state relating to that alternative. As an example,

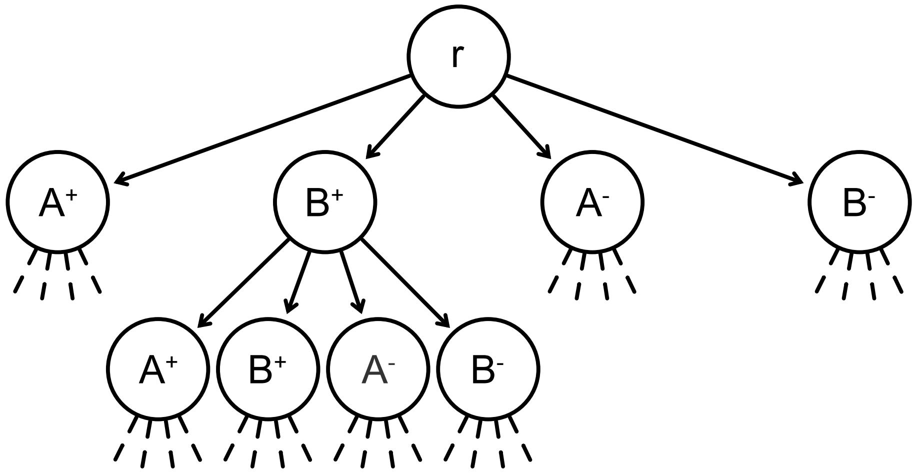

Figure 3.1: 4-ary tree for choice between two alternatives A and B. The root r has four outgoing edges directing to four vertices that represent the counters A^+, A^-, B^+ and B^-. Each of these vertices has four outgoing edges and thus four children itself, and so forth.

consider two alternatives A and B. The four counters are labeled A^+, A^-, B^+ and B^- and yield the preference states $Pref(A) = A^+ - A^-$ for alternative A and $Pref(B) = B^+ - B^-$ for alternative B. Beginning at a fixed point in time, the model chooses one counter and increases its state by one whenever a specific time interval h (e.g. one millisecond) has passed.The length h of the time interval can be chosen arbitrarily with a shorter time interval leading to more precision in the calculation of expected choice probabilities and choice response times. Due to limitations of recording devices, experimental data will be discrete as well and it is thus not necessary to aim for a continuous model. Note that increasing only one counter state at a time with a fixed amount of evidence equal to one is equivalent to increasing all counter states at the same time with an amount of evidence equal to the probability with which these counters are chosen and which also sum up to one (see below). Updating counters at discrete points in time creates a discrete structure of possible combinations of counter states which can be interpreted as graph or, more precisely, as (b-ary) tree[1].

Definition 1 (b-ary tree)**.** *A* b-ary tree *is a rooted tree* $T = (V, E, r)$ *with* vertices V, edges $E \subseteq V \times V$ *and* root $r \in V$ *where all vertices are directed away from* r *and each internal vertex has* b *children.*

For N choice alternatives, consider a $2N$-ary tree $T = (V, E, r)$. Figure 3.1 depicts the 4-ary tree for the two-alternative example. The topmost vertex is the root r with outgoing edges directing to four vertices that represent the counters A^+, A^-, B^+ and B^-. Each of these vertices has four outgoing edges and thus four children itself, and so forth. The information sampling process is mapped to this tree as a walk, i.e. a sequence of edges and vertices, beginning with the root r that takes one step, i.e. passes from one vertex through an edge to another vertex,

[1]Definitions of graph-related terms not defined here can be found in Korte and Vygen (2006).

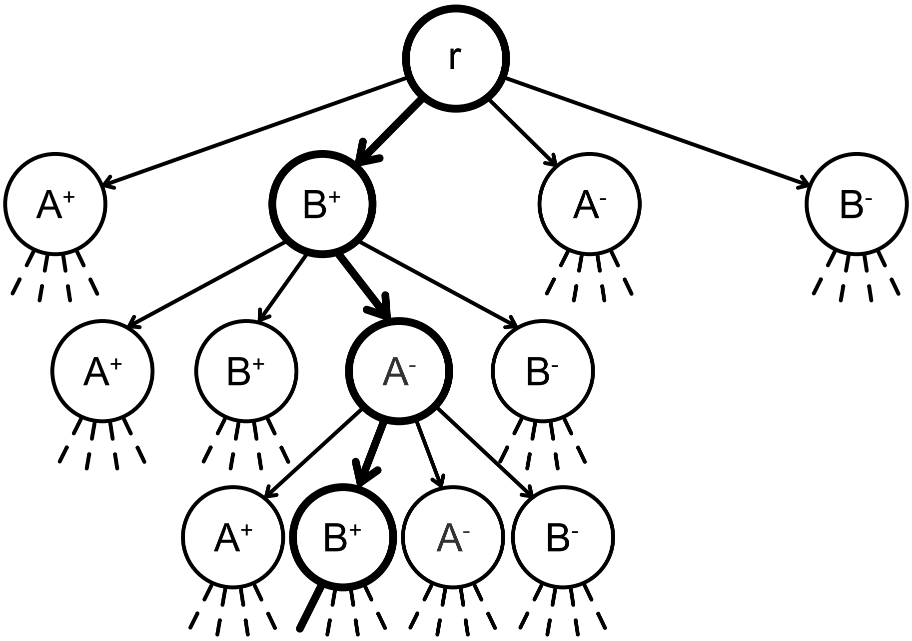

Figure 3.2: 4-ary tree for choice between two alternatives A and B with highlighted sample path $r \to B^+ \to A^- \to B^+ \to \ldots$

per time interval h. Whenever the walk reaches a vertex, the counter with the same label is updated by $+1$. Figure 3.2 shows an example for a walk on the 4-ary tree where first counter B^+ (information in favor for choice alternative B), then counter A^- (unfavorable information for choosing alternative A) and then again counter B^+ is updated.

Three features of the model are of specific interest: a) when and how the walk starts after presentation of choice alternatives (in an experimental trial), b) how the walk chooses the next edge to pass through in each step and c) when and how the walk stops. Without an a priori bias toward any of the choice alternatives, we assume that all counter states are set to zero at the outset of the information sampling process and hence, the process starts with presentation of the choice alternatives. Biases towards one or several of the alternatives can be implemented by either defining initial values unequal to zero for these alternatives or by independently sampling information for the alternatives from predefined distributions for some time before the actual information sampling process starts (cf. Diederich & Busemeyer, 2006; Diederich, 2008). For simplicity, we assume no biases here, i.e. initial values are set to zero for all alternatives. Note that for the $2N$-ary choice tree, initial values are counter states at the root r. Then the walk moves away from there step by step, choosing the next edge to pass through by means of so called transition probabilities $p_e, e \in E$. The transition probabilities are built up of the comparison of the alternatives the decision maker considers and supplemented with a counter-dependent component and a random component. For each vertex, the transition probabilities for all outgoing edges sum up to one, so that the walk stays still at any vertex it reaches throughout the information sampling process. We show the structure of the model first; a detailed description of the transition prob-

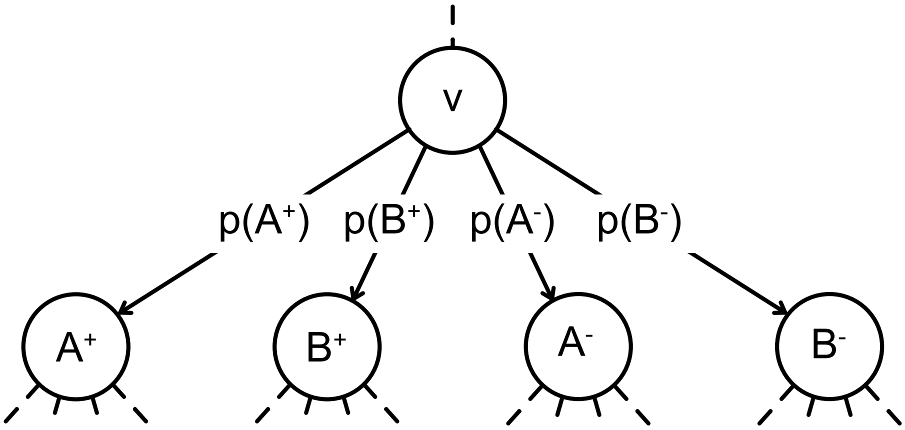

Figure 3.3: Transition probabilities for the two-alternative choice problem with no counter-dependent and random component. $p(A^+)$, $p(B^+)$, $p(A^-)$ and $p(B^-)$ sum up to one and are the same for the outgoing edges of each vertex $v \in V$.

abilities is presented in the next section. For simplicity consider a choice situation with two alternatives A and B; the counter-dependent component and the random component are set to zero. As shown in figure 3.3, transition probabilities are the same for the outgoing edges of each vertex $v \in V$, i.e. $p_{(v,v(A^+))} = p(A^+)$ for each edge $(v, v(A^+)) \in E$ leading to a vertex with label A^+, $p_{(v,v(B^+))} = p(B^+)$ for each edge leading to a vertex with label B^+ and so on for the other counters A^- and B^-. The probability for walking along a specific path is the product of transition probabilities of all edges on that path. In our example, the probability p for making the first three steps as shown in figure 3.2 is $p = p(B^+) \cdot p(A^-) \cdot p(B^+)$.

The third topic addresses the stopping rule, that is, when the decision maker stops sampling information and chooses a choice alternative. A specific stopping rule depends on the preference states associated with the alternatives, i.e. the differences of their respective two counters which are compared to certain thresholds θ. The thresholds can be defined in several ways, their suitability depending on the definition of transition probabilities and initial values. They are 1) one single positive threshold $\theta^+ > 0$ for all alternatives, 2) one positive and one negative threshold $\theta^+ > 0$ and $\theta^- < 0$ for all alternatives, 3) a positive threshold θ_i^+ for each alternative $i \in \{1, \ldots, N\}$ and 4) a positive and a negative threshold θ_i^+ and θ_i^- for each alternative $i \in \{1, \ldots, N\}$.

Obviously, the simplest setup is a single positive threshold $\theta^+ > 0$ for all alternatives, which is hit as soon as the information sampled in favor of any/one of the alternatives exceeds the information against it by θ^+ for the first time, i.e., when $Pref(i) = \theta^+$ for one alternative $i \in \{1, \ldots, N\}$. Sometimes, however, the probability for collecting negative information may be greater than the probability for sampling information in favor of these alternatives and reaching a positive threshold θ^+ is very unlikely. For those situations it is useful to introduce a second, negative threshold, $\theta^- < 0$, which is hit when negative information of one alternative exceeds the positive information of this alternative by $-\theta^-$, i.e. $Pref(i) = \theta^-$. In this case the respective alternative is not chosen but withdrawn from the choice set and the sampling process continues with one alternative less as described in

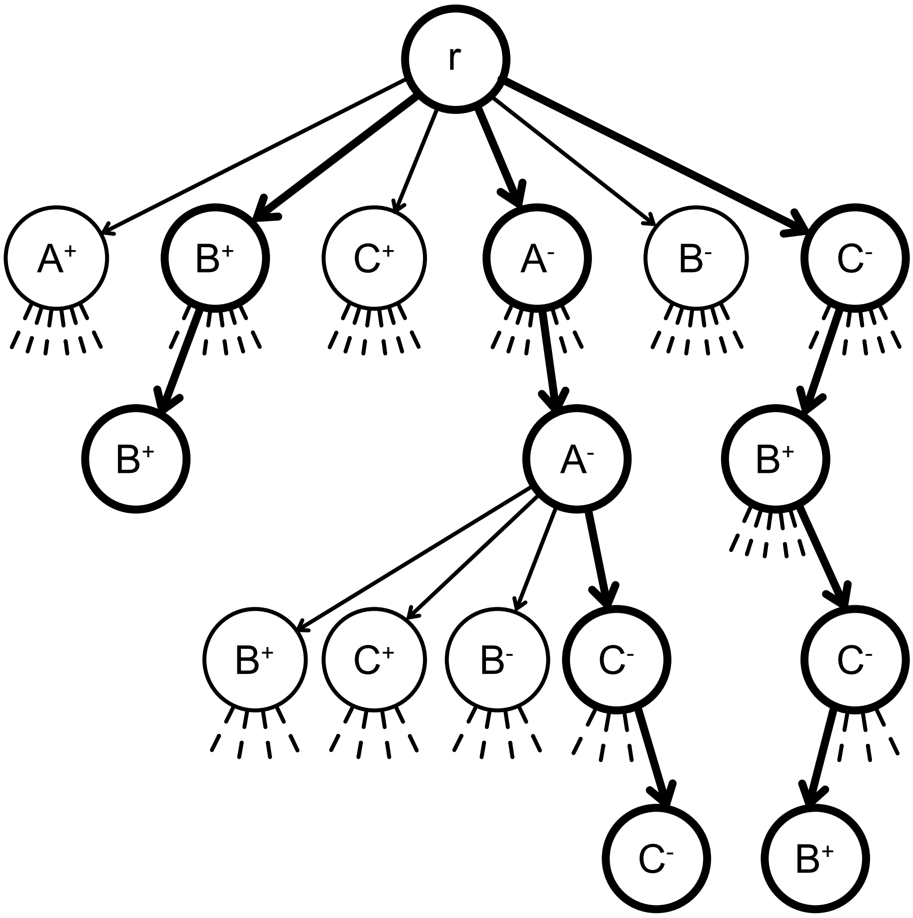

Figure 3.4: 6-ary tree for choice between three alternatives A, B and C with decision thresholds $\theta^+ = 2$ and $\theta^- = -2$ and three different sample paths that lead to choice of alternative B.

the next paragraph. Note, that in both cases the thresholds are global in the sense that the same thresholds apply for all choice alternatives. Finally, θ^+ (and θ^-) may vary from alternative to alternative, yielding one (or two) thresholds θ_i^+ (and θ_i^-) for each of N alternatives, $i \in \{1, \ldots, N\}$. Here the thresholds are local in the sense that each alternative has its own threshold(s). This is an alternative way to implement biases when the initial values are zero. That is, biases do not affect transition probabilities through the counter-dependent component and can thus be interpreted as the decision maker's stable opinion about the presented alternatives.

Withdrawal of alternatives from a choice set traces back to the model of elimination by aspects (EBA model, Tversky, 1972a). But whereas elimination is the only means to come to a decision in the EBA model, the $2N$-ary choice tree model like the multi-alternative DFT (Roe et al., 2001) provides several ways to reach a decision. An alternative i is chosen either if its preference state exceeds θ_i^+ or if all other alternatives have been withdrawn from the choice set or a combination of these two. Figure 3.4 shows three examples of walks that lead to the choice of alternative B from a set of three alternatives A, B and C with global thresholds $\theta^+ = 2$ and $\theta^- = -2$. The leftmost walk represents direct choice of alternative B, the rightmost withdrawal of alternative C and subsequent choice of option B and the middle walk illustrates withdrawal of alternative A first and then of option C. Note that after withdrawal of one alternative, there are two outgoing edges less from the respective vertex downwards. Transition probabilities change accordingly, i.e. the withdrawn alternative is removed from the comparison procedure and its counter states no longer contribute to the counter-dependent component (cf. next section). This corresponds to an anew started information sampling process between the remaining alternatives and their previous counter states as initial values.

For a choice set with N alternatives and given thresholds $\theta_i^\pm$ this defines the structure of the $2N$-ary choice tree. For each alternative $i \in \{1, 2, \ldots, N\}$ we can thus completely identify the set $V_i \subseteq V$ of vertices where alternative i is chosen. Defining $P_v := P_{\{r,v\}}$ to be the unique path from the root r to a vertex $v \in V$ and given transition probabilities p_e for all edges $e \in E$ we can identify the probability for walking along a path P_v as the product $p_v = \prod_{e \in P_v} p_e$ and therewith define:

Definition 2 (expected choice probability)**.** *The expected probability for choosing alternative $i \in \{1, 2, \ldots, N\}$ is the probability for reaching the set V_i:*

$$p_i = \sum_{v \in V_i} \prod_{e \in P_v} p_e. \tag{3.1}$$

The length $|P_v|$ of the path P_v from r to $v \in V$ indicates the number of steps that the random walk has to take to reach v. Multiplied by the length h of the time interval, this yields the time it takes to cover the distance from r to v. Thus $T_v = h \cdot |P_v|$.

Definition 3 (expected hitting time)**.** *The expected time for choosing alternative $i \in \{1, 2, \ldots, N\}$ is the sum of expected hitting times for each vertex $v \in V_i$ weighted by the probability for reaching v:*

$$\mathbb{E}[T_i] = h \cdot \sum_{v \in V_i} |P_v| \cdot p_v = h \cdot \sum_{v \in V_i} |P_v| \cdot \prod_{e \in P_v} p_e. \tag{3.2}$$

The expected choice probabilities and hitting times can be approximated up to absolute accuracy in finite time. See below for the formal statement and proof of this property.

Transition probabilities

Having defined the skeletal structure of our theory, we can now proceed to its heart, the transition probabilities. The main components are a) weighted comparison of alternatives, b) mutual or global inhibition, c) decay of already sampled information over time and c) a random part. The transition probabilities control the information sampling process and thus describe the development of human preferences in specific choice situations. Throughout this section we will consider such situations with N choice alternatives that are evaluated with respect to the same D attributes. For each alternative, the decision maker is provided with one nonnegative value per attribute, representing an objective evaluation of that alternative with respect to the attributes. The $N \cdot D$ values in total can be stored in a $N \times D$ matrix $L = (l_{ij})$ with $i = 1, \ldots, N$ and $j = 1, \ldots, D$.

The definition of transition probabilities is based on weighted integration of results of an attribute-wise comparison of alternatives. To ensure equally significant impact of the weight parameters, preprocessing of the values of the alternatives with respect to the attributes is necessary and we do so by separately normalizing them to one for each attribute. This yields a new matrix $M = (m_{ij}) = (l_{ij} / \sum_{k=1}^{N} l_{kj})$ with $i = 1, \ldots, N$ and $j = 1, \ldots, D$ and thus column sum $\sum_{i=1}^{N} m_{ij} = 1$ for all $j \in \{1, \ldots D\}$.

Comparison of alternatives

At first we focus on one attribute j only. The easiest way to define transition probabilities is simply to use the entries from the respective column j of M and assign them to the edges that affiliate with the counters for positive information. The counters for negative information get transition probabilities zero. This corresponds to a framework where the alternatives are compared to an inferior external reference point (for example cars A, B and C are compared to not having a car at all). Because the values for each attribute sum up to one already, no further normalization is needed.

We differentiate between external reference points that are not part of the choice set and internal reference points that are part of the choice set and available for the decision maker. For instance, if someone moves to a new city and has to choose between several available apartments, she will probably compare them to her old apartment which is no longer available in the new city and thus an example for an external reference point. Or consider a choice of dessert in a restaurant when the decision maker is told that the chocolate cake she ordered is no longer available because someone just had the last piece. An internal reference point, however, is part of the choice set, actually several or even all available alternatives can be used as internal reference points at the same time, possibly in combination with an external reference point.

Having decided which reference points to use, the alternatives $i \in \{1, \ldots, N\}$ are compared to them. For each alternative i, favorable and unfavorable comparisons are handled separately and the absolute values of their differences are summed up to obtain measures of evidence for and against alternative i respectively. This yields two nonnegative values per alternative and thus $2 \cdot N$ values in total that are then normalized to one in order to obtain probabilities. In the car example where three cars are compared to not having a car at all, probabilities associated with negative counters are set to zero as each car is better than, presumably, no car. Actually, whenever one single reference point is used, at least half of the probabilities are zero because each alternative is either favored over the reference point or not and hence there cannot be evidence for and against one alternative at the same time. However, our main focus is on situations where each of at least three alternatives in the choice set is used as internal reference point for all the other alternatives and thus there are at least two reference points.

In this case we obtain a vector

$$P_j = \begin{pmatrix} p_{1j} \\ \vdots \\ p_{Nj} \\ p_{(N+1)j} \\ \vdots \\ p_{(2N)j} \end{pmatrix} = \begin{pmatrix} p^+_{1j} \\ \vdots \\ p^+_{Nj} \\ p^-_{1j} \\ \vdots \\ p^-_{Nj} \end{pmatrix} / \sum_{i=1}^{N} (p^+_{ij} + p^-_{ij})$$

with

$$p^+_{ij} = \sum_{k \neq i} (m_{ij} - m_{kj}) \cdot \mathbb{I}(m_{ij} > m_{kj}),$$

$$p^-_{ij} = \sum_{k \neq i} (m_{kj} - m_{ij}) \cdot \mathbb{I}(m_{ij} < m_{kj})$$

for $k = 1, \dots N$ and

$$\mathbb{I}(x) = \begin{cases} 1, & \text{if x is true} \\ 0, & \text{else.} \end{cases}$$

Especially with an external reference point at hand, the actual choice may lead to a loss of some kind. For instance, in the apartment example above a loss could be a further way to the workplace or a smaller bathroom. People usually try to avoid losses more than they seek gains while overrating small losses and gains compared to larger ones (Kahneman & Tversky, 1991; Tversky & Simonson, 1993). The $2N-$ary choice tree model can account for the loss aversion principle (Kahneman & Tversky, 1979) with an asymmetric value function (Kahneman & Tversky, 1991; Tversky & Simonson, 1993) by increasing probabilities for sampling negative information compared to probabilities for gathering positive information.

In their LCA model, Usher and McClelland (2004) use an asymmetric value function

$$V(x) = \begin{cases} log(1+x), & \text{for } x > 0 \\ -[log(1+|x|) + log(1+|x|)^2], & \text{for } x < 0 \end{cases}$$

and apply it to the relative advantages ($x > 0$) and disadvantages ($x < 0$) of alternatives compared to each other on one dimension. $V(x)$ is steeper for losses than for gains but flattens for both advantages and disadvantages when they become bigger. This favors similar pairs of alternatives over dissimilar ones and allows the LCA model to account for attraction and compromise effects.

Adopting it to our $2N$-ary choice tree model this yields an asymmetric value function

$$A(x) = \begin{cases} log(1+x), & \text{for favorable comparisons} \\ log(1+x) + log(1+x)^2, & \text{for unfavorable comparisons,} \end{cases}$$

which can be applied to the absolute differences from the comparison process before normalizing them to one:

$$p^+_{ij} = \sum_{k \neq i} A(m_{ij} - m_{kj}) \cdot \mathbb{I}(m_{ij} > m_{kj}),$$

$$p_{ij}^{-} = \sum_{k \neq i} A(m_{kj} - m_{ij}) \cdot \mathbb{I}(m_{ij} < m_{kj}).$$

The asymmetric value function $A(x)$ is not necessary for explanation of similarity, attraction or compromise effects in the $2N$-ary choice tree model but moderates the strength of the compromise effect (see below). In cases where θ^+ and θ^- have the same order of magnitude, application of $A(x)$ leads to faster withdrawal of alternatives and hence, more decisions are based on withdrawal of all but one alternative.

In summary, the comparison of alternatives provides us with a set of $2 \cdot N$ transition probabilities for each attribute $j \in \{1, \ldots, D\}$ that form a vector P_j. Each of these vectors can be used to model an information sampling process based on a single attribute. As the probabilities are derived from comparison of alternatives only, they remain constant during the whole process.

Weighting of attributes

So far we have only focused on one attribute but choice alternatives in real life are most often described by several attributes and thus require more elaboration. In the following, we consider choice sets with N alternatives characterized by $D \geq 2$ attributes. Especially situations with three alternatives where similarity, attraction or compromise effects have been observed, require at least two attributes to distinguish the different alternatives from each other. Note that it is difficult to construct a choice set with two equally attractive but different alternatives due to the decision maker's individual salience. Diederich (1997) accounts for subjective salience by defining a Markov process on the attributes giving probabilities for switching attention from one attribute to the other. This process can be directly implemented into the transition probabilities by using a stationary distribution on the attributes. Each attribute $j \in \{1, \ldots, D\}$ is assigned a weight w_j that corresponds to the probability for considering this attribute during the information sampling process. For each alternative $i \in \{1, \ldots, N\}$ weighted positive and negative evidence is added up and normalized to obtain a proper probability distribution (the probabilities add up to one), that is,

$$P = \begin{pmatrix} p_1 \\ \vdots \\ p_N \\ p_{(N+1)} \\ \vdots \\ p_{(2N)} \end{pmatrix} = \begin{pmatrix} p_1^+ \\ \vdots \\ p_N^+ \\ p_1^- \\ \vdots \\ p_N^- \end{pmatrix} / \sum_{i=1}^{N} (p_i^+ + p_i^-)$$

with $p_i^+ = \sum_{j=1}^{D} (p_{ij}^+ \cdot w_j)$ and $p_i^- = \sum_{j=1}^{D} (p_{ij}^- \cdot w_j)$ for $i \in \{1, \ldots, N\}$.

The weights account for subjective salience that in turn may be influenced by several internal and external factors such as personal preferences, social influences, characteristics of the choice set or the experimenter's instructions. Personal preferences like, for instance, the preference of time over money or of tastiness to healthiness may be learned from friends, family or other people in our surrounding. They are generally independent from the choice situation and hence, their impact on the information sampling is indirect. On the contrary, the choice set itself has a direct influence on the subjective saliences. For example, the decision maker may primarily focus on those attributes where alternatives are very similar to each other, because this information may be crucial for the choice. Or she concentrates on attributes with somehow outstanding values. It is therefore important to normalize the values for each attribute as described before because this guarantees representation of these effects by the attention weights. People that are present during the deliberation process like sales people or immediately prior to it like the experimenter in a laboratory context can also have a direct influence on the saliences by drawing the decision maker's attention to a specific attribute. This can be used to verify influence of attention weights by instructing decision makers to focus on certain attributes while choosing between different cars, salad dressings, chocolate bars or shoes. Corresponding experiments are under way.

Noise

In order to account for random fluctuations in the decision maker's attention (cf. Busemeyer & Townsend, 1993) which are independent of the characteristics of the choice alternatives, we add a constant to each transition probability. This makes every outgoing edge of a vertex $v \in V$ available for the next (random) step because it guarantees non-zero transition probabilities for all of them. Let $\mathcal{N}$ be a vector of length $2 \cdot N$ with all entries equal to $\frac{1}{2N}$. Weighting the transition probabilities P from the weighted comparison of alternatives by $(1 - \xi)$ with $0 \leq \xi \leq 1$ and adding the product $\xi \cdot \mathcal{N}$ yields noisy transition probabilities where ξ moderates the strength of the uniformly distributed noise:

$$P_{\mathcal{N}} = (1 - \xi) \cdot P + \xi \cdot \mathcal{N}.$$

The vector $P_{\mathcal{N}}$ of noisy transition probabilities integrates comparison of alternatives on all present attributes. Related to the $2N$-ary choice tree, this information is global as it is independent of the local counter states and thus the transition probabilities are the same for the edges emanating from each vertex.

Leakage

During their development of Decision Field Theory, Busemeyer and Townsend (1993) introduce a factor s for serial positioning effects, called "growth-decay rate". It produces recency effects for $0 < s < 1$ and primacy effects for $s < 0$. In their multi-alternative version of Decision Field Theory (Roe et al., 2001) the reverse $(1-s)$ of this factor reappears as "self-feedback loop" and accounts for the memory of previous preference states. $(1 - s) = 1$ denotes perfect memory of the previous state, $(1 - s) = 0$ no memory at all. For their simulations, Roe et al. (2001) use $(1 - s) = 0.94$ or $(1 - s) = 0.95$. Usher and McClelland (2001, 2004) adopted the idea of the self-feedback loop, but call it "leakage" λ and – based on findings from neuroscience – interpret it as "neural decay".

In order to account for decay of already sampled information over time, we implement leakage $\mathcal{L}$ into our transition probabilities. Leakage obviously depends on already sampled information and thus we normalize the current states of our $2 \cdot N$ counters to $1 - \lambda$ and for each alternative $i \in \{1, \dots, N\}$ add the result for the positive (negative) counter of alternative i to the transition probability associated with the negative (positive) counter for alternative i weighted by λ. Like this, the overall sum of the transition probabilities remains 1 and only $100 \cdot \lambda\%$ of the sampled information is actually memorized. The greater λ, the longer it takes until the process reaches a threshold. Overall, this yields

$$P_{\mathcal{NL}} = (1-\lambda) \cdot [(1-\xi) \cdot P + \xi \cdot \mathcal{N}] + \lambda \cdot \mathcal{L}.$$

Inhibition

To account for the similarity, attraction and compromise effect, DFT (Roe et al., 2001) and the LCA model (Usher & McClelland, 2004) both rely on inhibition. Whereas distance-dependent inhibition enables DFT to account for the attraction and compromise effect, global inhibition produces the similarity effect in the LCA model. We can implement both types of inhibition into the $2N$-ary choice tree model to explore their impact on the aforementioned effects. We define weights for all pairs of alternatives by either using the same weight for all pairs like Usher and McClelland (2004) do with their "global inhibition" parameter β or different weights like Roe et al. (2001) do with their distance-dependent weights (i.e. higher weights for more similar alternatives). Those weights can be stored in a symmetric $N \times N$-matrix with zeros on the diagonal.

