Choice, Preference, and Utility: Proba...

A. A. J. Marley and M. Regenwetter

New Handbook of Mathematical Psychology

1A.A.J. Marley (corresponding author)
Department of Psychology
University of Victoria
PO Box 3050 STN CSC
Victoria
BC V8W 3P5
Canada
e-mail: ajmarley@uvic.ca

2Centre for the Study of Choice
University of Technology Sydney
PO Box 123 Broadway
Sydney, NSW 2007
Australia

3M. Regenwetter
Department of Psychology
University of Illinois at Urbana-Champaign
603, East Daniel Street
Champaign, IL 61820
USA
e-mail: regenwet@illinois.edu
PART ONE  CHAPTERS

1  Choice, Preference, and Utility: Probabilistic and Deterministic Representations  3
   1.1  General Remarks on the Study of Preference and Choice  4
       1.1.1  Comment on organization and assumed mathematical background knowledge  8
   1.2  Choice Probabilities  9
   1.3  Choice Probabilities for Binary Preference Relations  12
       1.3.1  Binary preference relations  12
       1.3.2  Choice probabilities induced by a single fixed binary preference relation  15
       1.3.3  Choice probabilities for varying or uncertain preferences  20
   1.4  Algebraic Theories and Their Real Valued Representations  31
       1.4.1  Numerical representations of binary preference relations  31
       1.4.2  Representations of simple lexicographic semiorders  32
       1.4.3  Numerical representations of weak orders over gambles  33
       1.4.4  Axiomatizations of representations of gambles  39
       1.4.5  Joint receipts  43
       1.4.6  Parametric forms for utility and weights  44
   1.5  Choice Probabilities for Numerical Representations  51
       1.5.1  Distribution-free random utility representations  51
Contents

1.5.2 Axiomatizations of expected utility representations in probabilistic choice 59
1.5.3 Horse race models of choice and response time 66
1.5.4 Context free linear accumulator models of choice and response time 71
1.5.5 Context dependent models of choice and response time 80
1.5.6 Context dependent models of choice 85
1.6 Discussion, Open Problems, and Future Work 88
References 91
Abstract

Historically, a significant proportion of the research on choice and preference has placed the emphasis on deterministic representations, most notably for utility and uncertainty. A notorious challenge has been the question of adequate representation for variability in behavior, both across decision makers and within a person. Here we emphasize probabilistic representations of choice, preference, and utility, and consider deterministic representations as the special case where all probability mass is concentrated on a single preference, a single utility function, or a single way of making a choice. We focus on developments during the past 10 years and, for the most part, restrict the content to certain or uncertain (multiattribute) options with no temporal or repeated component. Our focus is on theory, with data and model fitting introduced only when such is intimately connected to the theory. A challenge to much classic work on probabilistic choice is the assumption of context independence, which we address by summarizing and partially integrating several recent context dependent models of choice and response time.

1.1 General Remarks on the Study of Preference and Choice

There is a huge theoretical literature on preference and choice, with a corresponding large literature on testing theories. Recent important books, review articles, and handbooks include Barberá et al. (1999, 2004, 2012); Gilboa (2009); Luce (2000); Wakker (2010). We focus on the past 10 years in this chapter and, for the most part, restrict the content to certain or uncertain (multiattribute) options with no temporal or repeated component. Also, to keep the bibliography manageable, we often cite only one or two most recent articles on a topic.

Historically, research on preference and choice has placed an emphasis on deterministic representations, most notably, algebraic models of preference, utility, and choice. A notorious challenge, for over 50 years, has been the question of adequate representation for variability in behavior, both across decision makers and within a person. This formal modeling challenge is sometimes associated with Duncan Luce, whose 1959 choice axiom was a major milestone in probabilistic characterizations of utility. Many other leading scholars have also studied and discussed ways to generalize algebraic models into probability models (see Blavatskyy and Pogrebna, 2010; Stott, 2006; Wilcox, 2008, and their citations). We will emphasize probabilistic representations of preference, utility and choice, and we will consider deterministic representations as the special case where all probability mass is
concentrated on a single preference, a single utility function, or a single way of making a choice.

The focus is on theory, with data and model fitting introduced only when such is intimately connected to the theory. We take this approach mainly because of length restrictions but also because there are numerous works on empirical evaluation and because there are ongoing developments and debates about the appropriate frequentist and/or Bayesian analyses of such data; we cite these empirical works throughout the chapter. Another challenge to much classic work on probabilistic choice is the assumption of context independence, which we now begin to address by summarizing several recent context dependent choice models.

An important distinction in Economics is that of stated versus revealed preference. Stated preference refers to the choices that a person makes in, say, a typical psychology experiment, where frequently the participants are college students who either receive course credit or an hourly payment for participation; sometimes an additional payment is made that is contingent on the person’s choices in the experiment (see the example on ternary choice among gambles, below). Revealed preference refers to the actual choices made by a person in the real world, such as their grocery purchases or their participating, or not, in a weekly office lottery pool. As a result of scanner data, it is increasingly possible to study revealed preference for, say, food items, though much of those data are aggregated over consumers; recently, scanner data have become available for repeated choice occasions of a single individual. However, there remain areas where it is not possible to obtain revealed preferences, such as for new technology that is not yet generally available, or that is too expensive for general uptake (see the example on household renewable energy, below). Also, there are areas, such as medicine, where it is extremely difficult, if not impossible, to study revealed preferences in detail; end-of-life decisions is one such area where important work is being carried out using stated preferences (Flynn et al., 2013).

Lancsar and Swait (2013) summarize the literature on the external validity of stated preference with respect to revealed preference and other measures. We focus on models that are applied in experimental and consumer psychology to stated preference, though they are also frequently applied in economics to revealed preference.

We now present two examples of stated preference tasks, then give pointers to a more complex task involving both aspects of those tasks. The first task concerns stated ternary paired comparisons between gambles (lotteries); the second concerns multiple choice between possible programs for microgeneration of electricity by households; and the more complex task concerns choices
between risky or uncertain multiattribute health states. These examples are closely related to some of the formulations and models presented in Sections 1.3 and 1.4.

Many descriptive, prescriptive, and normative theories of decision making share structural assumption about the nature of individual preferences. Regenwetter and Davis-Stober (2012) investigated choices among pairs of lotteries where decision makers were permitted to express a lack of preference for either lottery (see the screen shot of an experimental trial in Figure 1.1). Their experimental study built on a seminal paper by Tversky (1969) and used lotteries that were contemporary dollar equivalents of lotteries in Tversky (1969), similar to Regenwetter et al. (2011a). The experiment was spread over three sessions of approximately one hour each. At the end of each session, the participant received a $5 flat payment plus played one of his/her chosen cash lotteries from one trial (chosen uniformly at random) for real money. If the respondent had selected ‘indifferent’ on that trial, then one of the two lotteries of that trial was selected, in turn with equal probability, for real play. Regenwetter and Davis-Stober investigated the behavior of 30 participants separately and secured high statistical power by collecting 45 trials per gamble pair and per respondent and by exposing each person to three distinct sets, each containing 10 gamble pairs. They found that variability in choices between and within person could be explained very well by a parsimonious model in which ternary choice probabilities are marginal probabilities of an unknown probability distribution over strict weak orders. Strict weak orders will be introduced in Section 1.3.1 and the model they tested will be discussed, e.g., in Lemma 1.31 and in Theorem 1.58.

Marley and Islam (2012) report the results of a stated preference survey that was conducted to examine trade-offs of features of solar panels for household level electricity generation; this is a potentially very significant energy source as individual households account for one third of all energy consumption in North America. Data for the survey were collected from 298 respondents by Pureprofile (a large online panel provider in Australia, North-America, and other countries), with study participants screened based on owning a house in Ontario, Canada. The attributes and attribute-levels studied were chosen based on extensive research, reviews of product/retailer ads/claims, websites, pilot study, etc. Each respondent was shown the list
of attributes and their ranges before the choice task, and informed that the
stated savings in energy cost and carbon emission were on an annual basis.

Figure 1.2 shows a sample screen shot of the first of the (20) choice sets
in the survey: each choice set of 4 options was constructed using recently
developed design theory that allows the researcher to efficiently “cover” the
set of possible options. The response task was framed as a sequential choice
process, with respondents instructed to choose the most preferred alternative
out of four (Q1), then the least preferred out of the remaining three (Q2),
and, finally, the most preferred out of the remaining two (Q3). Respondents
were also asked to indicate (Q4, Fig. 1) if they would choose none of the
4 options. This method of preference elicitation provides more information
than, say, a best choice in each choice set. Each time a respondent selected a
profile, that profile disappeared from the screen so as to restrict the respon-
dent’s next choice to the remaining profiles. The models studied are closely
related to those presented in Section 1.5, and gave parameter estimates that
had reasonable properties - for instance, the estimates showed that a shorter
payback time is preferred and that a monetary grant is preferred to a sales
tax refund.

Our third example combines the structures of the first two examples. Here,
each outcome in an option is a health state, which might be full health,
death, or some other health state defined in terms of various attributes;
and an option consists of one or more such outcomes with an associated
success (survival) rate and/or life expectancy. Figure 1.3 shows a pair of
such options that might appear in a stated preference (choice) task, where
the participant has to state which option is preferred. Matching tasks are also
studied in this domain: for example, in the time-tradeoff (TTO) procedure,
a respondent assigns a life expectancy to a reference health state (typically
full health) such that she is indifferent between that state and a target
health state with a specified life expectancy; and in the standard gamble
(SG) procedure, a respondent assigns a survival rate to a reference health
state (typically full health) with a specified life expectancy such that she
is indifferent between that state and a target health state with the same
life expectancy for sure (Bleichrodt and Pinto, 2006, review this literature).
Most of the theories in this domain are algebraic, paralleling those in Section
1.4 but with multiattribute outcomes; however, recent work combines such
representations with probabilistic models of the kind presented in Section 1.5 (Arons and Krabbe, 2013)

1.1.1 Comment on organization and assumed mathematical background knowledge

Different sections and subsections of this chapter rely on very different mathematical formalism and modeling tools. Section 1.2 discusses various classes of probability representations for response data using basic concepts of probability. These representations correspond to models of empirical data generating processes, on which the other work builds. Section 1.3.1 reviews binary preference relations, relying primarily on set theory and combinatorics, and Section 1.3.2 discusses a range of probability models of choice behavior. In particular, Section 1.3.2 discusses models of probabilistic choice when individual preference is deterministic and variability is caused by, for instance, random errors in responding. It relies on set theory, some combinatorics and basic probability theory. Section 1.3.3 discusses models in which preferences themselves are probabilistic. Besides set theory, combinatorics and basic probability, this subsection relies on basic tools of convex geometry for characterizing order-constraints on choice probabilities that are induced by probabilistic preferences. Section 1.4 studies algebraic structures for choice options (mainly gambles) and presents conditions under which real-valued (utility) representations capture those structures. It relies on set theory and general ideas about axiomatic structures. Section 1.5 proceeds to consider real-valued random variable representations to model random utility models of a variety of different kinds. It relies on basic concepts from real analysis, probability theory (including random variables), and elementary ideas about stochastic processes. The more abstract random utility models are directly connected to the convex geometry concepts of Section 1.3.3, whereas other models rely on parametric families of distributions defined on the real numbers; the latter models are well studied in consumer choice, statistics and econometrics.

All in all, we expect that readers not familiar with all the relevant areas of mathematics may find some parts of the chapter harder to follow than others. We assume familiarity with basic mathematical concepts in all areas. Many sections are interdependent. We cross-reference relevant and related results.
across sections in an effort to allow a reader to skip sections or subsections that she or he finds too specialized and then go back and only read parts that are referenced in parts that they want to know about in depth. When we state a result without an associated citation (usually as a Lemma), its proof is immediate.

1.2 Choice Probabilities

We will define a general concept of response probabilities to capture a variety of scenarios in which a decision maker is offered a subset $X$ of choice alternatives from a finite master set $\mathcal{A}$ with $|\mathcal{A}| \geq 2$, and is asked to provide information about the alternatives in that set $X$. The definition includes deterministic choice (and binary relations, as well as algebraic representations) as a special case. The information may range from choosing the better alternative in various pairs of alternatives, to indicating the decision maker’s favorite or least favorite alternative in offered sets of various sizes, to providing a complete rank ordering of the offered alternatives from most favorite to least favorite, etc. Different empirical paradigms differ in the types of choice sets $X$ that they offer and in the responses they solicit. The purpose of the first definition is to define response probabilities in a general fashion that can accommodate a variety of such paradigms. We denote by $P_{(X,R_X)}$ a probability distribution governing the response when a person is offered a set $X$ of choice alternatives and is asked to indicate the choice alternatives ranked at the rank positions listed in an index set $R_X$. For instance, if $R_X = \{1, 5\}$ and $X = \{a, b, c, d, e\}$, then the person is asked to report the option they prefer to all others (i.e., the ‘best’), as well as the option to which they prefer all others (i.e., the ‘worst’). The distribution $P_{(\{a,b,c,d,e\},\{1,5\})}$ describes the probabilities of all possible ‘best-worst’ responses when asked to choose from $\{a, b, c, d, e\}$.

Examples of possible queries are: all possible ways of identifying a single best alternative (if $R_X = \{1\}$); or all possible ways of identifying a single best and a single worst alternative (if $R_X = \{1, |X|\}$); all possible preference rankings (if $R_X = \{1, 2, \ldots, |X|\}$). Space limitations prevent us from presenting detailed results on models of ranking behavior; however, we do make extensive use of the concept of a “generalized rank” (Def. 1.9). Similarly, with the exception of “ternary paired comparisons,” we will not consider cases in which a person states indifference among choice options. The notation for ternary paired comparisons will be introduced when we discuss the relevant models.

\footnote{We use the standard convention that $\{1, 2, \ldots, 2\}$ denotes the set $\{1, 2\}$.}
Definition 1.1  Consider a finite master set $\mathcal{A}$ of two or more choice alternatives. Let $\Theta$ be a set containing elements $\theta$ of the form $\theta = (X, R_X)$ with each $X \subseteq \mathcal{A}$ indicating an available choice set and each $R_X \subseteq \{1, 2, \ldots, |X|\}$ indicating the rank positions queried with respect to the set $X$. Let $\Pi(R_X, X)$ denote the collection of all one-to-one mappings from $R_X$ into $X$. A collection $P = (P_\theta)_{\theta \in \Theta}$ is called a collection of response probability distributions if every $P_\theta$ with $\theta = (X, R_X)$ is a probability distribution over $\Pi(R_X, X)$, i.e.,

$$0 \leq P_{(X, R_X)}(\pi) \leq 1 \quad (\forall \pi \in \Pi(R_X, X))$$

and

$$\sum_{\pi \in \Pi(R_X, X)} P_{(X, R_X)}(\pi) = 1.$$

Each $\theta = (X, R_X)$ can be thought as a possible query to a decision maker: “Among the objects in $X$, indicate the objects at rank positions $R_X$ within $X$” and we can write $\Pi(R_X, X)$ more generically as $\Pi(\theta)$. The collection $\Pi(\theta)$ denotes all permissible responses that the decision maker may give to the query $\theta$. The response probability distributions (1.1)-(1.2) can also be stated more generically as,

$$0 \leq P_\theta(\pi) \leq 1 \quad (\forall \pi \in \Pi(\theta))$$

and

$$\sum_{\pi \in \Pi(\theta)} P_\theta(\pi) = 1.$$

Notice, in particular, that, for any query $\theta$, any response $\pi \in \Pi(\theta)$ satisfies $\pi(i) = x$ if and only if the response $\pi$ assigns $x$ rank position $i$. We now consider various special cases of this general framework, with $\mathcal{A}$ fixed, and $|\mathcal{A}| \geq 2$. Our first special case of interest concentrates on situations where the decision maker is offered two distinct choice alternatives $x$ and $y$ at a time and we wish to model the probability that the decision maker indicates, say, $x$ as the preferred alternative among the two.

Definition 1.2  A collection of response probabilities $(P_\theta)_{\theta \in \Theta}$ is a collection of binary choice probabilities if every $\theta \in \Theta$ is of the form $\theta = (X, \{1\})$, with $|X| = 2$. A complete collection of binary choice probabilities on $\mathcal{A}$ is a collection $(P_\theta)_{\theta \in \Theta}$ of binary choice probabilities in which every two-element set of alternatives is an available choice set: $(\{x, y\}, \{1\}) \in \Theta, \forall x, y \in \mathcal{A}$, with $x \neq y$. We also write $P_{xy}$ instead of $P_{(\{x, y\}, \{1\})}(\pi)$, where $\pi(1) = x$. $P_{xy}$ is the probability that a decision maker chooses $x$ over $y$ when instructed to choose the preferred among two choice options $x \neq y$ in $\mathcal{A}$.
The notation $P_{xy}$ is frequently used in the literature, as is $p(x, y)$.

We now move to more general situations, containing binary choice probabilities as special cases, where the decision maker is offered sets $X$ of potentially varying sizes and asked to indicate either the best, or the worst choice alternative among the available options in $X$, or both.

**Definition 1.3** Consider a collection of response probabilities $(P_\theta)_{\theta \in \Theta}$. The collection $(P_\theta)_{\theta \in \Theta}$ is a **collection of best choice probabilities** if every $\theta \in \Theta$ is of the form $\theta = (X, \{1\})$, with $|X| \geq 2$. The collection $(P_\theta)_{\theta \in \Theta}$ is a **collection of worst choice probabilities** if every $\theta \in \Theta$ is of the form $\theta = (X, \{|X|\})$, with $|X| \geq 2$. The collection $(P_\theta)_{\theta \in \Theta}$ is a **collection of best-worst choice probabilities** if every $\theta \in \Theta$ is of the form $\theta = (X, \{1, |X|\})$, with $|X| \geq 2$. A **complete collection of best choice probabilities** is a collection of best choice probabilities with $(X, \{1\}) \in \Theta$, for every subset $X \subseteq \mathcal{A}$ with $|X| \geq 2$. Likewise, in a complete collection of worst choice probabilities, $(X, \{|X|\}) \in \Theta$, for every subset $X \subseteq \mathcal{A}$ with $|X| \geq 2$ and, in a complete collection of best-worst choice probabilities, $(X, \{1, |X|\}) \in \Theta$, for every subset $X \subseteq \mathcal{A}$ with $|X| \geq 2$.

For the special cases that we considered in Definition 1.3, we can simplify the notation to make it more mnemonic and to agree with other publications. We write $B_X(x)$ to denote $P_{(X,\{1\})}(\pi)$, where $\pi(1) = x$, the probability that the decision maker identifies $x \in X$ as the best alternative out of subset $X \subseteq \mathcal{A}$. We will also write $W_X(x)$ to denote $P_{(X,\{|X|\})}(\pi)$, where $\pi(|X|) = x$, the probability that the decision maker identifies $x \in X$ as the worst alternative out of subset $X \subseteq \mathcal{A}$ (that is, an alternative ranked in position $|X|$ in $X$). We will also write $BW_X(x, y)$ to denote $P_{(X,\{1,|X|\})}(\pi)$, where $\pi(1) = x, \pi(|X|) = y$, the probability that the decision maker identifies $x$ as the best alternative and $y \neq x$ as the worst alternative out of subset $X \subseteq \mathcal{A}$.

**Definition 1.4** Let $n \geq 2$ be an integer. An $n$-dimensional unit hypercube is the set $[0, 1]^n$ of all mappings from $\{1, 2, \ldots, n\}$ into the real-valued interval $[0, 1]$. A **vertex** of an $n$-dimensional unit hypercube is a $0/1$-valued vector with $n$ coordinates, i.e., a mapping from $\{1, 2, \ldots, n\}$ into the set $\{0, 1\}$.

For example, a two-dimensional unit-hypercube $[0, 1]^2$ is a square with sides of length 1, and it’s vertices are the vectors $(0, 0), (0, 1), (1, 0), \text{ and } (1, 1)$. A three-dimensional unit-hypercube $[0, 1]^3$ is a cube with sides of length one. Its vertices are the eight ‘corners’ of the cube. Geometrically in higher dimensions, a vertex of a hypercube is a ‘corner’ of the hypercube.

**Remark** Suppose that $|\mathcal{A}| = n \geq 2$. A complete collection of binary choice
probabilities is contained in a \( n(n - 1) \) dimensional unit hypercube. A complete collection of binary choice probabilities can be reduced to a set of nonredundant parameters by retaining, for each \( x \neq y \) exactly one of \( P_{xy} \) or \( P_{yx} \). These parameters form exactly an \( \binom{n}{2} \)-dimensional unit hypercube. A complete collection of best choice or worst choice probabilities is contained in a unit hypercube of dimension \( \sum_{k=2}^{n} \binom{n}{k} k \) and, after we drop redundant probabilities, the remaining parameters form a unit hypercube of dimension \( \sum_{k=2}^{n} \binom{n}{k} (k - 1) \). A complete collection of best-worst multiple choice probabilities is contained in a unit hypercube of dimension \( \sum_{k=2}^{n} \binom{n}{k} k(k - 1) \) and, after we drop redundant probabilities, the remaining parameters form a unit hypercube of dimension \( \sum_{k=2}^{n} \binom{n}{k} [k(k - 1) - 1] \). The vertices of these hypercubes are the 0/1 vectors presenting collections of degenerate probabilities, and these can be thought of as the deterministic special cases where each deterministic choice translates into a choice probability of zero or one.

We will later see how various algebraic models of preference or even utility can be embedded in the space of choice probabilities by representing deterministic preferences as vertices of these hypercubes, and how some probabilistic models take the form of convex polytopes formed by the convex hull of those vertices (these terms will be defined later).

1.3 Choice Probabilities for Binary Preference Relations

1.3.1 Binary preference relations

We start with the classical definitions of several classes of binary relations, as they have been studied for more than a half century.

Except where explicitly stated - for instance, in the definition of a “complete” binary relation - the elements \( x, y, z \) need not be distinct.

Definition 1.5 A binary relation \( \succ \) on a set of choice alternatives \( \mathcal{A} \) is a subset of the Cartesian product of \( \mathcal{A} \) with itself, i.e., \( \succ \subseteq \mathcal{A} \times \mathcal{A} \). It is also standard to write \( (x, y) \in \succ \) as \( x \succ y \). We write \( \neg [x \succ y] \) to denote that \( x \succ y \) does not hold. We use \( \land \) for logical AND and \( \lor \) for logical OR. A binary relation \( \succ \) on \( \mathcal{A} \) is

- **complete** if \( [x \succ y] \lor [y \succ x] \) \( \forall x, y \in \mathcal{A} \) with \( x \neq y \),
- **asymmetric** if \( [x \succ y] \Rightarrow \neg [y \succ x] \) \( \forall x, y \in \mathcal{A} \),
- **transitive** if \( [x \succ y] \land [y \succ z] \Rightarrow [x \succ z] \) \( \forall x, y, z \in \mathcal{A} \),
- **negatively transitive** if \( \neg [x \succ y] \land \neg [y \succ z] \Rightarrow \neg [x \succ z] \) \( \forall x, y, z \in \mathcal{A} \).
We will sometimes use the symbol \(\succsim\) for a binary relation that has a specific property defined next.

**Definition 1.6** A binary relation \(\succsim\) on a set of choice alternatives \(A\) is strongly complete if
\[
[x \succsim y] \lor [y \succsim x] \quad (\forall x, y \in A).
\]

There are a number of binary relations that satisfy some, but not others, of these axioms. We now review some of the most prominent relations.

**Definition 1.7** A strict partial order is an asymmetric and transitive binary relation. We denote the collection of all strict partial orders on \(A\) by \(SPO_A\). An interval order is a strict partial order \(\succ\) with the property
\[
[w \succ x] \land [y \succ z] \Rightarrow [w \succ z] \lor [y \succ x] \quad (\forall w, x, y, z \in A).
\]
We denote the set of all interval orders on \(A\) by \(IO_A\). A semiorder is an interval order \(\succ\) with the property
\[
[w \succ x] \land [x \succ y] \Rightarrow [w \succ z] \lor [z \succ y] \quad (\forall w, x, y, z \in A).
\]
The collection of all semiorders on \(A\) is denoted by \(SO_A\). A strict weak order is an asymmetric and negatively transitive binary relation. The collection of all strict weak orders on \(A\) is denoted by \(SWO_A\). A strict linear order is a transitive, asymmetric, and complete binary relation. The collection of all strict linear orders on \(A\) is denoted by \(SLO_A\). A weak order is a strongly complete and transitive binary relation. The collection of all weak orders on \(A\) is denoted by \(WO_A\).

Section 1.4.1 reviews the standard numerical representations of the binary relations of Definition 1.7 in Theorem 1.32.

Next, we consider ways in which two or more binary relations can be combined, e.g., to form a new binary relation. One way to combine two binary relations is through a lexicographic process. Another is to make binary relations context-dependent.

**Definition 1.8** A lexicographic binary relation on \(A\) is an ordered list of binary relations \(\succ_i\), \(1 \leq i \leq k\) on \(A\).