Taking into account the basic concept of inhibition, we assume that the state of the positive counter for each alternative $i \in \{1, \ldots, N\}$ reduces sampling of positive information for all other alternatives $j \in \{1, \ldots, N\} - \{i\}$. Because this is equivalent to increasing sampling of negative information for these alternatives and vice versa for states of negative counters, we implement inhibition $\mathcal{I}$ into our model as follows: Multiplying the symmetric $N \times N$-matrix with both the vector of states of positive counters and negative counters yields two vectors with weighted sums of counter states. We concatenate them in inverted order and normalize the resulting vector of length $2N$ to μ before adding it to the vector of transition probabilities now weighted by $(1-\lambda-\mu)$. This completes the final definition of transition probabilities

$$P_{\mathcal{NLI}} = (1-\lambda-\mu) \cdot [(1-\xi) \cdot P + \xi \cdot \mathcal{N}] + \lambda \cdot \mathcal{L} + \mu \cdot \mathcal{I}.$$

In a nutshell, the transition probabilities consist of a global part that is independent from the current counter states of the random walk and a local part that depends on already sampled information. The global parts are weighted sums of comparative values P and noise $\mathcal{N}$ that remain constant during the whole process. They are complemented with leakage $\mathcal{L}$ and inhibition $\mathcal{I}$ which may change from step to step and hence, are local in the terminology of the $2N$-ary choice tree model.

3.3 Predictions of the $2N$-ary choice tree model

To show the predictions of the $2N$-ary choice tree model and how it accounts for similarity, attraction and compromise effects in choice settings with three alternatives characterized by two attributes, we run several simulations. An extension to more alternatives is straightforward. We will define values l_{ij} that range between 0 and 10. As values of choice alternatives are normalized to one on each dimension before comparison, only the relative amount of these values is of importance. Unless stated otherwise, we run 1000 trials per simulation with threshold $\theta = 20$, noise factor $\xi = 0.01$ and leakage factor $\lambda = 0.05$, but without inhibition (i.e. $\mu = 0$).

In order to meet the assumptions of the similarity, attraction and compromise effect, we constructed two equally attractive but dissimilar alternatives $A = (9, 1)$ and $B = (1, 9)$ that are both evaluated with respect to two attributes. The choice probabilities were 0.52, 0.51 and 0.47 for alternative A and 0.48, 0.49 and 0.53 for option B in three simulations with the above mentioned parameters and attribute weight 0.5 for both attributes.

To reproduce the similarity effect (Simonson, 1989) we add a third alternative to the choice set that is equal or similar to either option A or B, i.e. $C = (1, 9)$, $C_2 = (0.9, 9.1)$ or $C_3 = (1.1, 8.9)$. To prevent a combination of the similarity effect with a slight compromise effect (cf. Usher & McClelland, 2004), we will use only C for demonstration, but the results for options C_2 and C_3 are very similar to the ones presented here. The alternatives are put together in a 3×2-matrix L, whose columns are normalized to one, resulting in matrix M:

$$L = \begin{pmatrix} 9 & 1 \\ 1 & 9 \\ 1 & 9 \end{pmatrix} \qquad \text{and} \qquad M = \begin{pmatrix} 0.818 & 0.053 \\ 0.091 & 0.474 \\ 0.091 & 0.474 \end{pmatrix}.$$

M already shows smaller values for alternatives B and C on the second dimension than for alternative A on dimension one which characterizes the similarity effect. In the next step, the values on each dimension are compared to each other, resulting in a 6×2-matrix that is then multiplied by $W = \binom{0.4}{0.6}$ before being normalized to one again:

$$P' = \begin{pmatrix} 1.4545 & 0 \\ 0 & 0.4211 \\ 0 & 0.4211 \\ 0 & 0.8421 \\ 0.7273 & 0 \\ 0.7273 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix} = \begin{pmatrix} 0.5818 \\ 0.2526 \\ 0.2526 \\ 0.5053 \\ 0.2909 \\ 0.2909 \end{pmatrix}$$

and

$$P = \begin{pmatrix} 0.5818 \\ 0.2526 \\ 0.2526 \\ 0.5053 \\ 0.2909 \\ 0.2909 \end{pmatrix} /2.1742 = \begin{pmatrix} 0.2676 \\ 0.1162 \\ 0.1162 \\ 0.2324 \\ 0.1338 \\ 0.1338 \end{pmatrix}$$

Finally noise is added to this constant part of the transition probabilities. In contrast to leakage that depends on the respective counter states and has to be computed anew for every step, $P_{\mathcal{N}}$ remains constant over time. The only occasion where it changes is after withdrawal of one alternative from the choice set.

The most interesting parameters in this attempt to model a similarity effect are the attribute weights as they control the strength of the effect. Figure 3.5 demonstrates this by means of choice probabilities from simulations with different sets of attribute weights but otherwise unchanged parameters. It starts with $W = \binom{0.6}{0.4}$ and $W = \binom{0.55}{0.45}$ on the left side and gradually changes by 0.05 to $W = \binom{0.25}{0.75}$ on the right side. The relative frequency of choices for alternatives A, B, and C including the mean number of steps leading to these choices can be found in table 3.1.

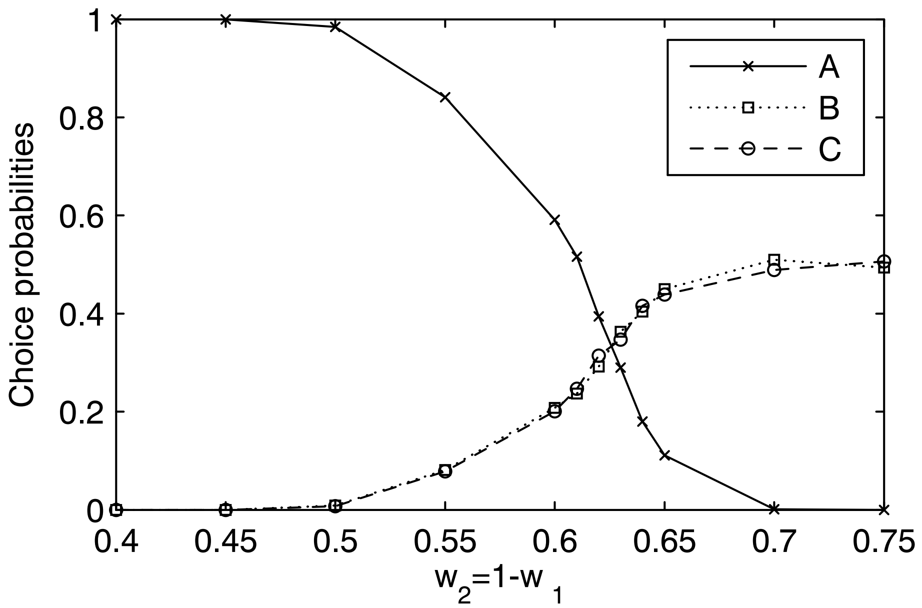

Figure 3.5: Choice probabilities for choice between three alternatives $A = (9, 1)$, $B = (1, 9)$ and $C = (1, 9)$ and different attention weights w_1 and w_2 for the two attributes. The abscissa is labeled with increasing values of w_2 corresponding to decreasing values of w_1. For $w_2 < 0.625$ a similarity effect can be observed.

w_2	0.4	0.45	0.5	0.55
A	1 (92.4)	1 (116.4)	0.984 (163.0)	0.841 (205.5)
B	0 (-)	0 (-)	0.009 (596.4)	0.081 (555.6)
C	0 (-)	0 (-)	0.007 (475.4)	0.078 (524.6)
w_2	0.6	0.61	0.62	0.63
A	0.591 (316.5)	0.516 (358.3)	0.394 (401.0)	0.29 (383.7)
B	0.208 (554.7)	0.237 (622.1)	0.292 (602.8)	0.363 (666.6)
C	0.201 (568.9)	0.247 (590.1)	0.314 (634.0)	0.347 (666.9)
w_2	0.64	0.65	0.7	0.75
A	0.180 (418.9)	0.111 (430.7)	0.001 (124.0)	0 (-)
B	0.404 (685.3)	0.450 (679.4)	0.510 (631.8)	0.494 (658.5)
C	0.416 (689.4)	0.439 (678.6)	0.489 (664.2)	0.506 (666.1)

Table 3.1: Relative number of choices and mean response times (arbitrary unit, in parentheses) for alternatives $A = (9, 1)$, $B = (1, 9)$ and $C = (1, 9)$ from 1000 simulations with $\theta = 20$, $\xi = 0.01$, $\lambda = 0.05$, $\mu = 0$ and $w_2 = 1 - w_1$ ranging from 0.4 to 0.75 as indicated in the first row.

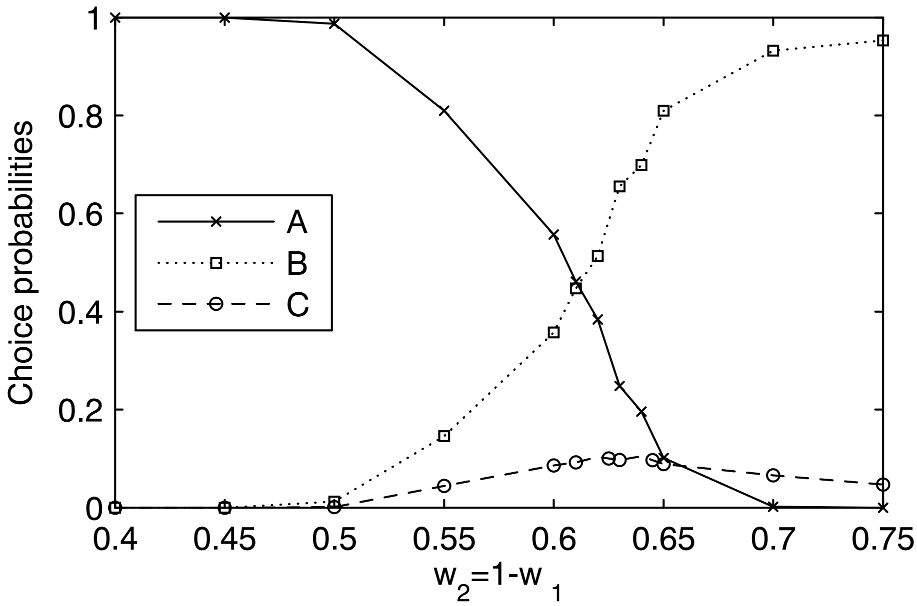

Figure 3.6: Choice probabilities for choice between three alternatives $A = (9, 1)$, $B = (1, 9)$ and $C = (1, 8.5)$ and different attention weights w_1 and w_2 for the two attributes. The abscissa is labeled with increasing values of w_2 corresponding to decreasing values of w_1. For $0.61 \leq w_2 < 0.65$ an attraction effect can be observed.

The same mechanisms account for the attraction effect (Huber et al., 1982) that occurs during choice between two equally attractive but dissimilar alternatives A and B and a third alternative C that is similar to one of these but slightly less attractive. For $A = (9, 1)$, $B = (1, 9)$ and $C = (1, 8.5)$ this yields

$$L = \begin{pmatrix} 9 & 1 \\ 1 & 9 \\ 1 & 8.5 \end{pmatrix} \qquad \text{and} \qquad M = \begin{pmatrix} 0.818 & 0.054 \\ 0.091 & 0.487 \\ 0.091 & 0.459 \end{pmatrix}.$$

As shown in figure 3.6, the attraction effect occurs between $W = \binom{0.39}{0.61}$ and $W = \binom{0.35}{0.65}$. Note that the deviation from weights $w_1 = w_2 = 0.5$ is due to a higher salience of attribute two because the values on this attribute differentiate between the alternatives. The relative frequency of choices for alternatives A, B, and C including the mean number of steps leading to these choices can be found in table 3.2.

For the compromise effect (Simonson, 1989), two equally attractive but dissimilar alternatives $A = (9, 1)$ and $B = (1, 9)$ compete against a compromise option $C = (5, 5)$. Note that the defined values for each alternative sum up to ten and thus all three alternatives objectively are equally attractive provided the attributes are equally weighted. We get

$$L = \begin{pmatrix} 9 & 1 \\ 1 & 9 \\ 5 & 5 \end{pmatrix} \qquad \text{and} \qquad M = \begin{pmatrix} 0.6 & 0.067 \\ 0.067 & 0.6 \\ 0.333 & 0.333 \end{pmatrix},$$

w_2	0.4	0.45	0.5	0.55
A	1 (93.5)	1 (118.4)	0.987 (168.7)	0.81 (204.7)
B	0 (-)	0 (-)	0.012 (593.3)	0.146 (523.4)
C	0 (-)	0 (-)	0.001 (499.0)	0.044 (601.7)
w_2	0.6	0.61	0.62	0.63
A	0.557 (313.1)	0.461 (343.6)	0.384 (391.2)	0.248 (396.5)
B	0.357 (466.2)	0.447 (476.1)	0.513 (473.6)	0.655 (487.3)
C	0.086 (558.7)	0.092 (568.1)	0.103 (569.2)	0.097 (605.3)
w_2	0.64	0.65	0.7	0.75
A	0.196 (395.5)	0.101 (342.5)	0.002 (155.5)	0 (-)
B	0.699 (448.9)	0.810 (414.6)	0.932 (237.2)	0.953 (159.5)
C	0.105 (483.7)	0.089 (459.0)	0.066 (243.7)	0.047 (174.6)

Table 3.2: Relative number of choices and mean response times (arbitrary unit, in parentheses) for alternatives $A = (9,1)$, $B = (1,9)$ and $C = (1,8.5)$ from 1000 simulations with $\theta = 20$, $\xi = 0.01$, $\lambda = 0.05$, $\mu = 0$ and $w_2 = 1 - w_1$ ranging from 0.4 to 0.75 as indicated in the first row.

and restricting $w_1 = w_2 = 0.5$ to be equal, this yields

$$P = \begin{pmatrix} 0.4 \\ 0.4 \\ 0.2667 \\ 0.4 \\ 0.4 \\ 0.2667 \end{pmatrix}.$$

So far, these transition probabilities do not seem to induce any compromise effect but as the probabilities for sampling negative information are comparatively high, withdrawal of one alternative from the choice set frequently occurs in that setting. After withdrawal of one alternative, comparison of the remaining alternatives is renewed. In the cases where alternative A or B are withdrawn, the new probabilities clearly favor the compromise option C, yielding an overall preference for that alternative:

$$M_{-A/-B} = \begin{pmatrix} 0.643 & 0.167 \\ 0.357 & 0.833 \end{pmatrix}, P_{-A/-B} = \begin{pmatrix} 0.15 \\ 0.35 \\ 0.35 \\ 0.15 \end{pmatrix}.$$

In 1000 trials with decision threshold $\theta = 20$, noise factor $\xi = 0.01$, leakage factor $\lambda = 0.05$ and no inhibition, alternatives A and B were chosen 247 (24.7 %) and 250 (25 %) times respectively and option C won 503 (50.3 %) decisions. Decreasing θ to 10 yields choice frequencies of 243 (24.3 %) for alternative A, 267 (26.7 %) for option B and 490 (49 %) for alternative C. $\theta = 5$ leads to 253 (36.9 %) choices with an average step number of 36.9 for alternative A, 269 (26.9 %) choices with 38.3 steps on average for option B and 478 (47.8 %) choices with 43.8 steps on average for alternative C. Figure 3.7 shows the response time distribution for alternative A for $\theta = 5$, $\xi = 0.01$, $\lambda = 0.05$ and $\mu = 0$. The expected response time, i.e. the mean of the distribution is 36.6. The magnitude

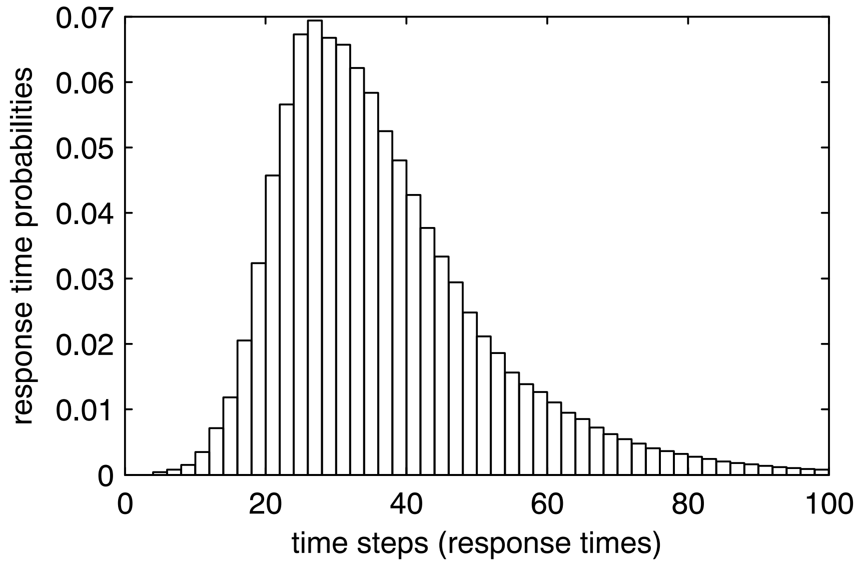

Figure 3.7: Response time distribution for alternative $A = (9, 1)$ in the compromise setup with $\theta = 5$, $\xi = 0.01$, $\lambda = 0.05$ and $\mu = 0$. The expected response time, i.e. the mean of the distribution is 36.6.

of the compromise effect can be influenced by application of an asymmetric value function after comparison of alternatives.

3.4 Comparison with other models

Multi-alternative DFT (Roe et al., 2001) and the LCA model (Usher & McClelland, 2004) both account for similarity, attraction and compromise effects in three-alternative preferential choice and thus build the theoretical background for the $2N$-ary choice tree model. Nevertheless there are some important differences and the first one to set the new model apart from the previous approaches is the attribute-wise normalization of the initially provided evaluations of alternatives. This preprocessing of input values makes them comparable over attributes. Effects that originate from differing orders of magnitude of the input values can thus be controlled by influencing the attention weights for the attributes. The comparison of alternatives on single attributes is basically the same in all three models but only the LCA model and the $2N$-ary choice tree model allow for external reference points that are not present in the choice set to influence the resulting values. Application of an asymmetric value function allows the LCA model to implement the loss-aversion principle (Kahneman & Tversky, 1979) and addition of a positive constant avoids negative activations and thus negated inhibition which was crucial for some of the results of multi-alternative DFT. Both concepts (asymmetric value function and positive constant) can be implemented into the $2N$-ary choice tree model as well but do not affect its ability to account for the aforementioned effects (except for the magnitude of the compromise effect). Whereas all three models use leakage to account for decay of already sampled information over time and have a random part that implements noise in human decision making, inhibition is another crucial difference between them. In multi-alternative DFT, local inhibition explains the attraction and compromise effect, the LCA model uses global inhibition to account for the similarity effect. Both types of inhibition can be implemented in the $2N$-ary choice tree model but are not necessary for explanation of the three effects.

Beside some similarities and dissimilarities between the models, in particular with respect to some underlying psychological concepts the $2N$-ary choice tree model is the first to provide expected choice probabilities and response times in closed form and thus allows for convenient estimation of the model parameters from the observed choice times and frequencies in experimental settings. Furthermore, it can be extended to more than three choice alternatives in a straightforward way to account for choice behavior in more complex, and possibly more realistic choice situations.

3.5 Concluding remarks

The $2N$-ary choice tree model provides alternative explanations for the similarity, attraction and compromise effect that can be experimentally tested as suggested before. Especially the manipulation of attention weights is of interest, because it differentiates the model on hand from former approaches and should allow to experimentally produce similarity and attraction effects which has been proven to be difficult in the past. One problem, however, we are currently encountering is limited machine accuracy which leads to accumulation of rounding errors during calculation of expected choice probabilities and response times.

3.6 Formal statement and proof

We can approximate the expected choice probabilities and hitting times up to absolute accuracy in finite time. This follows from theorem 1:

Theorem 1. *Each random walk Y_n on the above defined tree $T = (V, E, r)$ with transition probabilities p_e, ends in finite time with probability one.*

Corollary 1. *With probability one only finitely many addends in equations (1) and (2) are unequal zero.*

Considering expected values, i.e. limits of infinite sums, it is helpful to make use of a concept that allows for propositions about asymptotic behavior. For each alternative, the difference of the two counters that are associated with this alternative resemble a birth-death chain:

Definition 4 (birth-death chain)**.** *A sequence of random variables $X_1, X_2, \dots$ with values in a countable state space $\mathcal{S} \equiv \{0, 1, 2, \dots\} \subseteq \mathbb{N}$ is called* Markov chain*, if it satisfies the Markov property*

$$\mathbb{P}[X_{n+1} = x | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n] = \mathbb{P}[X_{n+1} = x | X_n = x_n].$$

A Markov chain is called time homogeneous*, if $\mathbb{P}[X_{n+1} = x | X_n = y] = p(x, y)$ for all n, i.e. the probability for going from x to y is independent from n. A* birth-death chain *is a time homogeneous Markov chain that does not skip any* *state. Its transition probabilities* $p(x, y)$ *are equal to zero for all* $x, y \in \mathcal{S}$ *with* $|y - x| > 1$*. A* non-homogeneous birth-death chain *is a birth-death chain that is not time homogeneous.*

Durrett (2010) proves the following theorem for (non-homogeneous) birth-death chains as special case of Markov chains:

Theorem 2. *Let* X_n *be a Markov chain and suppose*

$$\mathbb{P}[\cup_{m=n+1}^{\infty}\{X_m \in B_m\}|X_n] \geq \delta > 0 \text{ on } \{X_n \in A_n\}.$$

Then $\mathbb{P}[\{X_n \in A_n \text{ infinitely often }\} - \{X_n \in B_n \text{ infinitely often }\}] = 0.$

For each alternative $i \in \{1, 2, \ldots, N\}$, the above mentioned difference of its two counters can be interpreted as non-homogeneous birth-death chain X_n with absorbing states θ_i^- and θ_i^+ and state space $\mathcal{S} = \{\theta_i^-, (\theta_i^- + 1), (\theta_i^- + 2), \ldots, (\theta_i^+ - 2), (\theta_i^+ - 1), \theta_i^+\}$. Its transition probabilities are

$$\left.\begin{aligned} p^n(x, x+1) &= p_x^n = p_i^n(x), \\ p^n(x, x-1) &= q_x^n = p_{N+i}^n(x), \\ p^n(x, x) &= r_x^n = 1 - p_x^n - q_x^n, \end{aligned}\right\} \qquad \text{for } x \in \mathcal{S} - \{\theta_i^-, \theta_i^+\}$$

and

$$\left.\begin{aligned} p_x^n = q_x^n &= 0, \\ r_x^n &= 1, \end{aligned}\right\} \qquad \text{for } x \in \{\theta_i^-, \theta_i^+\}.$$

Due to the noise in the transition probabilities, $p_x^n > 0$ and $q_x^n > 0$ for all $x \in \mathcal{S} - \{\theta_i^-, \theta_i^+\}$. It follows that the probability for walking the direct way from $x \in \mathcal{S} - \{\theta_i^-, \theta_i^+\}$ to either θ_i^- or θ_i^+ is

$$\delta_x := \left(\prod_{y=\theta_i^-+1}^{x} q_i^{n+x-y}(y) + \prod_{z=x}^{\theta_i^+-1} p_i^{n+z-x}(z) \right) > 0$$

and thus

$$\delta := \min_{\theta_i^- < x < \theta_i^+} \delta_x > 0.$$

Define $T_{\theta_i^-} = \inf\{n : X_n = \theta_i^-\}$, $T_{\theta_i^+} = \inf\{n : X_n = \theta_i^+\}$, $T_i = T_{\theta_i^-} \wedge T_{\theta_i^+}$, $A_n = \mathcal{S}$ and $B_n = \{\theta_i^-, \theta_i^+\}$. Then $\{X_m \in B_m\}$ is equivalent to $T_i \leq m$ and

$$\mathbb{P}\left[\cup_{m=n+1}^{\infty}\{X_m \in B_m\}|X_n\right] = 1 > \delta$$

for $X_n \in B_n$. The probability for walking from any $x \in \mathcal{S}$ to either θ_i^- or θ_i^+ on every possible way is

$$\mathbb{P}\left[\cup_{m=n+1}^{\infty}\{X_m \in B_m\}|X_n\right] \geq \delta > 0$$

and thus fulfills the assumptions of theorem 2. It follows that

$$\mathbb{P}[\{X_n \in A_n \text{ infinitely often }\} - \{X_n \in B_n \text{ infinitely often }\}] = 0$$

which is equivalent to

$$\mathbb{P}\left[\{X_n \in A_n - B_n \text{ finitely often}\}\right] = \mathbb{P}\left[T_i < \infty\right] = 1$$

and as this is true for every alternative $i \in \{1, 2, \ldots, N\}$, $\mathbb{P}\left[T, \infty\right] = 1$ holds for $T := \min T_1, T_2, \ldots, T_N$. This proves theorem 1.

Chapter 4

The simple choice tree model

Lena M. Wollschlaeger, Adele Diederich

This chapter has been published in the Proceedings of the 39th annual conference of the Cognitive Science Society under the title "A computational model for constructing preferences for multiple choice options" on July 29, 2017.

Abstract

When choosing between multiple alternatives, people usually do not have ready-made preferences in their mind but rather construct them on the go. The 2N-ary Choice Tree Model (Wollschlaeger & Diederich, 2012) proposes a preference construction process for N choice options from description, which is based on attribute weights, differences between attribute values, and noise. It is able to produce similarity, attraction, and compromise effects, which have become a benchmark for multi-alternative choice models, but also several other context and reference point effects. Here, we present a new and mathematically tractable version of the model – the Simple Choice Tree Model – which also explains the above mentioned effects and additionally accounts for the positive correlation between the attraction and compromise effect, and the negative correlation between these two and the similarity effect as observed by Berkowitsch et al. (2014).

4.1 Introduction

The decision making process involves various steps such as setting and prioritizing objectives, identifying choice alternatives, searching for information, developing preferences, and eventually taking a course of action. Here, we focus on developing preferences in multi-alternative choice situations and use in the following decision making from description as basic paradigm. Given a set of at least three choice alternatives that are described by at least two attributes, which they have in common, how do people choose one of these options? Simon (1955) argues that preferences in this kind of situation are dynamically constructed over time due to limited processing capacities. The decision maker experiences preference uncertainty (cf. Simonson, 1989) and tries to overcome it by gradually integrating the given information (see Payne et al., 1992, for a review on constructive processing in decision making). The resulting preferences are stochastic and highly dependent on the context, i.e., on the alternatives in the choice set and on any external reference points. Naturally, a model describing multi-alternative decision making from description should be a context-sensitive cognitive process model. The recently proposed 2N-ary Choice Tree Model for preference construction for N choice options (2NCT; Wollschlaeger & Diederich, 2012) assumes that the decision maker compares attribute values within attributes and between alternatives in a pairwise manner. Attributes are selected for examination based on attribute weights that reflect salience. Within attributes, pairs of attribute values are selected for comparison based on so-called comparison values. In the 2NCT Model, the comparison values have a "global" component that remains constant over time during preference construction, a "local" component that depends on the outcomes of previous comparisons (reflecting leakage and inhibition, cf. Roe et al., 2001; Usher & McClelland, 2004), and a random component. Advantageous and disadvantageous comparison outcomes for each alternative are counted separately and the difference of these counters is compared to two thresholds: a positive choice criterion and a negative elimination criterion. Implementation of an asymmetric value function (emphasizing disadvantageous comparison outcomes, cf. Usher & McClelland, 2004) into the 2NCT Model is possible. Here, we present a revised and simpler version of the 2N-ary Choice Tree Model, the Simple Choice Tree (SCT) Model. Therein, the local component is omitted from the definition of comparison values, making the model mathematically tractable while maintaining its ability to account for similarity, attraction and compromise effects. Furthermore, a new parameter, the focus weight λ, is introduced. It replaces the asymmetric value function and allows the SCT Model to account for correlations between the effects.

Benchmark: Context effects

Three context effects, demonstrating the influence of choice set composition on preferences, have played a prominent role in the multi-alternative preference construction modeling literature: The similarity effect, the compromise effect, and the attraction effect. All three effects occur when adding a third alternative to a set of two equally attractive yet clearly distinguishable options described by two attributes. Let A_1 and A_2 be two choice alternatives with two common attributes, D_1 and D_2, describing them. We assume that D_1 is the unique strongest attribute for A_1 and D_2 is the unique strongest attribute for A_2, that is, A_1 scores high on D_1 but low on D_2 and vice versa for A_2. One can think of the alternatives as placed in a two-dimensional space with dimensions D_1 and D_2. We further assume that the probability for choosing alternative A_1 from the binary choice set is equal to the probability for choosing alternative A_2, $P(A_1|A_1, A_2) = P(A_2|A_1, A_2)$.

Similarity effect

The similarity effect was named and first studied systematically by Tversky (1972b). He observed the effect when comparing the binary choice set $\{A_1, A_2\}$ to the ternary choice set $\{A_1, A_2, A_3\}$ where A_3 is similar to one of the original alternatives, say A_1, in scoring high on attribute D_1 and low on attribute D_2 while overall being similarly attractive (i.e. $P(A_1|A_1, A_3) = P(A_3|A_1, A_3)$). The probability of choosing A_1 over A_2 decreases when the decision maker chooses from the ternary choice set as compared to the binary set:

$$\frac{P(A_1|A_1, A_2)}{P(A_2|A_1, A_2)} > \frac{P(A_1|A_1, A_2, A_3)}{P(A_2|A_1, A_2, A_3)}.$$

Attraction effect

The attraction effect (or decoy effect or asymmetric dominance effect) was introduced by Huber et al. (1982) as consistent violation of the regularity principle. This principle, as presumed for example by the theory of Elimination by Aspects (Tversky, 1972b), states that additional alternatives cannot increase the choice probabilities of the original options. However, Huber et al. (1982) claim that the relative probability for choosing alternative, say, A_1 can be increased by adding a third alternative A_3 to the choice set that is similar to but dominated by A_1 (and symmetrically for alternative A_2). A_3 then serves as a decoy for alternative A_1, drawing attention to it and therewith improving its evaluation and increasing its choice probability.

Compromise effect

Originally intended to explain the attraction effect, the theory of Reason-based Choice (Simonson, 1989) predicts an additional context effect, the compromise effect. It occurs when a third alternative A_3, equally attractive as the original alternatives A_1 and A_2, but more extreme with respect to the attribute values, is added to the choice set. If A_3 is more extreme than alternative A_1, that is, if it scores higher than A_1 on attribute D_1 but lower on attribute D_2, then it increases the choice share of A_1 as compared to the binary situation (and vice versa for alternative A_2): $P(A_1|A_1, A_2, A_3)/P(A_2|A_1, A_2, A_3) > P(A_1|A_1, A_2)/P(A_2|A_1, A_2)$. However, note that the more similar the additional extreme alternative A_3 is to its adjacent alternative A_1, the more shares it takes away from A_1 via the similarity effect.

Interrelations of the effects

Recently, several studies have explored similarity, attraction and compromise effects and their interrelations in different choice scenarios. In a within-subject consumer choice design, Berkowitsch et al. (2014) find that the similarity effect is negatively correlated with both the attraction and the compromise effect while the latter two are positively correlated. In a similar vein, Liew et al. (2016) criticize that most of the results regarding context effects are based on averages over participants, not taking into account individual differences. Before analyzing the data from their inference and consumer choice experiments, they cluster it according to the observed choice patterns. The differences between clusters are remarkable, some even show negative (reverse) context effects while positive effects are observed in the averaged data. Before explaining how the Simple Choice Tree (SCT) Model accounts for the similarity, attraction and compromise effects and their interrelations, we introduce the basic mechanisms of the model.

4.2 The simple choice tree model

Let n_a be the number of alternatives under consideration, $\{A_1, A_2, \ldots, A_{n_a}\}$, and n_d the number of attributes, $\{D_1, \ldots, D_{n_d}\}$, that characterize them. The decision maker is provided with one attribute value per alternative per attribute, that is, $n_a \cdot n_d$ attribute values in total. Let m_{ij} be the attribute value for alternative A_i with respect to attribute D_j. Attribute values within attributes and between alternatives are repeatedly compared and the resulting evidence is accumulated in two counters S_i^+ and S_i^- for each alternative $A_i, i \in \{1, \dots, n_a\}$. The positive counter S_i^+ accumulates evidence for choosing alternative A_i and the negative counter S_i^- accumulates evidence for rejecting it. Here, the initial counter states are set to zero, $S_i^+(0) = 0 = S_i^-(0)$. Definition of non-zero initial counter states accounting for prior knowledge about the choice alternatives is possible. However, these additional free parameters make the model less parsimonious and complicate parameter estimation. The counter states at time t, $S_i^+(t)$ and $S_i^-(t)$, are the initial counter states increased by the respective evidence accumulated until t. Their difference defines the momentary preference state for alternative A_i at time t: $Pref(A_i, t) = S_i^+(t) - S_i^-(t)$. We will now answer the following questions: (1) How is attention allocated between choice alternatives and attribute values? (2) How are alternatives evaluated and how is evidence accumulated? (3) When does evidence accumulation stop and which alternative is chosen?

Attention allocation

At the beginning of the process, when information about the alternatives and attributes is made available to the decision maker, each attribute D_j, $j \in \{1, \dots, n_d\}$, is assigned a weight ω_j, $0 \leq \omega_j \leq 1$, reflecting its salience. The attribute weights determine how much attention the decision maker gives to the respective attributes during the preference construction process. Attributes with higher weights get more attention than attributes with lower weights. To allow for at least some of the attention to be allocated randomly between attributes, we define a random component (see below) for which an additional weight $\omega_0, 0 \leq \omega_0 \leq 1$ is designated. Assuming that the weights sum up to one, $\sum_{j=0}^{n_d} \omega_j = 1$, they can be interpreted as attention probabilities for the attributes: At each points of the preference construction process, the decision maker concentrates on attribute $D_j, j \in \{1, \dots, n_d\}$ with probability ω_j.

Having selected an attribute D_j, the decision maker concentrates on the specific attribute values of two alternatives and compares them. Pairs of attribute values are selected for comparison according to their importance for the decision. The more diagnostic the attribute values are, i.e., the more they discriminate between the alternatives, the more important they become for the decision. Pair selection probabilities within attribute D_j are therefore defined to be proportional to the absolute differences $d_{ikj} = |m_{ij} - m_{kj}|$, $i \neq k \in \{1, \dots, n_a\}$. In order to obtain probabilities, we normalize these differences to sum up to one: The probability for selecting the pair $\{m_{ij}, m_{kj}\}$ for comparison is $p_{ikj} = d_{ikj} / \sum_{\{l,m\}} d_{lmj}$, $l \neq m \in \{1, \dots, n_a\}$. Note that the normalization of absolute differences balances out inequalities between attributes with – on average – bigger or smaller differences. Higher salience of an attribute D_j, $j \in \{1, \dots, n_d\}$, with, for example, higher absolute differences, is thus not hard-wired into the model but is reflected in a higher attribute weight ω_j instead.

Preference sampling

The actual comparison of the two selected attribute values m_{ij} and m_{kj} is ordinal and directional: Let $m_{ij} > m_{kj}$, then the comparison can be either positively phrased, e.g. "m_{ij} is greater than m_{kj}", or it can be negatively phrased, e.g. "m_{kj} is smaller than m_{ij}". For the positive phrasing, m_{ij} is called *focus value* and m_{kj}

is called *reference value*. The focus value determines the counter whose state is increased by $+1$, here S_i^+, since the comparison is advantageous for the associated alternative A_i. For the negative phrasing, m_{kj} is the focus value and m_{ij} is the reference value, leading to an increase by $+1$ of counter S_k^-, since the comparison is disadvantageous for alternative A_k. Which phrasing the decision maker uses for the comparison and therewith which counter is updated might, for example, depend on the wording of the task or the decision maker's attitude (cf. Choplin & Hummel, 2002). It is implemented into the model via the focus weight λ, $0 \leq \lambda \leq 1$. If $\lambda = 1 - \lambda = 0.5$, the decision maker uses the positive and negative phrasing both about equally often. If $\lambda > 0.5$, the decision maker has a tendency towards the negative phrasing and towards updating negative counters. If $\lambda < 0.5$, the decision maker has a tendency towards the positive phrasing and towards updating positive counters. The focus weight λ replaces the asymmetric value function that was applied to the absolute differences between attribute values in the original 2NCT Model (Wollschlaeger & Diederich, 2012). While the asymmetric value function hard-wired a tendency towards updating negative counters into the 2NCT Model, weighting with λ allows for flexible balancing of attention to positive versus negative aspects of the alternatives in the SCT Model. It is therefore especially useful in situations without a loss/gain-framing, e.g., in perceptual or preferential choice. Note that λ is a global weight and independent from the attributes and attribute values. However, it allows us to define counter updating probabilities for the positive and negative counter of alternative $A_i, i \in \{1, \ldots, n_a\}$ with respect to attribute $D_j, j \in \{1, \ldots, n_d\}$: $p_{ij}^+ = \sum_{k:(m_{ij} > m_{kj})} (1 - \lambda) \cdot p_{ikj}$ for updating S_i^+ and $p_{ij}^- = \sum_{k:(m_{ij} < m_{kj})} \lambda \cdot p_{ikj}$ for updating S_i^-.

Finally, the random component accounts for times where counter states are updated at random and without any connection to the actual attribute values (for instance due to inattention or misperception, cf. Busemeyer & Townsend, 1993). Technically, it is treated as an additional (phantom) attribute D_0. The counter updating probabilities $p_{i0}^+ = p_{i0}^- = 1/(2 \cdot n_a)$, $i \in \{1, \ldots, n_a\}$ with respect to D_0 depend on the number of available choice alternatives and therefore sum up to one: $\sum_{i=1}^{n_a} (p_{i0}^+ + p_{i0}^-) = 1$.

Combining attribute-wise counter updating probabilities $p_{ij}^{\pm}$ with attribute weights ω_j, we can now define weighted counter updating probabilities for the positive and negative counter of alternative A_i:

$$p_i^+ = \sum_{j=0}^{n_d} p_{ij}^+ \cdot \omega_j \quad \text{and} \quad p_i^- = \sum_{j=0}^{n_d} p_{ij}^- \cdot \omega_j. \tag{4.1}$$

Choice tree and stopping rules

Starting with the presentation of the choice alternatives and their attribute values, the preference construction process consists of a sequence of counter updates. In principle, every possible sequence of counter updates may occur and it is therefore of interest to have them conveniently summarized. For this purpose, we introduce the $(2 \cdot n_a)$-ary choice tree $T = (V, E, r)$ with vertices V, edges $E \subseteq V \times V$ and root $r \in V$, where all vertices are directed away from r and each internal vertex $v \in V$ has $2 \cdot n_a$ children that are associated with the $2 \cdot n_a$ counters. Figure 4.1 shows an example with three choice alternatives and six counters. The preference construction process is represented by a random walk on T, beginning at the root

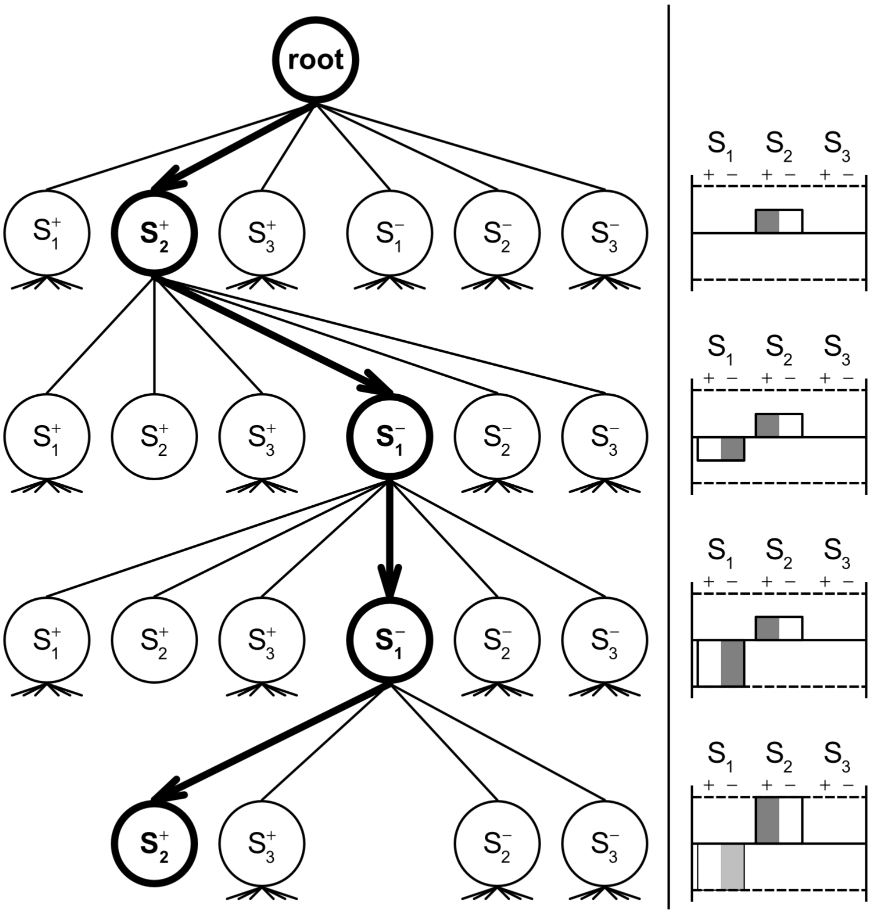

Figure 4.1: A random walk on the choice tree for three alternatives. The associated sequence of counter updates is $S_2^+, S_1^-, S_1^-, S_2^+$ and the probability for walking along this specific path is $p_2^+ \cdot p_1^- \cdot p_1^- \cdot p_2^+$. Supposing that the rejection threshold θ^- is equal to -2 and the choice threshold θ^+ is equal to 2, this sequence implicates first rejection of alternative A_1 and then choice of alternative A_2. When A_1 is eliminated from the choice set, the vertices associated with its counters no longer appear in the choice tree, as can be seen in the bottom row of vertices here.

and passing from there through an edge to another vertex, triggering the update (increase by $+1$) of the associated counter, moving on through another edge and so forth. The next edge to pass through is chosen according to the updating probability of the counter associated with its endpoint. Note that for each vertex the transition probabilities of all outgoing edges sum up to one. An example path of this random walk is pictured in bold in Figure 4.1.

The preference construction process stops when enough evidence has been accumulated to make the required choice. To this end, the preference states $Pref(A_i, t) = S_i^+(t) - S_i^-(t), i \in \{1, \ldots, n_a\}$ are constantly compared to two thresholds, a positive threshold θ^+ and a negative threshold $\theta^- = -\theta^+$. If the preference state for alternative A_i hits the positive threshold, the process stops and A_i is chosen. If, on the other hand, the preference state for alternative A_k hits the negative threshold, A_k is eliminated from the choice set and the process continues with the remaining alternatives until one of them is chosen or until all but one of them have been eliminated. Consider a simple example with three choice alternatives $\{A_1, A_2, A_3\}$ and thresholds $\theta^+ = 2$ and $\theta^- = -2$. The sample path in Figure 4.1 with its associated sequence of counter updates $S_2^+, S_1^-, S_1^-, S_2^+$, leads to elimination of alternative A_1 after three steps and choice of alternative A_2 after four steps. Other possible sequences resulting in choice of alternative A_2 include $S_3^+, S_1^-, S_2^+, S_2^+$ with direct choice of A_2 after four steps, and $S_1^-, S_3^-, S_3^-, S_1^-$ with elimination of alternatives A_3 after three steps and A_1 after four steps and therewith choice of the only remaining alternative A_2.

Choice probabilities and expected response times

The probability for walking along a specific path as, for example, shown in Figure 4.1, is the product of the transition probabilities along the respective edges. The choice probability for alternative $A_i, i \in \{1, \ldots, n_a\}$ is equal to the sum of the probabilities for walking along all the specific paths that lead to choice of alternative A_i. Since it is not feasible to calculate probabilities separately for each path and sum them up, we will analyze preference states, choice probabilities and response times instead by interpreting them as independent birth-death Markov chains with absorbing boundaries θ^+ and θ^-. The state space of these birth-death chains $Pref(A_i, t) = S_i^+(t) - S_i^-(t) =: S_i(t), i \in \{1, \ldots, n_a\}$ is $\mathcal{S} := \{\theta^-, \ldots, -1, 0, 1, \ldots, \theta^+\}$, with $|\mathcal{S}| = \theta^+ - \theta^- + 1$. The transition probabilities are

$$\left.\begin{array}{ll} p_i(x, x+1) & = p_i^+ > 0 \\ p_i(x, x-1) & = p_i^- > 0 \\ p_i(x, x) = 1 - p_i^+ - p_i^- & = p_i^0 > 0 \end{array}\right\} \text{ for } x \in \mathcal{S} - \{-\theta^-, \theta^+\},$$

where $p_i^{\pm}$ is defined in Eq. 4.1 above; $p_i(x, x+1) = p_i(x, x-1) = 0$, $p_i(x, x) = 1$, for $x \in \{-\theta^-, \theta^+\}$; and zero otherwise. They form a $|\mathcal{S}| \times |\mathcal{S}|$ transition probability matrix $P_i' = (p_{rs}')_{r,s=1,\ldots,|\mathcal{S}|}$, where p_{rs}' is the probability for the birth-death chain to transition from state x_r to state x_s in one step. P_i' can be written in its canonical form P_i by rearranging the rows and columns (changing the indices of the states such that the absorbing states $-\theta^-$ and θ come first). P_i can be decomposed into a 2×2 identity matrix I_2, a $2 \times n_t$ matrix 0 of zeros with $n_t = |\mathcal{S}| - 2$ (the number of transient states in $\mathcal{S}$), a $n_t \times 2$ matrix R_i, containing the probabilities for entering the absorbing states θ^+ and θ^-, that is, for hitting the elimination or choice threshold, and a $n_t \times n_t$ matrix Q_i, containing the transition probabilities between transient states (cf. Diederich, 1997): $P_i = \begin{pmatrix} I_2 & 0 \\ R_i & Q_i \end{pmatrix}$.

Given a row vector Z_i of length n_t which represents the initial preference state (e.g., $\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \end{pmatrix}$) or the initial distribution of preference over the transient states (e.g., $\begin{pmatrix} 0.05 & 0.10 & 0.70 & 0.10 & 0.05 \end{pmatrix}$, cf. Diederich & Busemeyer, 2003) for alternative A_i, the probability that the process is absorbed during the first step can be obtained by multiplying Z_i and R_i, yielding a vector of length 2: $Z_i \cdot R_i = [P(S_i(1) = \theta^+), P(S_i(1) = -\theta^-)]$. In the case that the process was not absorbed during the first step, the distribution of preference over the transient states after the first step is given by $Z_i \cdot Q_i$, a vector of length n_t. Multiplying the result with the matrix R_i yields the probabilities of absorption in the second step: $Z_i \cdot Q_i \cdot R_i = [P(S_i(2) = \theta^+), P(S_i(2) = -\theta^-)]$. The distribution of preference over the transient states is given by $(Z_i \cdot Q_i) \cdot Q_i = Z_i \cdot (Q_i \cdot Q_i) = Z_i \cdot (Q_i)^2$. The entries of the $n_t \times n_t$ matrix $(Q_i)^2$ are 2-step transition probabilities between the transient states, allowing for calculation of absorption in the third step: $Z_i \cdot (Q_i)^2 \cdot R_i = [P(S_i(3) = \theta^+), P(S_i(3) = -\theta^-)]$. Iterating these results indicates that all the relevant probabilities can be obtained from the vector Z_i, the matrix R_i and powers of the matrix Q_i. Since Q_i is a tridiagonal Toeplitz matrix (the entries on the main diagonal are all equal to p_i^0, the entries on the diagonal above the main diagonal are equal to p_i^+ and the entries on the diagonal below the main diagonal are equal to p_i^-), its eigenvalues, eigenvectors and its powers are known and given in closed form (Salkuyeh, 2006), making it easy to compute all the relevant quantities.

We are interested in the conditional probabilities and expected hitting times for each alternative A_i, $i \in \{1, \dots, n_a\}$, given that A_i is the first alternative to be chosen/eliminated. Therefore, we have to determine the probability that alternative A_k, $k \in \{1, \dots, n_a\}$ with $k \neq i$, has not been chosen/eliminated until time t. It is given by

$$\begin{aligned} P(-\theta^- < S_k(T) < \theta) &= 1 - \sum_{t=1}^{T} Z_k \cdot (Q_k)^{t-1} \cdot R_k \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= 1 - Z_k \cdot \left(\sum_{t=1}^{T} (Q_k)^{t-1} \right) \cdot R_k \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \end{aligned}$$

The choice and elimination probability for alternative A_i at time T is then equal to

$$\begin{aligned} &[P(S_i(T) = -\theta^-), P(S_i(T) = \theta)] \\ &= \left(Z_i \cdot \sum_{t=1}^{T} (Q_i)^{t-1} \cdot R_i \right) \cdot \prod_{k \neq i} \left(P(-\theta^- < S_k(T) < \theta) \right). \end{aligned}$$

Overall, this yields probabilities

$$\begin{aligned} &[P(chooseA_i), P(eliminateA_i)] \\ &= \sum_{T=1}^{\infty} \left([P(S_i(T) = -\theta^-), P(S_i(T) = \theta)] \right) \end{aligned}$$

and expected response times

$$\begin{aligned} &[E(T_i|chooseA_i), E(T_i|eliminateA_i)] \\ &= \sum_{T=1}^{\infty} T \cdot \left([P(S_i(T) = -\theta^-), P(S_i(T) = \theta)] \right). \end{aligned}$$

Note that the infinite sums over T have only a finite number of nonzero addends, since $P(N_i < \infty) = 1$ for all $i \in \{1, \dots, n_a\}$, thus the choice/elimination probabilities and expected response times can be easily computed.

4.3 Context effects explained

Three interacting mechanisms produce similarity, attraction, and compromise effects in the Simple Choice Tree Model: (1) selection of pairs of attribute values for comparison based on normalized differences, (2) the possibility to eliminate unwanted alternatives from the choice set, and (3) weighting of attributes based on salience. The first mechanism leads to a higher impact of dissimilar alternatives on the updating probabilities and thus faster evidence accumulation for alternatives with more distant competitors. In the similarity and attraction settings, this applies to the dissimilar alternative A_2, and in the compromise situation to the extreme alternatives A_2 and A_3. The second mechanism and the related focus weight λ determine whether choices are more likely to be based on eliminations or to be made directly. The greater λ, the more likely are the choices based on eliminations. In the similarity situation, greater λ leads to faster elimination of the dissimilar alternative A_2 and subsequent choice or elimination of either alternative A_1 or A_3, that is, a small or even negative similarity effect. On the other hand, smaller λ leads to more direct choices of alternative A_2 and thus a higher similarity effect. Regarding the dissimilar alternative A_2, the same is true in the attraction situation. Greater λ leads to faster elimination of A_2 while smaller λ leads to more direct choices of alternative A_2. However, the attraction effect is higher for greater λ, since after elimination of alternative A_2, either the dominating option A_1 is chosen directly or the dominated option A_3 is eliminated first. In the compromise setting, greater λ increases the probability for the extreme options to be eliminated from the choice set, leaving the decision maker with the compromise option. Smaller λ on the other hand more likely leads to choice of an extreme option and thus a smaller or even negative compromise effect. Attribute weights further moderate the strengths of the context effects, but as long as they are more or less balanced, they play a minor role in the explanation of the similarity, attraction, and compromise effects. However, a high attribute weight is able to bias choice towards the alternative that scores highest on that attribute, covering any context effect.

We ran several simulations to illustrate these mechanisms. The available choice alternatives were $A_1 = (70, 30)$, $A_2 = (30, 70)$ and $A_3 = (70, 30)$ for the similarity effect, $A_3 = (65, 25)$ for the attraction effect, $A_3 = (90, 10)$ for the asymmetric compromise effect, or $A_3 = (50, 50)$ for the symmetric compromise effect. The attribute weights were $\omega_0 = 0.1$ and $\omega_1 = \omega_2 = 0.45$, and the focus weight λ varied between 0 and 1 in steps of 0.1. For each data point we ran 10000 simulations and the resulting choice probabilities are presented in figure 4.2. According to the simulations, the similarity effect is opposed to the attraction and the compromise effect. The similarity effect is strongest for low λ, whereas the attraction and the compromise effect are strongest for high λ. This prediction is consistent with the finding that the attraction and the compromise effect are positively correlated with each other and negatively correlated with the similarity effect (Berkowitsch et al., 2014). Note that λ is assumed to be a global weight that does not change between trials but may vary between participants.

4.4 Conclusion

We propose a revised and simpler version of the $2N$-ary Choice Tree Model (Wollschlaeger & Diederich, 2012), the Simple Choice Tree (SCT) Model. It pre-

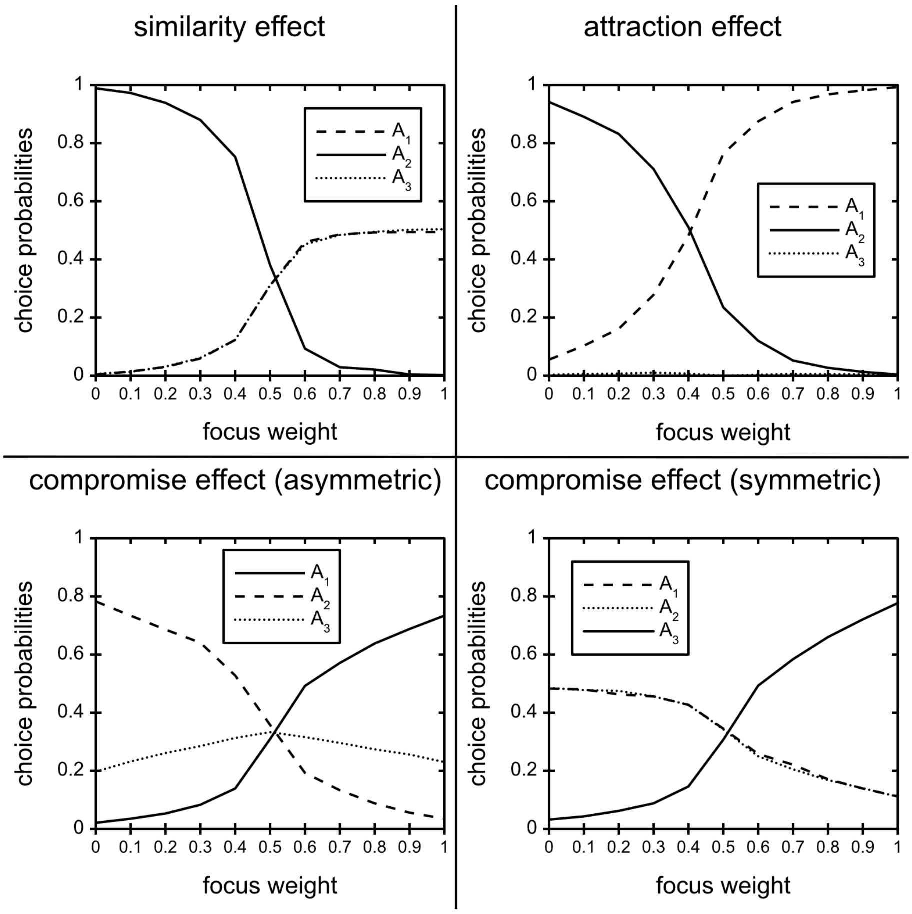

Figure 4.2: Simulations of choice probabilities for changing focus weight λ in the similarity, attraction, and compromise situation. There is a positive similarity effect for smaller λ and a negative similarity effect for larger λ (upper left) and vice versa for the attraction effect (upper right). The compromise effect (lower left and right) shows for larger λ and is reversed for smaller λ.

dicts choice probabilities and response times in multi-alternative multi-attribute preferential choice from description and accounts for several effects observed in these situations, including the similarity, attraction, and compromise effect. The SCT Model shares several aspects with existing models: Like Decision by Sampling (DbS; Stewart et al., 2006), it proposes binary ordinal comparisons and frequency accumulation as basic mechanisms. In DbS, however, pairs of attribute values are chosen at random and reference values may be sampled from long-term memory as well as from the given context. Only advantageous comparisons are counted and the model is not able to account for the above mentioned context effects, nor does it provide solutions for choice probabilities or choice response times. Multi-alternative Decision Field Theory (MDFT; Roe et al., 2001) and the Leaky Competing Accumulator (LCA) Model (Usher & McClelland, 2001, 2004) provide such solutions only for fixed stopping times. Both models, like the SCT Model, are based on pairwise differences of attribute values. To account for the similarity, attraction, and compromise effect simultaneously, however, additional non-linear mechanisms (among others leakage and inhibition, cf. the original 2NCT Model) are required, preventing the models from providing mathematically tractable solutions for optional stopping times. Elimination by Aspects (EBA; Tversky, 1972b) proposes "a covert elimination process based on sequential selection of aspects" (p. 296). As an early example for a cognitive process model, it does not make any predictions about choice response times and accounts only for the similarity effect. The SCT model mimics EBA for high values of the focus weight λ, where mostly disadvantageous comparison outcomes are considered and decisions are based on the elimination of choice alternatives. The Multi-attribute Linear Ballistic Accumulator Model (MLBA; Trueblood et al., 2014), basically a deterministic version of MDFT, provides analytic solutions for expected response times and choice probabilities like the SCT Model. However, it is unclear if and how the response times are related to the actual integration of information. Furthermore, the model has mostly been applied with fixed stopping times until now. Additional mechanisms allow the MLBA model to account for the compromise effect (a curved subjective value function) and the similarity effect (a higher weight on supportive information as compared to disconfirmatory evidence). The latter is comparable to low values of the focus weight λ in the SCT Model. To summarize, the SCT Model combines aspects of competing models in a new way, yielding qualitatively new explanations for the context effects and additionally predicting correlation patterns amongst the effects. It provides mathematically tractable solutions for both choice probabilities and expected choice response times for optional stopping times, by that outperforming existing models.