For example, a lexicographic semiorder on \(A\) is an ordered list of semiorders \(\succ_i\), \(1 \leq i \leq k\) on \(A\). A lexicographic binary relation can also be reduced down to a single binary relation \(\succ\), say, by setting \(x \succ y\) if \(\exists i\) such that \(x \succ_i y\) and, \(\forall j < i : \neg [x \succ_j y]\) and \(\neg [y \succ_j x]\). In other words, \((x, y)\) belongs to \(\succ\) if and only if the first semiorder, \(\succ_i\), in which either \(x \succ_i y\) or \(y \succ_i x\), actually satisfies \(x \succ_i y\).
In this fashion, we can, for instance, construct a binary relation
\[\succ = \{(a, c), (d, a), (e, a), (d, b), (e, b), (d, c), (e, c), (e, d)\}\]
(1.3) on \(A = \{a, b, c, d, e\}\), that forms a lexicographic semiorder derived from the following two semiorders
\[\succ_1 = \{(e, d), (e, c), (e, b), (e, a), (d, c), (d, b), (d, a)\},\]
\[\succ_2 = \{(a, c), (a, d), (a, e), (b, d), (b, e), (c, d), (c, e), (d, e)\}.\]

We now introduce the concept of generalized ranks (Regenwetter and Rykhlevskaia, 2004) that we need in the next section. This definition generalizes the concept of ranking from strict linear orders to general finite binary relations in that it assigns each object \(x\) in a finite set \(A\) a rank with respect to any binary preference relation \(\succ\) on that finite set \(A\). This will be useful when modeling responses to queries in which decision makers are asked to report which object is ranked at one or more rank positions within an available set \(X\), say the best (Rank 1) and worst (Rank \(|X|\)) object in \(X\).

**Definition 1.9** Let \(X \subseteq A\) and let \(\succ \subseteq A \times A\) be a binary relation on \(A\). The generalized rank, \(\text{Rank}_{X,\succ}(x)\), of \(x\) with respect to \(X\) and \(\succ\) is given by
\[
\text{Rank}_{X,\succ}(x) = \frac{|X| + 1 + |\{a \in X : a \succ x\}| - |\{b \in X : x \succ b\}|}{2}.
\]
When \(X = A\) we also write \(\text{Rank}_{\succ}(x)\) instead of \(\text{Rank}_{A,\succ}(x)\).

**Lemma 1.10** Generalized ranks are multiples of \(\frac{1}{2}\). For any binary relation \(\succ\) on \(A\), \(\text{Rank}_{X,\succ}(b) = 1\) if and only if \(b \succ x, \forall x \in X - \{b\}\) and \(\text{Rank}_{X,\succ}(w) = |X|\) if and only if \(x \succ w, \forall x \in X - \{w\}\). The object \(b\) is the (unambiguous) “best” and the object \(w\) is the (unambiguous) “worst” in \(X\). If \(\succ\) is a strict linear order on \(A\), then the generalized ranks of all \(x \in A\) are integers and reduce to the usual ranks associated with strict linear orders.

For example, consider the lexicographic semiorder \(\succ\) in Eq. 1.3.
\[\text{Rank}_{\succ}(a) = 3.5, \text{Rank}_{\succ}(b) = 4, \text{Rank}_{\succ}(c) = 4.5, \text{Rank}_{\succ}(d) = 2, \text{Rank}_{\succ}(e) = 1.\]

The following lemma is useful for many of the models we discuss next.

**Lemma 1.11** Let \(A\) be a finite set, \(\succ\) a binary relation on \(A\), \(X \subseteq A\), and \(R_X \subseteq \{1, 2, \ldots, |X|\}\). For any \(\pi \in \Pi(R_X, X)\) we have, for any \(x \in X\),
\[
\pi(i) = x \Leftrightarrow \text{Rank}_{X,\succ}(x) = i,
\]
and thus:
\[
\text{Rank}_{X,\succ}(\pi(i)) = i.
\]
If \( \succ \) is a binary relation on a finite set \( A \), and \( X \subseteq A \) a subset of \( A \) with \( |X| = n \geq 2 \), then there may be \( i \in \{1, 2, \ldots, n\} \) such that no element of \( X \) is ranked at position \( i \) in \( X \), i.e., \( \text{Rank}_{X,\succ}(x) = i \) may not hold for any \( x \in X \). Hence, when we query a respondent to indicate what object is at rank \( i \) in \( \succ \) with respect to \( X \), we need to assume that such an object exists. Hence, we often state the assumption that \( \forall i \in R_X, \exists x \in X, \text{Rank}_{X,\succ}(x) = i \). Whenever \( \succ \) is a strict linear order and \( i \) is integer-valued, this requirement is automatically met, and hence need not be stated.

We now proceed to discuss various kinds of probabilistic choice models that differ from each other regarding whether or not they model a fixed preference or varying preferences, and whether they model preferences as binary relations or, instead, are based on numerical utilities and/or utility functions. We start with the case of preferences that are binary relations.

### 1.3.2 Choice probabilities induced by a single fixed binary preference relation

The basic notion behind a constant error\(^2\) model for a binary preference relation is that a decision maker has a single “correct” response determined by a single fixed preference relation for each choice s/he needs to make, and that there is a constant probability \( \zeta \) of making an error in any given choice. Many authors also refer to this type of model as a tremble model. Whenever the decision maker is answering a question, there is a chance that she ‘trembles’ and, say, presses the wrong key, or clicks the wrong icon on a screen. Likewise, a constant correct model has a constant probability \( \nu \) of choosing correctly on any given choice. The natural interpretation of a constant correct model is that the decision maker has some fixed probability of being able to fully access her preference, and none of the erroneous responses are more likely than any of the others. A correct choice means that the “best” alternative has generalized rank 1, the “worst” alternative in \( X \) has generalized rank \( |X| \), etc. An incorrect choice occurs if one or more of the assignments are wrong - for example, in a best-worst choice, the response is incorrect if either the best or the worst or both alternatives are selected incorrectly. We will consider a more detailed example below.

For any available set \( X \) and any index set \( R_X \) of rank positions in \( X \) that are queried in query \( \theta = (X, R_X) \), recall that the collection of permissible responses to \( \theta \) is denoted as \( \Pi(\theta) \), which denotes the collection of all one-to-

---

\(^2\) To be precise, one should refer to a constant error probability or rate. We use the shorter terminology for convenience and to be consistent with the literature.
one mappings from $R_X$ into $X$. Hence, $|\Pi(\theta)|$ denotes the number of different permissible responses that are possible for query $\theta$.

Note that the permissible values of $\zeta$ in the following definition depend on $\Theta$.

**Definition 1.12** Consider a collection $P = (P_{\theta})_{\theta \in \Theta}$ of response probability distributions and a fixed binary relation $\succ$ that has the property that $\forall \theta = (X, R_X) \in \Theta$, and $\forall i \in R_X$, $\exists x \in X$, $\text{Rank}_{X,\succ}(x) = i$. $P$ satisfies a **constant error model** (with regard to $\succ$) with a constant error probability $\zeta$ if the error probability $\zeta < \min_{\theta \in \Theta} \frac{1}{|\Pi(\theta)| - 1}$ and $\forall \theta = (X, R_X) \in \Theta$, $\forall \pi \in \Pi(\theta)$

$$P_{\theta}(\pi) = \begin{cases} \zeta & \text{if } \text{Rank}_{X,\succ}(\pi(i)) \neq i, \text{for some } i \in R_X, \\ 1 - \zeta \times (|\Pi(\theta)| - 1) & \text{if } \text{Rank}_{X,\succ}(\pi(i)) = i, \forall i \in R_X. \end{cases}$$

When the respondent provides the correct object in $X$ for each rank position $i \in R_X$ that is queried, then the respondent gives a correct answer. The remaining $|\Pi(\theta)| - 1$ possible answers are incorrect because they match at least one rank position $i \in R_X$ with an incorrect member of $X$, i.e., an element that is not actually at rank $i$ in $X$. Each incorrect response has equal probability $\zeta$.

For example, consider the strict linear order on $A = \{a, b, \ldots, y, z\}$ given by the alphabetic ordering of letters, that is,

$$\succ = \{(a, b), (a, c), (a, d), \ldots (a, z), (b, c), (b, d), \ldots (y, z)\}.$$ 

A constant error model of best-worst choice for choice sets $X = \{a, f, g, w\}$ and $X' = \{b, c, d\}$ with this true preference $\succ$ states that there is an error probability $\zeta < \frac{1}{17} = \min \left(\frac{1}{12-1}, \frac{1}{5-1}\right)$, and

- $BW_X(w, a) = BW_X(a, g) = BW_X(g, a) = \cdots = BW_X(w, g) = \zeta$,
- $BW_X(a, w) = 1 - 11\zeta$,
- $BW_{X'}(b, c) = BW_{X'}(c, b) = BW_{X'}(c, d) = BW_{X'}(d, b) = BW_{X'}(d, c) = \zeta$,
- $BW_{X'}(b, d) = 1 - 5\zeta$.

**Definition 1.13** Consider a collection $P = (P_{\theta})_{\theta \in \Theta}$ of response probability distributions and a fixed binary relation $\succ$ that has the property that $\forall \theta = (X, R_X) \in \Theta$, and $\forall i \in R_X$, $\exists x \in X$, $\text{Rank}_{X,\succ}(x) = i$. $P$ satisfies a **constant correct model** with a constant probability $\nu$ of giving a correct response, with $0 \leq \nu \leq 1$, if $\forall \theta = (X, R_X) \in \Theta$, $\forall \pi \in \Pi(\theta)$,

$$P_{\theta}(\pi) = \begin{cases} \nu \frac{1 - \nu}{|\Pi(\theta)| - 1} & \text{if } \text{Rank}_{X,\succ}(\pi(i)) = i, \forall i \in R_X, \\ \frac{1 - \nu}{|\Pi(\theta)| - 1} & \text{if } \text{Rank}_{X,\succ}(\pi(i)) \neq i, \text{for some } i \in R_X. \end{cases}$$
For example, consider again the strict linear order on $A = \{a, b, \ldots, y, z\}$ that is given by the alphabetic ordering of letters,

$\succ = \{(a, b), (a, c), (a, d), \ldots (a, z), (b, c), (b, d), \ldots (y, z)\}$.

A constant correct model of best-worst choice for choice sets $X = \{a, f, g, w\}$ and $X' = \{b, c, d\}$ states that

$BW_X(a, w) = \nu, \quad BW_X(w, a) = BW_X(g, a) = BW_X(g, c) = \cdots = BW_X(g, z) = \frac{1 - \nu}{11}$,

$BW_{X'}(b, d) = \nu, \quad BW_{X'}(b, c) = BW_{X'}(c, b) = BW_{X'}(c, d) = BW_{X'}(d, b) = BW_{X'}(d, c) = \frac{1 - \nu}{5}$.

Comparing this example with the earlier constant error model example, we see that the two models differ in general, but there are natural circumstances in which the two models predict the same response probabilities.

**Lemma 1.14** Constant error models with $\zeta = 0$ are the same as constant correct models with $\nu = 1$. When, for each $(X, R_X), (X', R_{X'})$ in $\Theta$, we have $|X| = |X'|$ and $R_X = R_{X'}$, then a constant error model can be rewritten as constant correct model and vice-versa by setting $\zeta = \frac{1 - \nu}{|\Pi(R_X, X)| - 1}$. For example, for a complete collection of binary choice probabilities, a constant error model with error rate $\zeta$ is a constant correct model with correct rate $\nu = 1 - \zeta$.

Yet another simple approach is to assume that the person either accesses his/her preference successfully, or, otherwise, randomly picks a response using a uniform distribution over permissible responses.

**Definition 1.15** Consider a collection $P = (P_\theta)_{\theta \in \Theta}$ of response probability distributions and a fixed binary relation $\succ$ that has the property that $\forall \theta = (X, R_X) \in \Theta$, and $\forall i \in R_X$, $\exists x \in X$, $\text{Rank}_{X, \succ}(x) = i$. $P$ satisfies a uniform guessing model with probability $\gamma$ of guessing a uniformly distributed response among permissible responses and probability $1 - \gamma$ of reporting the true preference $\succ$ without guessing, if $\forall \theta = (X, R_X) \in \Theta$, $\forall \pi \in \Pi(\theta)$,

$$P_\theta(\pi) = \begin{cases} \gamma \times \frac{1}{|\Pi(\theta)|} & \text{if } \text{Rank}_{X, \succ}(\pi(i)) \neq i, \text{for some } i \in R_X, \\ 1 - \gamma \times \frac{|\Pi(\theta)| - 1}{|\Pi(\theta)|} & \text{if } \text{Rank}_{X, \succ}(\pi(i)) = i, \forall i \in R_X. \end{cases}$$

Historically, probabilistic choice models have been viewed as extensions of ‘core’ algebraic models. We emphasize that this makes deterministic models special cases of probabilistic ones. We can embed deterministic preferences
into choice probability models as special cases by considering the constant error and constant correct models where a person never trembles. Here, a deterministic, i.e., algebraic, model becomes a degenerate special case of a model of response probabilities.

We now consider more generally and formally how deterministic models are special cases of probabilistic ones. Given a binary preference relation $\succ$ on $\mathcal{A}$, a constant error model for $\succ$ with $\zeta = 0$, a constant correct model of $\succ$ with $\nu = 1$, and a uniform guessing model with $\gamma = 0$ are all the same and called a vertex representation of $\succ$ (defined next).

**Definition 1.16** Consider a finite set $\mathcal{A}$ of two or more choice alternatives and a collection $\Theta$ of elements of the form $\theta = (X, R_X)$ with each $X \subseteq \mathcal{A}$, and each $R_X \subseteq \{1, 2, \ldots, |X|\}$. Consider a fixed binary relation $\succ$ on $\mathcal{A}$ that has the property that $\forall (X, R_X) \in \Theta$, and $\forall i \in R_X, \exists x \in X, \text{Rank}_{X, \succ}(x) = i$.

A vertex representation $V_{\Theta}(\succ)$ of $\succ$ with respect to $\Theta$ is defined by the following collection of 0/1-coordinates: $\forall (X, \pi) \in \Pi(R_X, X)$,

$$V_{(X, \pi)}(\succ) = \begin{cases} 1 & \text{if } \text{Rank}_{X, \succ}(\pi(i)) = i, \forall i \in R_X, \\ 0 & \text{otherwise}. \end{cases}$$ \hfill (1.4)

The vertex representation in the definition uses a distinct coordinate for each $(X, \pi)$. We also call this a full-dimensional representation. In many cases it is natural to drop some of these coordinates because of redundancies.

There are several cases of special interest, for which we can simplify the notation substantially. In the vertex representation for binary choice with a fixed strict linear order $\succ \in \mathcal{SLO}_A$ one considers $X = \{x, y\}, R_X = \{1\}$, hence there are only two possible $\pi \in \Pi(R_X, X)$, namely

$$\pi: \begin{cases} 1 \mapsto x \\ 2 \mapsto y \end{cases}, \text{ and } \pi': \begin{cases} 1 \mapsto y \\ 2 \mapsto x \end{cases}.$$ 

Hence, with this choice of $\pi$, the vertex representation (1.4) for binary choice and a linear order preference $\succ$ has the following 0/1-coordinates:

$$V_{\{x, y\}, \pi}(\succ) = \begin{cases} 1 & \text{if } \text{Rank}_{\{x, y\}, \succ}(x) = 1, \text{Rank}_{\{x, y\}, \succ}(y) = 2, \\ 0 & \text{if } \text{Rank}_{\{x, y\}, \succ}(x) = 2, \text{Rank}_{\{x, y\}, \succ}(y) = 1, \\ 1 & \text{if } x \succ y, \\ 0 & \text{if } y \succ x. \end{cases}$$

Therefore, we rewrite the vertex representation of a preference $\succ \in \mathcal{SLO}_A$ for binary choice as, $\forall x, y \in \mathcal{A}, x \neq y$ with $(\{x, y\}, \{1\}) \in \Theta$,

$$V_{xy}(\succ) = \begin{cases} 1 & \text{if } x \succ y, \\ 0 & \text{if } y \succ x, \end{cases}, \quad V_{yx}(\succ) = \begin{cases} 1 & \text{if } y \succ x, \\ 0 & \text{if } x \succ y. \end{cases}$$
Because $V_{xy}(\succ) = 1 - V_{yx}(\succ)$, it is standard in applications to use only one coordinate per pair $x, y$.

A second case of special interest is the vertex representation for best choice with a fixed binary relation $\succ$ with the property that for every $X$ of interest, $\exists x \in X, \text{Rank}_{X,\succ}(x) = 1$. Here, $R_X = \{1\}$ and we can rewrite $V_{(X,\pi)}(\succ)$ for $\pi : 1 \mapsto x$ more mnemonically as a “vertex for best choice,” $VB(\succ)$, with coordinates

$$VB_{(X,x)}(\succ) = \begin{cases} 1 & \text{if } \text{Rank}_{X,\succ}(x) = 1, \\ 0 & \text{if } \text{Rank}_{X,\succ}(x) \neq 1. \end{cases}$$

In general, this gives $|X|$ many 0/1-coordinates per set $X$ under consideration. Since $\sum_{x \in X} VB_{(X,x)}(\succ) = 1$, for each each $X$, it is standard in vertex representations of best choice to leave out one coordinate for each set $X$.

A third case is the vertex representation for worst choice with a fixed binary relation $\succ$ that has the property that for every $X$ under consideration, $\exists x \in X, \text{Rank}_{X,\succ}(x) = |X|$. We can also write this vertex representation more mnemonically as a “vertex for worst choice,” $VW(\succ)$, with coordinates

$$VW_{(X,x)}(\succ) = \begin{cases} 1 & \text{if } \text{Rank}_{X,\succ}(x) = |X|, \\ 0 & \text{if } \text{Rank}_{X,\succ}(x) \neq |X|. \end{cases}$$

Since $\sum_{x \in X} VW_{(X,x)}(\succ) = 1$, we can also leave out one of these 0/1-coordinates per set $X$.

Likewise, we write the vertex representation of best-worst choice with a fixed binary relation $\succ$ that has the property that for every $X$ under consideration, $\exists (x,y) \in X \times X, \text{Rank}_{X,\succ}(x) = 1, \text{Rank}_{X,\succ}(y) = |X|$, as

$$V BW_{(X,x,y)}(\succ) = \begin{cases} 1 & \text{if } \text{Rank}_{X,\succ}(x) = 1 \land \text{Rank}_{X,\succ}(y) = |X|, \\ 0 & \text{if } \text{Rank}_{X,\succ}(x) \neq 1 \lor \text{Rank}_{X,\succ}(y) \neq |X|. \end{cases}$$

Since $\sum_{x,y \in X} V BW_{(X,x,y)}(\succ) = 1$, we can also leave out one of these 0/1-coordinates per set $X$.

**Lemma 1.17** The coordinates of a vertex representation form the coordinates of a vertex of a unit hypercube. In a full-dimensional representation, each such coordinate is a degenerate probability that equals zero or one, representing the deterministic preference $\succ$ as a special case of a constant error model with error rate $\zeta = 0$, a constant correct model with correct choice rate $\nu = 1$, or a uniform guessing model with $\gamma = 0$.

Using the vertex representations, we can generalize constant error probability and constant correct choice probability models by dropping the requirement of constant error rates $\zeta$ or constant correct choice rates $\nu$. In
these models, which are far more interesting than the constant error, constant correct, or uniform guessing model, we now proceed to place only upper bounds on error rates and lower bounds on correct response rates.

**Definition 1.18** (Generalization of Regenwetter et al., 2013) Consider a collection \( P = (P_\theta)_{\theta \in \Theta} \) of response probability distributions and a fixed binary relation \( \succ \) that has the property that \( \forall (X, R_X) \in \Theta, \) and \( \forall i \in R_X, \exists x \in X, \text{Rank}_{X, \succ}(x) = i \). Let \( d = |\Theta| \) and \( \Delta \) be a distance measure in \( \mathbb{R}^d \). A **distance-based probabilistic specification** of a deterministic preference \( \succ \), with distance \( \Delta \) and with an upper bound \( \tau > 0 \), states that the permissible response probabilities \( (P_\theta)_{\theta \in \Theta} \) that are allowable for fixed preference \( \succ \) must satisfy

\[
\Delta \left( (P_\theta)_{\theta \in \Theta}, \mathcal{V}_{\Theta}(\succ) \right) \leq \tau. \tag{1.5}
\]

This completes our discussion of probabilistic choice models for a single fixed binary preference relation. The binary preference relation \( \succ \) under consideration could, in an experiment, be a stated hypothesis. It could also, however, be an unknown. In the latter case, the experimenter could specify several binary preference relations and take the union of their probabilistic specifications, such as a union of constant error models, or a union of distance specifications of those preference relations, then find the best fitting model in that union of models. Even a union of such models with \( \succ \) an unknown ‘free parameter’ of the model, still represents the preference relation \( \succ \) as unique and fixed.

A well-known example of such a model is the weak utility model, also known as weak stochastic transitivity, for a complete collection of binary choice probabilities, according to which, writing \( u(x) \) for the utility of \( x \),

\[
P_{xy} \geq \frac{1}{2} \iff \neg[y \succ x] \iff u(x) \geq u(y). \tag{1.6}
\]

If the utility function \( u \) is one-to-one, but we do not otherwise fix \( u \), then the preference relation \( \succ \) can be any strict linear order on \( A \), and this is a special case of the distance based specification (1.5) where \( \Delta \) is the supremum-distance and \( \tau = \frac{1}{2} \) (Regenwetter et al., 2013).

We now move from a fixed, possibly unknown, preference to variable preferences.

### 1.3.3 Choice probabilities for varying or uncertain preferences

**Definition 1.19** Let \( A \) be a finite set. Consider a collection \( \succ_1, \succ_2, \ldots, \succ_k \) of binary preferences. A collection of probabilities \( 0 \leq P(\succ_j) \leq 1 \) for
(1 \leq j \leq k) \text{ with}
\sum_{j=1}^{k} P(\succ j) = 1,

is a mixture of preferences.

We can now define what it means for choices (e.g., in a given context) to be induced by a mixture of (i.e., probability distribution over) preferences.

**Definition 1.20** Let $\mathcal{A}$ be a finite set. Let $(P(\succ j))_{0 \leq j \leq k}$ be a mixture of preferences on $\mathcal{A}$. Let $\Theta$ be a set containing elements $\theta$ of the form $\theta = (X, R_X)$ with each $X \subseteq \mathcal{A}$ and each $R_X \subseteq \{1, 2, \ldots, |X|\}$. Assume also that $\forall (X, R_X) \in \Theta, \exists j \leq k, \exists \pi \in \Pi(R_X, X)$ with $\text{Rank}_{X, \succ j}(\pi(i)) = i, \forall i \in R_X$. Let $(P_{\theta})_{\theta \in \Theta}$ be a collection of response probability distributions. $(P_{\theta})_{\theta \in \Theta}$ is a mixture (synonymously, a random preference or a random relation) model of $(P(\succ j))_{0 \leq j \leq k}$ if $\forall \theta = (X, R_X) \in \Theta$ and $\forall \pi \in \Pi(R_X, X),$

$$P_{\theta}(\pi) = \sum_{j \in \{1, 2, \ldots, k\} \times \text{ s.t. } \text{Rank}_{X, \succ j}(\pi(i)) = i, \forall i \in R_X} P(\succ j). \quad (1.7)$$

The convex hull of a collection of points is the smallest convex set containing those points.

**Lemma 1.21** Let $\mathcal{A}$ be a finite set. Let $(P(\succ j))_{0 \leq j \leq k}$ be a mixture of preferences on $\mathcal{A}$. For any preference $\succ$, we obtain the vertex representation of $\succ$ by placing all probability mass on $\succ$ in the mixture model (1.7). Hence, the mixture model (1.7) of $(P(\succ j))_{0 \leq j \leq k}$ forms the convex hull of the vertices $V_{\Theta}(\succ j)_{0 \leq j \leq k}$, where $V_{\Theta}(\succ j)$, with $\theta = (X, \pi)$ has the 0/1-coordinates given by

$$V_{\Theta}(X, \pi)^{\succ j} = \begin{cases} 1 & \text{if } \text{Rank}_{X, \succ j}(\pi(i)) = i, \forall i \in R_X, \\ 0 & \text{otherwise.} \end{cases}$$

We can now utilize some helpful tools from convex geometry (Ziegler, 1995). In the definition below, and in all places where we use the concepts, we assume a finite-dimensional space.

**Definition 1.22** A convex polytope is the convex hull of a finite collection of points. Among all collections of points whose convex hull forms the given polytope, there is a (unique) minimal one; its elements are the vertices of the polytope. The set of vertices forms the vertex description of the polytope. A polytope is alternatively characterized as the intersection of finitely many (closed) half-spaces. Notice that a half-space is formed by
all the solutions to a given linear inequality. Hence, a polytope is the set of solutions of a finite system of linear inequalities. Assume from now on that the polytope is full-dimensional, i.e., that no strict subspace also contains the polytope. Among all finite collections of linear inequalities whose set of solutions is the polytope, there is a (unique) minimal one; its elements are the **facet-defining inequalities**. The set of points in the polytope that satisfy a fixed facet-defining inequality with equality forms a **facet** of the polytope. The facets are also the faces of maximal dimension. The polytope is the solution set of its facet-defining inequalities; the latter forms its **minimal linear description**, also called its **facet description**.

Next we consider some examples of mixture models and discuss their minimal descriptions briefly.

*Binary choice probabilities induced by rankings (strict linear orders).*

**Definition 1.23** A complete system of binary choice probabilities is **induced by rankings (strict linear orders)** if there exists a probability distribution on $SLO_{\mathcal{A}}$ with $P(\succ)$ denoting the probability of $\succ \in SLO_{\mathcal{A}}$, such that,

$$P_{xy} = \sum_{\succ \in SLO \atop a \succ b} P(\succ) \quad (\forall x, y \in \mathcal{A}, x \neq y). \quad (1.8)$$

This model is also called the **linear ordering model for binary choice probabilities**

Note that the left hand side of the model (1.8) denotes the probabilities of observable binary choices, whereas the right hand side refers to probabilities of latent (i.e., not directly observed) linear orders.