Chapter 5

The simple choice tree model: Additional simulations

In order to explore the capabilities of the simple choice tree model and the influence that the focus weight λ has on the choice probabilities and response times, I ran a series of simulations in Matlab. The code is given in appendix A. Some of the results are reported in sections 4.3 and 6.3. There, the attribute weight for dimension D_1 is the same as the attribute weight for dimension D_2, that is, $\omega_1 = \omega_2$ (and the weight for the random component is equal to $\omega_0 = 0.1$). Here, I report results for relative attribute weights ranging from 0 to 1 (and 1 to 0) that are scaled to ranges $[0, 0.9]$ and $[0.9, 0]$ because of the weight for the random component, $\omega_0 = 0.1$.

5.1 Similarity effect

Figures 5.1 and 5.2 show choice probabilities and response times (z-axis) for alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (solid mesh), and $A_3 = (70, 30)$ (dotted mesh). The proportion of direct choices (solid mesh), choiced after one elimination (dashed mesh) and choices after two eliminations (dotted mesh) is given in figure 5.3. Parameters $\omega_0 = 0.1$ and $\theta^+ = 10 = -\theta^-$ are fixed, focus weight λ varies from 0 to 1 in steps of 0.1 (x-axis), relative attribute weight ω_1 varies from 0 to 1 in steps of 0.1 (y-axis), and attribute weight ω_2 varies from 1 to 0 in steps of -0.1. For each data point, I ran 10,000 simulations. The upper left panel of figure 4.2 is the slice of this figure where the weight for dimension 1 is equal to 0.5.

Similarity effect: Choice probabilities

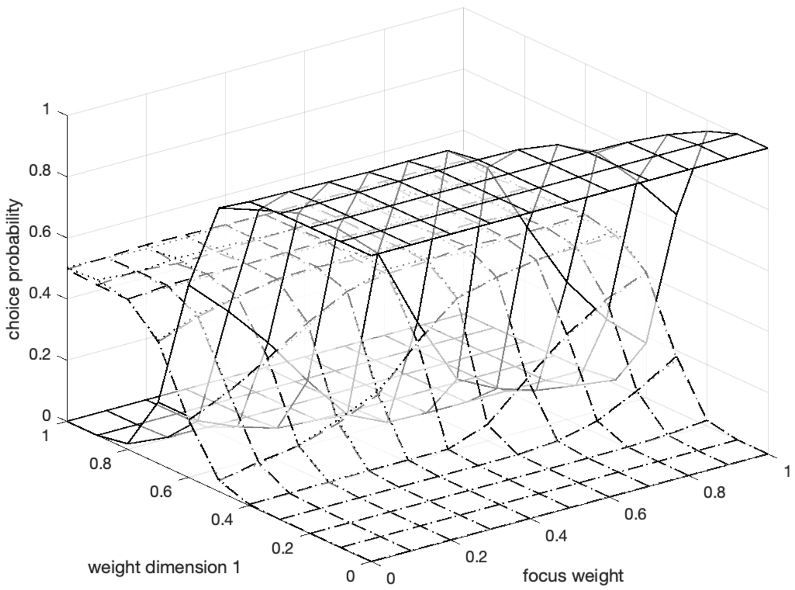

Figure 5.1: Choice probabilities for the similarity effect with alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (solid mesh), and $A_3 = (70, 30)$ (dotted mesh).

Similarity effect: Response times

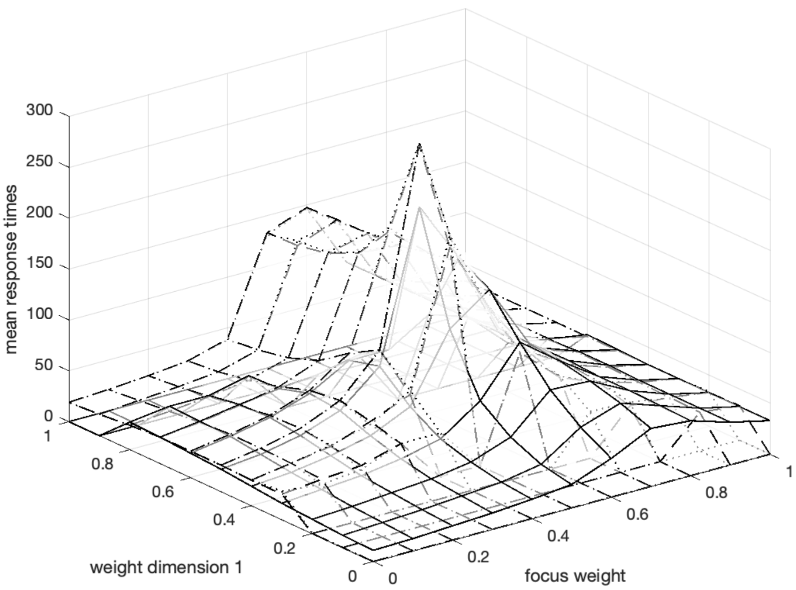

Figure 5.2: Response times for the similarity effect with alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (solid mesh), and $A_3 = (70, 30)$ (dotted mesh).

Similarity effect: Direct and indirect choices

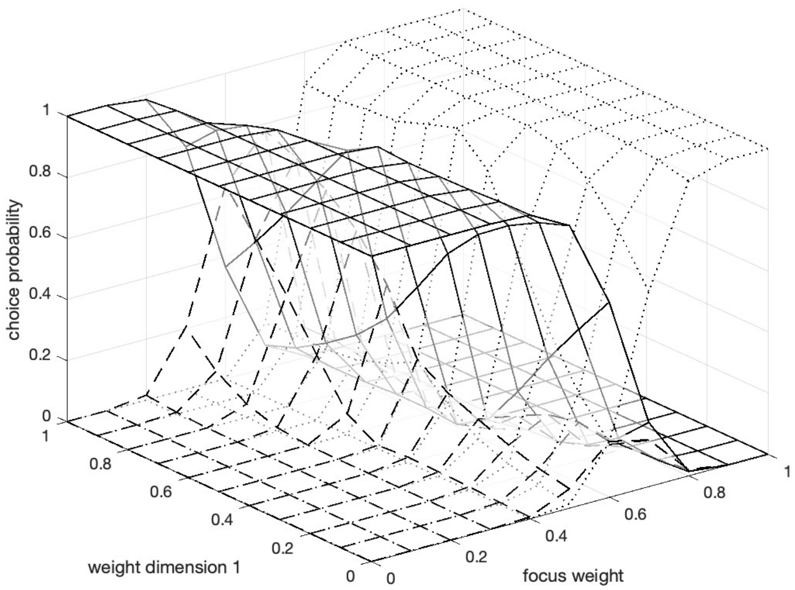

Figure 5.3: Direct choices (solid mesh), choices after one elimination (dashed mesh), and choices after two eliminations (dotted mesh) for the similarity effect with alternatives $A_1 = (70, 30)$, $A_2 = (30, 70)$, and $A_3 = (70, 30)$.

5.2 Range-frequency attraction effect

Figures 5.4 and 5.5 show choice probabilities and response times (z-axis) for alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (solid mesh), and $A_3 = (65, 25)$ (dotted mesh). The proportion of direct choices (solid mesh), choiced after one elimination (dashed mesh) and choices after two eliminations (dotted mesh) is given in figure 5.6. Parameters $\omega_0 = 0.1$ and $\theta^+ = 10 = -\theta^-$ are fixed, focus weight λ varies from 0 to 1 in steps of 0.1 (x-axis), relative attribute weight ω_1 varies from 0 to 1 in steps of 0.1 (y-axis), and relative attribute weight ω_2 varies from 1 to 0 in steps of -0.1. For each data point, I ran 10,000 simulations. The upper right panel of figure 4.2 is the slice of this figure where the weight for dimension 1 is equal to 0.5.

Range-frequency attraction effect: Choice probabilities

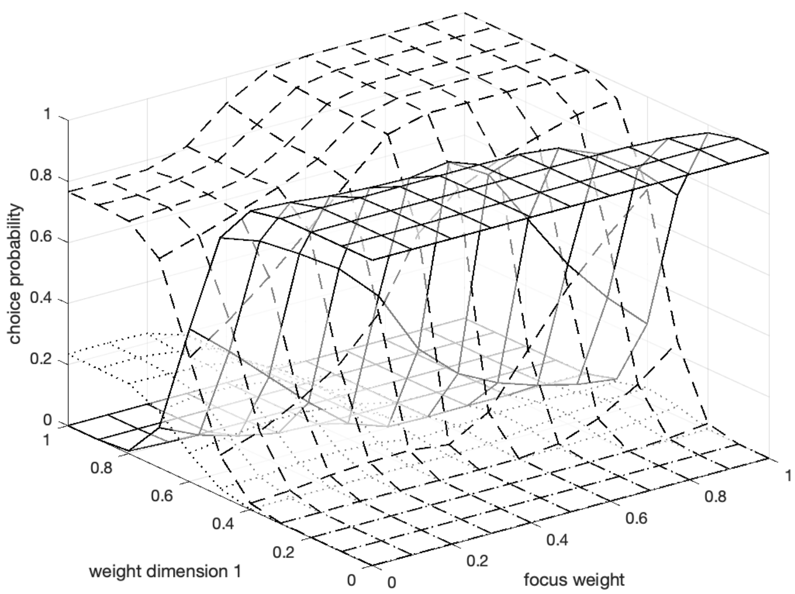

Figure 5.4: Choice probabilities for the range-frequency attraction effect with alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (solid mesh), and $A_3 = (65, 25)$ (dotted mesh).

Range-frequency attraction effect: Response times

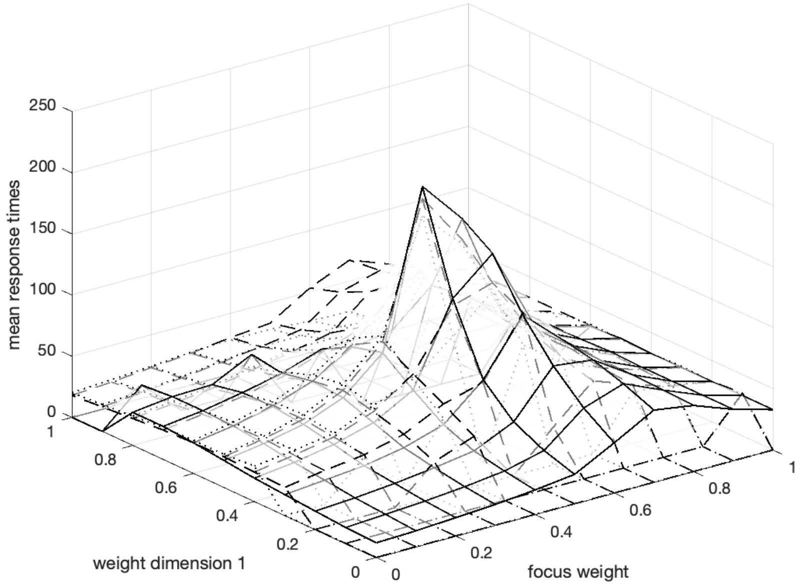

Figure 5.5: Response times for the range-frequency attraction effect with alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (solid mesh), and $A_3 = (65, 25)$ (dotted mesh).

Range-frequency attraction effect: Direct and indirect choices

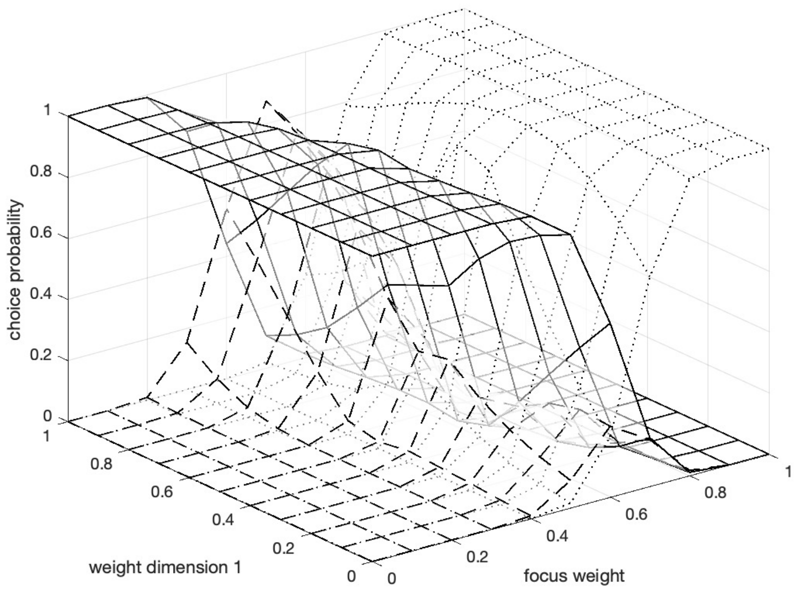

Figure 5.6: Direct choices (solid mesh), choices after one elimination (dashed mesh), and choices after two eliminations (dotted mesh) for the range-frequency attraction effect with alternatives $A_1 = (70, 30)$, $A_2 = (30, 70)$, and $A_3 = (65, 25)$.

5.3 Range attraction effect

Figures 5.7 and 5.8 show choice probabilities and response times (z-axis) for alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (solid mesh), and $A_3 = (70, 25)$ (dotted mesh). The proportion of direct choices (solid mesh), choiced after one elimination (dashed mesh) and choices after two eliminations (dotted mesh) is given in figure 5.9. Parameters $\omega_0 = 0.1$ and $\theta^+ = 10 = -\theta^-$ are fixed, focus weight λ varies from 0 to 1 in steps of 0.1 (x-axis), relative attribute weight ω_1 varies from 0 to 1 in steps of 0.1 (y-axis), and relative attribute weight ω_2 varies from 1 to 0 in steps of -0.1. For each data point, I ran 10,000 simulations.

Range attraction effect: Choice probabilities

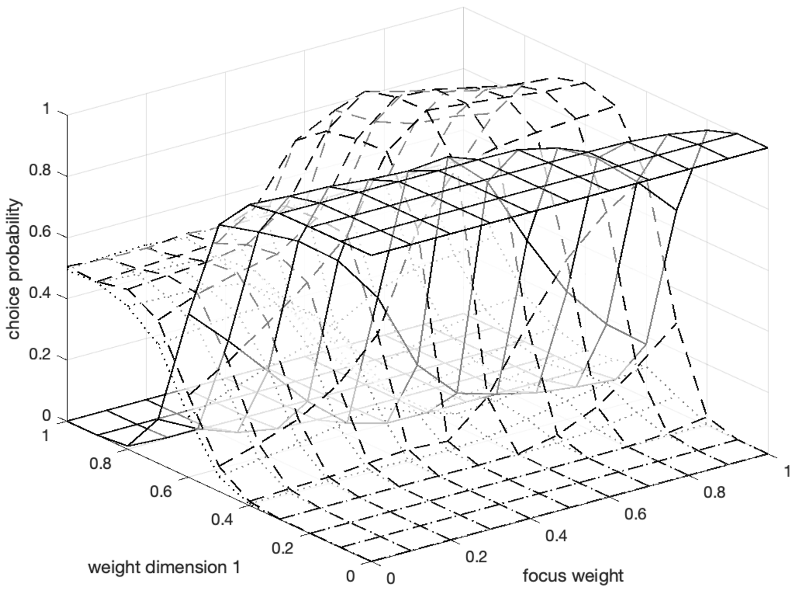

Figure 5.7: Choice probabilities for the range attraction effect with alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (solid mesh), and $A_3 = (70, 25)$ (dotted mesh).

Range attraction effect: Response times

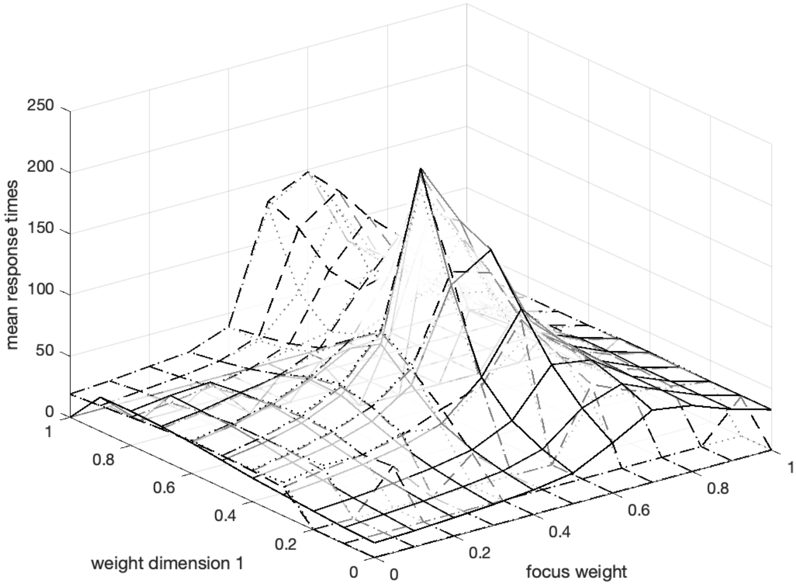

Figure 5.8: Choice probabilities for the range attraction effect with alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (solid mesh), and $A_3 = (70, 25)$ (dotted mesh).

Range attraction effect: Direct and indirect choices

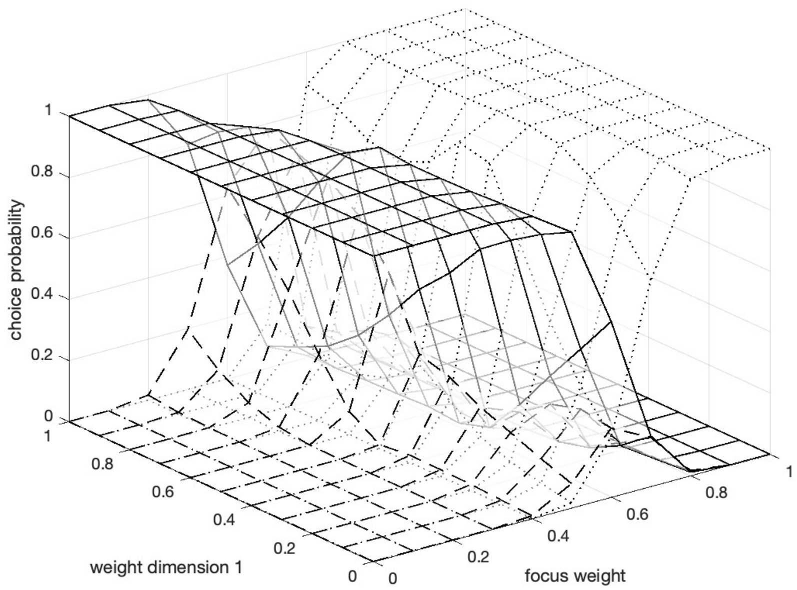

Figure 5.9: Direct choices (solid mesh), choices after one elimination (dashed mesh), and choices after two eliminations (dotted mesh) for the range attraction effect with alternatives $A_1 = (70, 30)$, $A_2 = (30, 70)$, and $A_3 = (70, 25)$.

5.4 Frequency attraction effect

Figures 5.10 and 5.11 show choice probabilities and response times (z-axis) for alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (solid mesh), and $A_3 = 65, 30$ (dotted mesh). The proportion of direct choices (solid mesh), choiced after one elimination (dashed mesh) and choices after two eliminations (dotted mesh) is given in figure 5.12. Parameters $\omega_0 = 0.1$ and $\theta^+ = 10 = -\theta^-$ are fixed, focus weight λ varies from 0 to 1 in steps of 0.1 (x-axis), relative attribute weight ω_1 varies from 0 to 1 in steps of 0.1 (y-axis), and relative attribute weight ω_2 varies from 1 to 0 in steps of -0.1. For each data point, I ran 10,000 simulations.

Frequency attraction effect: Choice probabilities

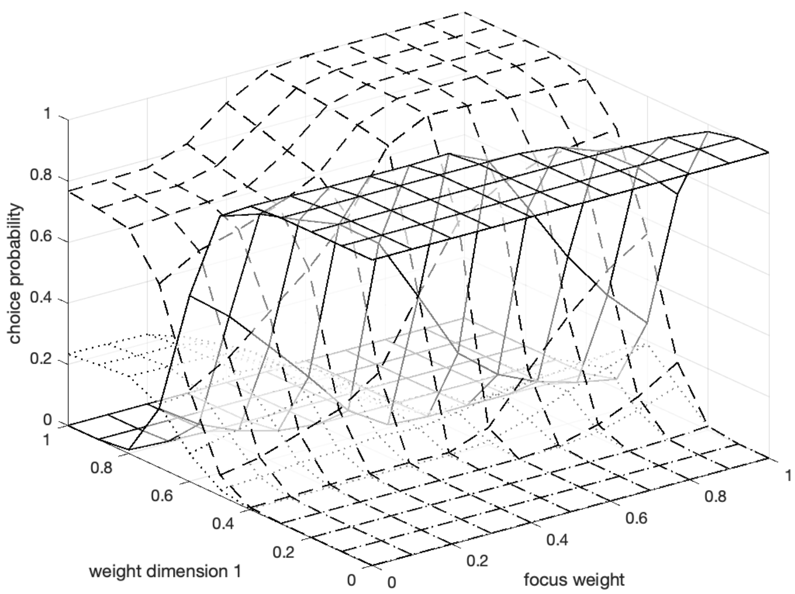

Figure 5.10: Choice probabilities for the frequency attraction effect with alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (solid mesh), and $A_3 = (65, 30)$ (dotted mesh).

Frequency attraction effect: Response times

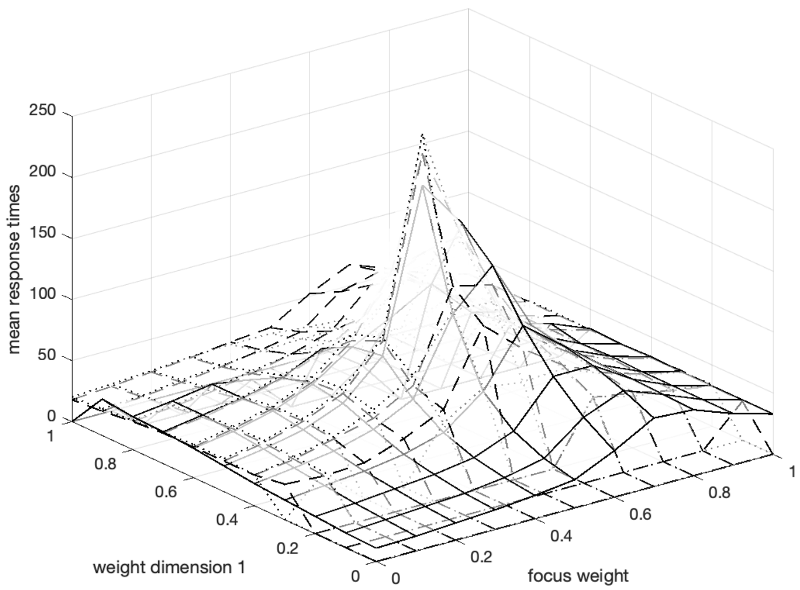

Figure 5.11: Response times for the frequency attraction effect with alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (solid mesh), and $A_3 = (65, 30)$ (dotted mesh).

Frequency attraction effect: Direct and indirect choices

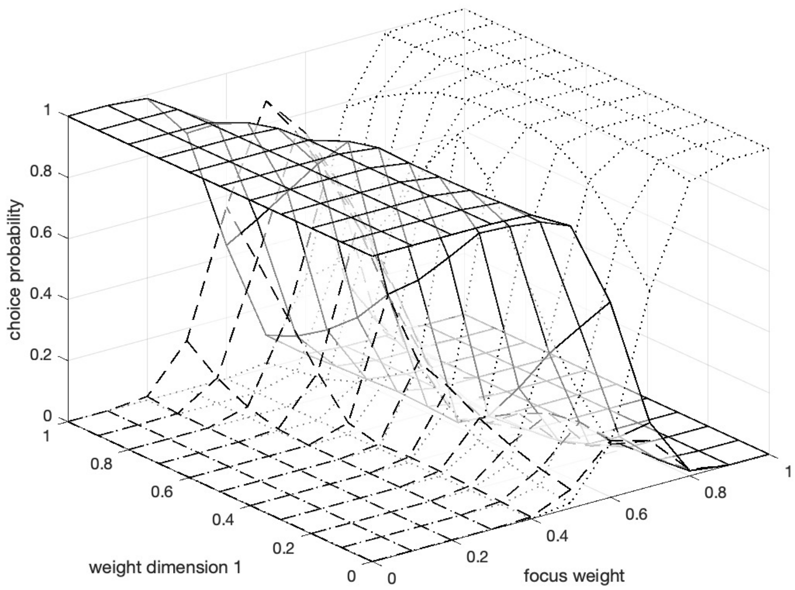

Figure 5.12: Direct choices (solid mesh), choices after one elimination (dashed mesh), and choices after two eliminations (dotted mesh) for the frequency attraction effect with alternatives $A_1 = (70, 30)$, $A_2 = (30, 70)$, and $A_3 = (65, 30)$.

5.5 Asymmetric compromise effect

Figures 5.13 and 5.14 show choice probabilities and response times (z-axis) for alternatives $A_1 = (70, 30)$ (solid mesh), $A_2 = (30, 70)$ (dashed mesh), and $A_3 = (90, 10)$ (dotted mesh). The proportion of direct choices (solid mesh), choiced after one elimination (dashed mesh) and choices after two eliminations (dotted mesh) is given in figure 5.15. Parameters $\omega_0 = 0.1$ and $\theta^+ = 10 = -\theta^-$ are fixed, focus weight λ varies from 0 to 1 in steps of 0.1 (x-axis), relative attribute weight ω_1 varies from 0 to 1 in steps of 0.1 (y-axis), and relative attribute weight ω_2 varies from 1 to 0 in steps of -0.1. For each data point, I ran 10,000 simulations. The lower left panel of figure 4.2 is the slice of this figure where the weight for dimension 1 is equal to 0.5.