Models like the linear ordering model have a long tradition in several disciplines. We now consider what constraints one can derive from such models, e.g., in order to test them empirically. The linear ordering model (1.8) states that the binary choice probability of choosing $x$ over $y$ is the marginal probability of all linear orders in which $x \succ y$. If we concentrate on $|\mathcal{A}| = 3$, say, $\mathcal{A} = \{a, b, c\}$, there are six strict linear orders, which we can also write as rankings $abc, bac, acb, cba, cab, bca$. If we write $P_{xyz}$ instead of $P(\{(x, y), (x, z), (y, z)\})$ for the probability of latent ranking $xyz$, then $P_{ab} = P_{abc} + P_{acb} + P_{cab}$, according to (1.8). No matter what probability distribution we use in the right side of (1.8), since $P_{abc} + P_{bac} + P_{acb} + P_{cba} + P_{cab} + P_{bca} = 1$, ...
it must be the case that

\[
P_{ab} + P_{bc} - P_{ac} = (P_{abc} + P_{acb} + P_{cab}) + (P_{abc} + P_{bac} + P_{bca}) - (P_{abc} + P_{bac} + P_{acb})
\]

\[
= P_{abc} + P_{cab} + P_{bca} - P_{acb} \leq 1.
\]

Similarly, \( P_{ac} + P_{db} - P_{ab} \leq 1 \). In general, it is well-known that the linear ordering model for binary choice probabilities for \( |A| \leq 5 \) holds if and only if the following **triangle inequalities** hold:

\[
P_{xy} + P_{yz} - P_{xz} \leq 1, \quad (\forall \text{ distinct } x, y, z \in A).
\] (1.9)

Consider all linear orders on \( A \), and for each linear order \( \succ \in SLO_A \), consider the vertex representation for \( \succ \) for a complete system of binary choice probabilities. The permissible binary choice probabilities under the linear ordering model (1.8) form the convex hull of the vertices \( V(\succ) \) over all \( \succ \in SLO \), where, for each \( \succ \) and each pair of distinct \( x, y \), we consider one of the coordinates

\[
V_{xy}(\succ) = \begin{cases} 1 & \text{if } x \succ y, \\ 0 & \text{if } y \succ x \end{cases}
\]

and not the other. The linear ordering model of binary choice probabilities forms a convex polytope, because it can be restated as the convex hull of those vertices. In the case of the linear ordering model, a minimal description of the associated **linear ordering polytope of binary choice probabilities**, \( SLOP_A(Binary) \), is currently known only for fairly small \( |A| \) and for complete collections of binary choice probabilities. In particular, the triangle inequalities form a complete minimal description via facet-defining inequalities of \( SLOP_A(Binary) \), for \( |A| \leq 5 \) (see, e.g., Fiorini, 2001a; Koppen, 1995). Reinelt (1993) reviews complete descriptions for \( |A| \leq 7 \). No such complete description is known for large \( |A| \), though many complicated necessary conditions are known (Charon and Hudry, 2010; Doignon et al., 2007).

Regenwetter et al. (2011a) provided and discussed the first published quantitative statistical test of the triangle inequalities (1.9) on empirical data (using order-constrained inference methods). They used three sets of five choice alternatives that were intermixed with each other in a single experiment; one of these sets was a replication of a seminal experiment by Tversky (1969)). The model fit extremely well on 17 of 18 participants, at a significance level of \( \alpha = 0.05 \). The one poor fit is consistent with the Type-I error rate of the test, if all participants were consistent with the model. Tversky’s (1969) original experiment aimed to demonstrate intransitive prefer-
ences by attempting to show violations of weak stochastic transitivity (1.6). Regenwetter et al. (2011a) concluded from their experiment that it is not so straightforward to demonstrate violations of transitivity when allowing for variability in preference, not just choice. Regenwetter et al. (2010) also found that weak stochastic transitivity (1.6), likewise, was descriptive of their data.

**Best choice probabilities induced by rankings (strict linear orders).**

**Definition 1.24** A complete collection of best choice probabilities is induced by rankings (strict linear orders) if there exists a probability distribution on $SLO_A$ with $P(\succ)$ denoting the probability of $\succ \in SLO_A$, such that,

$$B_X(x) = \sum_{\succ \in SLO_A \atop \text{Rank}_{X,\succ}(x)=1} P(\succ), \quad (\forall x \in X \subseteq A). \tag{1.10}$$

In words, the probability that a decision maker indicates that $x$ is best in $X$ is the total probability of all those rankings over $A$ (strict linear orders on $A$) in which $x$ is preferred to all other elements of $X$. The representation (1.10) is also called a linear ordering model of best choice probabilities. In Section 1.5, Theorem 1.6.1 presents a random utility representation of best choice probabilities induced by rankings and Section 1.5.4 includes discussion of a stochastic process model for best choice.

Falmagne (1978) provided a set of necessary and sufficient conditions for the model regardless of $|A|$, and Barberá and Pattanaik (1986) rediscovered the result. The linear ordering model of best choice probabilities forms a convex polytope, because it can be restated as the convex hull of the vertices $VB(\succ)$, for all $\succ \in SLO_A$. In the case of the linear ordering model, a minimal description of the associated linear ordering polytope of best choice probabilities, which we denote as $SLOP_A(\text{Best})$, was characterized by Fiorini (2004). Fiorini (2004) developed a much shorter proof of Falmagne’s (1978) result using conditions in the form

$$\sum_{U: X \subseteq U \subseteq A} (-1)^{|U \setminus X|} P_U(x) \geq 0, \quad (\forall X \subseteq A, X \neq \emptyset, \forall x \in X),$$

where $|U \setminus X|$ denotes the cardinality of the set $U \setminus X = \{z \in U \text{ with } z \notin X\}$. Provided that $|A| \geq 3$, and for a given $X$, the corresponding inequality is facet-defining if and only if $1 \leq |X| < |A|$.

Exactly parallel results can be stated for the analogous linear ordering
polytope of worst choice probabilities, i.e., where

$$W_X(y) = \sum_{\succ \in SLOA} Q(\succ), \quad (\forall x \in X \subseteq \mathcal{A}). \quad (1.11)$$

Note that now the sum is over final positions in the rank orders, i.e., over $\text{Rank}_{X,\succ}(y) = |X|$. de Palma et al. (2013) present very nice results relating best and worst choice probabilities to each other when they are induced by the same rankings - that is, when $P \equiv Q$ in (1.10) and (1.11).

Section 1.5, Theorem 1.61, discusses random utility formulations of the above material.

**Best-worst choice probabilities induced by rankings (strict linear orders).**

**Definition 1.25** A complete collection of best-worst choice probabilities is **induced by rankings (strict linear orders)** if there exists a probability distribution on $SLO_A$ with $P(\succ)$ the probability of $\succ \in SLO_A$, such that,

$$BW_X(x, y) = \sum_{\succ \in SLO_A} P(\succ), \quad (\forall x, y \in X \subseteq \mathcal{A}, x \neq y). \quad (1.12)$$

In words, the probability that a decision maker indicates that $x$ is best in $X$ and $y$ is worst in $X$ is the total probability of all those rankings over $\mathcal{A}$ (strict linear orders on $\mathcal{A}$) in which $x$ is preferred to all other elements of $X$ and in which also all other elements of $X$ are preferred to $y$. The representation (1.12) is also called a linear ordering model of best-worst choice probabilities. In Section 1.5, Theorem 1.62 presents a random utility representation of best-worst choice probabilities induced by rankings and Section 1.5.4 includes discussion of a stochastic process model for best-worst choice.

For $|\mathcal{A}| = 3$, the model (1.12) yields, for distinct $x, y, z$,

$$BW_{\{x,y,z\}}(x, y) = P(\{(x, y), (x, z), (y, z)\}),$$

$$BW_{\{x,y\}}(x, y) = BW_{\{x,y,z\}}(x, y) + BW_{\{x,y,z\}}(x, z) + BW_{\{x,y,z\}}(z, y).$$

We consider $|\mathcal{A}| = 4$, say, $\mathcal{A} = \{a, b, c, d\}$ for the remainder of this subsection. Writing $P_{wxyz}$ for the probability of the ranking $wxyz$, i.e.,

$$P_{wxyz} = P(\{(w, x), (w, y), (w, z), (x, y), (x, z), (y, z)\}),$$

the linear ordering model of best-worst choice probabilities states, for exam-
ple:

\[ BW_{\{a,b\}}(a,b) = P_{abcd} + P_{abdc} + P_{adbc} + P_{acbd} + \ldots + P_{dcab}, \]
\[ BW_{\{a,b,c\}}(a,c) = P_{abcd} + P_{abdc} + P_{adbc} + P_{dabc}, \]
\[ BW_{\{a,b,c,d\}}(a,d) = P_{abcd} + P_{acbd}. \]

It follows, e.g., that for all distinct \(x, y, z, w\) (relabelings of \(A\)),

\[ BW_{\{x,y\}}(x,y) = P_{wzxy} + P_{zwxw} + P_{zxyw} + P_{zyw} + P_{wzxy} + P_{zwxw} + P_{zxyw} + P_{zyw} \]
\[ = BW_{\{z,x,y\}}(z,y) + BW_{\{z,x,y\}}(x,y) + BW_{\{z,x,y\}}(x,z). \]

Likewise, for all distinct \(x, y, z, w\) (relabelings of \(A\)),

\[ BW_{\{w,x,z\}}(x,z) + BW_{\{w,x,z\}}(x,w) + BW_{\{w,x,z\}}(z,w) \]
\[ = BW_{\{w,y,z\}}(y,z) + BW_{\{w,y,z\}}(y,w) + BW_{\{w,y,z\}}(z,w), \]

which we state without proof.

Similarly, one can derive various inequality constraints. For example, the model directly implies that, for all distinct \(x, y, z, w\) (relabelings of \(A\)),

\[ BW_{\{z,w,x,y\}}(w,y) \leq BW_{\{z,w,y\}}(w,y), \]

and also

\[ BW_{\{x,y,w\}}(y,w) = P_{zyzw} + P_{yzxw} + P_{yzw} \]
\[ \leq P_{zyxw} + P_{yzw} \]
\[ = BW_{\{x,y,z,w\}}(z,w) + BW_{\{x,y,z,w\}}(y,w) + BW_{\{x,y,z,w\}}(y,z). \]

Consider all linear orders on \(A\), and for each linear order \(\succ \in SLO_A\), and recall the vertex representation for \(\succ\) for a complete system of best-worst choice probabilities, with coordinates

\[ VBW_{\{X,x,y\}}(\succ) = \begin{cases} 1 & \text{if } Rank_{X,\succ}(x) = 1 \land Rank_{X,\succ}(y) = |X|, \\ 0 & \text{if } Rank_{X,\succ}(x) \neq 1 \lor Rank_{X,\succ}(y) \neq |X|. \end{cases} \]

As we have seen in Lemma 1.2, there are redundant response probabilities in best-worst choice because the choice probabilities for each fixed available set \(X\) sum to one: \(\sum_{y \in X} BW_X(x, y) = 1, \forall X \subseteq A\). After dropping one redundant coordinate for each \(X\), the collections of all possible best-worst
choice probabilities on $A = \{a, b, c, d\}$ form a unit hypercube of dimension
\[
\sum_{k=2}^{4} \binom{4}{k} k(k-1) = \binom{4}{2} [2(2-1) - 1] \binom{3}{3} [3(3-1) - 1] \binom{2}{2} [4(4-1) - 1] = 37.
\]

Consider the vertex representations of the linear orders on $A = \{a, b, c, d\}$ in this coordinate system. The convex hull of the 24 vertices $VBW(\succ)$ (over all $\succ \in SLO_A$) forms the linear ordering model of best-worst choice probabilities. Even after leaving out one coordinate per set $X$, the resulting polytope is not full dimensional. This follows directly, e.g., from Equations 1.13-1.14. We determined this description using public domain software, PORTA, the POlyhedron Representation Transformation Algorithm, of T. Christof & A. Löbel, 1997. We found that there are 15 equations (including 5 nonredundant equations like Eq. 1.13 and 4 nonredundant equations like Eq. 1.14), yielding a polytope of dimension 22.

Since every convex polytope can be represented either as the convex hull of its vertices (here, 24 of the $2^{22}$ vertices of a 22 dimensional unit hypercube), or as the system of affine facet-defining inequalities, we can characterize all best-worst choice probabilities induced by rankings (on $A = \{a, b, c, d\}$) through such a minimal facet-defining system of affine inequalities. An analysis using PORTA yielded a system of 144 facet-defining inequalities, including 11 “trivial” inequalities of the form $BW_X(x, y) \geq 0$, 11 nonredundant inequalities of the form of Inequality 1.15 and 10 nonredundant inequalities of the form given in Inequality 1.16.

To our knowledge, a complete minimal description of the linear ordering polytope of best-worst choice probabilities, which we will denote $SLOP_A(\text{Best–Worst})$, is an open problem for $A > 4$. Doignon et al. (2013) provides some preliminary, but general, results.

We have presented a number of different probabilistic choice models for choices induced by rankings, i.e., by strict linear orders. It is natural to consider alternative models for the latent binary preferences, such as strict partial orders and various special cases, such as strict weak orders, semiorders, and interval orders. We consider one empirical paradigm for this, namely what we will refer to as “ternary paired comparison” probabilities.

**Ternary paired-comparison probabilities induced by strict partial orders, interval orders, semiorders, or strict weak orders.**

In this section, we discuss a choice paradigm that was not included in the, otherwise rather general, concept of a “collection of response probability distributions” of Definition 1.1. In this paradigm, unlike all cases we have

---

3 http://www.iwr.uni-heidelberg.de/groups/comopt/software/PORTA/. We thank Ying Guo for carrying out this analysis using PORTA.
discussed earlier, a decision maker is allowed to express “indifference” among choice alternatives (see Figure 1.1).

**Definition 1.26** Consider a finite set $\mathcal{A}$. A system of ternary paired comparison probabilities on $\mathcal{A}$ is a collection $(T_{x,y})_{x,y \in \mathcal{A}}$ with

$$0 \leq T_{xy}, T_{yx} \leq 1 \quad \text{and} \quad T_{xy} + T_{yx} \leq 1 \quad (\forall x, y \in \mathcal{A}, x \neq y).$$

The ternary paired comparison probability $T_{xy}$ denotes the probability that the person indicates $x$ as strictly preferable to $y$, $T_{yx}$ denotes the probability that the person indicates $y$ as strictly preferable to $x$, and $1 - T_{xy} - T_{yx}$ denotes the probability that the person indicates that neither alternative is strictly preferable to the other.

**Definition 1.27** Consider a system of ternary paired comparison probabilities on a finite set $\mathcal{A}$. The ternary paired comparison probabilities are **induced by strict partial orders** if there exists a probability distribution $P$ on $\mathcal{A}$ such that, $\forall x, y \in \mathcal{A}$, with $x \neq y$, and writing $P(\succ)$ for the probability of any strict partial order $\succ \in \mathcal{A}$,

$$T_{xy} = \sum_{\succ \in \mathcal{A} \atop x \succ y} P(\succ).$$

The ternary paired comparison probabilities are **induced by interval orders** if there exists a probability distribution $P$ on $\mathcal{A}$ such that, $\forall x, y \in \mathcal{A}$, with $x \neq y$, and writing $P(\succ)$ for the probability of any interval order $\succ \in \mathcal{A}$,

$$T_{xy} = \sum_{\succ \in \mathcal{A} \atop x \succ y} P(\succ).$$

The ternary paired comparison probabilities are **induced by semiorders** if there exists a probability distribution $P$ on $\mathcal{A}$ such that, $\forall x, y \in \mathcal{A}$, with $x \neq y$, and writing $P(\succ)$ for the probability of any semiorder $\succ \in \mathcal{A}$,

$$T_{xy} = \sum_{\succ \in \mathcal{A} \atop x \succ y} P(\succ).$$

The ternary paired comparison probabilities are **induced by strict weak orders** if there exists a probability distribution $P$ on $\mathcal{A}$ such that, $\forall x, y \in \mathcal{A}$, with $x \neq y$, and writing $P(\succ)$ for the probability of any strict weak order $\succ \in \mathcal{A}$,

$$T_{xy} = \sum_{\succ \in \mathcal{A} \atop x \succ y} P(\succ).$$

(1.17)
As we review next, these models have been studied for small $|A|$.

Lemma 1.28  (Fiorini, 2001b) Ternary paired comparison probabilities on $A = \{a, b, c, d\}$ are induced by strict partial orders if and only if the following list of facet-defining inequalities for the partial order polytope on four objects are satisfied (for all suitable distinct relabelings $w, x, y, z$ of $a, b, c, d$):

\begin{align*}
PO_1 : & \quad T_{xy} \geq 0, \\
PO_2 : & \quad T_{xy} + T_{yz} \leq 1, \\
PO_3 : & \quad T_{xy} + T_{yz} - T_{xz} \leq 1, \\
PO_4 : & \quad T_{xy} + T_{yz} + T_{zx} - T_{yz} - T_{xy} - T_{xz} \leq 1, \\
PO_5 : & \quad T_{xy} + T_{yz} + T_{zw} + T_{wx} - T_{yx} - T_{xz} - T_{wy} \leq 2, \\
PO_6 : & \quad T_{xy} + T_{yz} + T_{zx} + T_{yw} - T_{zy} - T_{xz} - T_{wy} - T_{wx} \leq 2, \\
PO_7 : & \quad T_{xy} + T_{yz} + T_{zx} + T_{yw} - T_{yx} - T_{xz} - T_{wy} \leq 2, \\
PO_8 : & \quad T_{xy} + T_{yz} + T_{zx} - 2T_{yz} - 2T_{xz} - 2T_{wy} \leq 3.
\end{align*}

As Fiorini (2001b) discusses in detail, the partial order polytope for four choice alternatives is a 12-dimensional polytope with 219 vertices and with 128 different facets. It has 12 facets of type $PO_1$, 6 of type $PO_2$, 24 of type $PO_3$, 8 of type $PO_4$, 24 of type $PO_5$, 24 of type $PO_6$, 24 of type $PO_7$, as well as 6 of type $PO_8$.

Lemma 1.29  (Regenwetter and Davis-Stober, 2011) Ternary paired comparison probabilities on $A = \{a, b, c, d\}$ are induced by interval orders if and only if the following list of facet-defining inequalities for the interval order polytope on four objects are satisfied (for all suitable distinct relabelings $w, x, y, z$ of $a, b, c, d$):

\begin{align*}
IO_1 : & \quad T_{xy} \geq 0, \\
IO_2 : & \quad T_{xy} + T_{yz} \leq 1, \\
IO_3 : & \quad T_{xy} + T_{yz} - T_{xz} \leq 1, \\
IO_4 : & \quad T_{xy} + T_{yz} + T_{zx} - T_{yz} - T_{xy} - T_{xz} \leq 1, \\
IO_5 : & \quad T_{xy} + T_{yz} + T_{zx} - T_{xz} - T_{xy} \leq 1, \\
IO_6 : & \quad T_{xy} + T_{yz} + T_{zx} - T_{yz} - T_{xy} - T_{xz} - T_{wy} \leq 1, \\
IO_7 : & \quad T_{xy} + T_{yz} + T_{vw} - T_{zx} - T_{yx} - T_{vy} \leq 1, \\
IO_8 : & \quad T_{xy} + T_{yz} + T_{vw} - T_{zx} - T_{yz} - T_{vy} - T_{yx} \\
& \quad - T_{xz} - T_{xx} - T_{yw} - T_{vy} \leq 1.
\end{align*}
Specifically, taking into account the omitted quantifiers, there are 12 facets

\[ T_{xy} + T_{yx} + T_{zv} + T_{xz} - T_{xx} - T_{yz} - T_{vy} \]

\[-T_{xx} - T_{zy} - T_{vx} - T_{vy} \leq 1, \]

\[ IO_{10} : \]

\[ T_{xy} + T_{yz} + T_{zv} + T_{zy} - T_{xx} - T_{yz} \]

\[-T_{xy} - T_{yx} - T_{vx} - T_{vy} \leq 2, \]

\[ IO_{11-12} : \]

\[-2 \leq 2T_{xy} + T_{yz} + T_{zv} + T_{vx} - T_{xy} - T_{yz} \]

\[-T_{xy} - T_{yx} - T_{vx} - T_{vy} \leq 2, \]

\[ IO_{13-14} : \]

\[-3 \leq 2T_{xy} + 2T_{yz} + 2T_{zv} + 2T_{vx} - T_{xy} - T_{yz} \]

\[-T_{xy} - T_{yx} - T_{vx} - T_{vy} - T_{vy} \leq 3. \]

This is a 12 dimensional polytope with 207 vertices and 191 facets. Taking
into account the omitted quantifiers, it has 12 facets of type \( IO_1 \), 6 of type
\( IO_2 \), 24 of type \( IO_3 \), 12 of type \( IO_4 \), 8 of type \( IO_5 \), 24 of type \( IO_6 \), 12 of type
\( IO_7 \), 6 of type \( IO_8 \), 3 of type \( IO_9 \), 24 of type \( IO_{10} \), 24 of type \( IO_{11} \), 24 of type
\( IO_{12} \), 6 of type \( IO_{13} \), and 6 of type \( IO_{14} \).

**Lemma 1.30** (Regenwetter and Davis-Stober, 2011) Ternary paired com-
parison probabilities on \( A = \{a, b, c, d\} \) that are induced by semiorders form
a 12 dimensional polytope with 183 vertices, characterized by 563 facets.
These include, e.g., (for all suitable distinct relabelings \( x, y, z, v \) of \( a, b, c, d \)):

\[ SO_1 : \]

\[ T_{xy} \geq 0, \]

\[ SO_2 : \]

\[ T_{xy} + T_{yx} \leq 1, \]

\[ SO_3 : \]

\[ T_{xy} + T_{yz} - T_{xz} \leq 1, \]

\[ SO_4 : \]

\[ T_{xy} + T_{zv} - T_{xz} - T_{zy} \leq 1, \]

\[ SO_5 : \]

\[ T_{xy} + T_{yz} - T_{xy} - T_{zy} \leq 1, \]

\[ SO_6 : \]

\[ T_{xy} + T_{yv} + T_{zx} - T_{xz} - T_{zy} - T_{xy} \leq 1, \]

\[ SO_7 : \]

\[ T_{xy} + T_{yz} + T_{zy} - T_{xy} - T_{zy} - T_{yz} \leq 1, \]

\[ SO_8 : \]

\[ T_{xy} + T_{yx} + T_{zx} - T_{xz} - T_{xy} - T_{yz} \leq 1, \]

\[ SO_9 : \]

\[ T_{xy} + T_{yz} + T_{zx} - T_{xz} - T_{zy} - T_{xy} - T_{yz} \leq 1, \]

\[ SO_{10} : \]

\[ T_{xy} + T_{yz} + T_{zv} - T_{xz} - T_{xy} - T_{yz} \leq 1, \]

\[ \vdots \]

\[ SO_{10} : \]

\[ 2T_{xy} + 2T_{yz} + 2T_{zv} + T_{vx} - T_{vx} - T_{xx} - 2T_{xy} \]

\[-T_{xy} - T_{yx} - T_{yx} - 2T_{xy} - T_{vy} \leq 2, \]

\[ SO_{11} : \]

\[ 2T_{xy} + 2T_{yz} + 2T_{zv} + T_{zx} + T_{vy} - 2T_{xz} - T_{xy} \]

\[-T_{xy} - 2T_{yx} - 4T_{xy} - T_{vy} \leq 3. \]

Specifically, taking into account the omitted quantifiers, there are 12 facets
The most extensively studied model of ternary paired comparison probabilities is the strict weak order polytope characterizing all ternary paired comparison probabilities induced by strict weak orders. This model has been studied for \(|A| \leq 5\).

**Lemma 1.31** (Regenwetter and Davis-Stober, 2012) The collection of all ternary paired comparison probabilities induced by strict weak orders on a five element set \(A = \{a, b, c, d, e\}\) is a convex polytope characterized by 541 distinct vertices of the 20-dimensional unit hypercube. The facet description providing a minimal nonredundant collection of affine inequalities consists of 75,834 distinct inequalities. These include, e.g., (for all suitable distinct relabelings \(v, w, x, y, z\) of \(a, b, c, d, e\)):

\[
\begin{align*}
T_{xy} & \geq 0, \\
T_{xy} + T_{yx} & \leq 1, \\
T_{xy} + T_{yz} - T_{xz} & \leq 1, \\
3(T_{xy} - T_{yx} + T_{yx} - T_{vy} + T_{yw} + T_{wz}) + T_{xz} + T_{yz} & + T_{xy} - T_{zy} - T_{yw} + T_{yw} \\
& + 3T_{yw} - T_{wy} - 3T_{wx} + T_{ux} - 3T_{wz} - T_{vx} - 3T_{vz} - T_{zu} & \leq 4.
\end{align*}
\]

Regenwetter and Davis-Stober (2012) provide all 75,834 facet-defining inequalities and they report a successful fit of this polytope to empirical ternary paired comparison data using order-constrained likelihood methods.