Asymmetric compromise effect: Choice probabilities

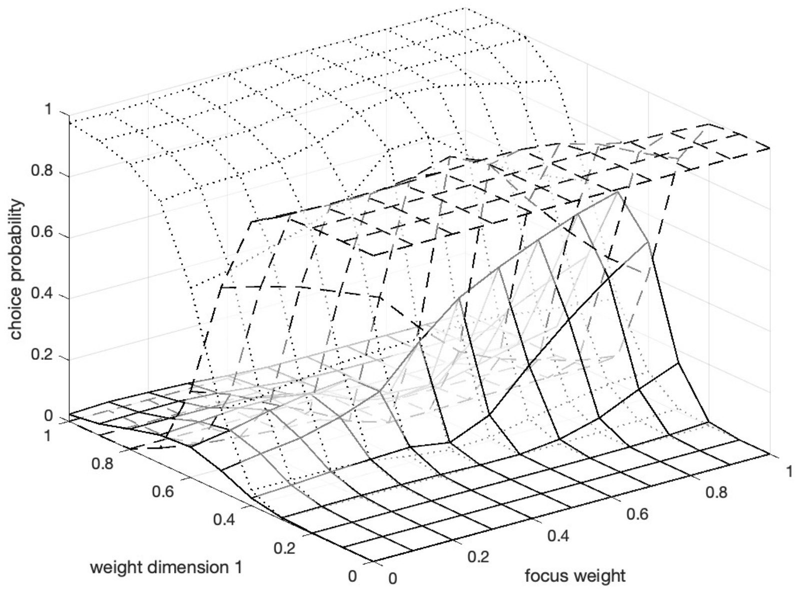

Figure 5.13: Choice probabilities for the asymmetric compromise effect with alternatives $A_1 = (70, 30)$ (compromise option, solid mesh), $A_2 = (30, 70)$ (dashed mesh), and $A_3 = (90, 10)$ (dotted mesh).

Asymmetric compromise effect: Response times

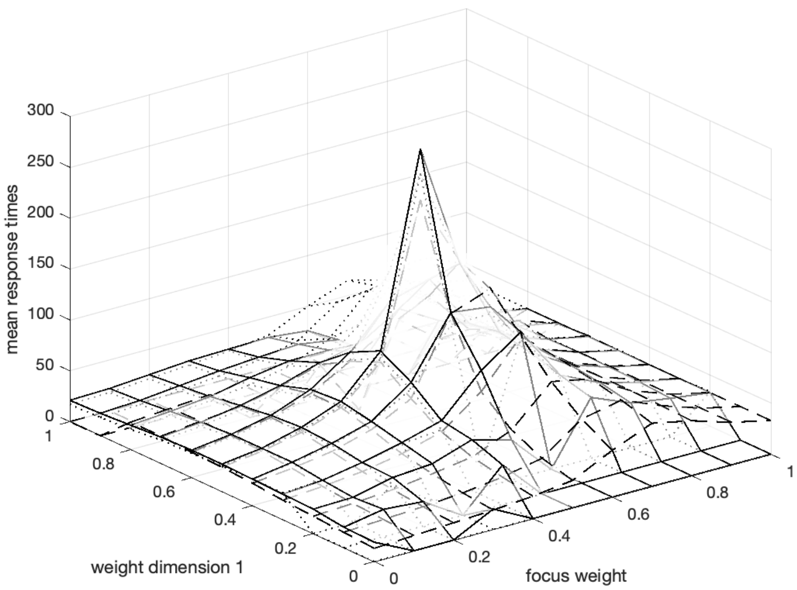

Figure 5.14: Response times for the asymmetric compromise effect with alternatives $A_1 = (70, 30)$ (compromise option, solid mesh), $A_2 = (30, 70)$ (dashed mesh), and $A_3 = (90, 10)$ (dotted mesh).

Asymmetric compromise effect: Direct and indirect choices

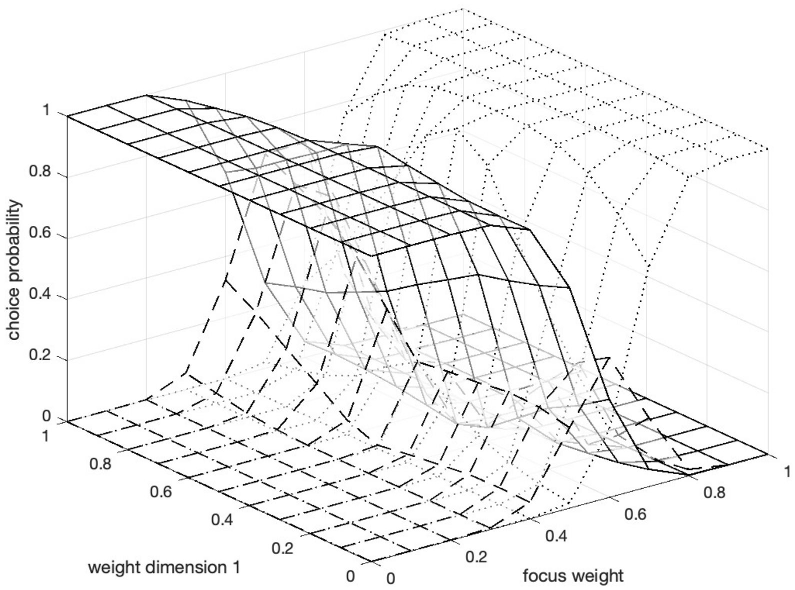

Figure 5.15: Direct choices (solid mesh), choices after one elimination (dashed mesh), and choices after two eliminations (dotted mesh) for the asymmetric compromise effect with alternatives $A_1 = (70, 30)$, $A_2 = (30, 70)$, and $A_3 = (90, 10)$.

5.6 Symmetric compromise effect

Figures 5.16 and 5.17 show choice probabilities and response times (z-axis) for alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (dotted mesh), and $A_3 = (50, 50)$ (solid mesh). The proportion of direct choices (solid mesh), choiced after one elimination (dashed mesh) and choices after two eliminations (dotted mesh) is given in figure 5.18. Parameters $\omega_0 = 0.1$ and $\theta^+ = 10 = -\theta^-$ are fixed, focus weight λ varies from 0 to 1 in steps of 0.1 (x-axis), relative attribute weight ω_1 varies from 0 to 1 in steps of 0.1 (y-axis), and relative attribute weight ω_2 varies from 1 to 0 in steps of -0.1. For each data point, I ran 10,000 simulations. The lower right panel of figure 4.2 is the slice of this figure where the weight for dimension 1 is equal to 0.5.

Symmetric compromise effect: Choice probabilities

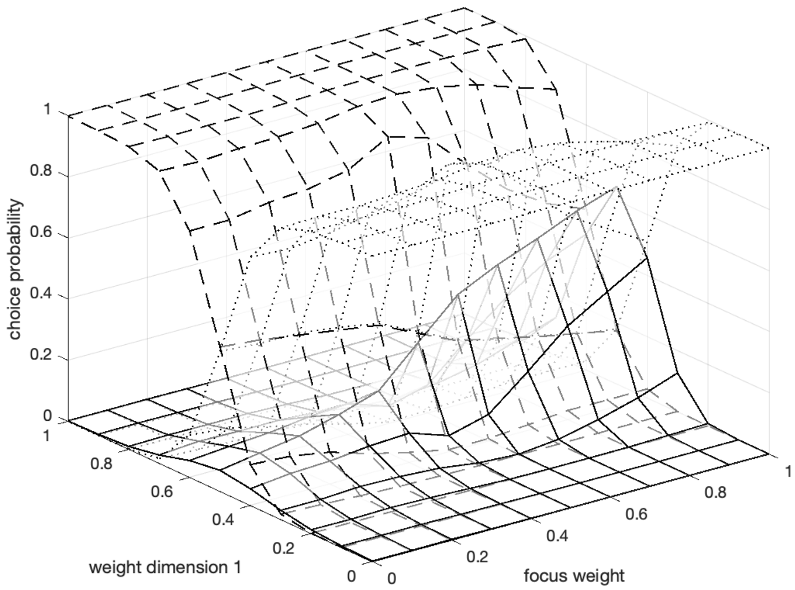

Figure 5.16: Choice probabilities for the symmetric compromise effect with alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (dotted mesh), and $A_3 = (50, 50)$ (compromise option, solid mesh).

Symmetric compromise effect: Response times

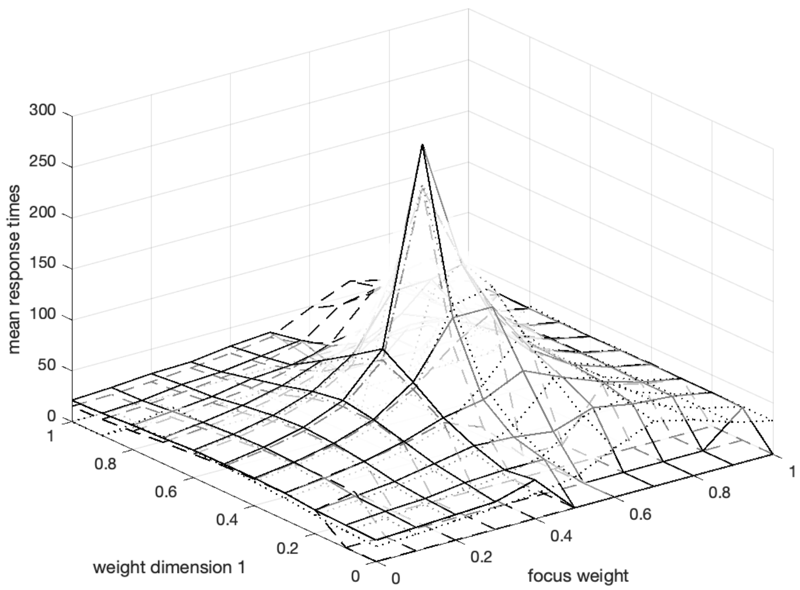

Figure 5.17: Response times for the symmetric compromise effect with alternatives $A_1 = (70, 30)$ (dashed mesh), $A_2 = (30, 70)$ (dotted mesh), and $A_3 = (50, 50)$ (compromise option, solid mesh).

Symmetric compromise effect: Direct and indirect choices

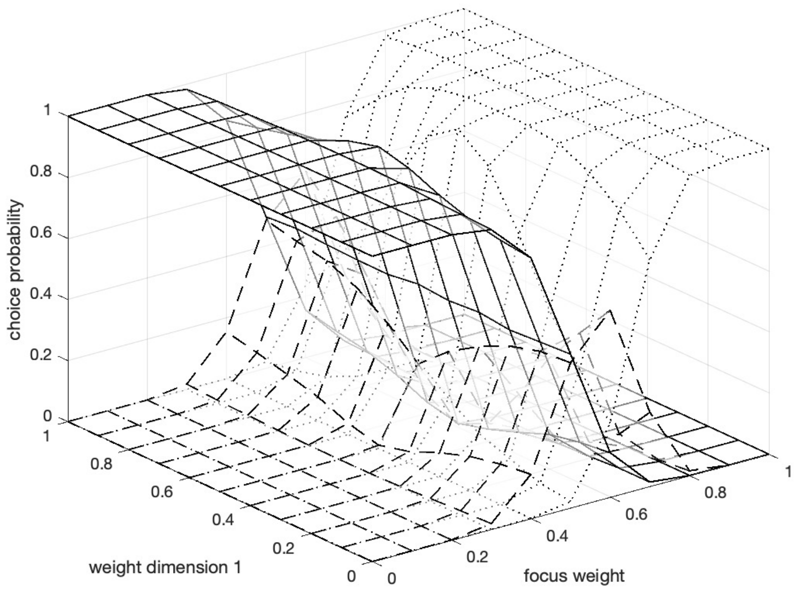

Figure 5.18: Direct choices (solid mesh), choices after one elimination (dashed mesh), and choices after two eliminations (dotted mesh) for the symmetric compromise effect with alternatives $A_1 = (70, 30)$, $A_2 = (30, 70)$, and $A_3 = (50, 50)$.

Chapter 6

A commentary on response times

Lena M. Wollschlaeger, Adele Diederich

This chapter has been submitted to Psychological Review on March 23, 2018 as a comment on Turner, Schley, Muller, and Tsetsos (2018). It was rejected on June 15, 2018.

Abstract

Turner et al. (2018) developed a taxonomy for theories of multi-alternative, multi-attribute preferential choice by comparing four such theories: Multi-alternative decision field theory, the leaky competing accumulator model, the associative accumulation model, and the multi-attribute linear ballistic accumulator model, and analyzed their performance in explaining similarity and attraction effects and a symmetric version of the compromise effect. This approach is highly valuable but should be extended to include dynamic aspects such as optional stopping times, and asymmetric compromise effects. Here, we expand the taxonomy accordingly and add the 2N-ary choice tree model (2NCT, Wollschlaeger & Diederich, 2012, 2017) as an alternative approach. In particular, the 2NCT is the only model that explains both versions of the compromise effect with the same mechanism.

6.1 Introduction

In his seminal paper "A behavioral model of rational choice", Simon (1955) laid the foundation for modern information processing theories of preferential choice, by stating that "the task is to replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist" (p.99). Simon argues that the decision maker, due to limited processing capacities and in order to make a satisfactory rather than optimal choice, employs simple mechanisms instead of effortful calculations. Among other mechanisms, Simon proposes information-gathering steps "for gradually improving the mapping of behavior alternatives upon possible outcomes" (p.108), that is, for overcoming uncertainty about the consequences of the decision.

About 20 years later, March (1978) and Bettman (1979) point towards a second kind of uncertainty – preference uncertainty – that decision makers may experience when faced with a choice problem. They suggest that, in order to overcome preference uncertainty, decision makers have to construct their preferences on the spot. Numerous information processing mechanisms have been proposed ever since to describe preference construction processes (for reviews, see Payne et al., 1992; Lichtenstein & Slovic, 2006). Mostly though, those mechanisms are tailored to a specific, empirically observed anomaly (cf. Tversky & Thaler, 1990). Studied anomalies range from violations of procedure invariance, that is, preference reversals due to changing elicitation procedures like discrete choice or matching (e.g. Tversky et al., 1988), and violations of description invariance, that is, framing effects (e.g. Tversky & Kahneman, 1981), to violations of independence assumptions, that is, effects of the context (e.g. Tversky, 1972b; Huber et al., 1982; Simonson, 1989).

The proposition of decision field theory (DFT, Busemeyer & Townsend, 1993) marks a turning point for theories of preference construction. Its dynamic and stochastic nature allows DFT to account for a range of preference reversals as well as response time phenomena like, e.g., time pressure effects. Furthermore, DFT's multi-alternative extension (MDFT, Roe et al., 2001) is the first theory to explain several context effect, that is, similarity (Tversky, 1972b), attraction (Huber et al., 1982), and compromise (Simonson, 1989) effects simultaneously, by means of a specific combination of information processing mechanisms. Three years later, Usher and McClelland (2004) proposed a modified version of the leaky competing accumulator model (LCA, Usher & McClelland, 2001) which also accounts for the three context effect, but with slightly different information processing mechanisms, including, most prominently, loss aversion (Tversky & Kahneman, 1991). More recent multi-alternative preference construction models include the (simple) 2N-ary choice tree model (Wollschlaeger & Diederich, 2012, 2017), the associative accumulation model (AAM, Bhatia, 2013), the multi-attribute linear ballistic accumulator model (MLBA, Trueblood et al., 2014), and multi-alternative decision by sampling (MDbS, Noguchi & Stewart, 2018). All of these models make different assumptions about attentional processes, evaluation of alternatives, evidence accumulation and stopping rules.

Turner et al. (2018) develop a taxonomy of such models by analyzing four of the above mentioned models, that is, MDFT, LCA, AAM, and MLBA. This taxonomy and a related so-called switchboard analysis allow them to evaluate single information processing mechanisms and their use for explaining the three context effects. In the first part of this article, we revisit the 2N-ary choice tree model (2NCT, Wollschlaeger & Diederich, 2012) and its recent version simple choice tree model (SCT, Wollschlaeger & Diederich, 2017) and class it with the taxonomy proposed by Turner et al. (2018). Then, we discuss the unique mechanisms that allow the 2NCT/SCT model to account for the three context effects and interactions between them and argue that they should be included in the taxonomy and switchboard analysis. Finally, we propose a revised taxonomy based on considerations about the time course of decision making processes, including stopping rules.

6.2 The (simple) 2N-ary choice tree model and Turner et al.'s (2018) taxonomy

The (simple) 2N-ary choice tree model (2NCT/SCT, Wollschlaeger & Diederich, 2012, 2017) assumes that the decision maker compares attribute values within attributes and between alternatives in a pairwise manner. Attributes are selected for examination based on attribute weights that reflect salience. Within attributes, pairs of attribute values are selected for comparison based on so-called comparison values. The pairwise comparisons are assumed to be of ordinal nature, that is, they are either perceived as advantageous for one of the choice alternatives (the "winning" alternative) or as disadvantageous for the other one (the "losing" alternative). Differently from the competing models, here the decision maker is assumed to separately count the instances in which each alternative is perceived as winning or losing a pairwise comparison. This requires two counters per alternative, a feature that distinguishes the current approach from the competing ones. The difference of the two counter states, that is, the preference state for each alternative, is constantly compared to two thresholds: A positive choice criterion and a negative elimination criterion. An alternative is chosen either if its preference state reaches the choice criterion or if the preference states of all the other alternatives have reached the elimination criterion. The individual counter updating probabilities can be calculated by multiplying attribute values, comparison values, and the focus weight specified in the following. An additional random component adds noise and, in the 2NCT model (Wollschlaeger & Diederich, 2012), a "local" component allows to implement lateral inhibition and leakage (cf. Roe et al., 2001; Usher & McClelland, 2004). The local component is omitted in the simpler version of the model (i.e. SCT) in order to make it mathematically tractable. Nevertheless the SCT model accounts for the three context effects as well.

The taxonomy proposed by Turner et al. (2018) divides the decision process into three processing stages with six sub-stages: (1) Subjective perception of the attribute space, with sub-stages (1.a) subjective mapping of attribute values, (1.b) absolute or relative representation of attributes, and (1.c) nonlinear filtering of representations, (2) allocation of attention between attributes, and (3) preference accumulation, with sub-stages (3.a) attribute integration, (3.b) competition through lateral inhibition, and (3.c) noise. For each of these (sub-)stages, we will now discuss the related mechanisms in the 2NCT/SCT model. Furthermore, we add choice and elimination criteria and stopping rules because they are not part of their taxonomy. We use the same switchboard analysis as proposed by Turner et al. (2018, Study 3) to describe the model's mechanisms. To enhance comparison with the alternative approaches, this analysis includes three choice alternatives with two attributes. Note however, that the 2NCT/SCT model is designed to account for situations with any number of alternatives and attributes (cf. Wollschlaeger & Diederich, 2012, 2017).

Let ***x***, ***y***, and ***z*** be the three choice options and *P* and *Q* the two attributes describing them. The specific values of each alternative with respect to a given attribute, denoted $x_P, y_P, \dots z_Q$ are presented in a matrix, M:

$$M = \begin{bmatrix} x_P & x_Q \\ y_P & y_Q \\ z_P & z_Q \end{bmatrix}.$$

Like in the other multi-alternative multi-attribute preferential choice models, they serve as input to the preference construction process. However, instead of using the values as such, in the 2NCT/SCT model they are transformed into probabilities for updating the counters by combining normalized differences of attribute values and the current counter states with appropriately restricted weights. With that, the components are the same as for MDFT, LCA, AAM, and MLBA: A "global" component, based on the attribute values and attribute weights, a "local" component, based on the current counter states, and a random component. Let $p_{\boldsymbol{a}}^{+}[t]$ and $p_{\boldsymbol{a}}^{-}[t]$ be the probabilities for updating the positive and negative counter for alternative $\boldsymbol{a}, \boldsymbol{a} \in \{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$ at time t, respectively. Then the preference states $P_{\boldsymbol{a}}[t], \boldsymbol{a} \in \{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$ for the three alternatives at time t are described by the following iterative equations:

$$P_{\boldsymbol{a}}[t] = \begin{cases} P_{\boldsymbol{a}}[t-1]+1 & \text{with probability } p_{\boldsymbol{a}}^{+}[t] \\ P_{\boldsymbol{a}}[t-1]-1 & \text{with probability } p_{\boldsymbol{a}}^{-}[t], \text{ and} \\ P_{\boldsymbol{a}}[t-1] & \text{with probability } 1-(p_{\boldsymbol{a}}^{+}[t]+p_{\boldsymbol{a}}^{-}[t]). \end{cases}$$

Note that a preference state may stay the same as before, that is, there may be no new evidence for or against choosing the alternative.

(1) Subjective perception of the attribute space

The first processing stage in the Turner et al. (2018) taxonomy addresses how the decision maker perceives the given choice set, which reference points the decision maker uses to evaluate the current alternatives, and whether or not the evaluations are (asymmetrically) transformed before accumulation.

(1.a) Subjective mapping of attribute values

Given the objective attribute values that describe the choice alternatives, each model has to define if and how these values are transformed into subjective representations. The latter are stored in a matrix N. While MDFT and LCA assume that the subjective representations are equal to the objective values, i.e. $N = M$, MLBA transforms the objective values by adding parameterized curvature to the attribute space, and AAM applies a parameterized power function to them. Both transformations make a distinction between intermediate and extreme values, allowing MLBA and AAM to account for (symmetric) compromise effects and extremity biases. The 2NCT/SCT model, like MDFT and LCA, makes the assumption that the subjective representations are equal to the objective values, $N = M$.

(1.b) Absolute or relative representation of attribute values

The pairwise comparison of attribute values or – more generally – the evaluation of attribute values against one or several reference points is interpreted as part of the perception process in the Turner et al. (2018) taxonomy. It results in absolute or relative representations of the attribute values. Absolute representations result from evaluations against a single neutral reference point, as for example in AAM. Relative representations result from evaluations of attribute values against each other, that is, from pairwise comparisons of attribute values. This is assumed in MDFT, LCA, MLBA and the 2NCT/SCT model but not for AAM. Each alternative serves as a reference point for the other alternatives in the choice set and differences between attribute values are used as relative representations of those values. Let $D'_A, A \in \{P, Q\}$ be the matrix in which these relative representations are stored:

$$D'_A = \begin{bmatrix} (x_A - x_A) & (x_A - y_A) & (x_A - z_A) \\ (y_A - x_A) & (y_A - y_A) & (y_A - z_A) \\ (z_A - x_A) & (z_A - y_A) & (z_A - z_A) \end{bmatrix}$$

Unlike in the other theories, those differences are normalized in the 2NCT/SCT model such that the sum of their absolute values is equal to one for each attribute. We store the normalized differences in a matrix $D_A, A \in \{P, Q\}$:

$$D_A = \frac{1}{2} \begin{bmatrix} \frac{(x_A - x_A)}{S_A} & \frac{(x_A - y_A)}{S_A} & \frac{(x_A - z_A)}{S_A} \\ \frac{(y_A - x_A)}{S_A} & \frac{(y_A - y_A)}{S_A} & \frac{(y_A - z_A)}{S_A} \\ \frac{(z_A - x_A)}{S_A} & \frac{(z_A - y_A)}{S_A} & \frac{(z_A - z_A)}{S_A} \end{bmatrix},$$

with $S_A = \sum_{\{a,b\} \subset \{x,y,z\}} |a_A - b_A|, A \in \{P, Q\}$. Note that each difference appears twice in D_A, once with a positive sign and once with a negative sign, but only once in S_A, $A \in \{P, Q\}$ (hence the multiplication with $\frac{1}{2}$). Negative and positive differences affect the input to two different counters and are possibly treated differently in the next sub-stage. Therefore, we split $D_A, A \in \{P, Q\}$ into two parts, $D_A^+ = h^+(D_A)$ and $D_A^- = h^-(D_A)$, by means of two functions

$$h^+(x) = x \cdot \mathbf{I}(x > 0), \text{ and } h^-(x) = |x| \cdot \mathbf{I}(x < 0), \text{ with}$$

$$\mathbf{I}(y) = \begin{cases} 1 & \text{if } y \text{ is true, and} \\ 0 & \text{otherwise.} \end{cases}$$

Note that all entries in D_A^+ and D_A^- are nonnegative and, for each attribute, sum up to one. Though normalization of the differences was primarily implemented in order to yield probabilities (for updating the counters), it has some advantages over simply using the absolute differences. First of all, it balances out inequalities between attributes with, on average, bigger or smaller differences. Higher salience for an attribute with, for example, higher absolute differences is thus not hardwired into the model but would instead be reflected in a higher attribute weight for this attribute. Furthermore, normalization makes model parameters comparable for varying choice situations, and even allows the model to make predictions for new situations.

(1.c) Nonlinear filtering of representations

Given the absolutely or relatively represented attribute values, that is, the results of the pairwise comparisons, each model has to determine whether they are accumulated as such or (asymmetrically and/or nonlinearly) transformed before accumulation – "filtered" in the Turner et al. (2018) taxonomy. No filtration is assumed by MDFT and AAM. The LCA model, on the other hand, applies a loss-averse asymmetric filtration function (cf. Tversky & Kahneman, 1991; Usher & McClelland, 2004), which is similar to prospect theory's reference-dependent, asymmetric and S-shaped value function (Kahneman & Tversky, 1979). MLBA uses a similarity-based filter mechanism that weights small values (i.e. differences between similar alternatives) higher than large values (i.e. differences between distant alternatives) by means of an exponentially decaying function. Two additional parameters allow the MLBA model to treat positive and negative values asymmetrically.

The 2NCT/SCT model assumes that decision makers perceive each pairwise comparison as advantageous for one of the choice alternatives (the "winning" alternative) or as disadvantageous for the other one (the "losing" alternative). In the 2NCT model, both perspectives are taken equally often. This is reflected in the factor $\frac{1}{2}$ in D_A, D_A^+ and D_A^-, $A \in \{P, Q\}$. Alternatively, it is possible to implement LCA's loss-averse asymmetric value function by applying it to the pairwise differences before normalizing them (cf. Wollschlaeger & Diederich, 2012). This would lead to overweighting of "losing" alternatives. In the SCT model, a focus weight $\lambda, 0 \leq \lambda \leq 1$ determines whether the decision maker focuses more on the "winning" or on the "losing" alternative. The possibly shifted focus is implemented by replacing the factor $\frac{1}{2}$ in D_A^- by λ and in D_A^+ by $(1 - \lambda)$. It follows that, if $\lambda > 0.5$, the decision maker focuses more on the "losing" alternative, similar to what is achieved with the loss-averse value function. On the other hand, if $\lambda < 0.5$, the decision maker focuses more on the "winning" alternative. For $\lambda = 0.5$, the decision maker is equally likely to focus on the "winning" and "losing" alternative, just like in the basic version of the 2NCT model. We store the filtered normalized differences in two matrices $V_A^+ = (1-\lambda) \cdot 2 \cdot D_A^+$ and $V_A^- = \lambda \cdot 2 \cdot D_A^-$, $A \in \{P, Q\}$. Weighting with λ allows the SCT model to flexibly balance focus on positive versus negative aspects of the alternatives relative to each other. It is therefore especially useful in situations without an explicit loss/gain-framing, e.g., in perceptual or preferential choice (cf. Trueblood et al., 2013).

(2) Allocation of attention between attributes

The second processing stage in the Turner et al. (2018) taxonomy addresses how the decision maker distributes attention between attributes. For that, each attribute is assigned a weight that reflects its salience. The weights are usually restricted such that their sum is less or equal one. In MDFT, LCA, 2NCT/SCT, and AAM they determine the proportion of time that the decision maker spends on evaluating alternatives based on the respective attribute. Technically speaking, the attribute weights are interpreted as attention probabilities. This goes hand in hand with the dynamic nature of these models. AAM additionally assumes that the attribute weights are proportional to the sum of attribute values – summed up separately for each attribute – in the choice problem at hand. In MLBA, which is basically a static model, the attribute weights are not related to the allocation of attention. Instead, they are simply multiplied with the relative representations resulting from the previous processing stage. The 2NCT/SCT model makes further assumptions about times where no specific attribute is attended: Let ω_P, $0 \leq \omega_P \leq 1$, and ω_Q, $0 \leq \omega_Q \leq 1$ be the attribute weights for the two attributes P and Q. The 2NCT/SCT model assumes that $\sum_{A \in \{P,Q\}} \omega_A \leq 1$, and interprets $\omega_0 := 1 - \sum_{A \in \{P,Q\}} \omega_A$ as weight for the random component, that is, as probability for not attending any specific attribute but updating counters at random.

(3) Preference accumulation

The third processing stage of the Turner et al. (2018) taxonomy addresses "how preferences dynamically evolve over time" (p. 20). For each model, it specifies how many counters accumulate evidence or preference, how the counters are updated over time, how they compete with each other, and how noise distorts the accumulation process. Since MLBA is a static model, it is only mentioned in the description of the third processing stage where applicable. MDFT, LCA, AAM, and MLBA all assume that there is one counter per alternative and that all of these counters are updated in parallel. In the 2NCT/SCT model on the other hand, counters are updated serially and evidence for and against choosing an alternative is counted in two separate counters per alternative. The two-counters-assumption is related to the distinction between "winning" and "losing" alternatives described in step (1.c) above. It is seized again in step (4) below, where we introduce thresholds for the preference states as both choice and elimination criteria. Note that the Turner et al. (2018) taxonomy does not include thresholds as choice or elimination criteria. In the switchboard analysis, the authors run simulations with a fixed amount of steps and assume that the accumulator with the highest state wins the race and the associated alternative is chosen by the decision maker. This corresponds to a fixed stopping rule rather than an optional stopping rule (Busemeyer & Diederich, 2002). We will argue that response time is an important aspect in modeling decision making and may even differentiate clearly between the competing models. But first, we describe the remaining mechanisms of the 2NCT/SCT model by means of the taxonomy at hand.