To our knowledge, a complete minimal description of the strict weak order polytope of ternary paired comparison probabilities, which we will denote \(SWOP_A(Ternary)\), is an open problem for \(A > 5\). For a recent discussion of the polytope’s known facet-structure see Doignon and Fiorini (2002).

### 1.4 Algebraic Theories and Their Real Valued Representations

#### 1.4.1 Numerical representations of binary preference relations

We now consider how some of the binary preference relations of Definition 1.7 are related to numerical representations of utility as they have been studied for more than a half century.

**Theorem 1.32** Let \(A\) be a finite set of choice alternatives. Consider a binary preference relation \(\succ\). 

of type \(SO_1\), 6 of type \(SO_2\), 24 of type \(SO_3\), 12 of type \(SO_4\), 24 of type \(SO_5\), 8 of type \(SO_6\), 24 of type \(SO_7\), 24 of type \(SO_8\), 24 of type \(SO_9\), 12 of type \(SO_{10}\), ..., 24 of type \(SO_{30}\), and 24 facets of type \(SO_{31}\).
Utility Representations of Strict Weak Orders: The relation $\succ$ is a strict weak order on $A$ if and only if there exists a real valued (utility) function $u : A \rightarrow \mathbb{R}$ which assigns values (utilities) to choice alternatives $x \in A$ via $x \mapsto u(x)$ such that
\[ x \succ y \iff u(x) > u(y) \quad (\forall x, y \in A, x \neq y). \tag{1.18} \]

Utility Representations of Strict Linear Orders: The relation $\succ$ is a strict linear order on $A$ if and only if there exists a one-to-one real valued (utility) function $u$ on $A$ satisfying Equivalence 1.18.

Utility Representations of Semiorders: The relation $\succ$ is a semiorder on $A$ if and only if there exists a real valued (utility) function $u : A \rightarrow \mathbb{R}$ which assigns real valued utilities to choice alternatives $x \in A$ via $x \mapsto u(x)$ and there exists a positive real number $\varepsilon \in \mathbb{R}^{++}$, such that
\[ x \succ y \iff u(x) > u(y) + \varepsilon \quad (\forall x, y \in A, x \neq y). \]

Utility Representations of Interval Orders: The relation $\succ$ is an interval order on $A$ if and only if there exists a real valued (lower utility) function $\ell : A \rightarrow \mathbb{R}$ which assigns real valued (lower bounds on) utilities to choice alternatives $x \in A$ via $x \mapsto \ell(x)$ and a positive real valued (threshold) function $\tau : A \rightarrow \mathbb{R}^{++}$ which assigns positive real values (thresholds of discrimination) to choice alternatives $x \in A$ via $x \mapsto \tau(x) > 0$, such that
\[ x \succ y \iff \ell(x) > \ell(y) + \tau(y) \quad (\forall x, y \in A, x \neq y), \]

\[ x \succ y \iff \ell(x) > \ell(y) \quad (\forall x, y \in A, x \neq y), \]

i.e.,
\[ x \succ y \iff \ell(x) > u(y) \quad (\forall x, y \in A, x \neq y), \]

with (upper utility) function $u$ defined by: $u(z) = \ell(z) + \tau(z), \forall z \in A$ that assigns choice alternatives to (upper bounds on) utilities.

1.4.2 Representations of simple lexicographic semiorders

The next definition builds on Definition 1.8 to consider special lexicographic binary relations on $A$ that define certain kinds of lexicographic semiorders. This definition is equivalent to an alternate definition of Davis-Stober (2012). The basic idea underlying this definition is that $g$ and $\ell$ may be two attributes that rank order the objects in $A$ from best to worst in opposite directions (e.g., in Tversky, 1969, and in the gambles a-e of Section 1.5.1, the probability of winning and the amount one can win trade off against each other), that the decision maker has a fixed perceptual threshold on each of
the two attributes $g$ and $\ell$, and considers the attributes in a lexicographic fashion.

**Definition 1.33** Let $\mathcal{A}$ be a finite set with $|\mathcal{A}| = n$, and let $\triangleright \in S\mathcal{L}O_{\mathcal{A}}$ be a single fixed strict linear order on $\mathcal{A}$. A **compatible simple lexicographic semiorder** is a lexicographic semiorder $\succ$ such that there exists real valued one-to-one functions $g$ and $\ell$ on $\mathcal{A}$ and positive constants $\varepsilon_g, \varepsilon_\ell \in \mathbb{R}^+$, with $\forall x, y \in \mathcal{A}$

$$\ell(x) < \ell(y) \Leftrightarrow x \triangleright y \Leftrightarrow g(x) > g(y),$$

and, furthermore, it holds that

either 1) $\forall x, y \in \mathcal{A},$

$$x \succ y \Leftrightarrow \begin{cases} \text{either} & g(x) > g(y) + \varepsilon_g \\ \text{or} & |g(x) - g(y)| \leq \varepsilon_g \land \ell(x) > \ell(y) + \varepsilon_\ell, \end{cases}$$

or 2) $\forall x, y \in \mathcal{A},$

$$x \succ y \Leftrightarrow \begin{cases} \text{either} & \ell(x) > \ell(y) + \varepsilon_\ell \\ \text{or} & |\ell(x) - \ell(y)| \leq \varepsilon_\ell \land g(x) > g(y) + \varepsilon_g, \end{cases}$$

For the fixed strict linear order $\triangleright$ on $\mathcal{A}$ we denote the collection of all compatible simple lexicographic semiorders as $S\mathcal{L}O_{\mathcal{A},\triangleright}$.

### 1.4.3 Numerical representations of weak orders over gambles

Luce and Marley (2005) develop extensive theoretical relations among a variety of representations over gambles of gains (equally, over losses); some of the representations and results are fairly classical, others involve relatively new axiomatizations. Included are the class of ranked weighted utility, rank-dependent utility (which includes cumulative prospect theory as a special case), gains-decomposition utility, simple utility with weights that depend only on the relevant event, as used in the original prospect theory of Kahneman and Tversky (1979), and subjective expected utility. Marley et al. (2008) and Bleichrodt et al. (2008) each solve a different open problem in that paper and Marley and Luce (2005) discuss various data relevant to the theoretical properties of the representations presented in Luce and Marley (2005). Here we begin with what is called a **ranked additive representation**, where a ranked gamble consists of a finite number of consequence-event
branches ordered by preferences among the consequences. The utility representation involves a sum of functions, one for each branch, that depends on two things: the utility of the consequence and the entire ordered vector of the event partition. This representation is developed using a form of ranked additive conjoint measurement that was axiomatized by Wakker (1991). We then summarize results for the rank weighted representation, which is the special case of the ranked additive representation where the utility of each consequence-event branch is a product of the utility of the outcome for that branch and of a weight depending on the event associated with that branch and the entire ordered vector of the event partition.

The thrust of Luce and Marley (2005), and the selected results below, is to add qualitative conditions that are necessary and sufficient to go from the ranked additive representation to each of the specializations, including the rank weighted representation. These conditions are all of a type that Luce (2000) has called accounting equivalences. Given an accounting equivalence, an accounting indifference asserts that when two gambles have the same bottom line, the decision maker is indifferent between them. Luce and Marley (2005) also describe what happens when one adds to the framework of preferences among gambles the concept of joint receipts of consequences and of gambles. Section 1.4.5 defines, using joint receipt, a number of additional properties. A brief history of this and related concepts appears in Luce (2000).

We restrict the presentation to uncertain gambles (defined in the next section). For results on risky and/or ambiguous gambles, see Wakker (2010) and Abdellaoui et al. (2011). It is also important to remember that all the results in this section are for gambles of gains, or, equivalently, for gambles of losses. Parallel results for mixed gambles, i.e., those with outcomes that can be gains and/or losses, can be derived in various ways. The representations for such mixed gambles depend on whether one considers the representations in both the gains and loss domains to be of the same form or of different forms; and, if the latter, how representations of mixed gambles are derived from those for gains/losses (Luce, 2000; Wakker, 2010).

Notation

Let $X$ denote the set of pure consequences for which chance or uncertainty plays no role. A distinguished element $e \in X$ is interpreted to mean no change from the status quo. We assume that there exists a preference.

\footnote{There is a slight inconsistency in the use here of $X$ to denote the set of pure consequences and its use in the previous sections to denote a set of options. However, we think the retention of this notation is warranted since much of the literature in both areas uses $X$ in this manner.}
order $\succsim$ over $X$ and that it is a weak order. Let $\sim$ denote the corresponding indifference relation. A typical first-order gamble of gains $g$ with $n$ consequences is of the form

$$g = (x_1, C_1; \cdots; x_i, C_i; \cdots; x_n, C_n) \equiv (\cdots; x_i, C_i; \cdots),$$

where $x_i$ are consequences and $C_n = (C_1, \cdots, C_i, \cdots, C_n)$ is a (exhaustive and exclusive) partition of some “universal event” $C(n) = \bigcup_{i=1}^{n} C_i$. The underlying event $C(n)$ is only “universal” for the purpose of this gamble.

A (consequence, event) pair $(x_i, C_i)$ is called a branch. Such a gamble is a 2n-tuple composed of $n$ branches. We deal mostly with cases where each $C_i \neq \emptyset$. The results are confined to all gains, i.e., where each $x_i \succsim e$ (or, equally, to all losses, i.e., where each $x_i \precsim e$).

We assume $X$ is so rich that for any first-order gamble $g$, there exists $CE(g) \in X$, called the certainty equivalent of $g$, such that $CE(g) \sim g$. Thus, the preference order $\succsim$ can be extended to the domain of gains $D_+$ that consists of $X$ and all first-order gambles. For some results we need to expand $D_+$ to include second-order gambles in which some of the $x_i$ are replaced by first-order gambles. Under the usual co-monotonicity assumption, i.e., monotonicity is preserved so long as the consequence ranking is preserved, indifferent alternatives may be substituted without altering the appraisal of the gamble. For example, if $f$ is a second-order gamble with just one consequence, say its $i$th, a gamble $g_i$, then $g_i$ can be replaced by the certainty equivalent $CE(g_i) \sim g_i$ to obtain a first order gamble equivalent to $f$. Thus, any second-order gamble can always be reduced to a first-order one by using certainty equivalents and preserving preferences.

We also assume that if $\rho$ is a permutation of the indices $\{1, 2, \ldots, n\}$, then

$$\left(x_{\rho(1)}, C_{\rho(1)}; \cdots; x_{\rho(i)}, C_{\rho(i)}; \cdots; x_{\rho(n)}, C_{\rho(n)}\right) \sim (x_1, C_1; \cdots; x_i, C_i; \cdots; x_n, C_n).$$

This an example of an accounting indifference. Each accounting indifference is a behavioral property, and hence testable. The basic feature of accounting indifferences is that the bottom lines on the two sides of $\sim$ are identical, and so, normatively, they should not be distinguished. One can view accounting indifferences as invariance of preference under reformulations of the gamble.

In stating numerous axioms and representations, it is convenient to assume that we have carried out a permutation of the indices such that the consequences are (rank) ordered from the most preferred to the least pre-
ferred, in which case we simply use the notation
\[ x_1 \succeq \cdots \succeq x_i \succeq \cdots \succeq x_n \succeq e. \] (1.20)
For the most part, this section is concerned with properties formulated for
the latter rank ordered case. In that case, there is a rank order induced
on the underlying partition \( \mathbf{C}_n \), which we emphasize by the vector nota-
tion \( \mathbf{C}_n \). We give the rank ordered representation the same name as the
corresponding unordered one, but with the prefix “ranked” added, and the
abbreviations also prefixed with R. We do not explicitly include the rele-
vant definitions for the unranked cases. In general, when researchers test
theories empirically involving ranked consequences, such as rank weighted
utility (below), they present the gambles in the ordered form; nonetheless,
so long as (1.19) is satisfied, the ranked form is nothing but a convenience
for writing the representation or for writing an axiom leading to a ranked
representation.

We explore utility representations \( u \) onto real intervals of the form \( I = [0, \kappa] \), where \( \kappa \in [0, \infty] \), that meet various, increasingly stronger, restrictions.
Two conditions that are common to all representations we consider are:
\[ g \succeq h \text{ iff } u(g) \geq u(h), \] (1.21)
\[ u(e) = 0. \] (1.22)
We refer to these as order-preserving representations. Note that be-
cause \( I \) is open on the right, there is no maximal element in the structure.

The framework above is for uncertain gambles, i.e., those where each
outcome occurs if a particular event occurs, and probabilities are not given.
We do discuss some representations for risky gambles - that is, those for
which outcome probabilities are given - in Sections 1.4.6 and 1.5.2. We do
not present any results for ambiguous gambles - those in which each
outcome is associated with either an event or a probability (but not both) -
Abdellaoui et al. (2011) present very nice empirical and (some) theoretical
results for this domain.

**Ranked additive utility representation**

The following is the most general order preserving representation of uncer-
tain gambles that we consider. Some history of its axiomatization is given
following the definition.

**Definition 1.34** An order-preserving representation \( u : \mathcal{D}_+ \rightarrow I \) is a
ranked additive utility (RAU) representation iff, for all \( x_i \in X \) satisfying (1.20) and for every corresponding ordered partition \( \mathbf{C}_n \) of \( C(n) \),
there exist strictly increasing functions $L_i(\cdot, \overrightarrow{C}_n) : I \onto I$, with the following properties:

$$u(x_1, C_1; \cdots; x_i, C_i; \cdots; x_n, C_n) = \sum_{i=1}^{n} L_i \left( u(x_i), \overrightarrow{C}_n \right), \quad (1.23)$$

$$L_i(0, \overrightarrow{C}_n) = 0,$$

$C_i = \emptyset$ implies $L_i(Z, \overrightarrow{C}_n) = 0 \quad (Z \in I)$.

It is an additive utility (AU) representation if the functions $L_i(\cdot, \overrightarrow{C}_n)$ are the same for every ordering of the partition $C_n$ of $C(n)$.

Note that in Definition 1.34, we assume that $C_i = \emptyset$ implies $L_i(Z, \overrightarrow{C}_n) = 0 \quad (Z \in I)$, whereas, for simplicity, the proofs of the relevant representation theorems are restricted to partitions with all non-null sets.

Wakker (1991) axiomatized a more general form than (1.23) where there exist functions $L(x_1, C_1; \cdots; x_i, C_i; \cdots; x_n, C_n) : D_+ \onto I$ and $L_i(\cdot, \overrightarrow{C}_n) : X \onto I$ such that

$$L(x_1, C_1; \cdots; x_i, C_i; \cdots; x_n, C_n) = \sum_{i=1}^{n} L_i (x_i) . \quad (1.24)$$

Bleichrodt et al. (2008) add additional conditions to those yielding the representation (1.24) that are sufficient for the representation (1.23); the major additional condition is rank additive utility tradeoff consistency across gambles (Bleichrodt et al., 2008, Def.4.4), a condition that is similar in form to the standard sequence invariance condition in additive conjoint measurement (Krantz et al., 1971, Section 6.11.2, p. 304). Bleichrodt et al.’s (2008) results hold for general gambles, not those restricted to be all gains (or, equally, all losses). As far as we know, none of the observable properties underlying (1.24) or (1.23) have been tested (though see Kobberling and Wakker, 2004, for a proposed similar tradeoff technique for testing nonexpected utility theories such as subjective expected utility).

Assuming (1.23) holds, Luce and Marley (2005) present specific behavioral constraints, usually accounting indifferences, that limit it to the particular forms found in the literature including ranked weighted utility, rank-dependent utility (which includes cumulative prospect theory as a special case), subjective expected utility, and the several descriptive configural weighted models$^6$, such as RAM and TAX of Birnbaum and his colleagues.

$^6$ Birnbaum has typically called them “configural weight models” but Luce and Marley (2005) believe their term is a better descriptor.
(for summaries of the latter, see Birnbaum, 2008). Here we illustrate, and extend, the results for ranked weighted utility and relate them to the properties of configural weight models.

**Ranked weighted utility representation**

The following concept of a ranked weighted utility (RWU) representation, is defined in Marley and Luce (2001).

**Definition 1.35** An order-preserving representation $u : \mathcal{D}_+ \xrightarrow{\text{only}} I \subseteq \mathbb{R}_+$ is a **ranked weighted utility (RWU)** representation iff there exist weights $S_i(\mathcal{C}_n)$ assigned to each index $i = 1, \ldots, n$ and possibly dependent on the entire ordered partition $\mathcal{C}_n$, where $0 \leq S_i(\mathcal{C}_n)$ and $S_i(\mathcal{C}_n) = 0$ iff $C_i = \emptyset$, such that, for (1.20) holding,

$$u(\cdots; x_i, C_i; \cdots) = \sum_{i=1}^{n} u(x_i)S_i(\mathcal{C}_n). \quad (1.25)$$

If the ranking is immaterial, it is called **weighted utility (WU)**.

Note that the multiplicative weights of (1.25) depend both on $i$ and on the entire ordered partition $\mathcal{C}_n$. We study several representations that place restrictions on the latter dependence. Also note that ranked weighted utility, (1.25), is the special case of ranked additive utility, (1.23), for which the functions $L_i(\cdot, \mathcal{C}_n)$ are all linear, i.e., $L_i(Z, \mathcal{C}_n) = ZS_i(\mathcal{C}_n)$, where $S_i(\mathcal{C}_n) = 0$ iff $C_i = \emptyset$. For $C_i \neq \emptyset$, then $S_i(\mathcal{C}_n) > 0, i = 1, \ldots, n$, which means that the representation satisfies co-monotonic consequence monotonicity, i.e., it is strictly increasing in each consequence, so long as the rank ordering is maintained.

We discuss an axiomatization of the RWU form in Section 1.4.4.

Consider a gamble with $x_1 = \cdots = x_i = \cdots = x_n = x$ and consider the following property:

**Definition 1.36** **Idempotence** of gambles is satisfied iff, for every $x \in X$ and every ordered event partition $(C_1, \cdots, C_i, \cdots, C_n)$,

$$(x, C_1; \cdots; x_i; C_i; \cdots; x; C_n) \sim x. \quad (1.26)$$

If we assume idempotence along with RWU, we see that

$$\sum_{i=1}^{n} S_i(\mathcal{C}_n) = 1. \quad (1.27)$$

Most theories of utility, including that for the RWU representation of Def.
1.35, have either explicitly or implicitly assumed idempotence. A major exception is the extensive series of papers on the utility of gambling that develops representations involving a relatively standard idempotent representation, plus a weighted entropy-type term; these results are summarized in Luce et al. (2009).

1.4.4 Axiomatizations of representations of gambles

We now explore the additional behavioral conditions that are needed to reduce the ranked additive utility representation of Definition 1.34 to the rank weighted utility representation of Definition 1.35, and the relation of the latter to the configural weighted models of Birnbaum and his colleagues.

**Ranked weighted utility**

Let us suppose that the domain $D_+$ is extended to include second-order compound gambles. Within that extended domain, one major property of binary gambles that has played a fairly key theoretical role and, to a lesser extent, an empirical one, is the concept of binary event commutativity:

For all events $C, C', D, D'$ with $C \cap C' = D \cap D' = \emptyset$, and $x \succeq y \succeq e$

$((x, D; y, D'), C; y, C') \sim ((x, C; y, C'), D; y, D'). \quad (1.28)$

One sees that if one reduces this to a first-order gamble, it amounts to saying that on each side $x$ is the consequence if $C$ and $D$ both occur, and otherwise $y$ is the consequence, the only difference being the order in which the experiments $(C, C')$ and $(D, D')$ are conducted; the binary RDU model satisfies this property. If one has $y = e$ in (1.28) it is called status-quo, binary event commutativity, which seems to be supported by data (Luce, 2000, pp.74-76). We explore one possible generalization of this concept for gambles with more than two consequences, which gives rise to the RWU representation.

**Definition 1.37** Suppose that $\overrightarrow{C}_n, \overrightarrow{D}_m$ are any two ordered event partitions and that $x_1 \succeq x_2 \succeq ... \succeq x_n \succeq e$. Then event commutativity is satisfied iff

$((x_1, D_1; e, D_2; ...), C_1; (x_2, D_1; e, D_2; ...), C_2; ...; (x_n, D_1; e, D_2; ...), C_n) \sim ((x_1, C_1; x_2, C_2; ...; x_n, C_n), D_1; e, D_2; ...; e, D_m). \quad (1.29)$

where all of the events $D_k$, $k = 2, ..., m$, have $e$ attached to them.

As in the binary case, both sides give rise to $x_i$ iff both $C_i$ and $D_1$ occur. This
seems just as rational as binary event commutativity. This general property has not been studied empirically.

**Theorem 1.38**  (*Luce and Marley, 2005, Th. 9*) Consider a structure \(<D_+, \succeq>\) for gambles with \(n \geq 2\) which has a ranked additive utility (RAU) representation, Def. 1.34, with the functions \(L_i(\cdot, \mathcal{C}_n)\), \(i = 1, 2, \ldots, n\), having derivatives at 0. Then the following statements are equivalent:

1. Event commutativity, (1.29), holds.
2. RWU, Def. 1.35, holds.

**Rank-dependent utility**

The RWU representation, (1.25), is of interest because it encompasses several models in the literature including the standard rank-dependent model (which includes cumulative prospect theory as a special case).

Define

\[
C(i) := \bigcup_{j=1}^{i} C_j. \tag{1.30}
\]

Consider a set of positive weights \(W_{C(n)}(C(i)), i = 0, \ldots, n\), and define

\[
W_i(\mathcal{C}_n) = \begin{cases} 
0, & i = 0 \\
W_{C(n)}(C(i)), & 0 < i < n \\
1, & i = n
\end{cases}.
\]

**Definition 1.39**  *Rank-dependent utility* (RDU) is the special case of RWU, (1.25), with weights of the form: for \(i = 1, \ldots, n\),

\[
S_i(\mathcal{C}_n) = W_i(\mathcal{C}_n) - W_{i-1}(\mathcal{C}_n).
\]

Luce and Marley (2005, Footnote 2) summarize the history of this representation and its various names. With certain additional assumptions it has also been called *cumulative prospect theory for gains (losses)* (Tversky and Kahneman, 1992).

The RDU representation exhibits the following property:

**Definition 1.40**  *Coalescing* is satisfied iff for all ordered consequences, (1.20), and corresponding ordered partitions with \(n > 2\) and with \(x_{k+1} = x_k\), \(k < n\),

\[
(x_1, C_1; \cdots; x_k, C_k; x_k, C_{k+1}; \cdots; x_n, C_n)
\sim (x_1, C_1; \cdots; x_k, C_k \cup C_{k+1}; \cdots; x_n, C_n) \quad (k = 1, \ldots, n - 1). \tag{1.31}
\]
Note that the gamble on the left has \( n \) branches with \( n - 1 \) distinct consequences whereas the one on the right has \( n - 1 \) branches as well as \( n - 1 \) distinct consequences. The next result shows that coalescing is the key to RDU.

**Theorem 1.41** (Luce and Marley, 2005, Th. 11) For \( n > 2 \), the following statements are equivalent:

1. RWU, Def. 1.35, idempotence, Def. (1.26), and coalescing, (1.31), all hold.
2. RDU, Def. 1.39, holds.

There is a large literature on ways to arrive at idempotent RDU (see, e.g., Luce, 2010, for further discussion and references). None of these approaches seem simpler or more straightforward than the assumptions used here.

**Configural weighted utility**

Birnbaum, over many years with various collaborators, has explored a class of representations that Marley and Luce (2005) called configural weighted utility; Birnbaum has typically called them configural weight models. In contrast to utility models that have been axiomatized in terms of behavioral properties, Birnbaum and his collaborators have shown that certain special configural weight models do or do not exhibit certain behavioral properties, and have compared how various of those properties fare relative to data. In contrast to this approach, Proposition 1.43, below, shows the close relation between the most frequently studied configural weighted utility model, called TAX, and the ranked weighted utility representation. According to Birnbaum and Navarrete (1998), TAX has the following form\(^7\):

**Definition 1.42** Let \( u \) be a utility function over ranked gambles and pure consequences, \( T \) a function from events into the non-negative real numbers, and \( \omega_{i,j}(\widehat{C}_n) \) mappings from ordered event partitions to real numbers. Then, TAX is the following representation over gambles in ranked order:

\[
u(g_{\widehat{C}_n}) = \frac{\sum_{i=1}^{n} u(x_i)T(C_i) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [u(x_i) - u(x_j)]\omega_{i,j}(\widehat{C}_n)}{T(\widehat{C}_n)}, \tag{1.32}
\]

where \( T(\widehat{C}_n) := \sum_{i=1}^{n} T(C_i) \).

\(^7\) The notation is that of Marley and Luce (2005), and differs from that of Birnbaum and his collaborators in two, non-substantive, ways.
The name TAX arises because Birnbaum describes the term on the right as imposing a tax from (resp., to) lower ranked consequences to (resp., from) higher ranked ones depending on whether the relevant weight is positive (resp., negative); he usually assumes a particular form for the weights \( \omega_{i,j}(\overrightarrow{C}_n) \). Marley and Luce (2005, Prop. 6) proved the following result when the weights are unrestricted in form\(^8\).