(3.a) Attribute integration

Like MDFT, LCA, and AAM, the 2NCT/SCT model assumes a stochastic integration of attributes in the overall evaluation. The decision maker pays attention to attribute P with probability ω_P, to attribute Q with probability ω_Q, and to none of them with probability ω_0, with $\omega_P + \omega_Q + \omega_0 = 1$. Note that we do not include MLBA here since it weights attributes deterministically during a preprocessing stage. In MDFT, LCA, and AAM, the counters for all alternatives are updated in parallel, and their states are increased by the perceived attribute values (described in processing stage (1) above) with respect to the currently attended attribute. In the 2NCT/SCT model, however, the counters are updated serially and thus the model makes further assumptions about allocation of attention and integration of information: Within attributes, attention is allocated between pairs of attribute values proportionally to their absolute differences, that is, attention is fully on a pair $\{a_A, b_A\}$, $\{\boldsymbol{a},\boldsymbol{b}\} \subset \{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\}$, $A \in \{P, Q\}$, with probability $\omega_A \cdot \frac{|a_A - b_A|}{S_A}$. Within pairs of attribute values, however, the decision maker focuses on the "winning" alternative with probability $1-\lambda$, or on the "losing" alternative with probability λ. In the first case, the counter state of the "positive" counter for the winning alternative is increased by one, in the second case, the counter state of the "negative" counter for the losing alternative is increased by one. For example, let $a_A > b_A$, then the probability for focusing on a_A and updating the positive counter for alternative $\boldsymbol{a}$ is $(1-\lambda) \cdot \omega_A \cdot \frac{|a_A - b_A|}{S_A}$ and the probability for focusing on b_A and updating the negative counter for alternative $\boldsymbol{b}$ is $\lambda \cdot \omega_A \cdot \frac{|a_A - b_A|}{S_A}$.

Note that updating counters by increasing their states by one (i.e., simply counting won and lost comparisons) is related to another mechanism that Simon (1955) proposes as leading to "substantial computational simplifications in the making of a choice" (p.104): Simple pay-off functions with two (1, 0) or three (1, 0, -1) output values. Vlaev, Chater, Stewart, and Brown (2011) classify models of decision making that implement such a simple pay-off function as "comparison-based theories without value calculation". Multi-alternative decision by sampling theory (Noguchi & Stewart, 2018) is another example from this category. In the Vlaev et al. (2011) terminology, MDFT, LCA, AAM, and MLBA belong to the category of "comparison-based theories with value calculation" as opposed to the category of "value-based theories" like multi-attribute utility theory (Keeney & Raiffa, 1967/1993).

Overall, the global component of the weighted counter updating probabilities can be calculated by multiplying the attribute weights with the comparison values and the focus weight. Together with the random component, this yields two probabilities $p^+_{\boldsymbol{a}}$ and $p^-_{\boldsymbol{a}}$ for each alternative $\boldsymbol{a}, \boldsymbol{a} \in \{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$:

$$p^+_{\boldsymbol{a}} = \omega_P \cdot V^+_{P,\boldsymbol{a}} + \omega_Q \cdot V^+_{Q,\boldsymbol{a}} + \omega_0 \cdot V_0, \text{ and} \tag{6.1}$$

$$p^-_{\boldsymbol{a}} = \omega_P \cdot V^-_{P,\boldsymbol{a}} + \omega_Q \cdot V^-_{Q,\boldsymbol{a}} + \omega_0 \cdot V_0, \tag{6.2}$$

with $V_0 = \begin{bmatrix} \frac{1}{2\cdot 3} & \frac{1}{2\cdot 3} & \frac{1}{2\cdot 3} \end{bmatrix}^T$. Note that the sum of all six counter updating probabilities is equal to 1, such that on average one counter per time step is updated.

(3.b) Competition through lateral inhibition

Competition between choice alternatives is crucial for explaining violations of independence principles (e.g. context effects) and all descriptive models of multi-alternative multi-attribute decision making processes implement it in one or the other way. MDFT and LCA both implement competition on the counter state level, that is, through lateral inhibition. In MDFT, lateral inhibition is distance-dependent, that is, if two alternatives are "close" to each other in the attribute space, positive activation in one of the counters leads to decreasing activation in the other and negative activation in one of the counters leads to increasing activation in the other. These mechanisms explain the attraction effect and - for a suitable definition of "closeness" (cf. Tsetsos et al., 2010; Hotaling, Busemeyer, & Li, 2010) - the compromise effect. MLBA also uses distance-dependence to explain the attraction effect. However, the effect is hard-wired into the model since MLBA is a static model and it is not possible to implement dynamic mechanisms like competition. In the LCA model, lateral inhibition is defined globally, promoting single alternatives (or groups of very similar alternatives) once they gain an advantage over the other alternatives. Leakage or decay of accumulated information over time offsets the effects of lateral inhibition to some degree, allowing the models to produce primacy or recency effects. Turner et al. (2018) implement global inhibition and leakage also into AAM, where choice alternatives originally compete by means of context-dependent associations between alternatives and attributes.

In the 2NCT/SCT model, competition stems from serially updating the counters with updating probabilities based on normalized differences. However, in order to compare the 2NCT model to MDFT and LCA, lateral inhibition and leakage are implemented there as well (Wollschlaeger & Diederich, 2012). For that, a "local" component, that depends on the current counter states, is added to the counter updating probabilities. Let $S^+_{\boldsymbol{a}}$ and $S^-_{\boldsymbol{a}}$ be the two counters for alternative $\boldsymbol{a}, \boldsymbol{a} \in \{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$, $S^+_{\boldsymbol{a}}[t]$ and $S^-_{\boldsymbol{a}}[t]$ their counter states at time t, and $S[t]$ the sum of the counter states at time t. Then lateral inhibition is implemented by increasing the probability for updating the negative counter of alternative $\boldsymbol{a}$, $p^-_{\boldsymbol{a}}$, with an amount proportional to the sum of the counter states of the positive counters of the other alternatives, $\frac{S^+_{\boldsymbol{b}}[t-1]+S^+_{\boldsymbol{c}}[t-1]}{2\cdot S[t-1]}$, $\boldsymbol{b}, \boldsymbol{c} \in \{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$, and vice versa for the positive counter of alternative $\boldsymbol{a}$. Leakage is implemented by increasing the probability for updating the negative counter of alternative $\boldsymbol{a}$, with an amount proportionally to its own positive counter state, $\frac{S^+_{\boldsymbol{a}}[t-1]}{S[t-1]}$, and vice versa for the positive counter. Weights $L, 0 \leq L \leq 1$ for lateral inhibition and $k, 0 \leq k \leq 1$ for leakage, with $L + k \leq 1$, determine the proportion of the counter updating probabilities that is defined by the local component. Altogether, this yields

$$(p^+_{\boldsymbol{a}})' = (1-k-L)\cdot p^+_{\boldsymbol{a}} + k\cdot \frac{S^-_{\boldsymbol{a}}[t-1]}{S[t-1]} + L\cdot \frac{S^-_{\boldsymbol{b}}[t-1]+S^-_{\boldsymbol{c}}[t-1]}{2\cdot S[t-1]}, \text{ and}$$

$$(p^-_{\boldsymbol{a}})' = (1-k-L)\cdot p^-_{\boldsymbol{a}} + k\cdot \frac{S^+_{\boldsymbol{a}}[t-1]}{S[t-1]} + L\cdot \frac{S^+_{\boldsymbol{b}}[t-1]+S^+_{\boldsymbol{c}}[t-1]}{2\cdot S[t-1]}.$$

Other implementations of lateral inhibition and/or leakage, e.g., based on preference states instead of counter states, are conceivable. But since competition based on serial updating of counters allows the 2NCT model to explain the three context effects and other reference points effects without lateral inhibition and leakage, we did not explore them further. As a matter of fact, the simple choice tree model (Wollschlaeger & Diederich, 2017) completely omits lateral inhibition and leakage in order to achieve mathematical tractability.

(3.c) Noise

Probabilistic-static models of decision making (Busemeyer & Townsend, 1993), e.g., random utility models, use noise to account for inconsistent behavior between trials. Moving from static to dynamic models (Busemeyer & Townsend, 1993), e.g., to (discrete) sequential sampling models, each sample resembles a static trial. In this case, noise can be either kept between trials or between samples, that is, within trials of the sequential sampling model. MDFT, for example, implements noise within trials and interprets it as influence of irrelevant attributes on the preference construction process. Similar assumptions are made for LCA and AAM. MLBA, as a static model, implements noise between trials. In the 2NCT/SCT model, noise cannot be added to the counter states as such, since they are restricted to nonnegative integer values. However, the random component as described above in sub-stage (3.a) allows for random counter updates from time to time during the preference construction process. It has a similar effect on the counter states as the within-trial noise implemented in MDFT, LCA, and AAM.

(4) Choice tree and stopping rules

Here we expand the taxonomy proposed by Turner et al. (2018) because it does not include decision criteria for accepting or rejecting a choice option. In sequential sampling models, decision criteria are related to the time it takes to make a decision and reflect the underlying deliberation process, i.e., the preference construction over time (Lichtenstein & Slovic, 2006). Note that the time to make a decision is a random variable. The switchboard analysis, however, uses a simple fixed stopping rule, that is, the time is fixed. We argue, that including the time to come to a decision is crucial because it may provide essential information about the preference construction process. Most of the models considered here are dynamic

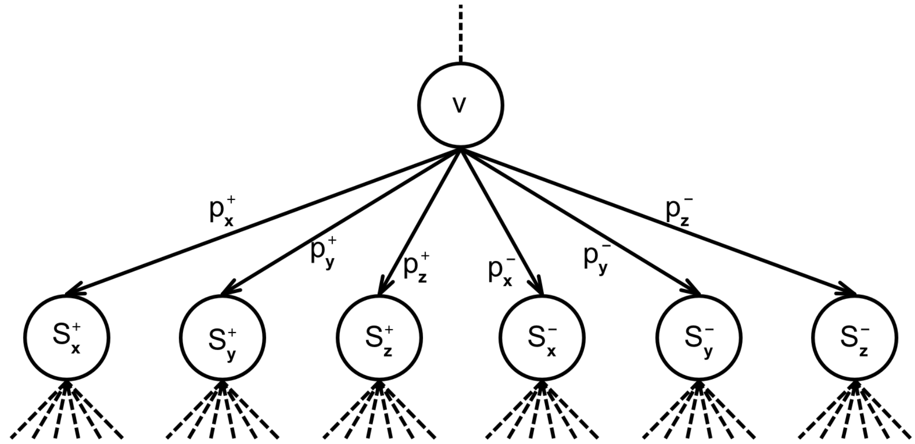

Figure 6.1: One section of the tree T for three choice alternatives. The vertex v has six outgoing edges leading to its children associated with the six counters. Each edge is associated with a transition probability $p_{\boldsymbol{a}}^{+}$ or $p_{\boldsymbol{a}}^{-}$, $\boldsymbol{a} \in \{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$ as defined in equation 6.2.

and stochastic in nature. If however, the dynamic part is basically neglected, this information is not available. For the 2NCT/SCT model it naturally follows from its structure. We illustrate it by referring to the eponymous choice tree $T = (V, E, r)$ with vertices V, edges $E \subseteq V \times V$ and root $r \in V$. All vertices of T are directed away from r and each internal vertex $v \in V$ has six (that is, two times the number of choice alternatives) children that are associated with the six counters. The edges leading from v to its children are associated with the probabilities for updating the respective counters (defined in equation 6.2 above). Figure 6.1 shows one section of the tree for three choice alternatives. The vertex v has six outgoing edges leading to its children associated with the six counters.

The preference construction process is represented by a random walk on T, beginning at the root and passing from there through an edge to another vertex, triggering the update (increase by $+1$) of the associated counter, moving on through another edge and so forth. The next edge to pass through is chosen according to the probabilities associated with the edges. Note that for each vertex the transition probabilities of all outgoing edges sum up to one. An example path of this random walk is pictured in bold in Figure 6.2.

The preference construction process stops when enough evidence has been accumulated to make the required choice. To this end, the preference states $P_{\boldsymbol{a}}[t] = S_{\boldsymbol{a}}^{+}[t] - S_{\boldsymbol{a}}^{-}[t], \boldsymbol{a} \in \{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$ are constantly compared to two thresholds, a positive choice threshold θ^{+} and a negative elimination threshold $\theta^{-} = -\theta^{+}$. If the preference state for alternative $\boldsymbol{a}$ hits the positive threshold, the process stops and $\boldsymbol{a}$ is chosen. If, on the other hand, the preference state for alternative $\boldsymbol{b}$ hits the negative threshold, $\boldsymbol{b}$ is eliminated from the choice set and the process continues with the remaining alternatives until one of them is chosen or until all but one of them have been eliminated. Consider a simple example with three choice alternatives $\{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$ and thresholds $\theta^{+} = 2$ and $\theta^{-} = -2$. The sample path in Figure 6.2 with its associated sequence of counter updates $S_{\boldsymbol{y}}^{+}, S_{\boldsymbol{x}}^{-}, S_{\boldsymbol{x}}^{-}, S_{\boldsymbol{y}}^{+}$, leads to elimination of alternative $\boldsymbol{x}$ after three steps and choice of alternative $\boldsymbol{y}$ after

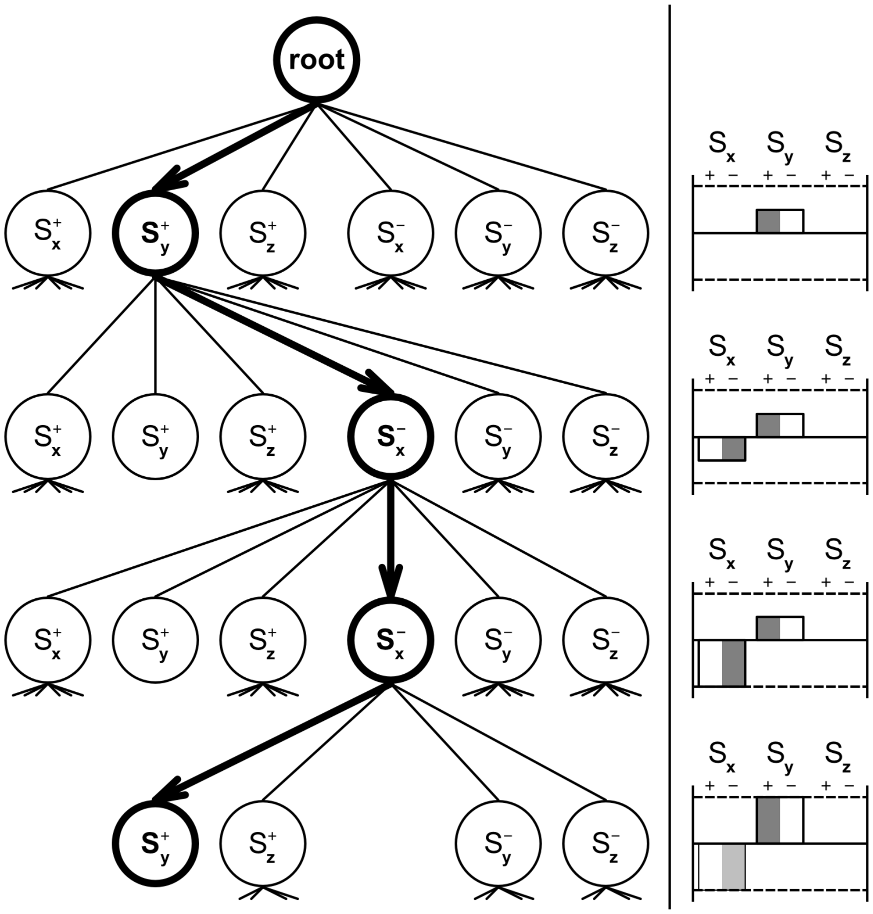

Figure 6.2: A random walk on the choice tree for three alternatives. The associated sequence of counter updates is $S_{\boldsymbol{y}}^{+}, S_{\boldsymbol{x}}^{-}, S_{\boldsymbol{x}}^{-}, S_{\boldsymbol{y}}^{+}$ and the probability for walking along this specific path is $p_{\boldsymbol{y}}^{+} \cdot p_{\boldsymbol{x}}^{-} \cdot p_{\boldsymbol{x}}^{-} \cdot p_{\boldsymbol{y}}^{+}$. Supposing that the rejection threshold θ^{-} is equal to -2 and the choice threshold θ^{+} is equal to 2, this sequence implicates first rejection of alternative ***x*** and then choice of alternative ***y***. When ***x*** is eliminated from the choice set, the vertices associated with its counters no longer appear in the choice tree, as can be seen in the bottom row of vertices here.

four steps. Note that there are only four vertices in the bottom row of Figure 6.2, since alternative ***x*** has been eliminated before. Other possible sequences resulting in choice of alternative ***y*** include $S_{\boldsymbol{z}}^{+}, S_{\boldsymbol{x}}^{-}, S_{\boldsymbol{y}}^{+}, S_{\boldsymbol{y}}^{+}$ with direct choice of ***y*** after four steps, and $S_{\boldsymbol{x}}^{-}, S_{\boldsymbol{z}}^{-}, S_{\boldsymbol{z}}^{-}, S_{\boldsymbol{x}}^{-}$ with elimination of alternatives ***z*** after three steps and ***x*** after four steps and therewith choice of the only remaining alternative ***y***, see figure 6.3.

6.3 The three (four) context effects

Similarity, attraction, and compromise effects have become a benchmark for multi-alternative multi-attribute preferential choice models. However, as for the compro-

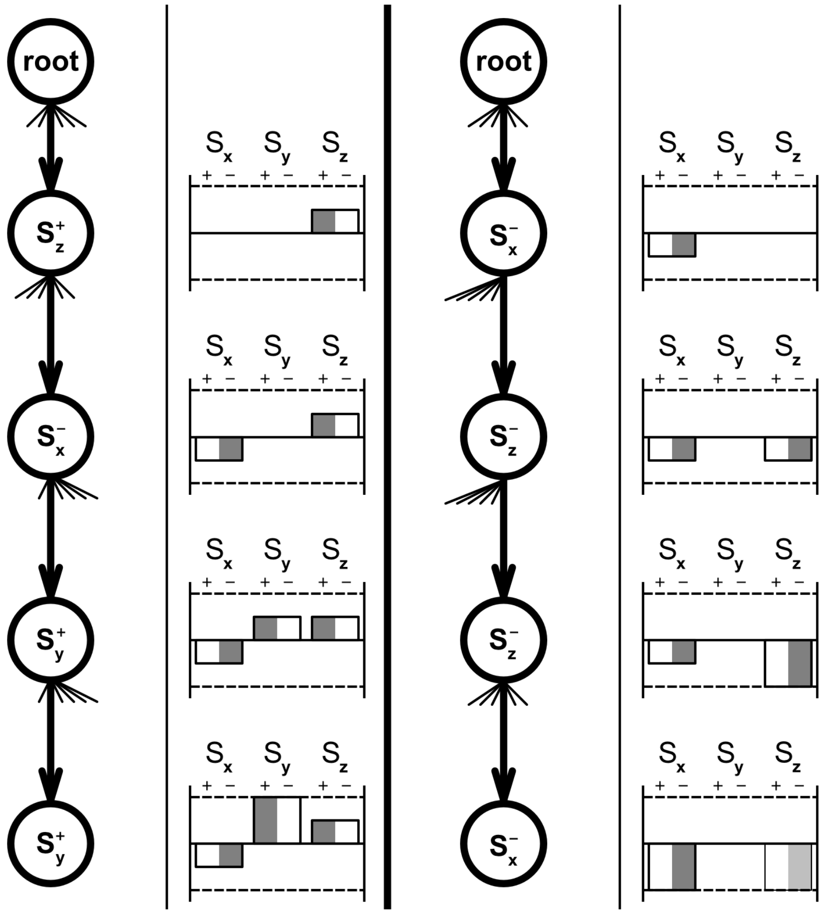

Figure 6.3: Two sequences of counter updates, both leading to choice of alternative ***y***, given choice and rejection thresholds equal to two. The right panel features two rejections, leaving alternative ***y*** as the only available option. The left panel features a choice of ***y*** without any rejections.

mise effect, some studies examine an asymmetric version of the effect and others a symmetric version. The 2NCT/SCT model is the first theory to explain, in addition to the similarity and attraction effect, both the asymmetric and the symmetric version of the compromise effect with the same set of mechanisms. Before we turn to the SCT model's explanation of the effects, we shortly summarize the findings. All three effects have been observed after adding a third alternative to a set of two equally attractive yet clearly distinguishable options described by two attributes. Let A_1 and A_2 be two choice alternatives with two common attributes, D_1 and D_2, describing them. We assume that D_1 is the unique strongest attribute for A_1, and D_2 is the unique strongest attribute for A_2, that is, A_1 scores high on D_1 but low on D_2 and vice versa for A_2. One can think of the alternatives as placed in a two-dimensional space with dimensions D_1 and D_2. We further assume that the probability for choosing alternative A_1 from the binary choice set is equal to the probability for choosing alternative A_2:

$$P(A_1|A_1, A_2) = P(A_2|A_1, A_2).$$

Note that this last assumption is relatively strict and therefore hard to meet experimentally (but see Berkowitsch et al., 2014, for an example of how to approach it by means of a matching task). In order to avoid this problem, some studies use relative choice shares to define the context effects (e.g., Trueblood et al., 2014, 2015). However, dimensional biases are covered by the relative choice shares and potentially distort the effects of interest (cf. Liew et al., 2016). Here, we maintain the equal probability assumptions, since they do not constrain our demonstration of the 2NCT/SCT model's ability to explain the three context effects.

The *similarity effect* was introduced by Tversky (1972b). He claims that the probability of choosing, say, A_1 over A_2 decreases after adding a third alternative A_3 that is similar to A_1 to the choice set. Let A_3 score high on attribute D_1 and low on attribute D_2 like alternative A_1 while overall being similarly attractive (i.e., $P(A_1|A_1, A_3) = P(A_3|A_1, A_3)$). Then the similarity effect is observed if

$$P(A_1|A_1, A_2, A_3) < P(A_2|A_1, A_2, A_3).$$

The *attraction effect* or *decoy effect* or *asymmetric dominance effect* was introduced by Huber et al. (1982). They claim that the probability for choosing alternative, say, A_1 can be increased by adding a third alternative A_3 to the choice set that is similar to but dominated by A_1. A_3 then serves as a decoy for alternative A_1, drawing attention to it and therewith improving its evaluation and increasing its choice probability. That is, the attraction effect is observed if

$$P(A_1|A_1, A_2, A_3) > P(A_2|A_1, A_2, A_3).$$

The *(asymmetric) compromise effect* was introduced by Simonson (1989). He claims that adding a third alternative A_3 to the choice set can increase the choice share of alternative, say, A_1, if A_3 is more extreme than alternative A_1 (i.e., A_3 scores even higher than A_1 on attribute D_1 and lower than A_1 on attribute D_2), but is overall similarly attractive as both A_1 and A_2 (i.e., $P(A_1|A_1, A_3) = P(A_3|A_1, A_3)$ and $P(A_2|A_2, A_3) = P(A_3|A_2, A_3)$). The compromise effect is observed if

$$\begin{aligned} &P(A_1|A_1, A_2, A_3) > P(A_2|A_1, A_2, A_3), \text{ and} \\ &P(A_1|A_1, A_2, A_3) > P(A_3|A_1, A_2, A_3). \end{aligned}$$

Note that, the more similar the additional extreme alternative A_3 is to its adjacent alternative A_1, the more share it takes away from A_1 via the similarity effect.

When Roe et al. (2001) proposed Multi-alternative Decision Field Theory, they used it to explain similarity, attraction, and compromise effects. However, when simulating choices, instead of adding an extreme alternative to the original choice set $\{A_1, A_2\}$, they add a compromise alternative A_3 in between A_1 and A_2. Again supposing that all binary choice probabilities are equal (i.e., $P(A_1|A_1, A_3) = P(A_3|A_1, A_3)$, and $P(A_2|A_2, A_3) = P(A_3|A_2, A_3)$), a *(symmetric) compromise effect* is observed if

$$\begin{aligned} &P(A_3|A_1, A_2, A_3) > P(A_1|A_1, A_2, A_3), \text{ and} \\ &P(A_3|A_1, A_2, A_3) > P(A_3|A_1, A_2, A_3). \end{aligned}$$

The symmetric version of the compromise effect challenges theories that explain context effects by changing attribute weights like, for example, the associative accumulation model (Bhatia, 2013). The asymmetric version on the other hand cannot be explained by introducing symmetric curvature to the attribute space like, for example, the multi-attribute linear ballistic accumulator model (Trueblood et al., 2014) does. Loss-averse value functions (e.g., Usher & McClelland, 2004; Wollschlaeger & Diederich, 2012) seem most promising for explaining both versions of the compromise effect, but let us now review the 2NCT/SCT model's explanations for the three context effects.

The 2NCT/SCT model's account for the context effects

In the 2NCT/SCT model, mainly two interacting mechanisms produce similarity, attraction, and compromise effects: (1) selection of pairs of attribute values for comparison based on normalized differences, and (2) the possibility to eliminate unwanted alternatives from the choice set. The first mechanism leads to a higher impact of dissimilar alternatives on the updating probabilities and thus faster evidence accumulation for alternatives with more distant competitors. In the similarity and attraction settings, this applies to the dissimilar alternative A_2, in the asymmetric compromise situation to the extreme alternatives A_2 and A_3, and in the symmetric compromise situation to the extreme alternatives A_1 and A_2. The second mechanism and the related focus weight λ determine whether choices are more likely to be based on eliminations or to be made directly. The greater λ, the more likely are the choices based on eliminations. In the similarity situation, greater λ leads to faster elimination of the dissimilar alternative A_2 and subsequent choice or elimination of either alternative A_1 or A_3, that is, a small or even negative similarity effect. On the other hand, smaller λ leads to more direct choices of alternative A_2 and thus a higher similarity effect. Regarding the dissimilar alternative A_2, the same is true in the attraction situation. Greater λ leads to faster elimination of A_2 while smaller λ leads to more direct choices of alternative A_2. However, the attraction effect is higher for greater λ, since after elimination of alternative A_2, either the dominating option A_1 is chosen directly or the dominated option A_3 is eliminated first. In the compromise setting, greater λ increases the probability for the extreme options to be eliminated from the choice set, leaving the decision maker with the compromise option. Smaller λ on the other hand more likely leads to choice of an extreme option and thus a smaller or even negative compromise effect. Attribute weights further moderate the strengths of the context effects, but as long as they are more or less balanced, they play a minor role in the explanation of the similarity, attraction, and compromise effects. However, a high attribute weight is able to bias choice towards the alternative that scores highest on that attribute, covering any context effect.