**Proposition 1.43**

(i) A TAX representation, (1.32), with \( S_i(\overrightarrow{C}_n), i = 1, \ldots, n, \) defined by

\[
S_i(\overrightarrow{C}_n) = \frac{T(C_i) + \sum_{j=i+1}^{n+1} \omega_{i,j}(\overrightarrow{C}_n) - \sum_{j=0}^{i-1} \omega_{j,i}(\overrightarrow{C}_n)}{T(\overrightarrow{C}_n)}, \tag{1.33}
\]

where

\[
\omega_{0,i}(\overrightarrow{C}_n) := 0, \\
\omega_{i,n+1}(\overrightarrow{C}_n) := 0,
\]

is an idempotent RWU representation, (1.25), provided each \( S_i(\overrightarrow{C}_n) \geq 0 \) and \( S_i(\overrightarrow{C}_n) = 0 \) iff \( C_i = \emptyset \).

(ii) Any idempotent RWU representation, (1.25), can be put (in many ways) in the form of a TAX representation, (1.32). One such has

\[
T(C_i) > 0
\]

and

\[
\omega_{i,j}(\overrightarrow{C}_n) = \begin{cases} 
T(\overrightarrow{C}_n) \sum_{k=1}^i S_k(\overrightarrow{C}_n) - \sum_{k=1}^i T(C_k), & i = 1, \ldots, n-1, \cr 
0, & i = 0 \text{ or } j = n + 1, \cr 
or j \neq i + 1, &
\end{cases}
\]

Thus, the general TAX model is equivalent to the idempotent ranked weighted utility representation (Section 1.4.3) when the TAX model satisfies the following two conditions: \( S_i(\overrightarrow{C}_n) \) in (1.33) are nonnegative and \( S_i(\overrightarrow{C}_n) = 0 \) iff \( C_i = \emptyset \). The latter constraints correspond to requiring the TAX model to satisfy co-monotonic consequence monotonicity, i.e.,

\(^8\) Marley and Luce failed to include the conditions \( S_i(\overrightarrow{C}_n) \geq 0 \) and \( S_i(\overrightarrow{C}_n) = 0 \) iff \( C_i = \emptyset \) in their statement of part (i). There is also a typographic error in the final sum in the numerator of their (19) - each term \( \omega_{i,j}(\overrightarrow{C}_n) \) in that term should be \( \omega_{j,i}(\overrightarrow{C}_n) \), as it is in our (1.33). The results in Marley and Luce that use the proposition remain valid as all those results assume a RWU representation, hence that \( S_i(\overrightarrow{C}_n) \geq 0 \) and \( S_i(\overrightarrow{C}_n) = 0 \) iff \( C_i = \emptyset \).
the representation is strictly increasing in each consequence so long as the rank ordering is maintained; it is not known what constraints on the $\omega_{i,j}(\mathcal{C}_n)$ in (1.33) are sufficient for TAX to satisfy that condition. Marley and Luce (2005) explore various independence properties that are implied by the idempotent ranked weighted utility representation (Section 1.4.3), and hence by any TAX model that satisfies co-monotonic consequence monotonicity. Birnbaum (2008) re-summarizes various of the data included in Marley and Luce (2005), plus more recent data, and concludes that rank-dependent utility is not a satisfactory representation of his data, whereas TAX can handle the majority of the data. However, recent discussions of Birnbaum’s analyses (often based either on descriptive modal choice or his pattern based true-and-error models) suggest that further detailed analyses of data of individual decision makers are needed to place such a conclusion on a firm basis (see, e.g., Birnbaum, 2011; Regenwetter et al., 2011b).

1.4.5 Joint receipts

In this section we extend the domain $D_+$ of gambles to include the joint receipt of pure consequences and gambles. With $X$ the set of pure consequences, for $x, y \in X, x \oplus y \in X$ represents receiving both $x$ and $y$. When $X$ denotes money, many authors assume that $x \oplus y = x + y$, but as discussed in Luce (2000) this is certainly not necessary and may well be false. When $f$ and $g$ are gambles, $f \oplus g$ means having or receiving both gambles. Here, we assume $\oplus$ to be a commutative operator, strictly increasing in each variable, with $e$ its identity.

The following concept of generalized additivity is familiar from the functional equation literature:

**Definition 1.44** The operation $\oplus$ has a generalized additive representation $u : D_+ \rightarrow \mathbb{R}_+ := [0, \infty[$ iff (1.21), (1.22), and there exists a strictly increasing function $\varphi$ such that

$$u(f \oplus g) = \varphi^{-1} (\varphi (u(f)) + \varphi (u(g))).$$

(1.34)

It is called additive if $\varphi$ is the identity.

Note that $V = \varphi(u)$ is additive and therefore one cannot distinguish

---

9 As already noted, Marley and Luce (2005, Proposition 8) failed to recognize that the constraints $S_i(\mathcal{C}_n) \geq 0$ are necessary for a TAX model to be an idempotent weighted utility representation. However, they assume the latter representation in all the later relevant results, and hence that $S_i(\mathcal{C}_n) \geq 0$. Therefore, their results are true for any TAX model satisfying the latter constraints.
between a generalized additive representation $u$ and an additive representation $V$ without considering additional structural assumptions about the properties of the utility representation, such as separability; we consider such structural conditions in Section 1.4.6.

An important special case is when $\varphi$ is the identity in which case $u$ is additive over $\oplus$. This is, of course, a strong property. For example, if for money consequences $x \oplus y = x + y$, then additive $u$ implies $u(x) = \alpha x$ (with $\alpha > 0$ for a monotonic increasing relation). For at least modest amounts of money—“pocket money”—this may not be unrealistic, as Birnbaum and collaborators have argued by fitting data (for example, Birnbaum and Beeghley, 1997, for judgment data, as well as Birnbaum and Navarrete, 1998, for choice).

Another important special case of generalized additivity is: for some $\delta \neq 0$,

$$u(f \oplus g) = u(f) + u(g) + \delta u(f)u(g),$$

which form has been termed $p$-additivity (Ng et al., 2009). This corresponds to the mapping $\varphi(z) = \text{sgn}(\delta) \ln(1 + \delta z)$ in (1.34).

### 1.4.6 Parametric forms for utility and weights

**Parametric utility forms**

As pointed out after the definition of a generalized additive representation, Definition 1.44, one cannot distinguish between a generalized additive representation $u$ and an additive representation $V$ without considering additional structural assumptions about the properties of the utility representation; we now discuss such properties and the resulting representations.

**Definition 1.45  Binary segregation** holds for gains iff for all gambles $f, g$ of gains,

$$(f \oplus g, C_1, g, C_2) \sim (f, C_1, ; g, C_2) \oplus g.$$  

(1.36)

Luce (2000, Th. 4.4.4, Th 4.4.6) shows the following:

**Theorem 1.46** Let $u$ be an order preserving representation of gambles over gains and $W_{C(2)}$ a weighting function. Then any two of the following imply the third:

1. Binary segregation, (1.36).
2. Binary rank dependent utility:

$$u(f, C_1, ; g, C_2) = u(f)W_{C(2)}(C_1) + u(g)[1 - W_{C(2)}(C_1)].$$

3. $p$-additivity: For some real constant $\delta$ that has the unit of $1/u$,

$$u(f \oplus g) = u(f) + u(g) + \delta u(f)u(g),$$
and for some event $K$ there is a weighting function $W_K$ onto $[0, 1]$ such that $(u, W_K)$ forms a separable representation of the gambles $(x, C, ; e, K \setminus C)$, i.e.,

$$u(x, C, ; e, K \setminus C) = u(x)W_K(C).$$

This result thus gives a quite natural motivation for the assumption of a p-additive representation over gains, with a parallel result for losses. We are not aware of a parallel theorem over mixed gambles (gains and losses). Ng et al. (2009), Theorem 21, present an axiomatization of the rank-dependent form over both gains and losses in which the utility function is assumed to be p-additive over gains (over losses). They also assume segregation and derive binary RDU. A careful check would be needed to see whether their result can be derived by assuming binary RDU (which has been axiomatized by Marley and Luce (2002)), without assuming p-additivity, and then using the above result to derive p-additivity over gains (over losses).

Given the p-additive form, with $V(r) := 1 + \delta u(r)$, we obtain the additive representation $V(f \oplus g) = V(f) + V(g)$. It is routine to show that the following representations are possible for mixed gambles (gains and losses) under the conditions of Ng et al. (2009, Th. 21) and arguments paralleling those of Luce (2000, Corollary, p. 152) show that they are the most general solutions:

(i) If $\delta = 0$, then for some $\alpha > 0$,

$$u(f) = \alpha V(f).$$  \hfill (1.37)

(ii) If $\delta > 0$, then $u$ is superadditive, i.e., $u(f \oplus g) > u(f) + u(g)$, unbounded, and for some $\kappa > 0$,

$$u(f) = \frac{1}{\delta} [e^{\kappa V(f)} - 1].$$  \hfill (1.38)

(iii) If $\delta < 0$, then $u$ is subadditive, i.e., $u(f \oplus g) < u(f) + u(g)$, bounded by $1/|\delta|$, and for some $\kappa > 0$,

$$u(f) = \frac{1}{|\delta|} [1 - e^{-\kappa V(f)}].$$  \hfill (1.39)

When the pure consequences are money, many economists assume that $x \oplus y = x + y$, in which case there is a constant $A > 0$ such that $V(x) = Ax$. Note that then, for $\delta > 0$, where we have (1.38), $u$ is strictly increasing and convex, which many identify as corresponding to risk seeking behavior; and for $\delta < 0$, where we have (1.38), $u$ is strictly increasing and concave, which many identify as corresponding to risk averse behavior. Luce (2010a,b) includes
Further discussion of these representations, including their implications for interpersonal comparisons of utility.

The above representations have a common value for $\delta$ for gains and losses, which, in the special case of money just discussed gives either risk seeking behavior for both gains and losses, or risk averse behavior for both gains and losses. However, it is more customary to assume that people are risk seeking on gains and risk averse on losses. Such behavior can be accommodated by having a $p$-additive function for pure consequences that has a $\delta^+ > 0$ and $\kappa^+ > 0$ for gains and a $\delta^- < 0$ and $\kappa^- > 0$ for losses - that is,

$$u(x) = \begin{cases} \frac{1}{\delta^+}(e^{\kappa^+x} - 1), & x \geq 0, \\ \frac{1}{\kappa^-}(1 - e^{-\kappa^-x}), & x < 0. \end{cases}$$

Another possible form has the utility of gains as well as the utility for losses bounded: there are positive constants $a, b, A,$ and $B$ such that

$$u(x) = \begin{cases} A(1 - e^{-ax}), & x \geq 0, \\ B(e^{bx} - 1), & x < 0. \end{cases}$$

Finally, the above representations assume that risk seeking or risk aversion is modelled by the shape of the utility function; in particular, risk aversion is modelled by a concave utility function. However, rank-dependent utility (RDU) with $u(x) = x^2$ and $W(p) = p^2$ produces risk aversion, even though the utility function is strictly convex (Chateauneuf and Cohen, 1994, Corollary 2).

Luce (2010a) proposed a condition that discriminates between the representations (1.37)-(1.39): Consider the case where the pure outcomes are amounts of money, assume that $r \oplus s = r + s$, and let $x, x', y, y'$ be such that $x \succ x' \succ y \succ y'$. Then Luce’s condition is:

$$\delta = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \iff (x + x', E; y + y', \Omega \setminus E) \begin{bmatrix} > \\ \sim \\ < \end{bmatrix} (x + y, E; x' + y', \Omega \setminus E),$$

where $E$ is a chance event (with complement $\Omega \setminus E$, where $\Omega$ is the “universal” event) such that for all amounts of money $x, y, (x, E; y, \Omega \setminus E) \sim (y, E; x, \Omega \setminus E)$; that is, $E$ has (subjective) probability $1/2$.

Davis-Stober and Brown (2013) extend Luce’s original criterion to accommodate decision makers with risk preferences that vary as a function of the monetary environment (e.g., gambles of all gains, all losses, or mixed gains and losses); to a significant extent, their extension corresponds to testing both the $\delta = -1, 0, +1$ cases covered by Luce’s conditions and the mixed
cases of $\delta$ introduced above. Using a Bayesian modeling approach, they evaluated the repeated choices of 24 participants and found that one participant was best described as risk seeking; 6 as risk averse; 2 as risk neutral; 6 as risk seeking for losses and risk averse for gains; 6 could not be described by a $p$-additive representation; and 1 of what they call stakes sensitive and 2 of what they call mixed gamble.

**Parametric weight forms**

A risky gamble is one where the events (of an uncertain gamble) are replaced by probabilities. For notational simplicity, let $(x, p)$ denote a risky gamble in which the consequence $x$ occurs with probability $p$ and nothing otherwise. A consequence can be either a pure one (no risk), such as a sum of money, or a gamble such as $(y, q)$ where $y$ is a pure consequence. In this section, the only gambles considered are of the (first-order) form $(x, p)$ or the (second-order) form $((y, q), p)$ where $x$ and $y$ range over the pure consequences and $p$ and $q$ can be any probabilities. We assume that there is a weak order preference relation $\succsim$ over first- and second-order gambles of the above form; and that $\succsim$ is represented by an order-preserving representation $u$.

**Definition 1.47** An order-preserving representation $u$ is separable if there is a strictly increasing weighting function $W : [0, 1] \rightarrow [0, 1]$ such that for each first-order gamble or pure consequence $z$ and probability $p$,

$$u(z, p) = u(z)W(p).$$

(1.40)

where $u(x) =: u(x, 1)$. Thus, given a a pure consequence $x$, for a first-order gamble we have $u(x, p) = u(x)W(p)$ and for a second order gamble we have $u((x, q), p) = u(x)W(q)W(p)$.

We now present some results on parametric forms for the weights in first- and/or second-order binary risky gambles of the above forms. We focus on the case where there is a common representation over all the gambles, which can be either over gains, over losses, or over gains and losses (Prelec, 1998, extends the results discussed below to cases where the representations can differ for gains and losses).

Prelec (1998) was the first to axiomatize weighting functions assuming a separable representation (with the weights onto the interval $[0, 1]$). He axiomatized three different forms; we summarize the results for one case.

The following definition is a variant of Prelec (1998, Def. 1) and Luce (2001, Def. 1):
Definition 1.48  Let $N \geq 1$ be any integer. Then **N-compound invariance** holds iff, for consequences $x, y, u, v$, and probabilities $p, q, r, s \in [0, 1]$ with $p < q, r < s$,

$$(x, p) \sim (y, q), (x, r) \sim (y, s) \text{ and } (u, p^N) \sim (v, q^N)$$

imply

$$(u, r^N) \sim (v, s^N).$$

**Compound invariance** holds when N-compound invariance holds for all integers $N \geq 1$.

Prelec’s (1998, Prop. 1) then says, in essence, that in the presence of separability, (1.40), and a suitable density of consequences, compound invariance is equivalent to the following **compound-invariance** family of weighting functions: there are constants $\alpha > 0, \beta > 0$ such that

$$W(p) = \exp[-\beta(\ln p)^\alpha].$$

(1.41)

The necessity of compound invariance for this representation is seen by noting that the representation satisfies the property: for any real number $\lambda \geq 1, W(p^\lambda) = W(p)^{\lambda \alpha}$. Note that when $\alpha = 1$ the representation reduces to a power function with positive exponent. Also, for $\alpha < 1$, the general representation has an inverse-S shape, which is the form found for many estimated weighting functions (Prelec, 1998; Wakker, 2010, Section 7.7).

Compound invariance involves first-order gambles and a relatively complicated antecedent condition, plus the assertion that it holds for all integers $N \geq 1$. We now present a condition due to Luce (2001, Def. 2) that is equivalent to a weighting function belonging to the compound invariance family. This condition has the advantages that it has a simpler antecedent condition and only requires the assertion of the condition holds for the integers $N = 1, 2$, but has the (possible) disadvantage that it involves compound gambles. Note that the following Definition and Theorem include conditions stated for the doubly open unit interval, denoted $]0, 1[$.  

**Definition 1.49**  (Luce, 2001, Def.1) Let $N \geq 1$ be any integer. Then **N-reduction invariance** holds iff, for any consequence $x$, and probabilities $p, q, r \in [0, 1]$,

$$(x, p), q \sim (x, r) \sim (y, s)$$

implies

$$(x, p^N), q^N \sim (x, r^N).$$

$Luce (2001, Footnote 1)$ notes that Prelec focussed on the special case $\beta = 1$. 

---

10 Luce (2001, Footnote 1) notes that Prelec focussed on the special case $\beta = 1$. 

Reduction invariance holds when $N$-reduction invariance holds for all integers $N \geq 1$.

**Theorem 1.50** *(Luce, 2001, Prop.1)* Suppose that a structure of binary gambles of the form $(x, p)$ and $((y, p), q)$ is weakly ordered in preference and has a separable representation (1.40) with $W : [0, 1] \xrightarrow{onto} [0, 1]$, where $W$ is strictly increasing in $p$. Then the following two conditions are equivalent:

(i) $N$-reduction invariance (Definition 1.49), holds for $N = 2, 3$.

(ii) $W$ satisfies (1.41).

Both Prelec (1998) and Luce (2001) generalize their condition to include other weight forms, including the exponential-power and hyperbolic-logarithm. However, all cases essentially assume a separable representation. The next section presents an extremely elegant joint derivation of a separable representation with a weighting function satisfying (1.41).

Diecidue et al. (2009) give additional preference foundations for parametric weighting functions under rank-dependent utility, namely power functions $p^a$ with $a > 0$; exponential functions $e^{cp-1}$ with $c \neq 0$; and switch-power functions (to separate power functions on a lower- and an upper-range); and Chechile and Barch (2013) review recent theoretical and empirical work on (risky) weighting function and present new evidence in support of the Prelec function and their own exponential odds function.

**Multiplicative separability of utility and weight**

We now summarize the very nice result by Bleichrodt et al. (2013) that, under weak regularity conditions, a weaker version of compound invariance implies that an order-preserving multiplicative representation exists for first-order gambles of the form $(x, p)$ with the weight form satisfying (1.41).

**Definition 1.51** For real $\lambda > 0$, $\lambda$-compound invariance holds iff, for nonzero consequences $x, y, z$, and nonzero probabilities $p, q, r$,

$$(x, p) \sim (y, q), (x, q) \sim (y, r) \text{ and } (y, p^\lambda) \sim (z, q^\lambda)$$

imply

$$(y, q^\lambda) \sim (z, r^\lambda).$$

The main results of Bleichrodt et al. (2013) only use $\lambda$-compound invariance for the integer cases $\lambda = 1, 2, 3$, in which cases it agrees with the special case of $N$-compound invariance, Definition 1.48, with $N = 1, 2, 3$, and with $x, y, u, v$ (respectively, $p, q, r, s$) in $N$-compound invariance replaced by the special case $x, y, y, z$ (respectively, $p, q, q, r$) in $\lambda$-compound invariance.

Figure 1.4 illustrates 1-compound invariance.
We need various assumptions and terminology from Bleichrodt et al. (2013). We continue to use the representation \((x,p)\) for a gamble where the pure consequence \(x\) is received with probability \(p\), otherwise nothing. Assume that the set of such gambles is \(X \times [0, 1]\) with \(X\) a nonpoint interval within \(\mathbb{R}^+\) containing 0, and that \((0,p) \sim (x,0) \sim (0,0)\) for all \(p\) and \(x\). The preference order \(\succeq\) on \(X \times [0, 1]\) is continuous if, for each gamble \((x,p)\), \(\{(y,q) : (y,q) \succeq (x,p)\}\) and \(\{(y,q) : (y,q) \preceq (x,p)\}\) are closed subsets of the gamble space \(X \times [0, 1]\). Strict stochastic dominance holds if \((x,p) \succ (y,p)\) whenever \(x > y\) and \(p > 0\) and \((x,p) \succ (x,q)\) whenever \(x > 0\) and \(p > q\). Finally, the representation \(u\) on the set of gambles \((x,p)\) is multiplicative if it is separable, Definiton 1.47, with \(u(0) = 0\), \(W(0) = 0\), and each of \(u\) and \(W\) is continuous and strictly increasing. Clearly, multiplicative representability implies strict stochastic dominance.

Bleichrodt et al.’s (2013) very nice main result shows that the multiplicative relation is a consequence when \(\lambda\)-compound invariance holds for \(\lambda = 1, 2, 3\).

**Theorem 1.52** (Bleichrodt et al., 2013, Th.3.4). The following statements are equivalent for \(\succeq\) on \(X \times [0, 1]\):

1. A multiplicative representation exists.
2. \(\succeq\) is a continuous weak order that satisfies strict stochastic dominance, \(\lambda\)-compound invariance, Def. 1.51, for \(\lambda = 1, 2, 3\), and \((0,p) \sim (x,0) \sim (0,0)\) for all \(p\) and \(x\).

**Other parametric and weight forms**

The previous two sections presented axiomatizations for utility functions derived from \(p\)-additive representations, and weighting functions derived from compound invariance. A tremendous number of other utility and weighting functions have been suggested, though most have not been axiomatized in the way presented above. Stott (2006) summarizes many of the suggested utility functions for gambles over monetary gains and weighting functions over probabilities and then considers various combinations of utility and weighting function (in the cumulative prospect theory form) in conjunction with three standard functional forms for the (probabilistic) choice function. He fit these combinations to data on binary choices between gambles of the form discussed in the previous section, and concluded that the best fitting model had a power utility function, a Prelec weighting function, (1.41), and
a multinomial logit (MNL) form, (1.59), for the (best) choice probabilities; Blavatskyy and Pogrebna (2010) discuss related work using a larger variety of econometric specifications and conclude that the particular probabilistic response function one uses can be pivotal for the relative performance of competing core deterministic theories. The next section presents recent axiomatizations of some of these (probabilistic) choice functions when expected utility holds.

1.5 Choice Probabilities for Numerical Representations

This section begins with definitions and results on distribution-free random utility representations (i.e., those without a specific parametric form, such as the Normal or Extreme Value), then focuses on various of the parametric models that have been developed and applied in discrete choice surveys, mainly by Economists and Marketing scientists, and those developed and studied in controlled experiments, mainly by Experimental Economists and Cognitive Psychologists, and attempts to integrate them in a common theoretical framework.

We begin with the classic models in psychology and economics, which were formulated mainly to account for choice probabilities. We then present extensions of those models to deal with more complicated phenomena, including the time to make a choice (response time). All of these models are context free - that is, in a sense to be defined, the representation of a choice option is independent of the other available choice options. We then present context dependent models - that is, where the representation of a choice option depends on the other available options.

1.5.1 Distribution-free random utility representations

Definition 1.53 Let $A$ be a finite collection of choice alternatives. A (distribution-free) random utility model for $A$ is a family of jointly distributed real random variables $U = (U_{x,i})_{x \in A, i \in I}$ with $I$ some finite index set. If $U = (U_{x,i})$ is a family of jointly independent random variables, the model is a independently distributed random utility model.

The realization of a random utility model at some sample point $\omega$, given by the real-valued vector $(U_{x,i}(\omega))_{x \in A, i \in I}$, assigns to alternative $x \in A$ the utility vector $(U_{x,i}(\omega))_{i \in I}$. One possible interpretation of such a utility vector is that $I$ is a collection of attributes, and $U_{x,i}(\omega)$ is the utility of choice alternative $x$ on attribute $i$ at sample point $\omega$. 
Definition 1.54 Let \( \mathcal{A} \) be a finite collection of choice alternatives. A (distribution-free) \textbf{unidimensional, noncoincident random utility model} on \( \mathcal{A} \) is a family of jointly distributed real random variables \( U = (U_x)_{x \in \mathcal{A}} \) with \( \Pr(U_y = U_z) = 0, \forall y \neq z \in \mathcal{A} \).

The most common use of the term “random utility model” in the discrete choice literature (Train, 2003) refers to context-independent, noncoincident, and unidimensional parametric models, where the random variables \( (U_x)_{x \in \mathcal{A}} \) can be decomposed as follows:

\[
(U_x)_{x \in \mathcal{A}} = (u(x))_{x \in \mathcal{A}} + (\epsilon_x)_{x \in \mathcal{A}},
\]

and where \( u(x) \) is the deterministic real-valued utility of option \( x \) and \( (\epsilon_x)_{x \in \mathcal{A}} \) is multivariate normal or multivariate extreme value noise. We present such models in some detail in Section 1.5.3.

We now consider the probabilistic generalizations of the results in Theorem 1.32.

Theorem 1.55 (Regenwetter and Marley, 2001) Let \( \mathcal{A} \) be a finite set with \( |\mathcal{A}| = N \). Consider a family of \( k \times N \) many jointly distributed real valued utility random variables \( U = (U_{x,i})_{x \in \mathcal{A}; i = 1, \ldots, k} \). Then the collection \( (U) \) satisfies the following properties (with all \( w, x, y, z \in \mathcal{A} \)):

**Random Utility Representations of Strict Linear Orders:** A unidimensional (i.e., \( k = 1 \)), noncoincident random utility model \( U \) on \( \mathcal{A} \) induces a probability distribution \( \succ \mapsto P(\succ) \) on SLO as follows. For any given \( \succ \in SLO \), writing \( x_i \) for the object at rank \( i \) in \( \succ \), i.e., \( \text{Rank}_{\mathcal{A}, \succ}(x_i) = i, \forall i = 1, 2, \ldots, N \), let

\[
P(\succ) = \Pr(U_{x_1} > U_{x_2} \cdots > U_{x_N}).
\]

**Random Utility Representations of Strict Weak Orders:** A unidimensional (i.e., \( k = 1 \)) random utility model on \( \mathcal{A} \) induces a probability distribution \( \succ \mapsto P(\succ) \) on SWO, regardless of the joint distribution of \( U \), through

\[
P(\succ) = \Pr \left( \bigcap_{w \succ x} (U_w > U_x) \bigcap_{\neg (y \succ z)} (U_y \leq U_z) \right).
\]

**Random Utility Representations of Weak Orders:** A unidimensional (i.e., \( k = 1 \)) random utility model on \( \mathcal{A} \) induces a probability distribution \( \succsim \mapsto P(\succsim) \) on WO, regardless of the joint distribution of \( U \), through

\[
P(\succsim) = \Pr \left( \bigcap_{w \succsim x} (U_w \geq U_x) \bigcap_{\neg (y \succsim z)} (U_y < U_z) \right).
\]
Random Utility Representations of Semiorders: A unidimensional (i.e., $k = 1$) random utility model on $\mathcal{A}$ induces a probability distribution $\succ \mapsto P(\succ)$ on $\text{SO}$, regardless of the joint distribution of $U$ through, given a strictly positive real valued (constant) threshold $\varepsilon \in \mathbb{R}^+$,

$$P(\succ) = \Pr \left( \bigcap_{w \succ x} (U_w > U_x + \varepsilon) \bigcap \bigcap_{[y \succ z]} (U_y - U_z \leq \varepsilon) \right).$$

Random Utility Representations of Interval Orders: If $k = 2$ and $\Pr(U_{x,1} \leq U_{x,2}) = 1, \forall x \in A$, then, writing $L_x$ for $U_{x,1}$ (lower utility) and $U_x$ for $U_{x,2}$ (upper utility), $U$ induces a probability distribution $\succ \mapsto P(\succ)$ on $\text{IO}$, through,

$$P(\succ) = \Pr \left( \bigcap_{w \succ x} (L_w > U_x) \bigcap \bigcap_{[y \succ z]} (L_y \leq U_z) \right).$$

Conversely, each probability distribution on $\text{SLO, SWO, WO, SO, or IO}$ can be represented in the above fashion (nonuniquely) by an appropriately chosen family of jointly distributed random variables.

By a function on $\mathcal{A}$, we mean a mapping from $\mathcal{A}$ into the real numbers $\mathbb{R}$. The collection of all functions on $\mathcal{A}$ is the space $\mathbb{R}^A$. When $\mathcal{A}$ contains $N$ elements, this is $\mathbb{R}^N$, the $N$-dimensional reals. Let $\mathcal{B}(\mathbb{R}^A)$ denote the sigma-algebra of Borel sets in $\mathbb{R}^A$.

**Definition 1.56** Let $\mathcal{A}$ be a finite set of choice alternatives. A random function model for $\mathcal{A}$ is a probability space $\langle \mathbb{R}^A, \mathcal{B}(\mathbb{R}^A), \mathbb{P} \rangle$.

The idea behind a random function model is to define a probability measure $\mathbb{P}$ on the space of (e.g., utility) functions on $\mathcal{A}$. This space, of course, contains all conceivable unidimensional, real-valued, utility functions on $\mathcal{A}$. We can now summarize key results about binary choice probabilities induced by linear orders, as given in Equation (1.8).

**Theorem 1.57** Consider a finite set $\mathcal{A}$ of choice alternatives and a complete collection $(P_{xy})_{x,y \in \mathcal{A}, x \neq y}$ of binary choice probabilities. The binary choice probabilities are induced by strict linear orders (1.8) if and only if they are induced by a (distribution-free) unidimensional, noncoincident random utility model (Block and Marschak, 1960). Furthermore, this holds if and only if they are induced by a (distribution-free) random function model with one-to-one functions (Regenwetter and Marley, 2001). Formally, there exists a probability distribution on $\text{SLO}_A$ with $P(\succ)$ the probability of $\succ \in \text{SLO}_A$,
such that Equation 1.8 holds, i.e.,

\[ P_{xy} = \sum_{\succ \in SLOA_{xy}} P(\succ), \quad (\forall x, y \in \mathcal{A}, x \neq y), \]  

(see Eq. 1.8)

if and only if there exists a family of jointly distributed random variables \((U_x)_{x \in \mathcal{A}}\) that are noncoincident, i.e., \(\Pr(U_x = U_y) = 0, \forall x, y \in \mathcal{A}, x \neq y\), such that

\[ P_{xy} = \Pr(U_x > U_y), \quad (\forall x, y \in \mathcal{A}, x \neq y), \]  

(1.42)

if and only if there exists a probability space \((\mathbb{R}^A, \mathcal{B}(\mathbb{R}^A), \mathbb{P})\), such that

\[ P_{xy} = \mathbb{P} \left( \{ u \in \mathbb{R}^A \mid u(x) > u(y), \text{ where } u \text{ is a one-to-one function} \} \right) \]  

\((\forall x, y \in \mathcal{A}, x \neq y)\).  

(1.43)

A complete collection \((P_{xy})_{x,y \in \mathcal{A}, x \neq y}\) of binary choice probabilities satisfies any one of these three conditions if and only if

\[(P_{xy})_{x,y \in \mathcal{A}, x \neq y} \in \mathcal{SLOP}_A(\text{Binary}).\]

Suck (2002) presents a set of complicated necessary conditions for the random variables to be independent for a finite set \(\mathcal{A}\), including necessary and sufficient conditions in the case when \(|\mathcal{A}| = 3\).

If we want to drop the requirement that the random utilities be noncoincident or that the utility functions be one-to-one functions, then we obtain the following theorem.

**Theorem 1.58** (Regenwetter and Marley, 2001; Regenwetter and Davis-Stober, 2012) Consider a finite set \(\mathcal{A}\) of choice alternatives and a complete collection \((T_{xy})_{x,y \in \mathcal{A}, x \neq y}\) of ternary paired comparison probabilities. The ternary paired comparison probabilities are induced by strict weak orders (1.17) if and only if they are induced by a (distribution-free) unidimensional random utility model. Furthermore, this holds if and only if they are induced by a (distribution-free) random function model. Formally, there exists a probability distribution on \(SWO_{\mathcal{A}}\) with \(P(\succ)\) the probability of \(\succ \in SWO_{\mathcal{A}}\), such that Equation 1.17 holds, i.e.,

\[ T_{xy} = \sum_{\succ \in SWO_{\mathcal{A}} \backslash \succ = y} P(\succ), \quad (\forall x, y \in \mathcal{A}, x \neq y), \]  

(see Eq. 1.17)

if and only if there exists a family of jointly distributed random variables \((U_x)_{x \in \mathcal{A}}\) such that

\[ T_{xy} = \Pr(U_x > U_y), \quad (\forall x, y \in \mathcal{A}, x \neq y), \]
if and only if there exists a probability space \((\mathbb{R}^A, \mathcal{B}(\mathbb{R}^A), \mathbb{P})\), such that

\[
T_{xy} = \mathbb{P}\left(\{u \in \mathbb{R}^A \mid u(x) > u(y)\}\right), \quad (\forall x, y \in A, x \neq y).
\]  

(1.44)

A complete collection \((T_{xy})_{x,y \in A, x \neq y}\) of ternary paired comparison probabilities satisfies any one of these three conditions if and only if

\[
(T_{xy})_{x,y \in A, x \neq y} \in \text{SWOP}_A(\text{Ternary}).
\]

The equivalences in these two theorems state that whenever one of three mathematical representations for binary choice or ternary paired comparison probabilities exists, the others exist as well. It is important to note, however, that each of the three representations can have different special cases of particular interest. For example, the linear ordering model specification can be specialized by considering parametric ranking models of a particular type, such as “Mallows’ Φ-models,” in the right term of Eq. 1.8 (Critchlow et al., 1993). The random utility formulation can be specialized by considering parametric families of random utilities, such as logit or probit models, and many others in the right term of Eq. 1.42 (Böckenholt, 2006; Train, 2003). Last but not least, the random function model can be specialized by considering only certain (utility) functions \(u\) that must satisfy a particular form of cumulative prospect theory (Stott, 2006) in the right hand side of Eq. 1.43 or Eq. 1.44 (Regenwetter et al., 2013). We will discuss these issues in more detail below. In particular, we give an example of a random function model containing cumulative prospect theory with Goldstein-Einhorn weighting functions, and with power utility functions, as the deterministic special case.

We now briefly consider an alternative type of ternary choice probabilities, namely we build on Definition 1.26 to define ternary-paired comparison probabilities that are consistent with compatible simple lexicographic semiorders and state a theorem about this model.

**Definition 1.59** Consider a finite set \(A\) and a system of ternary paired comparison probabilities \((T_{x,y})_{x,y \in A, x \neq y}\) on \(A\). Let \(\succ \in \text{SLO}_A\). The ternary paired comparison probabilities are induced by (compatible) simple lexicographic semiorders if there exists a probability distribution \(P\) on \(\text{SLSO}_A,\succ\) such that, \(\forall x, y \in A\), with \(x \neq y\), and writing \(P(\succ)\) for the probability of any compatible simple lexicographic semiorder \(\succ \in \text{SLSO}_A,\succ\),

\[
T_{xy} = \sum_{\succ \in \text{SLSO}_A,\succ} P(\succ).
\]  

(1.45)
Davis-Stober (2012, Theorems 1-3, Corollary 1) showed the following, using the famous “Catalan numbers” $C_n$.

**Theorem 1.60** Let $A = \{x_1, x_2, \ldots, x_n\}$, and let $\succ \in \text{SLO}_A$ be a single fixed strict linear order on $A$. Suppose that $x_i \succ x_j \iff i < j$. Let $C_n = \frac{1}{n+1} \binom{2n}{n}$. Then

$$|\text{SLSO}_A, \succ| = C_nC_{n+2} - C_{n+1}^2.$$  

In particular, ternary paired comparison probabilities induced by compatible simple lexicographic semiorders form a convex polytope of dimension $n(n-1)$ with precisely $C_nC_{n+2} - C_{n+1}^2$ many distinct vertices. This polytope has precisely $\frac{5n^2-11n+8}{2}$ many facets. The facet-defining inequalities form seven distinct families and, for any finite $n$, are given by

\begin{align*}
T_{x_ix_j} - T_{x_i x_j+1} &\leq 0, \quad (\forall i, j \in \{1,2,\ldots,n\}, i < j < n), \\
T_{x_{i+1}x_j} - T_{x_ix_j} &\leq 0, \quad (\forall i, j \in \{1,2,\ldots,n\}, i + 1 < j), \\
T_{x_j x_i} + T_{x_i x_j} - T_{x_j x_{i+1}} &\leq 0, \quad (\forall i, j \in \{1,2,\ldots,n\}, i < j < n), \\
T_{x_{i+1}x_{j+1}} + T_{x_j x_i} - T_{x_{i+1}x_i} &\leq 0, \quad (\forall i, j \in \{1,2,\ldots,n\}, i + 1 < j), \\
- T_{x_i x_{i+1}} &\leq 0, \quad (\forall i \in \{1,2,\ldots,n-1\}), \\
- T_{x_ix_j} &\leq 0, \quad (\forall i, j \in \{1,2,\ldots,n\}, i > j), \\
T_{x_1x_n} + T_{x_n x_1} &\leq 1.
\end{align*}

Similar results to those of Theorem 1.57 also hold for other response probabilities induced by rankings. We state these in the cases of best choice and best-worst choice probabilities.

**Theorem 1.61** Consider a finite set $A$ of choice alternatives, and a complete collection $(B_X(x))_{x \in X \subseteq A}$ of best choice probabilities on $A$. The best choice probabilities are induced by strict linear orders if and only if they are induced by a (distribution-free) unidimensional, noncoincident random utility model (Block and Marschak, 1960). Furthermore, this holds if and only if they are induced by a (distribution-free) random function model with one-to-one functions (Falmagne, 1978). Formally, there exists a probability distribution on $\text{SLO}_A$ with $P(\succ)$ the probability of $\succ \in \text{SLO}_A$, such that

$$B_X(x) = \sum_{\succ \in \text{SLO}_A \atop \text{Rank}_X, \succ[1]=x} P(\succ), \quad (\forall x \in X \subseteq A),$$

if and only if there exists a family of jointly distributed random variables $(U_x)_{x \in A}$ that are noncoincident (i.e., $\forall x, y \in A, x \neq y, \Pr(U_x = U_y) = 0$)
such that
\[ B_X(x) = \Pr(U_x = \max_{y \in X} U_y), \quad (\forall x \in X \subseteq A) \]
if and only if
\[ B_X(x) = \mathbb{P}\left( \left\{ u \in \mathbb{R}^A \mid u(x) = \max_{y \in X} u(y), \text{ where } u \text{ is a 1-to-1 function} \right\} \right), \]
with suitable choices of probabilities and random variables on the right hand side. Any one these three conditions holds if and only if
\[ (BW_X(x))_{x \in X \subseteq A} \in SL\mathcal{OP}_A(Best). \]

**Theorem 1.62** Consider a finite set \( A \) of choice alternatives, and a complete collection \((BW_X(x,y))_{x,y \in X, x \neq y, X \subseteq A}\) of best-worst choice probabilities on \( A \). The best-worst choice probabilities are induced by strict linear orders if and only if they are induced by a (distribution-free) unidimensional, noncoincident random utility model paralleling Block and Marschak (1960). Furthermore, this holds if and only if they are induced by a (distribution-free) random function model with one-to-one functions. Mathematically,
\[ BW_X(x, y) = \sum_{\succ \in SLOP_A} P(\succ) \]
if and only if
\[ BW_X(x, y) = P\left( U_x = \max_{w \in X} U_w, U_y = \min_{z \in X} U_z \right), \quad \text{with } \Pr(U_x = U_y) = 0 \]
if and only if
\[ BW_X(x, y) = \mathbb{P}\left( \left\{ u \in \mathbb{R}^A \mid u(x) = \max_{w \in X} u(w), u(y) = \min_{z \in X} u(z), \text{ where } u \text{ is a one-to-one function} \right\} \right), \]
with suitable choices of probabilities and random variables on the right hand side. Any one these three conditions holds if and only if
\[ (BW_X(x, y))_{x,y \in X \subseteq A} \in SLOP_A(Best - Worst). \]

**Distribution-free random cumulative prospect theory: An example with Goldstein-Einhorn weighting and power utility**

We briefly discuss an example of a random function model for binary choice, as stated in Equation 1.43, in which we only consider certain functions as permissible (Regenwetter et al., 2013). In fact, we will consider only finitely many functions and an unknown probability distribution over them.
For simplicity, consider the following five two-outcome gambles (Regenwetter et al., 2011a), each of which offers some probability of winning a positive gain, and otherwise neither a gain nor loss. Gamble \(a\) offers a \(7/24\) probability of winning \$28, gamble \(b\) offers a \(8/24\) probability of winning \$26.60, gamble \(c\) offers a \(9/24\) probability of winning \$25.20, gamble \(d\) offers a \(10/24\) probability of winning \$23.80, gamble \(e\) offers a \(11/24\) probability of winning \$22.40. These lotteries were 2007 dollar equivalents of those used by Tversky (1969).

Consider cumulative prospect theory with a Goldstein-Einhorn weighting function (Stott, 2006) using weighting parameters \(\gamma \in [0,1]\) and \(s \in [0,10]\) and a power utility function using a parameter \(\alpha \in [0,1]\). Denote these assumptions by \(\text{CPT} - \text{GE}\). According to this \(\text{CPT} - \text{GE}\), a gamble \((x,p)\) with probability \(p\) of winning \(x\) (and nothing otherwise) has a subjective numerical value of

\[
u(x,p) = \frac{sp^\gamma}{sp^\gamma + (1-p)^\gamma}x^\alpha.
\]

We consider only finitely many such functions \(u\), all of which are 1-to-1 on the domain \(\mathcal{A} = \{a,b,c,d,e\}\). More specifically, we only consider those values \(\alpha, \gamma \in [0.001, 0.991]\) that are multiples of 0.01 and those values of \(s \in [0.01, 9.96]\) that are multiples of 0.05 (Regenwetter et al., 2013). The resulting random function model turns out to be equivalent to a linear ordering model (1.8), in which only 11 of the 120 strict linear orders in \(\text{SLOP}_{\{a,b,c,d,e\}}\) have positive probability. Hence, the random function model for \(\text{CPT} - \text{GE}\) on \(\mathcal{A} = \{a,b,c,d,e\}\) forms a convex polytope with 11 vertices. Regenwetter et al. (2013) report that the facet-defining inequalities for this model are

\[
0 \leq P_{ab} \leq P_{ac} \leq P_{ad} \leq P_{ae} \leq P_{bc} \leq P_{bd} \leq P_{be} \leq P_{cd} \leq P_{ce} \leq P_{de} \leq 1.
\]

Regenwetter et al. (2013) also characterize three other random function models. For instance, the random function model for \(\text{CPT} - \text{GE}\) on a different set of two-outcome gambles has 487 distinct facet-defining inequalities. Regenwetter et al. (2013) find that these models perform poorly in quantitative order-constrained statistical tests on laboratory data, even though the full linear ordering model fits the same data extremely well (Regenwetter et al., 2011a).

Random function models can be highly restrictive, hence make strong testable predictions. For example, suppose that there are two distinct elements \(x, y \in \mathcal{A}\) such that every function \(u\) under consideration satisfies \(u(x) > u(y)\), i.e., \(y\) is dominated by \(x\) no matter what utility function a decision maker uses. Then the random function model predicts that \(P_{xy} = 1\).
and $P_{yx} = 1 - P_{xy} = 0$, i.e., a decision maker satisfying this random function model has probability zero of choosing the dominated option $y$. Guo and Regenwetter (2013) discuss a similar set of conditions for the perceived relative argument model (PRAM) (Loomes, 2010). We now proceed to distributional random utility models that do not share such restrictive features.

**1.5.2 Axiomatizations of expected utility representations in probabilistic choice**

Fechnerian models for binary choice

Now we turn to recent axiomatic results on probabilistic choice between risky (first-order) gambles; we are not aware of similar results for uncertain gambles. A risky gamble $g$ with $n$ consequences is of the form

$$g = (x_1, p_1; \cdots; x_i, p_i; \cdots; x_n, p_n) \equiv (\cdots; x_i, p_i; \cdots),$$

(1.46)

where $p_i$ is the probability of outcome $x_i$ occurring; therefore $0 \leq p_i \leq 1$, and $\sum_{i=1}^{n} p_i = 1$. We also need compound risky gambles $(g_1, \alpha; g_2, 1 - \alpha)$ which yield the risky gamble $g_1$ with probability $\alpha \in [0, 1]$ and the risky gamble $g_2$ with probability $1 - \alpha$. Finally, we focus on expected utility - that is, there is an order preserving utility function $u$ such that

$$u(x_1, p_1; \cdots; x_i, p_i; \cdots; x_n, p_n) = \sum_{i=1}^{n} p_i u(x_i).$$

Both restrictions - to risky choice and expected utility - are because these are the only cases where probabilistic choice between gambles has been axiomatized. Clearly, much more research is needed in this area for uncertain (and ambiguous) gambles and representations other than expected utility.

In the following, we focus on binary choice - that is, for gambles $g, h$, $p(g, h)$ denotes\(^\text{11}\) the probability that $g$ is preferred to $h$. We assume that indifference is not allowed - that is, $p(g, h) + p(h, g) = 1$; Sections 1.3.3 and 1.4.2 discuss models and data for ternary (binary) choice - i.e., where indifference is allowed.

We begin with results from Blavatskyy (2008), then present related results by Dagsvik (2008) and Blavatskyy (2012), finishing with those of Gul and Pesendorfer (2006).

Assume that the set $X$ of outcomes is finite. As in Blavatskyy (2012), $L$ denotes the set of all risky gambles\(^\text{12}\).

\(^{11}\) Previously we used the notation $P_{xy}$. Our present notation is that used in the literature that we now summarize. It is also better for the relatively complex gamble structures we consider.

\(^{12}\) Blavatskyy (2008, 2012) uses the term lottery and Blavatskyy (2008) uses $\Lambda$ where we use $L$. 


Blavatskyy assumes five axioms.

**Axiom 1.63** *(Completeness).* For any two risky gambles, \( g_1, g_2 \in \mathcal{L} \),
\[ p(g_1, g_2) + p(g_2, g_1) = 1. \]

**Axiom 1.64** *(Strong stochastic transitivity).* For any three risky gambles, \( g_1, g_2, g_3 \in \mathcal{L} \), if \( p(g_1, g_2) \geq 1/2 \) and \( p(g_2, g_3) \geq 1/2 \), then
\[ p(g_1, g_3) \geq \max[p(g_1, g_2), p(g_2, g_3)]. \]

**Axiom 1.65** *(Continuity).* For any three risky gambles, \( g_1, g_2, g_3 \in \mathcal{L} \), the sets
\[ \{ \alpha \in [0,1] \mid p((g_1, \alpha; g_2, 1-\alpha), g_3) \geq 1/2 \} \]
and
\[ \{ \alpha \in [0,1] \mid p((g_1, \alpha; g_2, 1-\alpha), g_3) \leq 1/2 \} \]
are closed.

Intuitively, continuity ensures that a small change in the probability distribution over outcomes does not result in a significant change in the choice probabilities. In particular, continuity rules out lexicographic choices (e.g., the participant first eliminates all the gambles that do not have a nonzero probability of occurrence of the largest outcome across the set of available gambles).

**Axiom 1.66** *(Common consequence independence).* For any four risky gambles, \( g_1, g_2, g_3, g_4 \in \mathcal{L} \), and any probability \( \alpha \in [0,1] \),
\[ p[(g_1, \alpha; g_3, 1-\alpha), (g_2, \alpha; g_3, 1-\alpha)] = p[(g_1, \alpha; g_4, 1-\alpha), (g_2, \alpha; g_4, 1-\alpha)]. \]

This axiom says that if two risky gambles are identical in one component, then the choice probability is not affected by that component. Note that Axiom 1.66 is weaker than the condition
\[ p[(g_1, \alpha; g_3, 1-\alpha), (g_2, \alpha; g_3, 1-\alpha)] = p(g_1, g_2). \]

**Axiom 1.67** *(Interchangeability).* For any three risky gambles, \( g_1, g_2, g_3 \in \mathcal{L} \), if \( p(g_1, g_2) = p(g_2, g_1) = 1/2 \), then \( p(g_1, g_3) = p(g_2, g_3) \).

**Theorem 1.68** *(Blavatskyy, 2008, Th.1).* A function \( p : \mathcal{L} \times \mathcal{L} \rightarrow [0,1] \) satisfies Axioms 1.63-1.67 if there exist an assignment of a real number \( u_i \) to each outcome \( x_i, i = 1, \ldots, n \), and there exists a non-decreasing function \( \Psi : \mathbb{R} \rightarrow [0,1] \) such that for any two risky gambles \( g = (x_1, p_1; \cdots; x_i, p_i; \cdots; x_n, p_n) \)}
and \( h = (x_1, q_1; \cdots; x_i, q_i; \cdots; x_n, q_n) \) with \( g, h \in \mathcal{L} \),

\[
p(g, h) = \Psi \left( \sum_{i=1}^{n} u_i p_i - \sum_{i=1}^{n} u_i q_i \right).
\]

Luce and Suppes (1965, p. 360) call this representation a strong expected utility model, and present some evidence in favor of the special case where

\[
\Psi(a) = \begin{cases} 
\frac{1}{2} e^a & \text{if } a \leq 0 \\
1 - \frac{1}{2} e^a & \text{if } a \geq 0
\end{cases}.
\]

Dagsvik (2008) arrives at a more general representation by a somewhat related set of axioms, with the above representation a result of assuming a stronger version of his independence axiom. Dagsvik’s general representation (given in his Th. 2) is

\[
p(g, h) = \Psi [hV(g) - hV(h)],
\]

where

\[
V(g) = \sum_{i=1}^{n} u_i p_i \text{ and } V(h) = \sum_{i=1}^{n} u_i q_i,
\]

\( \Psi \) is a continuous and strictly increasing cumulative distribution function defined on \( \mathbb{R} \) with \( \Psi(r) + \Psi(-r) = 1 \), and \( h : \mathbb{R} \rightarrow \mathbb{R} \) is strictly increasing. Dagsvik (Th. 1) also states the uniqueness properties of \( \Psi, h, \) and \( V \).

Dagsvik and Hoff (2011) build on the extensive axiomatic work by Falmagne (1985, Chap. 14) (and the many other researchers that they mention) in the use of the methods of dimensional analysis in justifying specific functional forms in probabilistic models of choice. They illustrate their methods with the estimation of the utility of income using data from a Stated Preference (SP) survey.

Blavatskyy (2012) axiomatizes a representation related to that in Theorem 1.68 that includes natural endogenous reference points. Three of Blavatskyy’s (2012) axioms agree with his earlier axioms (Blavatskyy, 2008), with a minor change. The axioms are Axioms 1.63, 1.65, 1.66, with the minor change being that now the axioms are restricted to non-identical gambles - that is, in the new axioms \( p(g, g) \) is not defined for any gamble \( g \), whereas the 2008 axioms set \( p(g, g) = 1/2 \). Thus, in the following theorem statement, we implicitly assume that: Axiom 1.63 has \( g_1 \neq g_2 \); 1.65 has \((g_1, \alpha; g_2, 1 - \alpha) \neq g_3 \) for some \( \alpha \in [0, 1] \); and 1.66 has \( g_1 \neq g_2 \). For the following new axioms, we explicitly included the relevant non-identity constraints.
Axiom 1.69  (Weak stochastic transitivity). For any three risky gambles, \( g_1, g_2, g_3 \in \mathcal{L} \) with \( g_1 \neq g_2, g_2 \neq g_3, g_1 \neq g_3 \), if \( p(g_1, g_2) \geq 1/2 \) and \( p(g_2, g_3) \geq 1/2 \), then \( p(g_1, g_3) \geq 1/2 \).