Simulations

We ran several simulations to illustrate these mechanisms. The available choice alternatives were $A_1 = (70, 30)$, $A_2 = (30, 70)$ and $A_3 = (70, 30)$ for the similarity effect, $A_3 = (65, 25)$ for the attraction effect, $A_3 = (90, 10)$ for the asymmetric compromise effect, or $A_3 = (50, 50)$ for the symmetric compromise effect. The attribute weights were $\omega_0 = 0.1$ and $\omega_1 = \omega_2 = 0.45$, and the focus weight λ varied between 0 and 1 in steps of 0.1. For each data point we ran 10000 simulations

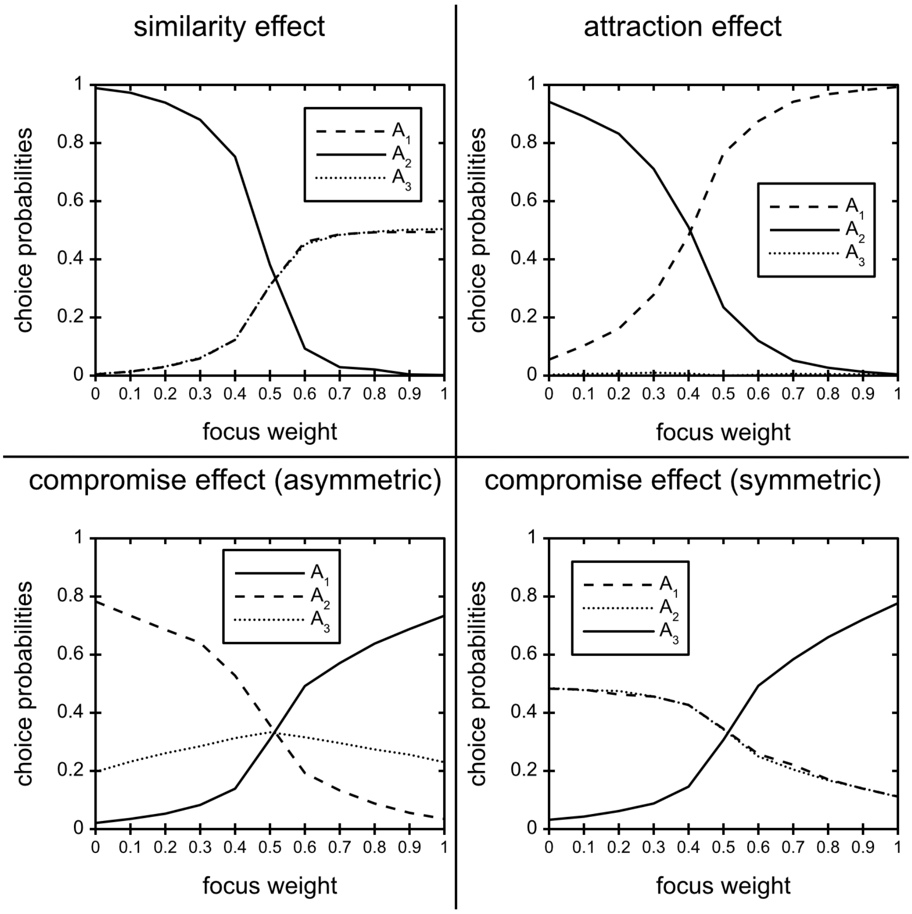

Figure 6.4: Simulations of choice probabilities for changing focus weight λ in the similarity, attraction, and (a)symmetric compromise situation. There is a positive similarity effect for smaller λ and a negative similarity effect for larger λ (upper left) and vice versa for the attraction effect (upper right). The compromise effect (lower left and right) shows for larger λ and is reversed for smaller λ.

and the resulting choice probabilities are presented in figure 6.4. According to the simulations, the similarity effect is opposed to the attraction and the compromise effect. The similarity effect is strongest for low λ, whereas the attraction and the compromise effect are strongest for high λ. This prediction is consistent with the finding that the attraction and the compromise effect are positively correlated with each other and negatively correlated with the similarity effect (Berkowitsch et al., 2014). Note that λ is assumed to be a global weight that does not change between trials but may vary between participants.

As the simulations show, the possibility to eliminate unwanted alternatives from the choice set during deliberation is crucial for the 2NCT/SCT Model's explanation of the three context effects (and their interrelations). However, this mechanism goes beyond the Turner et al. (2018) taxonomy, which does not include any stopping rules. Therefore, in the following section, we propose a revised taxonomy based on considerations about the time course of decision making processes.

6.4 A revised taxonomy

In their motivation for developing a taxonomy for theories of multi-alternative multi-attribute preferential choice, (Turner et al., 2018) state that "while the three processing stages do not necessarily define the temporal structure of the decision process, such a temporal distinction among the processing stages seems psychologically plausible" (p. 20). Here, we revise the Turner et al. (2018) taxonomy based on considerations about the time course of decision making processes. Where applicable, we take into account available process tracing data (e.g. Noguchi & Stewart, 2014).

(1) Subjective mapping of attribute values. The first processing stage in our revised taxonomy is equal to sub-stage (1.a) of the (Turner et al., 2018) taxonomy. Given the objective attribute values that describe the choice alternatives, each model has to define if and how these values are transformed into subjective representations.

(2) Attention allocation. The second processing stage of our revised taxonomy addresses how the decision maker distributes attention **(2.a) between attributes**, and – within attributes – **(2.b) between pairs of attribute values**. Sub-stage (2.a) here is equal to processing stage (2) of the Turner et al. (2018) taxonomy, that is, each model has to define attribute weights that may serve as attention probabilities. Sub-stage (2.b) is based on an eye-tracking study by Noguchi and Stewart (2014), who find that in multi-alternative choice, "alternatives are repeatedly compared in pairs on single dimensions" (p.44). In order to allow for such comparisons to take place, each model has to define weights that may serve as attention probabilities for pairs of attribute values. In the 2NCT/SCT model, these weights are called comparison values. MLBA similarly defines weights for pairs of attribute values, but uses them only during preprocessing and not as attention weights. For the other models, uniform attention probabilities may be assumed.

(3) Evaluation of alternatives. The third processing stage of our revised taxonomy addresses how alternatives are evaluated by the decision maker. In the Turner et al. (2018) taxonomy this is part of the perception process as described in sub-stages (1.b) and (1.c). Here, we divide the evaluation of alternatives into **(3.a) selecting a focus value and a reference value** and the actual **(3.b) comparison**. Given a pair of attribute values, each model has to determine for which alternative evidence will be accumulated by selecting the associated attribute value as focus value. Note that this could be interpreted as part of the attention allocation stage. However, in models with some kind of asymmetry (nonlinear filtering, i.e. sub-stage (1.c) of the Turner et al. (2018) taxonomy), calculation of pairwise differences may be required. We therefore interpret it as part of the evaluation stage. In models that evaluate alternatives against each other (cf. sub-stage (1.b) of the Turner et al. (2018) taxonomy), like MDFT, LCA, 2NCT/SCT, and MLBA, the second attribute value of the pair is automatically selected as reference value. Alternatively, an external value may be selected as reference value, e.g., the neutral reference point in AAM. Once the focus value and reference value are selected, the resulting evidence is determined by comparing the two values. MDFT, LCA, AAM, and MLBA explicitly calculate the difference of the two values while the 2NCT/SCT model only determines whether it is positive or negative.

(4) Evidence accumulation. The fourth processing stage of our revised taxonomy addresses how evidence is accumulated. It is divided into **(4.a) setup of counters**, **(4.b) counter updates including competition and noise**, and **(4.c) stopping rules**. All of these sub-stages have been discussed above in stages (3) and (4): Sub-stage (4.b) is equal to processing stage (3) of the Turner et al. (2018) taxonomy and sub-stages (4.a) and (4.c) correspond to the extensions that we made to the Turner et al. (2018) taxonomy in order to describe the 2NCT/SCT model. To summarize, each model has to specify how many counters accumulate evidence or preference (for each alternative), how these counters are updated over time, i.e., the amounts they are updated with, and when and why updating stops and a decision is made, i.e., which decision criteria are used.

6.5 Summary and Discussion

In this commentary, we discuss the 2N-ary choice tree model (2NCT, Wollschlaeger & Diederich, 2012) and its recent version simple choice tree model (SCT, Wollschlaeger & Diederich, 2017) and classify it by means of a taxonomy for computational models of multi-alternative multi-attribute preferential choice proposed by Turner et al. (2018). Where necessary, we expand the Turner et al. (2018) taxonomy to capture a unique feature of the 2NCT/SCT model: Separate and possibly asymmetric accumulation of evidence against and in favor of choosing each option. This affects the setup of counters (only the 2NCT/SCT model assumes two counters per alternative), the asymmetric weighting of positive and negative information (the free parameter λ allows the SCT model to mimic loss aversion as well as a bias for positive evidence), and stopping rules (the 2NCT/SCT model defines decision criteria for accepting and rejecting choice options). We then show how the 2NCT/SCT model accounts for the similarity, attraction, and compromise effects – an ability it shares with multi-alternative decision field theory (MDFT, Roe et al., 2001), the leaky competing accumulator model (LCA, Usher & McClelland, 2001, 2004), the associative accumulation model (AAM, Bhatia, 2013), the linear ballistic accumulator model (MLBA Trueblood et al., 2014), and multi-alternative decision by sampling theory (MDbS Noguchi & Stewart, 2018). All of these models employ different information processing mechanisms to explain the three context effect.

Turner et al. (2018) compare several of these mechanisms, that is, the ones employed in MDFT, LCA, AAM, and MLBA, based on their usefulness for explaining the three effects in a so-called switchboard analysis. They find that filtering of evidence before accumulation, comparison of attribute values, competition between counters, and the kind of noise that is added during accumulation of evidence have the largest effects on model performance. With respect to filtering mechanisms, the switchboard analysis revealed that no filtration of evidence before accumulation and loss-aversion both outperform the similarity-based filter mechanism proposed by MLBA. The 2NCT model allows for either no filtration or loss-aversion, while the SCT model with its focus weight λ additionally accounts for filtration biased towards confirmatory evidence. Weighting filtration with a free parameter (the focus weight λ) was not part of the Turner et al. (2018) switchboard analysis and will have to be compared to the other filtering mechanisms in the future. With respect to the comparison of attribute values, the switchboard analysis revealed that pairwise comparison of attribute values outperforms no comparison of attribute values, that is, using each alternative as reference point for the other alternatives as in MDFT, LCA, 2NCT/SCT, and MLBA outperforms evaluation against a single neutral reference point as in AAM. In the Turner et al. (2018) taxonomy, the comparison is part of the first processing stage that addresses subjective perception of the attribute space. However, we believe that selection of reference points at least partly depends on the task at hand and a general model of multi-alternative multi-attribute preferential choice should go beyond similarity, attraction, and compromise effects, and be able to also explain other reference point effects, e.g., the endowment effect (Kahneman, Knetsch, & Thaler, 1991).

We propose a revised taxonomy where the evaluation of alternatives takes place in a separate processing stage with two sub-stages that require selection of reference values and comparison of focus and reference values, respectively. With respect to competition, the switchboard analysis revealed that no competition between counter states and competition weighted by a free parameter (that is, global lateral inhibition) outperform the distance-dependent lateral inhibition proposed by MDFT. The 2NCT/SCT model proposes a different kind of competition, based on serial updating of counters instead of lateral inhibition between counter states. Future analysis will have to show whether this kind of competition can compete with the previously proposed ones. With respect to noise, the switchboard analysis revealed that between-trial variability in valuation noise outperforms within-trial variability. In the 2NCT/SCT model, the accumulators count instances in which the associated alternatives are perceived as winning or losing a comparison of attribute value pairs, that is, the counter states are restricted to positive integer values. It does therefore not make sense to add noise to the updating amounts. Instead, we add noise to the updating probabilities, leading to random counter updates once in a while. Again, this mechanism will have to be compared to the other implementations of noise into the accumulation process in the future. The revised taxonomy that we propose takes into account considerations about the time course of decision making processes. Where applicable, it is based on process tracing data (e.g. Noguchi & Stewart, 2014), but most importantly, it adds choice and elimination criteria and thereby optional instead of fixed stopping rules to the choice process.

Chapter 7

Discussion

In this thesis, a computational cognitive process model of multi-alternative multi-attribute preferential choice is proposed, revised, tested for its ability to simulate three benchmark context effects and interactions between them, and compared with earlier and more recent theories. The 2N-ary choice tree model assumes that the decision maker, given a set of N choice alternatives that are described by the same attributes, repeatedly compares pairs of attribute values and counts how often each alternative wins and loses a comparison. The number of favorable and unfavorable comparisons is stored in two separate counters per alternative and the difference of the counter states forms the preference state for the respective alternative. If the preference state for an alternative hits a negative threshold, this alternative is eliminated from the choice set and the comparison process continues without it. On the other hand, if the preference state for an alternative hits a positive threshold, this alternative is chosen and the whole process stops. The counter updating process is modeled as a random walk on a 2N-ary choice tree, a rooted tree with 2N children at each vertex that are associated with the 2N counters. Transition probabilities, that is, counter updating probabilities, are composed of attribute weights, comparison values, a local component, and a random component.

The simple choice tree model, a revised version of the 2N-ary choice tree model, differs from its predecessor in mainly two points. First, it omits the local component of the counter updating probabilities, which proved to be unnecessary for explaining the desired effects. This ensures improved mathematical tractability of the model. And second, it introduces an additional parameter for regulating the focus on the winning or losing alternative in a comparison, which has an effect on the proportion of choices and eliminations that take place. The 2N-ary choice tree model and the simple choice tree model are both able to explain similarity, attraction, and compromise effects, three context effects that have been observed after adding a third option to a set of two choice alternatives. With its additional parameter, the simple choice tree model beyond that accounts for the positive correlation between attraction and compromise effects and the negative correlation between these two and the similarity effect, that Berkowitsch et al. (2014) found in their recent study. To my knowledge, the simple choice tree model is the only model that accounts for the whole range of related findings, including negative similarity, attraction, and compromise effects (cf. section 2.6 and see Liew et al., 2016, for a commentary on averaging across participants and negative context effects).

The next steps will be to further explore the capabilities of the simple (2N-ary) choice tree model, both theoretically and in application to data. Its competitor models have been applied, for example, to phantom decoy effects (Usher et al., 2008; Bhatia, 2013; Trueblood & Pettibone, 2015), to choice deferral (Busemeyer et al., 2006; Bhatia & Mullett, 2016), and to best-worst scaling (Hawkins et al., 2013, 2014). The simple choice tree model includes mechanisms for considering external reference points like phantom decoys and naturally accounts for choice deferral and best-worst scaling with its elimination threshold. Parameter estimations could be conducted for data from the context effects experiments discussed in section 2.4. However, to exploit the simple choice tree model's full potential, not only choice frequencies but also response times should be taken into account. Most applications of computational cognitive process models of context effect so far ignore response times altogether, neglecting the dynamic nature of these models (cf. section 2.6).

The simple (2N-ary) choice tree model makes novel predictions with respect to the focus weight and the related elimination mechanism that require new experiments and/or extensions of current experimental paradigms. For example, the simple choice tree model predicts different choice probabilities for the extreme alternatives in symmetric and asymmetric compromise situations, dependent on the focus weight. Symmetric and asymmetric versions of the compromise effect have never been tested in the same experiment so far. Another interesting prediction affects the different types of decoys used in attraction effect studies. Huber et al. (1982) reported descending magnitudes of the effect for range, range-frequency, and frequency decoys. In the simple choice tree model, the same order is predicted for higher values of the focus weight, but the reversed order is predicted for lower values of the focus weight. Furthermore, it would be interesting to examine if the focus weight is related to the concept of maximizing versus satisficing (Schwartz et al., 2002). Mao (2016) examined the compromise effect under maximizing tendencies and found that maximizers choose a compromise option more often, indicating that the focus weight of a maximizer might be higher than that of a satisficer.

Regarding the elimination mechanism, eye tracking possibly provides insights about ignored alternatives (see, for example, Stewart, Hermens, & Matthews, 2016, for a discussion of gaze bias in risky choice). Data from the context effects study by Noguchi and Stewart (2014) could be re-analyzed with respect to predicted eliminations. Additionally, new experiments should be designed that provide the possibility to explicitly eliminate alternatives from the choice set. Best-worst scaling methods could inform the experimental design, however, alternatives marked as the worst option in this kind of task usually remain on the screen until the best option is chosen as well.

On a final note, I want to emphasize a remark made in section 2.6. It is argued there, that computational cognitive process models should be compared on the level of the microprocesses they are based on instead of on the level of whole theories. The switchboard analysis conducted by Turner et al. (2018) is a first step in this direction. Adding optional stopping times and an elimination mechanism, as proposed in chapter 6, could be the next step, extending the benchmark beyond context effects the step after next.

References

Ashby, N. J. S., Johnson, J. G., Krajbich, I., & Wedel, M. (2016). Applications and Innovations of Eye-movement Research in Judgment and Decision Making. *Journal of Behavioral Decision Making*, *29*(2-3), 96–102. https://doi.org/10.1002/bdm.1956

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Probabilities of choices among very similar objects: an experiment to decide between two models. *Behavioral science*, *8*(4), 306–311. https://doi.org/10.1002/bs.3830080403

Berkowitsch, N. A. J., Scheibehenne, B., & Rieskamp, J. (2014). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology: General*, *143*(3), 1331–1348. https://doi.org/10.1037/a0035159

Bettman, J. (1979). *An information processing theory of consumer choice*. Reading, MA: Addison-Wesley Publishing Company, Inc.

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, *120*(3), 522–543. https://doi.org/10.1037/a0032457

Bhatia, S., & Mullett, T. L. (2016). The dynamics of deferred decision. *Cognitive Psychology*, *86*(C), 112–151. https://doi.org/10.1016/j.cogpsych.2016.02.002

Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences*, *43*(3), 345–370. https://doi.org/10.1016/S0165-4896(02)00016-1

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive Modeling*. Los Angeles, CA: SAGE Publications.

Busemeyer, J. R., Johnson, J. G., & Jessup, R. K. (2006). Preferences constructed from dynamic microprocessing mechanisms. In S. Lichtenstein & P. Slovic (Eds.), *The construction of preference* (p. 220-234). New York, NY: Cambridge University Press.

Busemeyer, J. R., & Townsend, J. T. (1992). Fundamental derivations from decision field theory. *Mathematical Social Sciences*, *23*, 255–282. https://doi.org/10.1016/0165-4896(92)90043-5

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459. https://doi.org/10.1037/0033-295X.100.3.432

Cataldo, A. M., & Cohen, A. L. (2018). Reversing the similarity effect: The effect of presentation format. *Cognition*, *175*, 141–156. https://doi.org/10.1016/j.cognition.2018.02.003

Chang, C.-C., & Liu, H.-H. (2008). Which is the compromise option? Information format and task format as determinants. *Journal of Behavioral Decision Making*, *21*(1), 59–75. https://doi.org/10.1002/bdm.574

Choplin, J. M., & Hummel, J. E. (2002). Magnitude comparisons distort mental representations of magnitude. *Journal of Experimental Psychology: General*, *131*(2), 270–286. https://doi.org/10.1037/0096-3445.131.2.270

Clifton Jr, C., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016). Eye movements in reading and information processing: Keith Rayner's 40 year legacy. *Journal of Memory and Language*, *86*(C), 1–19. https://doi.org/10.1016/j.jml.2015.07.004

Cohen, A. L., Kang, N., & Leise, T. L. (2017). Multi-attribute, multi-alternative models of choice: Choice, reaction time, and process tracing. *Cognitive Psychology*, *98*, 45–72. https://doi.org/10.1016/j.cogpsych.2017.08.001

Debreu, G. (1960). Review: Individual choice behavior: A theoretical analysis. By R. Duncan Luce. *The American Economic Review*, *50*(1), 186–188.

Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, *39*, 197–215. https://doi.org/10.1006/jmps.1995.1020

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, *41*(3), 260–274. https://doi.org/10.1006/jmps.1997.1167

Diederich, A. (2008). A further test of sequential-sampling models that account for payoff effects on response bias in perceptual decision tasks. *Perception & Psychophysics*, *70*(2), 229–256. https://doi.org/10.3758/PP.70.2.229

Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, *47*(3), 304–322. https://doi.org/10.1016/S0022-2496(03)00003-8

Diederich, A., & Busemeyer, J. R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: Bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception & Psychophysics*, *68*(2), 194–207. https://doi.org/10.3758/BF03193669

Durrett, R. (2010). *Probability: Theory and Examples* (4th ed.). New York, NY: Cambridge University Press.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.

Hawkins, G. E., Marley, A. A. J., Heathcote, A., Flynn, T. N., Louviere, J. J., & Brown, S. D. (2013). Integrating cognitive process and descriptive models of attitudes and preferences. *Cognitive Science*, *38*(4), 701–735. https://doi.org/10.1111/cogs.12094

Hawkins, G. E., Marley, A. A. J., Heathcote, A., Flynn, T. N., Louviere, J. J., & Brown, S. D. (2014). The best of times and the worst of times are interchangeable. *Decision*, *1*(3), 192–214. https://doi.org/10.1037/dec0000012

Hick, W. E. (1952). On the rate of gain of information. *The Quarterly Journal of Experimental Psychology*, *4*(1), 11–26. https://doi.org/10.1080/17470215208416600

Highhouse, S. (1996). Context-Dependent Selection: The Effects of Decoy and Phantom Job Candidates. *Organizational Behavior and Human Decision Processes*, *65*(1), 68–76. https://doi.org/10.1006/obhd.1996.0006

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin*, *138*(2), 211–237. https://doi.org/10.1037/a0025940

Hotaling, J. M., Busemeyer, J. R., & Li, J. (2010). Theoretical developments in decision field theory: Comment on Tsetsos, Usher, and Chater (2010). *Psychological Review*, *117*(4), 1294–1298. https://doi.org/10.1037/a0020401

Houpt, J. W., Yang, C.-T., & Townsend, J. T. (2016). Editorial: Modeling individual differences in perceptual decision making. *Frontiers in Psychology*, *7*, 1602. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2016.01602

Huber, J., Payne, J. W., & Puto, C. P. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*(1), 90–98. https://doi.org/10.1086/208899

Huber, J., & Puto, C. P. (1983). Market boundaries and product choice: Illustrating attraction and substitution effects. *Journal of Consumer Research*. https://doi.org/10.1086/208943

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*(3), 188–196. https://doi.org/10.1037/h0056940

Johnson, J. G., & Busemeyer, J. R. (2005). A dynamic, stochastic, computational model of preference reversal phenomena. *Psychological Review*, *112*(4), 841–861. https://doi.org/10.1037/0033-295X.112.4.841

Johnson, J. G., & Busemeyer, J. R. (2016). A computational model of the attention process in risky choice. *Decision*, *3*(4), 254–280. https://doi.org/10.1037/dec0000050

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives*, *5*(1), 193–206. https://doi.org/10.1257/jep.5.1.193

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), 341–350. https://doi.org/10.1037/0003-066X.39.4.341

Kahneman, D., & Tversky, A. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061. https://doi.org/10.2307/2937956

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value tradeoffs*. Cambridge, UK: Cambridge University Press. (Original work published 1967)

Koehler, D. J., & Harvey, N. (2007). *Blackwell Handbook of Judgment and Decision Making* (2nd ed.). Oxford, UK: Blackwell Publishing.

Korte, B., & Vygen, J. (2006). *Combinatorial optimization: theory and algorithms* (3rd ed.). Berlin, Germany: Springer.

Kuhn, T. S. (2012). *The structure of scientific revolutions* (4th ed.). Chicago, IL: The University of Chicago Press. (Original work published 1962)

LaBerge, D. (1962). A recruitment theory of simple behavior. *Psychometrika*, *27*(4), 375–396. https://doi.org/10.1007/BF02289645

LaBerge, D. (1994). Quantitative Models of Attention and Response Processes in Shape Identification Tasks. *Journal of Mathematical Psychology*, *38*(2), 198–243. https://doi.org/10.1006/jmps.1994.1015

Laming, D. R. J. (1968). *Information theory of choice reaction times*. London, UK: Academic Press.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of experimental psychology*, *89*(1), 46. https://doi.org/10.1037/h0031207

Lichtenstein, S., & Slovic, P. (1973). Response-induced reversals of preference in gambling: An extended replication in Las Vegas. *Journal of Experimental Psychology*, *101*(1), 16–20. https://doi.org/10.1037/h0035472

Lichtenstein, S., & Slovic, P. (2006). The construction of preference: An overview. In S. Lichtenstein & P. Slovic (Eds.), *The construction of preference* (p. 1-40). New York, NY: Cambridge University Press.

Liew, S. X., Howe, P. D. L., & Little, D. R. (2016). The appropriacy of averaging in the study of context effects. *Psychonomic Bulletin & Review*, *23*(5), 1639–1646. https://doi.org/10.3758/s13423-016-1032-7

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, *40*(1), 77–105. https://doi.org/10.1007/BF02291481

Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling: Theory, methods and applications*. Cambridge, UK: Cambridge University Press.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, *15*(3), 215–233. https://doi.org/10.1016/0022-2496(77)90032-3

Luce, R. D. (1986). *: Their role in inferring elementary mental organization*. New York, NY: Oxford University Press.

Luce, R. D. (2012). *Individual choice behavior: A theoretical analysis*. Mineola, NY: Dover Publications. (Original work published 1959)

Mao, W. (2016). When one desires too much of a good thing: The compromise effect under maximizing tendencies. *Journal of Consumer Psychology*, *26*(1), 66–80. https://doi.org/10.1016/j.jcps.2015.04.007

March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of ehoice. *The Bell Journal of Economics*, *9*(2), 587–608. https://doi.org/10.2307/3003600

Markman, A. B., & Medin, D. L. (1995). Similarity and alignment in choice. *Organizational Behavior and Human Decision Processes*, *63*(2), 117–130. https://doi.org/10.1006/obhd.1995.1067

Noguchi, T., & Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition*, *132*(1), 44–56. https://doi.org/10.1016/j.cognition.2014.03.006

Noguchi, T., & Stewart, N. (2018). Multialternative decision by sampling: A model of decision making constrained by process data. *Psychological Review*, *125*(4), 512–544. https://doi.org/10.1037/rev0000102

Oppenheimer, D. M., & Kelso, E. (2015). Information processing as a paradigm for decision making. *Annual Review of Psychology*, *66*, 277–294. https://doi.org/10.1146/annurev-psych-010814-015148

Otworowska, M., Blokpoel, M., Sweers, M., Wareham, T., & van Rooij, I. (2017). Demons of ecological rationality. *Cognitive Science*, *42*(3), 1057–1066. https://doi.org/10.1111/cogs.12530

Payne, J. W., & Bettman, J. R. (2008). Walking with the scarecrow: The information-processing approach to decision research. In *Blackwell handbook of judgment and decision making* (pp. 110–132). Malden, MA: John Wiley & Sons. https://doi.org/10.1002/9780470752937.ch6

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, *43*, 87–131. https://doi.org/10.1146/annurev.ps.43.020192.000511

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press.

Pettibone, J. C., & Wedell, D. H. (2000). Examining models of nondominated decoy effects across judgment and choice. *Organizational Behavior and Human Decision Processes*, *81*(2), 300–328. https://doi.org/10.1006/obhd.1999.2880

Pettibone, J. C., & Wedell, D. H. (2007). Testing alternative explanations of phantom decoy effects. *Journal of Behavioral Decision Making*, *20*(3), 323–341. https://doi.org/10.1002/bdm.557

Pike, A. R. (1966). Stochastic models of choice behaviour: Response probabilities and latencies of finite markov chain systems. *British Journal of Mathematical and Statistical Psychology*, *19*(1), 15–32. https://doi.org/10.1111/j.2044-8317.1966.tb00351.x

Pratkanis, A. R., & Farquhar, P. H. (1992). A brief history of research on phantom alternatives: Evidence for seven empirical generalizations about phantoms. *Basic and Applied Social Psychology*, *13*(1), 103–122. https://doi.org/10.1207/s15324834basp1301_9

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-Choice reaction time. *Psychological Review*, *111*(2), 333–367. https://doi.org/10.1037/0033-295X.111.2.333

Ratneshwar, S., Shocker, A. D., & Stewart, D. W. (1987). Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity. *Journal of Consumer Research*, *13*(4), 520–533. https://doi.org/10.1086/209085

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological Review*, *108*(2), 370–392. https://doi.org/10.1037/0033-295X.108.2.370

Ronayne, D., & Brown, G. D. A. (2017). Multi-attribute decision by sampling: An account of the attraction, compromise and similarity effects. *Journal of Mathematical Psychology*, *81*, 11–27. https://doi.org/10.1016/j.jmp.2017.08.005

Salkuyeh, D. K. (2006). Positive integer powers of the tridiagonal Toeplitz matrices. *International Mathematical Forum*, *1*(22), 1061–1065. https://doi.org/10.12988/imf.2006.06086

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59. https://doi.org/10.1007/BF00055564

Schulte-Mecklenbeck, M., Kuehberger, A., & Ranyard, R. (2011). The role of process data in the development and testing of process models of judgment and decision making. *Judgment and Decision Making*, *6*(8), 733–739. Retrieved from http://journal.sjdm.org/11/m36/m36.pdf

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology*, *83*(5), 1178–1197. https://doi.org/10.1037/0022-3514.83.5.1178

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*(2), 207–222. https://doi.org/10.1037/0033-2909.134.2.207

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. https://doi.org/10.2307/1884852

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*(2), 158–174. https://doi.org/10.1086/209205

Simonson, I., Carmon, Z., & O'Curry, S. (1994). Experimental evidence on the negative effect of product features and sales promotions on brand choice. *Marketing Science*, *13*(1), 23–40. https://doi.org/10.1287/mksc.13.1.23

Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, *29*(3), 281–295. https://doi.org/10.2307/3172740

Smith, P. L., & Van Zandt, T. (2000). Time-dependent Poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology*, *53*(2), 293–315. https://doi.org/10.1348/000711000159349

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1–26. https://doi.org/10.1016/j.cogpsych.2005.10.003

Stewart, N., Hermens, F., & Matthews, W. J. (2016). Eye Movements in Risky Choice. *Journal of Behavioral Decision Making*, *29*(2-3), 116–136. https://doi.org/10.1002/bdm.1854

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, *1*(1), 39–60. https://doi.org/10.1016/0167-2681(80)90051-7

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge, UK: Cambridge University Press.