Axiom 1.70  (Outcome monotonicity). For any pure consequences \( x_i, x_j \), \( i, j \in \{1, \ldots, n\} \), \( i \neq j \), and any \( \alpha \in [0, 1) \), \( p(x_i, (x_i, \alpha; x_j, 1 - \alpha)) \in \{0, 1\} \).

When \( \alpha = 0 \), Axiom 1.70 states that choice under certainty is deterministic. When \( \alpha \in (0, 1) \), the idea is that if \( x_i \succ x_j \), then \( p(x_i, (x_i, \alpha; x_j)) = 1 \) and if \( x_j \succ x_i \), then \( p(x_i, (x_i, \alpha; x_j, 1 - \alpha)) = 0 \).

Before stating the final axiom, we need the following. First, we assume that there is a weak order \( \succeq \) such that the finite set of pure outcomes satisfies\(^{13} \) \( x_1 \succeq x_2 \succeq \cdots \succeq x_n \). Also, we use the abbreviated notion of (1.46), i.e.,

\[
g = (x_1, p_1; \cdots; x_i, p_i; \cdots; x_n, p_n) \equiv (\cdots; x_i, p_i; \cdots).
\]

Definition 1.71  For risky gambles \( g = (\ldots; x_i, p_i; \ldots) \), \( h = (\ldots; x_i, q_i; \ldots) \), with \( g, h \in \mathcal{L} \), \( g \) weakly dominates \( h \) if, for every \( j \in \{1, \ldots, n\} \), \( \sum_{i=1}^j p_i \geq \sum_{i=1}^j q_i \).

Roughly speaking, \( g \) weakly dominates \( h \) if \( g \) is at least as likely as \( h \) to take on equal or larger outcomes with respect to the order \( \succeq \). We write \( g \succeq h \) if \( g \) weakly dominates \( h \) and \( g \not\succeq h \) if \( g \) does not weakly dominate \( h \).

Definition 1.72  For risky gambles \( g, h \in \mathcal{L} \),

1. A gamble \( k \in \mathcal{L} \) is a least upper bound of \( g \) and \( h \) if (i) \( k \succeq g \) and \( k \succeq h \), and (ii) if \( l \in \mathcal{L} \) is another gamble with \( l \succeq g \) and \( l \succeq h \), then \( l = k \);
2. A gamble \( k \in \mathcal{L} \) is a greatest lower bound of \( g \) and \( h \) if (i) \( g \succeq k \) and \( h \succeq k \), and (ii) if \( l \in \mathcal{L} \) is another gamble with \( g \succeq l \) and \( h \succeq l \), then \( l = k \).

When they exist, we denote a least upper bound of \( g, h \in \mathcal{L} \) by \( g \land h \) and a greatest lower bound by \( g \lor h \). Blavatskyy (2012, Propn. 3) shows that, when Axioms 1.63, 1.65, 1.66 (all restricted to non-identical gambles), 1.69, and 1.70 hold, \( (\mathcal{L}, \succeq) \) is a lattice - that is, for \( g, h \in \mathcal{L} \), there exists a least upper bound \( g \land h \) and a greatest lower bound by \( g \lor h \); the form of these bounds is given in the proof of his Proposition 3; also, he shows that, for \( g, h \in \mathcal{L} \),

\[
(g, 1/2; h, 1/2) \sim (g \lor h, 1/2; g \land h, 1/2).
\]

The final axiom corresponds to a weakening of Luce’s (1959) choice axiom.

\(^{13}\) Blavatskyy (2012) assumes the opposite ordering: \( x_n \succeq x_{n-1} \succeq \cdots \succeq x_1 \).
Axiom 1.73 (Odds ratio independence). Consider any three risky gambles, \( g_1, g_2, g_3 \in \mathcal{L} \) with \( g_1 \neq g_2, g_2 \neq g_3, g_1 \neq g_3 \), such that \( g_1 \preceq g_2, g_1 \preceq g_3, g_3 \preceq g_1 \) and \( g_1 \wedge g_2 = g_1 \wedge g_3 \). Then the odds ratio \( \frac{p(g_1, g_2)}{p(g_2, g_1)} / \frac{p(g_1, g_3)}{p(g_3, g_1)} \) is independent of \( g_1 \).

Blavatskyy (2012) illustrates Axiom 6 with the following example. Consider an individual choosing where to go for lunch. There are three restaurants close to his office: an Egyptian restaurant \((g_1)\) around the corner, and a Thai \((g_2)\) and a Swiss \((g_3)\) restaurant, each two blocks away. Then the relative probability that the Egyptian restaurant \((g_1)\) is chosen over the Thai restaurant \((g_2)\) compared to the relative chance that the Egyptian restaurant \((g_1)\) is chosen over the Swiss restaurant \((g_3)\) depends only on the characteristics of the Thai restaurant \((g_2)\) and the Swiss restaurant \((g_3)\), and hence not the characteristics of the Egyptian restaurant \((g_1)\). Axiom 1.73 is weaker than Luce’s (1959) choice axiom, with the latter as a special case: For the above example, when all the choice probabilities are nonzero, Luce’s axiom gives

\[
\frac{p(g_1, g_2)}{p(g_2, g_1)} / \frac{p(g_1, g_3)}{p(g_3, g_1)} = \frac{p(g_3, g_2)}{p(g_2, g_3)}.
\]

Thus, both Luce’s axiom and axiom 1.73 require the right hand term to be independent of \( g_1 \), with Luce’s axiom specifying the specific form of that term. Also, axiom 1.73 only applies to triples where the greatest lower bound on \( g_1 \) and \( g_2 \) is the same as that on \( g_1 \) and \( g_3 \).

Theorem 1.74 (Blavatskyy, 2012, Th.1) The binary choice probability function \( p : \mathcal{L} \times \mathcal{L} \rightarrow [0, 1] \) satisfies Axioms 1.63, 1.65, 1.66 and 1.69-1.73 (all restricted to non-identical gambles) iff for any two risky gambles \( g = (\ldots; x_i, p_i; \ldots), h = (\ldots; x_i, q_i; \ldots) \) with \( g, h \in \mathcal{L}, g \neq h \), there exists \( u : \mathcal{L} \rightarrow \mathbb{R} \) such that for \( k = (\ldots; x_i, r_i; \ldots) \in \mathcal{L} \)

\[
u(\ldots; x_i, r_i; \ldots) = \sum_{i=1}^{j} u(x_i) r_i
\]

with \( u(x_i) \geq u(x_j) \) iff \( x_i \succeq x_j \) and \( \varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) such that

\[
p(g, h) = \frac{\varphi(u(g) - u(g \wedge h))}{\varphi(u(g) - u(g \wedge h) + \varphi(u(h) - u(g \wedge h))} = \frac{\varphi(u(g \vee h) - u(h))}{\varphi(u(g \vee h) - u(h)) + \varphi(u(g \vee h) - u(g))}.
\]
This representation has several interesting properties, including: \( p(g, h) \geq 1/2 \) iff the expected utility of \( g \) is greater than the expected utility of \( h \); \( p(g, h) = 1 \) if \( g \) dominates \( h \); and the closer \( g \) is to the endogenous lower bound reference lottery \( g \land h \), the closer is \( p(g, h) \) to 0. Blavatskyy (2012) discusses, but does not axiomatize, extensions where the axioms would lead to \( u \) being a non-expected utility (for instance, a ranked weighted utility form as in Section 1.4.3) which would extend the description of additional classic (deterministic) phenomena to the probabilistic case.

A second application of this model would be to binary choice between multiattribute options (see Section 1.5.6 for work on choice between more than two such options for the above model and others). A further extension would involve choices between risky (or uncertain) gambles where the pure outcomes are multiattribute options. Bleichrodt et al. (2009) axiomatize a (deterministic) representation for uncertain gambles with multiattribute outcomes, obtaining a version of prospect theory (i.e., with referents on each attribute or option) with additive representation of the outcomes.

**Random choice rules for multiple choice**

We now present the axioms for **random expected utility** of Gul and Pesendorfer (2006). We slightly adapt the notation and terms in Gul and Pesendorfer (2006) to be more similar to that in the remainder of this chapter; however, for ease of cross-references to their paper, we include some of Gul and Pesendorfer’s terms in bold, with our parallel terms mentioned parenthetically. We have a finite set \( X \) of outcomes \( x_i, i = 1, ..., n + 1 \), and \( G \) denotes the set of risky gambles over those outcomes. Each nonempty finite set \( D \subseteq G \) represents the set of currently available options; the participant has to choose a single option from \( D \) (this is a decision problem). Let \( D \) denote the set of all decision problems. The probability that the participant chooses \( g \in D \) is denoted \( P_D(g) \); these probabilities across all \( D \in D \) form a **random choice rule** (equivalently, a collection of best choice probabilities, Definition 1.3). A **random utility function** is a probability measure \( \mu \) on some set of utility functions \( U \subseteq \{ u : X \rightarrow \mathbb{R} \} \). The set of choice probabilities \( P_D, D \in D \), is **representable by a** (distribution-free) **random utility function** \( U \) if for all \( g \in D \in D \),

\[
P_D(x) = \mu(u(g) > u(h) \forall h \in D, h \neq g).
\]

(see the corresponding definition in Definition 1.56 and related results in Theorem 1.61).
Gul and Pesendorfer (2006) pose and solve the question of when a set of subsets of choice probabilities over risky gambles are **representable by a random expected utility function**, i.e., representable by a random utility function \( U \) where each \( u \in U \) has the expected utility form; Luce and Suppes (1965, p. 361) present some evidence against such a representation. Gul and Pesendorfer (2006) show that the following properties of a random choice rule ensure that it is representable by a random expected utility function:

i. (Monotonicity). The probability of choosing \( g \) from \( D \) is at least as high as the probability of choosing \( g \) from \( D \cup \{h\} \).

In the psychology literature, this property is referred to as **regularity** (Luce and Suppes, 1965), and has been reported to fail in various contexts (Rieskamp et al. (2006)).

For any \( D, D' \in \mathcal{D} \) and \( \lambda \in [0, 1] \), let

\[
\lambda D + (1 - \lambda)D' := \{ \lambda g + (1 - \lambda)h | g \in D, h \in D' \}.
\]

Note that if \( D, D' \in \mathcal{D} \), then \( \lambda D + (1 - \lambda)D' \in \mathcal{D} \); using our earlier notation for gambles, we would define

\[
(D, \lambda; D', (1 - \lambda)) := \{(g, \lambda; g, (1 - \lambda)) | g \in D, h \in D' \}.
\]

ii. (Mixture continuous). For all \( D, D' \in \mathcal{D} \), \( P_{\alpha D + (1 - \alpha)D'} \) is continuous in \( \alpha \).

See Gul and Pesendorfer (2006) for the technical details on the definition of continuity in this framework.

iii. (Linearity). For each \( \lambda \in [0, 1] \), the probability of choosing \( g \) from \( D \) is the same as the probability of choosing \( \lambda g + (1 - \lambda)h \) from \( \lambda D + (1 - \lambda)\{h\} \).

The **extreme points of a decision problem** are those options (risky gambles) in that problem that are the unique maximum for some expected utility function.

iv. (Extreme). For each decision problem, the chosen lottery is an extreme point.

Finally, before stating the result, a **regular** random utility function is one where in each decision problem, with probability 1, the realized utility function has a unique maximizer. In the psychology literature, this property is commonly referred to as **noncoincident** (see, e.g., Regenwetter and Marley, 2001).

We then have (Gul and Pesendorfer, 2006, Th.2):
Theorem 1.75 A random choice rule is representable by some regular (finitely additive) random expected utility function if and only if the random choice rule is monotone, mixture continuous, linear, and extreme.

Gul and Pesendorfer (2006, Supplement) extend the result to nonregular random utility functions.

1.5.3 Horse race models of choice and response time

We develop general results on horse race models of choice and response time for best choices, only, for three reasons: first, the developments for worst choices exactly parallel those for best choices; second, those for best-worst choices are much more complicated; third, although special cases yield numerous standard representations for choice probabilities, the models are unable to fit various context effects in choice and have restrictive relations between choice and response times. Thus, we later both restrict the models in certain ways, and extend them in others, to make them tractable and realistic as models of both choice and response time.

Distribution-free horse race models

Assume a master set $\mathcal{A}$ of options, and let $X \subseteq \mathcal{A}$ be any subset with two or more options. For some purposes, it is useful to enumerate $\mathcal{A}$ as $\mathcal{A} = \{c_1, c_2, \ldots, c_n\}$ where $n = |\mathcal{A}|$, and a typical $X \subseteq \mathcal{A}$ as $X = \{x_1, x_2, \ldots, x_m\}$ where $n \geq m = |X|$.

For each $x \in X \subseteq \mathcal{A}$, $T_X$ is a nonnegative (real) random variable denoting the time at which a choice is made, $B_X$ is a random variable denoting the (single “best”) option chosen, and $B_X(x; t), t \geq 0$, is the probability that option $x$ is chosen as best in $X$ after time $t$. Thus

$$B_X(x; t) = \Pr[B_X = x \text{ and } T_X > t]$$

may be thought of as a survival function for option $x$ when presented in the set $X$. A collection of survival functions $\{B_X(x; t) : x \in X \subseteq \mathcal{A}\}$ for a fixed $X \subseteq \mathcal{A}$ is a joint structure of choice probabilities and response times. A joint structure is complete if the collection of survival functions ranges over all subsets of $\mathcal{A}$; unless otherwise specified, we assume that the joint structure is complete.

We now investigate a class of models generating such a joint structure, all of which belong to the class of (possibly context dependent) horse race

---

14 The notation has to be generalized for worst and best-worst choice, and for other designs, such as where the person can select a subset of options to consider (a consideration set, as in Swait, 2001).
random utility models (cf. Marley, 1989; Marley and Colonius, 1992, those papers focus on the context independent case - see below). We will see that this class of models is sufficiently broad to include many of the standard models of choice used in the discrete choice literature, though they are inadequate for context effects on choice and the details of response time (see Busemeyer and Rieskamp, 2013; Rieskamp et al., 2006, for summaries of the limitations of the choice predictions of these models). Section 1.5.5 discusses context dependent extensions of these models that overcome numerous of their limitations.

For each \( z \in X \subseteq A \), let \( T_X(z) \) be the time for a process (e.g., an “accumulator,” as detailed in Section 1.5.4) associated with \( z \) to produce a specified event (e.g., to reach a “threshold”) when \( X \) is set of available choice options. When we give details for specific models, \( T_X(z) \) will depend on various parameters such as a threshold; drift rates; drift rate variability; start point variability; etc. There is no assumption that the \( T_X(z) \) are independent of each other; however, we will see that the models become vacuous when their dependence on \( X \) is not restricted in some manner.

Consider a generic set \( X = \{x_1, ..., x_m\} \subseteq A \) with \( |X| = m \) and assume that there is a multivariate (survival) distribution \( S_X \) such that: for \( t_i \geq 0, i = 1, ..., m \),

\[
S_X(t_1, ..., t_n) = \Pr(T_X(x_1) > t_1, ..., T_X(x_m) > t_m). \quad (1.47)
\]

In particular, there can be (complicated) dependencies between the \( T_X(x_i) \), \( i = 1, ..., m \). We use the survival form as various results are easier to state and prove in that form, plus that is the way the main results were stated in earlier papers (Marley, 1989; Marley and Colonius, 1992).

Given \( X \subseteq A \) as the currently available choice set, let \( B_X \) be the option chosen and \( T_X \) the time it is chosen. Then we assume that, for \( x \in X \) and \( t > 0 \),

\[
B_X(x; t) = \Pr(B_X = x \text{ and } T_X > t) = \Pr(t < T_X(x) = \min_{z \in X} T_X(z)), \quad (1.48)
\]

where the distributions are given by the \( S_X \) in (1.47) The above expressions give the joint choice and (tail) response time (RT) distribution. The choice probabilities are given by (1.48) with \( t = 0 \), i.e.,

\[
B_X(x) = \Pr(B_X = x) = \Pr(T_X(x) = \min_{z \in X} T_X(z)). \quad (1.49)
\]

In the general form (1.48), the \( T_X(z) \) can depend on \( X \), i.e., they can be context dependent, and/or they can be dependent on each other. When
none of the $T_X(z)$ in (1.48) depend on $X$, we have a **context free** horse race random utility model (cf. Marley and Colonius, 1992), and when they are independent of each other, we have an **independent horse race random utility model**; note that all four combinations of context free and independent are possible. Parallel terms apply to the choice probabilities in (1.49).

There are two major limitations of context free horse random utility models, namely that:

1. $\Pr(\min_{z \in X} T_X(z) > t) \geq \Pr(\min_{z \in X \cup \{y\}} T_X(z) > t)$,
2. $B_X(x; t) \geq B_{X \cup \{y\}}(x; t)$.

The first relation means that response times do not increase with set size (Colonius and Vorberg, 1995, Lemma 2) and the second that the choice probabilities satisfy **regularity**: the probability of choosing an option cannot increase when an additional option is added to the choice set. Both of these properties have been reported to fail in various experiments (Busemeyer and Rieskamp, 2013; Teoderescu and Usher, 2013). These and other limitations are addressed when we consider context dependent horse race models (Section 1.5.5).

The following standard results hold for the class of horse race random utility models for best choice; parallel results can be developed for worst, and best-worst, choice. For simplicity, we assume that $B_X(x; t) > 0$ for all $t \geq 0$, though the results are easily generalized to various weaker conditions (see Marley, 1992); and that all distributions are absolutely continuous, though, again, various of the results are valid when this is not the case (see the study by Marley, 1989, of Tversky’s elimination by aspects (EBA) model).

**Theorem 1.76** (Marley and Colonius, 1992, Th. 1) (a) Consider a joint structure of (best) choice probabilities and response times $\{B_X(x; t) : x \in X\}$ on a fixed finite set $X$. If each $B_X(x; t)$, $x \in X$, is absolutely continuous and positive for all $t \geq 0$, then there exist unique independent random variables $t_X(x)$, $x \in X$, such that (1.48) holds with those random variables.

(b) A (complete) joint structure of choice probabilities and response times $\{B_X(x; t) : x \in X \subseteq A\}$ can be uniquely represented by an independent horse race random utility model if the conditions of (a) hold and

$$\frac{(d/dt)B_X(x; t)}{\sum_{z \in X} B_X(z; t)}$$

is independent of $X \subseteq A$ for all $t \geq 0$. 
Essentially, (a) says that any set of (best) choice probabilities and response times on a fixed finite set $X$ can be represented by an independent horse race random utility model, and thus, indirectly, that any (complete) set of (best) choice probabilities and response times on (all) subsets of finite set $A$ can be fit by a context dependent horse race random utility model with the random variables independent within each choice set; though the random variables may differ across subsets. Part (b) gives a set of conditions under which a (complete) joint structure of choice probabilities and response times can be written as a context independent horse race random utility model with a common set of independent random variables across all the choice sets. These results can be interpreted as saying that the assumption that a horse race random utility model with independent random variables holds for a single choice set $X$ is a descriptive theoretical language rather than an empirically falsifiable model; Dzhafarov (1993) and Jones and Dzhafarov (2013) discuss related issues when response time distributions are modeled by a deterministic stimulation-dependent process that terminates when it crosses a randomly preset criterion, a framework that is closely related to that of the linear ballistic accumulators (LBAs) presented in Section 1.5.4.

Given the above results, one can fit “any” choice and response time data using a context dependent horse race random utility model with the random variables independent within each choice set; in particular, one can fit the classic context effects in choice (see Section 1.5.5 for those effects). Thus, the goal of context dependent models must be to motivate plausible constraints on the nature of the context dependencies - see Section 1.5.5 for recent models with such constraints.

**Luce’s choice model derived from a horse race model**

The following results are stated for best choices; exactly parallel results hold for worst and best-worst.

Before proceeding, we need a definition:

**Definition 1.77** A complete set of choice probabilities on the subsets of a finite set $A$ satisfy **Luce’s choice model** if all the choice probabilities are nonzero and there exists a ratio scale $b$ such that for every $x \in X \subseteq A$ with $|X| \geq 2$,

$$B_X(x) = \frac{b(x)}{\sum_{z \in X} b(z)}.$$  \hspace{1cm} (1.50)

This model is equivalent to the **multinomial logit (MNL)** model,
which is usually written with scale values \( u(z) = \log b(z) \), i.e., \( b(z) = \exp u(z) \).

For simplicity in the following theorem, we assume that all the choice probabilities are nonzero. The result can be generalized when this is not so by adding a connectivity and a transitivity condition (Luce, 1959, Th.4., p.25).

**Theorem 1.78** (Marley and Colonius, 1992, Th.2). Consider an independent, context independent horse race random utility model where for each \( x \in X \subseteq A \), \( B_X(x; t) \) is absolutely continuous and positive for all \( t \geq 0 \) and \( B_X(x) \) is nonzero. If the option chosen, \( C_X \), is independent of the time of choice, \( T_X \), then the choice probabilities satisfy Luce’s choice model, (1.50).

This derivation of Luce’s choice model is theoretical interesting. However, the resultant model of choice and response times is, in general, unsatisfactory, as it implies that the response time distribution conditional on a particular choice is independent of that choice. That is, the assumption of Theorem 1.78 is that

\[
\Pr(B_X = x \text{ and } T_X > t) = \Pr(B_X = x) \Pr(T_X > t).
\]  

From this, we obtain that

\[
\Pr(T_X > t | B(X = x)) = \frac{\Pr(B_X = x \text{ and } T_X > t)}{\Pr(B_X = x)} = \frac{\Pr(B_X = x) \Pr(T_X > t)}{\Pr(B_X = x)} = \Pr(T_X > t).
\]

This result says that the distribution of response times is independent of the option chosen, which is generally incorrect - for example, the most preferred (“correct”) option is often chosen faster than less preferred options (though experimental manipulations can give the opposite). Since there are considerable data showing that the Luce choice model (also known as the multinomial logit (MNL) model) does not provide a satisfactory description of various data (Busemeyer and Rieskamp, 2013; Rieskamp et al., 2006), the above result should not concern us if it does not hold when the choice probabilities have a more general form than the Luce choice model. When the MNL model was found to be inadequate, it was generalized to a class of dependent random utility models that gave more general closed form representations for choice probabilities - namely, the generalized extreme value (GEV) class (Marley, 1989; McFadden, 1978). Marley (1989)
showed that there is a class of dependent horse race random utility models for choice and responses times that generates that class of models (and related representations) for choice probabilities. However, regrettably, that class continues to have the property that the choices made and the time to make them are independent. Although models in the more general class have the major negative property of the probability of the choice made being independent of the time to make it - that is, (1.51) holds - Marley (1989) also shows that we can overcome that limitation by considering mixtures of models of the above form; to date, the properties of the latter class of models have not been studied in detail. A further extension would be to add context dependence, which is relatively easily done in the parametric representations of the class of models presented in Marley (1989). Rather than pursue that direction, we return to the most basic models in this class; reinterpret them in terms of accumulators with thresholds; summarize recent context dependent models in that framework; and extend them to worst and best-worst choice.

1.5.4 Context free linear accumulator models of choice and response time

The class of models that we studied in Section 1.5.3 have the property that the option chosen is independent of the time of choice, (1.51). Note that if this condition holds, then if we let \( B_X(x, t) \) be the probability that option \( x \) is chosen at or before time \( t \), then we have

\[
B_X(x, t) = \Pr(B_X = x \text{ and } T_X \leq t)
\]

\[
= \Pr(B_X = x) - \Pr(B_X = x \text{ and } T_X > t)
\]

\[
= \Pr(B_X = x) - \Pr(B_X = x) \Pr(T_X > t)
\]

\[
= \Pr(B_X = x)[1 - \Pr(T_X > t)]
\]

\[
= \Pr(B_X = x) \Pr(T_X \leq t).
\]

Thus we have the independence property in terms of the cumulative distribution, as well as for the survival distribution. Since most response time models are written in terms of cumulative distributions, we will use that form in the present section, even though we study models that do not satisfy the above independence property.

**Context free accumulator models** are essentially horse race random utility models with additional structure imposed on the properties of the random variables; Section 1.5.5 presents context dependent accumulator.
models. Here, we discuss the linear ballistic accumulator (LBA) model (Brown and Heathcote, 2008). Like other multiple accumulator models, the LBA is based on the idea that the decision maker accumulates evidence in favor of each choice, and makes a decision as soon as the evidence for any choice reaches a threshold amount. The time to accumulate evidence to threshold is the predicted decision time, and the response time is the decision time plus a fixed offset ($t_0$), the latter accounting for processes such as response production. The LBA shares this general evidence accumulation framework with many models (Busemeyer and Rieskamp, 2013) but has a practical advantage - namely, the computational tractability afforded by having an easily computed expression for the joint likelihood of a given response time and response choice amongst any number of options.