Trueblood, J. S. (2012). Multialternative context effects obtained using an inference task. *Psychonomic Bulletin & Review*, *19*(5), 962–968. https://doi.org/10.3758/s13423-012-0288-9

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice.

Psychological Review, *121*(2), 179–205. https://doi.org/10.1037/a0036137

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2015). The fragile nature of contextual preference reversals: Reply to Tsetsos, Chater, and Usher (2015). *Psychological Review*, *122*(4), 848–853. https://doi.org/10.1037/a0039656

Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013). Not just for consumers: context effects are fundamental to decision making. *Psychological Science*, *24*(6), 901–908. https://doi.org/10.1177/0956797612464241

Trueblood, J. S., & Pettibone, J. C. (2015). The phantom decoy effect in perceptual decision making. *Journal of Behavioral Decision Making*, *30*(2), 157–167. https://doi.org/10.1002/bdm.1930

Trukenbrod, H. A., & Engbert, R. (2014). ICAT: a computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review*, *21*(4), 907–934. https://doi.org/10.3758/s13423-013-0575-0

Tsetsos, K., Chater, N., & Usher, M. (2015). Examining the mechanisms underlying contextual preference reversal: Comment on Trueblood, Brown, and Heathcote (2014). *Psychological Review*, *122*(4), 838–847. https://doi.org/10.1037/a0038953

Tsetsos, K., Usher, M., & Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological Review*, *117*(4), 1275–1291. https://doi.org/10.1037/a0020580

Turner, B. M., Schley, D. R., Muller, C., & Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. *Psychological Review*, *125*(3), 329–362. https://doi.org/10.1037/rev0000089

Tversky, A. (1972a). Choice by elimination. *Journal of Mathematical Psychology*, *9*(4), 341–367. https://doi.org/10.1016/0022-2496(72)90011-9

Tversky, A. (1972b). Elimination by aspects: A theory of choice. *Psychological Review*, *79*(4), 281–299. https://doi.org/10.1037/h0032955

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458. https://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, *59*(4), 251–278. https://doi.org/10.1086/296365

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061. https://doi.org/10.2307/2937956

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Tversky, A., & Russo, J. E. (1969). Substitutability and similarity in binary choices. *Journal of Mathematical Psychology*, *6*(1), 1–12. https://doi.org/10.1016/0022-2496(69)90027-3

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, *95*(3), 371–384. https://doi.org/

10.1037/0033-295X.95.3.371

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, *39*(10), 1179–1189. https://doi.org/10.1287/mnsc.39.10.1179

Tversky, A., & Thaler, R. H. (1990). Anomalies: Preference reversals. *The Journal of Economic Perspectives*, *4*(2), 201–211. https://doi.org/10.1257/jep.4.2.201

Usher, M., Elhalal, A., & McClelland, J. L. (2008). The neurodynamics of choice, value-based decisions, and preference reversal. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for bayesian cognitive science* (pp. 277–300). Oxford, UK: Oxford University Press.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. https://doi.org/10.1037/0033-295X.108.3.550

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*(3), 757–769. https://doi.org/10.1037/0033-295X.111.3.757

Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychnomic Bulletin & Review*, *7*(2), 208–256. https://doi.org/10.3758/BF03212980

Vlaev, I., Chater, N., Stewart, N., & Brown, G. D. A. (2011). Does the brain calculate value? *Trends in Cognitive Sciences*, *15*(11), 546–554. https://doi.org/10.1016/j.tics.2011.09.008

von Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behavior* (3rd ed.). Princeton, NJ: Princeton University Press.

Wedell, D. H. (1991). Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 767–778. https://doi.org/10.1037/0278-7393.17.4.767

Wollschlaeger, L. M., & Diederich, A. (2012). The 2N-ary choice tree model for N-alternative preferential choice. *Frontiers in psychology*, *3*(189). https://doi.org/10.3389/fpsyg.2012.00189

Wollschlaeger, L. M., & Diederich, A. (2017). A computational model for constructing preferences for multiple choice options. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th annual conference of the cognitive science society* (pp. 1351–1356). Austin, TX: Cognitive Science Society. Retrieved from https://mindmodeling.org/cogsci2017/papers/0259/paper0259.pdf

Appendix A

Matlab code for simulations

A.1 The 2N-ary choice tree model

```
% 2N-ary choice tree model for N-alternative preferential choice
% Simulations

clear all;
%close all;

startTime=clock;


%% choice setting

% number of alternatives
N=3;

% number of dimensions
D=2;

% evaluations
X=[0.7,0.3;0.3,0.7;0.7,0.3]
M=scalingc(X)

% number of walks during simulation
K=100;


%% free parameters

% weight of dimensions
W=[0.5;0.5]
%W=(1/D).*ones(D,1);

% decision criterion
theta=5;

% noise factor
xi=0.01;

% leakage factor
lambda=0.05;

% inhibition factor
mu=0;
```

```
%% model constants


compWIS=cell(2^N,4);

% comparison of alternatives
compWIS{1,1}=compdiff(N,M);

% weighted comparison
tempWcomp=compWIS{1,1}*W;
    if sum(tempWcomp)==0
        compWIS{1,2}=0;
    else
        compWIS{1,2}=scaling(tempWcomp);
    end

% inhibition matrix
compWIS{1,4}=dftdist(N,M);

% counter update matrix
compWIS{1,3}=eye(2*N);

% prepare combinations, source:
% https://www.mathworks.com/matlabcentral/answers/
%   uploaded_files/8064/combn.m
comb=combn([1,0],N);

for i=2:(2^N)

    index0=find(comb(i,:)==0);
    index1=find(comb(i,:));

    tempX=X;
    tempX(index0,:)=0;
    tempM=scalingc(tempX);
    compWIS{i,1}=compdft2(index1,N,D,tempM);

    tempWcomp=compWIS{i,1}*W;
    if sum(tempWcomp)==0
        compWIS{i,2}=0;
    else
        compWIS{i,2}=scaling(tempWcomp);
    end

    tempI=eye(2*N);
    tempI(index0+N,:)=0;
    tempI(index0,:)=0;
    compWIS{i,3}=tempI;

    tempS=compWIS{1,4};
    tempS(index0,:)=0;
    tempS(:,index0)=0;
    compWIS{i,4}=tempS;

end


%% main part

data=zeros(K,2);
```

```
for k=1:K

    %initializations
    % timestep counter
    h=0;
    % state of random walk
    % [A,B,...,-A,-B,hits/rejections,stop indicator,probability]
    counter=[zeros(1,2*N),ones(1,N),0,1];

    while counter(end-1)==0

        %specification of row to pick from compWIS
        rowInd = ...
            (2^N) - ...
            bin2dec(sprintf('%i',counter((2*N+1):(3*N))));
        %indices of withdrawn elements
        index=find(counter((2*N+1):(3*N))==0);


        %computation of leakage
        normdecay=decaydft2(counter(1:(2*N)),N,index);

        %computation of inhibition
        norminhib=inhibdft(counter(1:(2*N)),N,compWIS{rowInd,4});

        %computation of transition probabilities
        % for the current state/position,
        probabilities=transitdft2(...
            index,sum(counter((2*N+1):(3*N))),N,...
            compWIS{rowInd,2},...
            norminhib,normdecay,...
            xi,lambda,mu);


        %next step of random walk
        counter=stepdft2(probabilities,counter,N,theta);

        h=h+1;
    end

    data(k,:)=[counter(end-1),h];
end


endTime=clock;
usedTime=etime(endTime,startTime)

%% plot
m=1:N;
for i=1:N
    m(2,i)=sum(data(:,1)==m(1,i));
    m(3,i)=sum(data(data(:,1)==m(1,i),2))/m(2,i);
end
m
bar(m(1,:),m(2,:)./K)
```

```
function[csm]=scalingc(X)

scale=sum(X,1);

for i=1:size(X,2)
```

```
    csm(:,i)=X(:,i)/scale(i);
end
```

```
function[val]=compdiff(N,X)

val=repmat(X,2,1);

for i=1:N
    for j=1:size(X,2)
        val(i,j)=sum((X(i,j)>X(:,j)).*(X(i,j)-X(:,j)));
        val(i+N,j)=sum((X(i,j)<X(:,j)).*(X(:,j)-X(i,j)));
    end
end
```

```
function[sv]=scaling(v)

scale=sum(v);

sv=v/scale;
```

```
function[S]=dftdist(N,M)

S=zeros(N);

D=pdist(M);

F = @(x) exp(-x);

for i=1:(N-1)
    for j=1:i
        S(i+1,j)=F(D(i-1+j));
        S(j,i+1)=S(i+1,j);
    end
end
```

```
function[val]=compdft2(index1,N,D,M)

val=repmat(M,2,1);

for i=index1
    for j=1:D
        val(i,j) = ...
            sum((M(i,j)>M(index1,j)).*(M(i,j)-M(index1,j)));
        val(i+N,j) = ...
            sum((M(i,j)<M(index1,j)).*(M(index1,j)-M(i,j)));
    end
end
```

```
function[normdecay]=decaydft2(state,N,index)

counter=zeros(2*N,1);

for i=1:N
    counterdiff=state(i)-state(i+N);
    counter(i)=max([counterdiff,0]);
    counter(i+N)=-min([counterdiff,0]);
end
```

```
decay=[(counter((N+1):(2*N)));(counter(1:N))];

decay(index)=0;
decay(index+N)=0;

if sum(decay)==0;
    normdecay=0;
else
    normdecay=decay/sum(decay);
end
```

```
function[norminhib]=inhibdft(state,N,S)

counter=zeros(2*N,1);

for i=1:N
    counterdiff=state(i)-state(i+N);
    counter(i)=max([counterdiff,0]);
    counter(i+N)=-min([counterdiff,0]);
end


inhib=[S*(counter((N+1):(2*N)));S*(counter(1:N))];

if sum(inhib)==0;
    norminhib=0;
else
    norminhib=inhib/sum(inhib);
end
```

```
function[probabilities]=transitdft2( ...
    index,m,N,normcomp,norminhib,normdecay,xi,lambda,mu)

probabilities=...
    ((1-lambda-mu)*(...
    (1-xi)*normcomp+...
    xi*(ones(2*N,1)/(2*m)))+...
    lambda*norminhib+...
    mu*normdecay);

probabilities(index+N)=0;
probabilities(index)=0;
```

```
function[state]=stepdft2(probabilities,oldstate,N,theta)

state=oldstate;

%the vector of cumulative sums of the
% probabilities divides the unit interval
% into intervals with lengths equal to
% the transition probabilities. An uniformly
% distributed random variable x lies within
% each of these intervals with the associated
% probability
cumprobabilities=cumsum(probabilities);

x=rand(1);

%update of counter associated to the
% interval that was hit by x in that step
% state(end) contains the probability
```

```
% for having reached that state, test
% tests, whether a threshold is hit.
% If so, state(end-1) is set to the number
% of the corresponding alternative
if x<=cumprobabilities(1)
    state(1)=oldstate(1)+1;
    state(end)=oldstate(end)*probabilities(1);
    test=state(1)-state(N+1);
    state(end-1)=1*(test==theta);
elseif ((cumprobabilities(N)<x)*...
        (x<=cumprobabilities(N+1)))==1
    state(N+1)=oldstate(N+1)+1;
    state(end)=oldstate(end)*probabilities(N+1);
    test=state(1)-state(N+1);
    state(1+(2*N))=oldstate(1+(2*N))-(test==-theta);
else
    for i=2:N
        if ((cumprobabilities(i-1)<x)*...
                (x<=cumprobabilities(i)))==1
            state(i)=oldstate(i)+1;
            state(end)=oldstate(end)*probabilities(i);
            test=state(i)-state(N+i);
            state(end-1)=i*(test==theta);
        elseif ((cumprobabilities(N+i-1)<x)*...
                (x<=cumprobabilities(N+i)))==1
            state(N+i)=oldstate(N+i)+1;
            state(end)=oldstate(end)*probabilities(N+i);
            test=state(i)-state(N+i);
            state(i+(2*N))=oldstate(i+(2*N))-(test==-theta);
        end
    end
end

if sum(state((2*N+1):(3*N)))==1
    state(end-1)=find(state((2*N+1):(3*N)));
end
```

A.2 The simple choice tree model

```
function[] = Simulations()
% The Simple (2N-ary) Choice Tree Model
% SIMULATIONS simulates choices and response times
% based on the simple choice tree model

%% define choice alternatives and external reference points
choiceAlternatives = [ 70,  30; ... % Alternative 1
                       30,  70; ... % Alternative 2
                       70,  25 ];   % Alternative 3
% extract number of alternatives and attributes
[nAlternatives, nDimensions] = size(choiceAlternatives);
%define external reference points
externalReferencePoints = [ ];


%% define fixed parameters and file names
randomWeight = 0.1;
decisionThreshold = 10; % (theta^+ = theta^-)
nSimulations = 10000; % number of simulations
dataFile = false;
dataFileName = 'SCT_simulations.dat';
```

```
summaryFile = true;
summaryFileName = 'SCT_simulations_summary.dat';


%% write headers
if dataFile
    dataFileID = fopen(dataFileName, 'w'); % discard content
    headerSpec = ['m11 m12 m21 m22 m31 m32  '...
        'fw   w1   w2   rw   ' ...
        'a1  a2  a3  t1  t2  t3 ' ...
        's11 s12 s13 s14 s15 s16 '... % counter states at time t1
        's21 s22 s23 s24 s25 s26 '... % counter states at time t2
        's31 s32 s33 s34 s35 s36\n']; % counter states at time t3
    header = [];
    fprintf(dataFileID, headerSpec, header);
    fclose(dataFileID);
end

if summaryFile
    summaryFileID = fopen(summaryFileName, 'w'); % discard content
    headerSpec = ['m11 m12 m21 m22 m31 m32  '...
        'fw   w1   w2   rw   ' ...
        'a1  a2  a3    t1    t2    t3 ' ...
        'a11 a12 a13   t11   t12   t13 '...
        'a21 a22 a23   t21   t22   t23 '...
        'a31 a32 a33   t31   t32   t33\n'];
    header = [];
    fprintf(summaryFileID, headerSpec, header);
    fclose(summaryFileID);
end


%% compute comparison values and random component
% compute internal comparison values
[internalComparisonValues, randomComponent] = ...
    computecomparisonvalues(choiceAlternatives);
% check if there are external reference points
if isempty(externalReferencePoints)
    % if not, set externalComparisonValues to zero
    externalComparisonValues = 0;
else
    % extract number of external reference points and attributes
    nExternalDimensions = size(externalReferencePoints,2);
        % check if number of attribute match
        if nExternalDimensions ~= nDimensions
            msg = ['Error. '...
            'Number of attributes ' ...
            'of external reference points ' ...
            'does not match number of attributes ' ...
            'of choice alternatives.'];
            error(msg)
        end
    % compute external comparison values
    [externalComparisonValues, ~] = ...
        computecomparisonvalues( ...
        choiceAlternatives, externalReferencePoints);
end


%% define varying weights and compute probabilities
for focusWeight = 0:0.1:1
    for attributeWeight = 0:0.1:1
```

```
        attributeWeights = [attributeWeight, (1-attributeWeight)];


        % multiply negative comparison values with focusWeight and
        % positive comparison values with (1-focusWeight)
        focusedComparisonValues = [ ...
            (1-focusWeight) .* ...
            internalComparisonValues(1:nAlternatives,:,:);
            focusWeight .* ...
            internalComparisonValues((nAlternatives+1):end,:,:)];
        % add comparison values
        comparisonValues = ...
            focusedComparisonValues + ...
            externalComparisonValues;
        % weight dimensions and add up comparison values
        repeatedWeights = permute( ...
            repmat(attributeWeights, ...
            2*nAlternatives, 1, 2^nAlternatives), ...
            [1 3 2]);
        weightedComparisonValues = ...
            repeatedWeights .* comparisonValues;
        probabilities = (randomWeight * randomComponent) + ...
            (1-randomWeight) * sum(weightedComparisonValues, 3);


        %% define decision threshold and simulate
        %   choices and response times.

        simulatedData = simulatechoices( ...
            nAlternatives, nSimulations, ...
            probabilities, decisionThreshold );

        summarizedData = summarizedata( ...
            nAlternatives, simulatedData );

        alternativeInfo = [choiceAlternatives(1,:), ...
            choiceAlternatives(2,:), ...
            choiceAlternatives(3,:), ...
            focusWeight, ...
            attributeWeights, ...
            randomWeight];

        data = [repmat(alternativeInfo,nSimulations,1), ...
        simulatedData];
        summary = [ alternativeInfo, summarizedData ];

        %% write data and/or summary into files
        if dataFile
            dataFileID = fopen(dataFileName, 'a');
            formatSpec = ['%3i %3i %3i %3i %3i %3i ' ...
                '%1.2f %1.2f %1.2f %1.2f ' ...
                '%3i %3i %3i %3i %3i %3i ' ...
                '%3i %3i %3i %3i %3i %3i ' ...
                '%3i %3i %3i %3i %3i %3i ' ...
                '%3i %3i %3i %3i %3i %3i\n'];
            fprintf(dataFileID, formatSpec, data');
            fclose(dataFileID);
        end

        if summaryFile
            summaryFileID = fopen(summaryFileName, 'a');
            formatSpec = ['%3i %3i %3i %3i %3i %3i ' ...
```

```
                '%1.2f %1.2f %1.2f %1.2f ' ...
                '%3i %3i %3i %5.0f %5.0f %5.0f ' ...
                '%3i %3i %3i %5.0f %5.0f %5.0f ' ...
                '%3i %3i %3i %5.0f %5.0f %5.0f ' ...
                '%3i %3i %3i %5.0f %5.0f %5.0f\n'];
            fprintf(summaryFileID, formatSpec, summary');
            fclose(summaryFileID);
        end
    end
end

fprintf('\nDone.\n');
```

```
function [ comparisonValues, randomComponent ] = ...
    computecomparisonvalues( ...
    choiceAlternatives, externalReferencePoints )
% COMPUTECOMPARISONVALUES
% compares attribute values of the available
% alternatives with each other and with attribute
% values of the external reference points,
% if applicable.

% check if number of input arguments is one or two
narginchk(1,2);
% extract number of alternatives and attributes
[nAlternatives, nDimensions] = size(choiceAlternatives);
% preallocate space for the comparison values
comparisonValues = ...
    zeros(2*nAlternatives, 2^nAlternatives, nDimensions);
% prepare combinations, source:
% https://www.mathworks.com/matlabcentral/answers/
%   uploaded_files/8064/combn.m
combinations = combn([1 0], nAlternatives);

%% define random component
randomComponent = repmat(combinations',2,1);
% sum columns
columnSum = sum(randomComponent);
% replace zeros by ones (to avoid NaNs due to division by zero)
columnSum(columnSum == 0) = 1;
% normalize (divide random component by column sum)
randomComponent = randomComponent./columnSum;


%% compute comparison values
% if there is only one input value,
%   only internal comparison values are computed
%   if there are two input values,
%   only external comparison values are computed
switch nargin
    case 1
        nReferencePoints = nAlternatives;
        referenceValues = choiceAlternatives;
        for iAlternative = 1:nAlternatives
            focusValues = ...
                repmat(choiceAlternatives(iAlternative,:), ...
                nReferencePoints,1);
            differences = focusValues - referenceValues;
            positiveDifferences = max(differences, 0);
            negativeDifferences = -min(differences, 0);
            for iCombination = (combinations(:,iAlternative) == 1)
                comparisonValues( ...
```

```
                    iAlternative , iCombination , :) ...
                    = combinations(iCombination , :) * ...
                    positiveDifferences;
                comparisonValues( iAlternative + ...
                    nAlternatives , iCombination , :) = ...
                    combinations(iCombination , :) * ...
                    negativeDifferences;
            end
        end
        for iDimension = 1:nDimensions
            % sum columns
            columnSum = sum(comparisonValues(:,:,iDimension));
            % find zeros
            iZeros = (columnSum == 0);
            % replace zeros by ones
            columnSum(iZeros) = 1;
            % replace zero columns by columns from the
            %   normalized random component
            comparisonValues(:,iZeros,iDimension) = ...
                randomComponent(:,iZeros);
            % divide comparison values by 0.5 * column sum
            % (or normalize values separately for positive and
            %   negative comparison values)
            comparisonValues(:,:,iDimension) = ...
                comparisonValues(:,:,iDimension) ./ ...
                (0.5 .* columnSum);
        end
    case 2
        nReferencePoints = size(externalReferencePoints , 1);
        referenceValues = externalReferencePoints;
        for iAlternative = 1:nAlternatives
            focusValues = ...
                repmat(choiceAlternatives(iAlternative,:), ...
                nReferencePoints,1);
            differences = focusValues - referenceValues;
            positiveDifferences = max(differences , 0);
            negativeDifferences = -min(differences , 0);
            for iCombination = (combinations(:,iAlternative) == 1)
                comparisonValues( ...
                    iAlternative , iCombination , :) ...
                    = combinations(1, :) * positiveDifferences;
                comparisonValues( ...
                    iAlternative + nAlternatives , ...
                    iCombination , :) = ...
                    combinations(1, :) * negativeDifferences;
            end
        end
        for iDimension = 1:nDimensions
            % sum columns
            columnSum = sum(comparisonValues(:,:,iDimension));
            % find zeros
            iZeros = (columnSum == 0);
            % replace zeros by ones
            columnSum(iZeros) = 1;
            % replace zero columns by columns from the
            %   normalized random component
            comparisonValues(:,iZeros,iDimension) = ...
                randomComponent(:,iZeros);
            % divide comparison values by column sum
            comparisonValues(:,:,iDimension) = ...
                comparisonValues(:,:,iDimension) ./ columnSum;
        end
```

```
end
```

```
function[ simulatedData ] = simulatechoices( nAlternatives, ...
    nSimulations, probabilities, decisionThreshold )
%SIMULATECHOICES simulates choices from a given choice set with
%   the defined parameter setting.

% preallocate space for simulated data (one row per simulation):
%   nAlternatives columns for choices/eliminations
%   nAlternatives columns for response times
%   2*nAlternatives*nAlternatives columns for
%   counter states at response times
simulatedData = ...
    zeros(nSimulations, (2 * nAlternatives + 2) * nAlternatives);

% prepare probabilities
cumulatedProbabilities = cumsum(probabilities)';
% prepare combinations
%
combinations = combn([1 0], 3);


% define number of steps per simulation
%   set to a high value to make sure that most of the choices are
%   made before that time
nSteps = 10000;

for iSimulation = 1:nSimulations

    % draw random values
    % randomValues = repmat(rand(nSteps, 1),1,2*nAlternatives)
    randomValues = rand(nSteps, 2*nAlternatives);
    % define initial counter states
    counterStates = zeros(1,2*nAlternatives);
    % and initial combinations/probabilities index
    %   (start with all alternatives: row 1)
    iCombination = 1;
    availableAlternatives = combinations(iCombination,:);
    nAvailableAlternatives = nAlternatives;

    upperEndpoints = cumulatedProbabilities(iCombination,:);
    lowerEndpoints = [0, upperEndpoints(1:(end-1))];

    for iStep = 1:nSteps

        % testProbabilities = randomValues(iStep,1); % dependent
        testProbabilities = randomValues(iStep,:); % independent

        counterUpdates = ...
            (lowerEndpoints < testProbabilities) .* ...
            (testProbabilities <= upperEndpoints);

        counterStates = counterStates + counterUpdates;

        testPreferences = counterStates(1:nAlternatives) - ...
            counterStates((nAlternatives+1):end);



        index = find(availableAlternatives);
        permuteIndex = randperm(nAvailableAlternatives);
        testIndex = index(permuteIndex);
```

```
        for iTest = testIndex
            if testPreferences(iTest)>=decisionThreshold
                simulatedData( ...
                    iSimulation, ...
                    nAlternatives - ...
                    nAvailableAlternatives + 1) = ...
                    iTest;
                simulatedData( ...
                    iSimulation, ...
                    2 * nAlternatives - ...
                    nAvailableAlternatives + 1) = ...
                    sum(counterStates);
                simulatedData( ...
                    iSimulation, (2 + 2 * ...
                    (nAlternatives-nAvailableAlternatives)) * ...
                    nAlternatives + 1 : (4 + 2 * ...
                    (nAlternatives-nAvailableAlternatives)) * ...
                    nAlternatives) = ...
                    counterStates;

                availableAlternatives(iTest) = 0;
                iCombination = ...
                    2^nAlternatives - ...
                    availableAlternatives * ...
                    (2.^((nAlternatives-1):-1:0))';
                nAvailableAlternatives = ...
                    nAvailableAlternatives-1;
                upperEndpoints = ...
                    cumulatedProbabilities(iCombination,:);
                lowerEndpoints = ...
                    [0, upperEndpoints(1:(end-1))];
                break

            elseif testPreferences(iTest)<=(-decisionThreshold)
                simulatedData( ...
                    iSimulation,...
                    nAlternatives - ...
                    nAvailableAlternatives + 1) = ...
                    -iTest;
                simulatedData( ...
                    iSimulation, ...
                    2 * nAlternatives - ...
                    nAvailableAlternatives + 1) = ...
                    sum(counterStates);
                simulatedData( ...
                    iSimulation, (2 + 2 * ...
                    (nAlternatives - ...
                    nAvailableAlternatives)) * ...
                    nAlternatives + 1:(4 + 2 * ...
                    (nAlternatives - ...
                    nAvailableAlternatives)) * ...
                    nAlternatives) = ...
                    counterStates;

                availableAlternatives(iTest) = 0;
                iCombination = ...
                    2^nAlternatives - ...
                    availableAlternatives * ...
                    (2.^((nAlternatives-1):-1:0))';
                nAvailableAlternatives = nAvailableAlternatives-1;
                upperEndpoints = ...
```

```
                    cumulatedProbabilities(iCombination,:);
                lowerEndpoints = [0, upperEndpoints(1:(end-1))];
                break

            end
        end

        if nAvailableAlternatives==0
            break
        end
    end
end
```

```
function[ summarizedData ] = summarizedata( ...
    nAlternatives, simulatedData )
% SUMMARIZEDATA counts for each alternative how often it
%   has been chosen in simulatedData

% preallocate space
summarizedData = zeros(1,2*nAlternatives*(nAlternatives+1));

for iAlternative = 1:nAlternatives
    for iPosition = 1:nAlternatives
        if iPosition == 1
            choiceData = ...
                simulatedData(simulatedData(:,1) == ...
                iAlternative,:);
        elseif iPosition == nAlternatives
            choiceData = ...
                simulatedData( ...
                (abs(simulatedData(:,iPosition)) == ...
                iAlternative & ...
                sum(simulatedData(:,1:(iPosition-1))<0,2) == ...
                (iPosition-1)),:);
        else
            choiceData = ...
            simulatedData((simulatedData(:,iPosition) == ...
            iAlternative & ...
            sum(simulatedData(:,1:(iPosition-1))<0,2) == ...
            (iPosition-1)),:);
        end
        nChoices = size(choiceData,1);
        summarizedData( ...
            iAlternative * 2 * nAlternatives + iPosition) = ...
            nChoices;
        summarizedData(iAlternative) = ...
            summarizedData(iAlternative) + nChoices;
        if sum(choiceData(:,nAlternatives+iPosition)) == 0
            meanResponseTime = 0;
        else
            summedResponseTimes = ...
                sum(choiceData(:,nAlternatives+iPosition));
            meanResponseTime = ...
                round(summedResponseTimes/nChoices);
            summarizedData(iAlternative+nAlternatives) = ...
                summarizedData(iAlternative+nAlternatives) + ...
                summedResponseTimes;
        end
        summarizedData( ...
            (1+iAlternative*2)*nAlternatives+iPosition) = ...
            meanResponseTime;
    end
```

```
    if summarizedData(iAlternative + nAlternatives) > 0
        summarizedData(iAlternative+nAlternatives) = ...
            round(summarizedData(iAlternative+nAlternatives) / ...
            summarizedData(iAlternative));
    end
end
```