Figure 1.5 gives an example of an LBA decision between two options A and B. The A and B response processes are represented by separate accumulators that race against each other. The vertical axes represent the amount of accumulated evidence, and the horizontal axes the passage of time. Response thresholds ($b$) are shown as dashed lines in each accumulator, indicating the quantity of evidence required to make a choice. The amount of evidence in each accumulator at the beginning of a decision (the “start point”) varies independently between accumulators and randomly from choice to choice, usually assumed sampled from a uniform distribution: $U(0, A)$, with $A < b$. Evidence accumulation is linear, as illustrated by the arrows in each accumulator of Figure 1.5. The speed of accumulation is traditionally referred to as the “drift rate”, and this is assumed to vary randomly from accumulator to accumulator and decision to decision according to an independent distribution for each accumulator, reflecting choice-to-choice changes in factors such as attention and motivation. Rate means, $d$, and standard deviations, $s$, can differ between accumulators and options, although at least one value of $s$ is usually fixed as a scaling constant.

---

In the following, we consider two classes of LBA - both assume subtractive start point variability, but one model assumes multiplicative drift rate variability, the second assumes additive drift rate variability; both classes assume the random variables $T_X(x)$ are independent. The first class relates in a natural way to the models in Section 1.5.3, the second class agrees with the assumptions in the original development of the LBA.

This section is confined to context independent models; Section 1.5.5
presents context dependent models. Also, to agree with published papers, we use $B_x$ to denote the distribution of time for option $x$ to reach threshold, when the task is to choose the best of the available options.

**Multiplicative drift rate variability**

In the following summaries, we ignore the fixed offset ($t_0$). Let $\Delta$ be a random variable on the positive reals; this will be the distribution of the (multiplicative) drift rate variability. Let, for $r > 0$,

$$\Pr(\Delta < r) = G(r), \quad (1.52)$$

where $G$ is a cumulative distribution function (CDF). Let $b$ be the threshold and let $x$ be a typical choice option with drift rate $d(x)$, and let $\Sigma$ be a second random variable on $[0, A]$ with $A \leq b$; this will be the distribution of the start point. Let, for $h \in [0, A]$,

$$\Pr(\Sigma < h) = H(r).$$

For option $x$, the start point has distribution $\Sigma$ and the drift rate has distribution $\Delta d(x)$; also, assume the samples of $\Sigma$ and $\Delta$ for different $x$’s are independent, which we denote by $\Sigma_x$ and $\Delta_x$; thus, the subscripts do not imply dependence on $x$ beyond the assumption of independent samples for different $x$’s. Then, with $B_x = \frac{b - \Sigma_x}{\Delta_x d(x)}$ the random variable for the time taken for the LBA with (mean) drift rate $d(x)$ to reach threshold, we have

$$\Pr(B_x < t) = \Pr(\frac{b - \Sigma_x}{\Delta_x d(x)} \leq t). \quad (1.53)$$

The major LBA applications to date have assumed that the starting point variability is given by a random variable $p$ with uniform distribution on $[b - A, b]$, in which case (1.53) can be rewritten as

$$\Pr(B_x < t) = \Pr(\frac{p_x}{\Delta_x d(x)} \leq t). \quad (1.54)$$

The above representation is a perfectly plausible multiplicative LBA model for any cumulative distributions $G$ on the nonnegative reals and $H$ on $[0, A]$ with $A \leq b$. However, a major argument advanced for LBA models is their computational tractability in terms of probability density functions (PDF) and cumulative density functions (CDFs). In particular, it is usually assumed that the accumulators are independent, and therefore the main need is for the PDF and CDF of, say, (1.53) to be tractable. We write out the CDF corresponding to that expression, then summarize what is known about such forms.
With \( p \) the uniform distribution on (1.54), and the CDF \( G \) given by (1.52) with PDF \( g \), we have

\[
\Pr(B_x < t) = \Pr(\frac{p_x}{\Delta_x d(x)} < t) = \int_{-\infty}^\infty U(u|b - A, b) \Pr(\Delta_x d(x) = \frac{u}{t}) \, du
\]

\[
= \int_{b-A}^b \frac{u - b + A}{A} g\left(\frac{u}{td(x)}\right) \, du + [1 - G\left(\frac{b}{td(x)}\right)]
\]

\[
= \frac{1}{A} \int_{b-A}^b u g\left(\frac{u}{td(x)}\right) \, du + \frac{A - b}{A} \int_{b-A}^b g\left(\frac{u}{td(x)}\right) \, du + [1 - G\left(\frac{b}{td(x)}\right)]
\]

\[
= \frac{1}{Atd(x)} \int_{\frac{b}{td(x)}}^{\frac{b-A}{td(x)}} s g(s) \, ds + 1 + \frac{b}{A} \left[ G\left(\frac{b}{td(x)}\right) - G\left(\frac{b - A}{td(x)}\right) \right] + [1 - G\left(\frac{b}{td(x)}\right)]
\]

The integral in the first term is the mean of the distribution \( G \) truncated to the interval \([\frac{b}{td(x)}, \frac{b-A}{td(x)}]\). Closed forms are known for this quantity for various CDFs on the nonnegative reals, including for Gamma, inverted Beta, Fréchet, and Levy distributions (Nadarajah, 2009). However, one of the major motivations for the LBA framework was to achieve easily computable forms for the joint distribution of choice probabilities and choice times. In addition to independence, this requires the PDFs corresponding to the CDFs, above, to be easily computable. Ongoing research is exploring the extent to which this is the case for various distributional assumptions. For example, Heathcote and Love (2012) develop and test the lognormal race (LNR), which has the form: for each option \( x \),

\[
\Pr(B_x < t) = \Pr(\frac{S_x}{D_x} < t),
\]

where \( S_x \) is the distribution to the boundary and \( D_x \) is the drift rate distribution, with each being lognormally distributed - that is, with \( \Phi \) the cumulative normal, there are (mean) parameters \( \beta(x) \) and \( \delta(x) \), and positive (scale) parameters \( s(x) \) and \( \sigma(x) \) such that, for \( r > 0 \),

\[
\Pr(S_x < r) = \Phi\left(\frac{\ln r - \beta(x)}{s(x)}\right),
\]
and
\[ \Pr(D_x < r) = \Phi \left( \frac{\ln r - \delta(x)}{\sigma(x)} \right). \]

Note that when \( \sigma(x) = 1 \) for all \( x \), the drift rate random variable can be written in the multiplicative form \( D_x \sim \Delta e^{\delta(x)} \), with, for \( t > 0 \)
\[ \Pr(\Delta < t) = \Phi (\ln t). \]

The PDF's of this model are easily computable and the model does not imply that the choice made is independent of the time of choice.

Later, we describe the original LBA model, which has additive drift rate variability, in which case the PDFs are also easily computable and the model does not imply that the choice made is independent of the time of choice. However, before proceeding to that material, we take a step back to present a model with multiplicative drift rate variability that is equivalent (for choices) to the most basic, and frequently applied, random utility models of best, worst, and best-worst choice.

Relation of multiplicative LBA models with no start point variability and Fréchet drift rate variability to MNL models

We now present a multiplicative LBA with no start point variability and Fréchet drift scale variability that leads to a standard set of models for best, worst, and best-worst choice. Consider the special case of (1.54) where there is no start point variability, i.e., \( A = 0 \), and so \( p \) has a constant value \( b \); in this case, without loss of generality, we can set \( b = 1 \). Then (1.54) reduces to
\[ \Pr(B_x < t) = \Pr(\frac{1}{\Delta d(x)} \leq t). \]

We also assume that \( \Delta \) has a Fréchet distribution, i.e., there are constants \( \alpha, \beta > 0 \) such that, for \( r \geq 0 \),
\[ \Pr(\Delta < r) = G(r) = e^{-(\alpha r)^{-\beta}}. \] (1.55)

The following is then an obvious manner to obtain a set of models for best, worst, and best-worst choice, respectively, and the corresponding response times.\(^15\) As above, \( B_X(x, t) \) denotes the probability of choosing \( x \) as the best option in \( X \) before time \( t \). The corresponding notation for worst is \( W_X(y, t) \) and for best-worst is \( BW_X(x, t; y, t) \).

We then assume:

\(^{15}\) Remember that the subscript on \( \Delta_z, \Delta_{x,y}, \) etc., indicates independent samples, not other dependence on \( z \) or \( x, y \).
i. Best, with drift rates \(d(z), z \in X\):

\[
B_X(x, t) = \Pr \left(\frac{1}{\Delta_x d(x)} = \min_{z \in X} \frac{1}{\Delta_z d(z)} \leq t\right),
\]

(1.56)

ii. Worst, with drift rates \(1/d(z), z \in X\):

\[
W_X(y, t) = \Pr \left(\frac{1}{\Delta_y d(y)} = \min_{z \in X} \frac{1}{\Delta_z d(z)} \leq t\right),
\]

(1.57)

iii. Best-worst, with drift rates \(d(p)/d(x)\) and for all \(p, q \in X, p \neq q\),

\[
BW_X(x, t; y, t) = \Pr \left(\frac{1}{\Delta_{x,y} d(x)} = \min_{p,q \in Y} \frac{1}{\Delta_{p,q} d(p)} \leq t\right) (x \neq y).
\]

(1.58)

Note that (1.58) implies that the best and the worst option are chosen at the same time; this is an unreasonably strong assumption, which we weaken later.

For \(t \geq 0\), let \(F(t) = 1 - \exp - \left(\sum_{z \in X} d(z)^\beta\right) (\alpha t)^\beta\). Then routine calculations (paralleling those in Marley (1989)) give: for \(X \subseteq A\) and \(x, y \in X\),

\[
B_X(x, t) = \frac{d(x)^\beta}{\sum_{z \in X} d(z)^\beta} F(t),
\]

(1.59)

\[
W_X(y, t) = \frac{1/d(y)^\beta}{\sum_{z \in X} 1/d(z)^\beta} F(t),
\]

(1.60)

and

\[
BW_X(x, t; y, t) = \frac{d(x)^\beta/d(y)^\beta}{\sum_{r,s \in X, r \neq s} d(r)^\beta/d(s)^\beta} F(t) (x \neq y).
\]

(1.61)

The corresponding choice probabilities \(B_X(x), W_X(y), BW_X(x, y)\) are given by these formula in the limit as \(t \rightarrow \infty\), i.e., as \(F(t) \rightarrow 1\). Now, for \(z, p, q \in A, p \neq q\), let \(u(z) = \ln d(z), \epsilon_z = \ln \Delta_z, \epsilon_{p,q} = \ln \Delta_{p,q}\). Then the choice probabilities can equally well be written as: for all \(x, y \in X \in D(A)\),

\[
B_X(x) = \Pr \left(\epsilon_y + \epsilon_z = \max_{z \in X} [u(z) + \epsilon_z]\right),
\]

(1.62)
\[ W_X(y) = \Pr \left( -u(y) + \epsilon_y = \max_{z \in X} [-u(z) + \epsilon_z] \right), \quad (1.63) \]

and for all \( x, y \in X \in D(A), x \neq y, \)

\[ BW_X(x, y) = \Pr \left( u(x) - u(y) + \epsilon_{x,y} = \max_{p \neq q \in X} [u(p) - u(q) + \epsilon_{p,q}] \right). \quad (1.64) \]

However, given that \( \Delta_z \) and \( \Delta_{p,q} \) are generated by Fréchet drift rate variability, i.e., \( (1.55) \), we have that \( \epsilon_z = \ln \Delta_z \) and \( \epsilon_{p,q} = \ln \Delta_{p,q} \) satisfy extreme value distributions\(^{16}\). When treated as a single model, the three models \( (1.62), (1.63), \) and \( (1.64) \), then satisfy an inverse extreme value maximum\(^{17}\) random utility model (Marley and Louviere, 2005, Def.11). Standard results (summarized by Marley & Louviere, 2005\(^{18}\), and the derivations above) show that the expression for the choice probabilities given by \( (1.62) \) (respectively, \( (1.63), (1.64) \)) agrees with standard multinomial logit (MNL) forms. Note that \( \beta \) is not identifiable from the choice probabilities, but it is, in general, identifiable when response times are also available (though the form predicted for the response time distribution is not suitable for data).

In particular, when \( (1.59) \) and \( (1.60) \) both hold with \( F(t) \rightarrow 1 \), we have that for all \( x, y \in X, x \neq y, B_{(x,y)}(x) = W_{(x,y)}(y) \); empirically, this relation may not always hold (Shafir, 1993).

As already noted, each of these models (for best, worst, and best-worst choice) has the property that the option chosen is independent of the time of choice; that property would not hold for (most) distributions other than the Fréchet for drift scale variability. Also, though the underlying utility maximization process may be cognitively plausible for best (or worst) choices, that for best-worst choice in \( (1.58) \) appears to require the participant to “simultaneously” compare all possible discrete pairs of options in the choice set, leading to a significant cognitive load; as already noted, it also implies that the best and the worst option are chosen at the same time. Fortunately, Marley and Louviere (2005) present the following plausible process involving separate best and worst choice processes that generates the same choice probabilities: the person chooses the best option in \( X \) and, independently, chooses the worst option in \( X \); if the resulting choices differ, then they are given as the responses; otherwise, the person repeats the process until the

\(^{16}\) With, in this case: for \(-\infty < t < \infty \) \( \Pr(\epsilon_z \leq t) = \exp - (\alpha e^t)^{-\beta} \) and \( \Pr(\epsilon_{p,q} \leq t) = \exp - (\alpha e^t)^{-\beta} \).

\(^{17}\) We have added maximum to Marley & Louviere’s definition to emphasize that the random utility models of choice are written in terms of maxima, whereas the equivalent (“horse race”, accumulator) models of response time are written in terms of minima.

\(^{18}\) Marley & Louviere used the case \( \alpha = \beta = 1 \).
selected pair of options do differ. Then Marley and Louviere (2005, Sect. 4.1.2, Case 2, and above) show that if the best (resp., worst) choices satisfy the MNL for best, (1.59) with $F(t) \rightarrow 1$ (resp., MNL for worst, (1.60) with $F(t) \rightarrow 1$) choices, with $v = -u$, then the best-worst choices given by the above process satisfy the maxdiff model, (1.61) with $F(t) \rightarrow 1$; in particular, we have a more cognitively plausible model than (1.64) for the best-worst choices.

**Additive drift rate variability**

In contrast to the multiplicative drift rate model presented above, the linear ballistic accumulator (LBA) model assumes additive drift rate variability generated by independent normal random variates, truncated at zero - i.e., constrained to be nonnegative\(^\text{19}\). This model is a horse race random utility model, but does not have the undesirable property that the choices made and the time to make them are independent. Nonetheless, it has been shown to make extremely similar predictions for best, worst, and best-worst choice probabilities to those made by the multiplicative LBA model with no start point variability and Fréchet drift scale variability (Hawkins et al., 2013a,b). The derivations for this model exactly parallel those described above with $\Delta d(x)$ replaced by $\text{trunc}(Dz + d(z))$, with a major advantage of the additive model being that all formulae for cumulative distribution functions (CDFs) and probability distribution functions (PDFs) are computationally very tractable when the start point distribution is uniform. We now briefly summarize recent applications of this model to some standard data on best, worst, and best-worst choice and in Section 1.5.5 summarize its extension to handle context effects in best choice and response times.

Hawkins et al. (2013a,b) fit four best-worst LBA models to each of three sets of data - two sets of best-worst choice data obtained in DCEs, the first involving choice between aspects of dermatology appointments, the second between mobile phones; and one set of best-worst choice and response time data in a perceptual judgment task: choosing the rectangle with the largest (resp., smallest) area in a set of four rectangles presented on a trial, with multiple presentations of various sets of rectangles. Each additive LBA model is based on processes that follow naturally from a corresponding MNL model, with the latter as described above. In each fit of the models to only the best and/or worst choices, an essentially linear relationship was found between log drift rate and utility estimates for the multinomial logit (MNL)

\(^{19}\) In their data analyses, Brown and Heathcote (2008) ignored the small percentage of times that the normal distributions could take on negative values; Heathcote and Love (2012) present the correct analysis in terms of truncated normals.
model; and goodness of fit, as determined by log-likelihood and root mean-square error, was comparable. However, when both choice and response time data are fit, three of the models are inappropriate: one because it implies that best choices are always made before worst choices; the second because it implies the reverse; and the third because it implies that the best and worst choices are made simultaneously. The relevant one of the first two models might be appropriate in experiments where participants are forced, by the design, to respond in the order best, then worst (respectively, worst, then best); however, the third is implausible (for response times) under any reasonable circumstances. Participants in the perceptual judgment task were free to make the best and worst choice in whatever order they wished. The data show: large differences between participants in best-then-worst versus worst-then-best-responding, but normally with both orders for each participant; changes in the proportion of response order as a function of choice difficulty; and changes in inter-response time due to choice difficulty. These data are incompatible with three of the models, and, as Hawkins et al. show, they cannot be satisfactorily fit by mixtures of those models. Therefore, we restrict our (brief) presentation to the fourth model, called the parallel (additive) LBA model, which is an extension of Marley and Louviere’s (2005, Sect. 4.1.2, Case 2) process model for choice to response time.

The parallel (additive) LBA model for best-worst choice assumes concurrent best and worst races, with, in general, different (drift rate) parameters in the two sets of races; Hawkins et al. (2013b) data analysis assumed that the worst drift rate is the reciprocal of the corresponding best race. The option chosen as best (resp., worst) is that associated with the first accumulator to reach threshold in the best (resp., worst) race. Hawkins et al. (2013b) presented, and tested, an approximation to the full model, in that they allow the same option to be selected as both best and worst, which was not allowed in their experiment; their model also allows for vanishingly small inter-response times, which are not physically possible. These predictions affect a sufficiently small proportion of their data that they were acceptable. The parallel model overcomes the drawbacks of the other three models described above by accounting for all general choice and response time trends observed in data - for instance, the model is able to capture inter- and intra-individual differences in response style, reflected by the data of those participants who prefer to respond first with the best option, or first with the worst option.

The fact that data on choice could be well-fit by each the four best-worst LBA models (Hawkins et al., 2013a) but data on both choice and response
time could only be well-fit that the parallel LBA model (Hawkins et al., 2013b) gives support to the trend in consumer research to collect both the choices made and the time to make them.

1.5.5 Context dependent models of choice and response time

As far as we know, there is no formal definition of what it means for a set of (theoretical or empirical) choice probabilities to demonstrate a context effect. The general usage of that term appears to be that the choice probabilities are not consistent with either a random utility or a constant utility representation as defined in Chapter 10 of Suppes et al. (1989). We do not attempt such a formal definition here, but simply present some examples of standard (choice) context effects and of models that explain them.

We discuss three (context) effects, stated in terms of theoretical choice probabilities but which have also been demonstrated in data, that are incompatible with either a random utility model, or a constant utility model, or both. We then present the multiattribute LBA (MLBA) model (Trueblood et al., 2013a), for which the three context effects hold simultaneously for various sets of parameter values, and summarize the 2N-ary tree model for N-alternative preferential choice (Wollschläger and Diederich, 2012); both of these models also predict response times. We then summarize three models of choice, only, that predict some, or all, of the context effects.

Busemeyer and Rieskamp (2013); Rieskamp et al. (2006) summarize the following three context effects, and others, and the extant models that are, or are not, compatible with those effects holding; they do not discuss the MLBA or the 2N-ary tree model.

Three standard context effects

Following Trueblood et al. (2013a), we present the context effects in terms of choice sets with three multiattribute options, though often some of them are stated in terms of a combination of two- and three-option choice sets. These effects have traditionally been observed in between subject designs; however, recent work demonstrates them within subjects in an inference task (Trueblood, 2012) and in a perceptual and a preference task (Trueblood et al., 2013b). We use slightly more general notation for psychological (subjective) representations than Trueblood et al. (2013a) and describe their specific assumptions, as needed.

We restrict attention to options with two attributes; the theoretical formulation generalizes naturally to $m \geq 2$ attributes.
Let \( r = (r_1, r_2) \) denote a typical option, where \( r_i \) is an indicator for its value on attribute \( i \); this value could be quantitative (e.g., price, miles per gallon) or qualitative (e.g., camera/no camera on a mobile phone). There are preference orders \( \succeq_i \) and real valued functions \( u_i \) such that for options \( x \) and \( z \), \( x_i \succeq z_i \) iff \( u_i(x_i) \geq u_i(z_i) \). We say that \( x \) strictly dominates \( z \) iff \( x_i \succ z_i \) for \( i = 1, 2 \), and, equivalently say that \( z \) is (strictly) inferior to \( x \).

We now give examples of three of the main context effects.

The attraction effect: Consider \( \{x, z\} \) and a decoy \( a_x \) (respectively, \( a_z \)) that is similar, but inferior, to option \( x \) (respectively, \( z \)). Then the attraction effect occurs when the probability of choosing \( x \) is greater when the decoy is similar to \( x \) than when the decoy is similar to \( z \) and vice versa:

\[
B_{\{x, z, a_x\}}(x) > B_{\{x, z, a_z\}}(x) \text{ and } B_{\{x, z, a_z\}}(z) > B_{\{x, z, a_x\}}(z)
\]

In obtaining this effect, \( x \) and \( y \) are usually chosen so that they lie, approximately, on a common indifference curve. Early (Huber et al., 1982) and recent (Trueblood, 2012; Trueblood et al., 2013b) experiments demonstrate different magnitudes of the attraction effect for three different placement of decoys in the attribute space (so-called range, frequency, and range-frequency decoys).

The similarity effect: Consider \( \{x, y\} \) and a decoy \( s_x \) (respectively, \( s_y \)) that is similar to \( x \) (respectively, \( y \)). The similarity effect occurs when the probability of choosing \( x \) is greater when the decoy is similar to \( y \) as compared to when it is similar to \( x \), with the parallel result for \( y \):

\[
B_{\{x, y, s_x\}}(x) < B_{\{x, y, s_y\}}(x) \text{ and } B_{\{x, y, s_y\}}(y) > B_{\{x, y, s_x\}}(y).
\]

In obtaining this effect, \( x \) and \( y \) are usually chosen so that they lie, approximately, on a common indifference curve.

The compromise effect: Consider a options \( x, y, z \) where \( y \) is “between” \( x \) and \( z \), and an option \( c_z \) that makes \( z \) “between” \( c_z \) and \( y \). The compromise effect occurs when the probability of choosing \( y \) is greater when \( y \) is a compromise (“between”) alternative than when it is an extreme alternative, with a parallel result for \( z \).

\[
B_{\{x, y, z\}}(y) > B_{\{y, z, c_z\}}(y) \text{ and } B_{\{x, y, z\}}(z) < B_{\{y, z, c_z\}}(z).
\]

In obtaining this effect, \( x, y, z \), and are usually chosen so that they lie, approximately, on a common indifference curve.

\[20\] For the most part, Trueblood et al. (2013a) develop and test the MLBA using values of the options on quantitative attributes.
Multiattribute linear ballistic accumulator (MLBA) model

The multiattribute linear ballistic accumulator (MLBA) model (Trueblood et al., 2013a) has five major components, which we summarize in turn: context free subjective value functions, weighted differences, indifference/dominance coordinates, relative comparisons, and context dependent drift rates. As above, we deal with various sets of three distinct options $x, y, z$, and we let $X = \{x, y, z\}$.

Context free subjective value functions

In the example of options with two attributes, we assume that each option $r = (r_1, r_2)$ has a real-valued context free psychological (subjective) representation $(u_1(r_1), u_2(r_2))$ of its attribute values. As Trueblood et al. (2013a) work with attributes that are measured by nonnegative real numbers, they consider the set of nonnegative valued vectors $r = (r_1, r_2)$ in this attribute space, and map each “indifference” line $r_1 + r_2 = c$, a nonnegative constant, into a curve $(u_1(r_1), u_2(r_2))$; that curve can be convex, concave, or linear, depending on the parameters of the mapping.

There are attribute weights $w_i \geq 0, i = 1, 2$, with $w_1 + w_2 = 1$, that indicate the amount of importance allocated to each attribute of the psychological representation, which leads to the context free representation

$$u(r) = w_1 u_1(r_1) + w_2 u_2(r_2).$$

Using that representation, option $p$ context free dominates option $q$ if $u_1(p_1) \geq u_1(q_1)$ and $u_2(p_2) \geq u_2(q_2)$, and option $p$ is context free indifferent to $q$ iff

$$w_1 u_1(p_1) + w_2 u_2(p_2) = w_1 u_1(q_1) + w_2 u_2(q_2),$$

i.e., iff

$$w_2[u_2(p_2) - u_2(q_2)] = w_1[u_1(q_1) - u_1(p_1)].$$

Weighted differences

For options $r$ and $s$, the weighted difference in psychological value between the options on attribute $i, i = 1, 2$, denoted $\Delta_i(r, s)$, is defined by

$$\Delta_i(r, s) = w_i[u_i(r_i) - u_i(s_i)].$$

Indifference/dominance coordinates

The vector of weighted differences $[\Delta_1(r, s), \Delta_2(r, s)]$, is converted to indifference/dominance coordinates $(\Delta I_{rs}, \Delta D_{rs})$ by the form

$$(\Delta I_{rs}, \Delta D_{rs}) = 1/\sqrt{2} \cdot [\Delta_2(r, s) - \Delta_1(r, s)], [\Delta_2(r, s) + \Delta_1(r, s)].$$
Trueblood et al. (2013a) call $\Delta I_{rs}$ the difference along the indifference dimension and $\Delta D_{rs}$ the difference along the dominance dimension. The motivation for this terminology is that $\Delta I_{rs}$ can be considered as a similarity measure along the indifference dimension - the smaller it is, the more similar the options are on that dimension. And $\Delta D_{rs}$ is nonnegative if $r$ dominates $s$ and larger the larger $r$ is relative to $s$ on the dominance dimension.

**Relative comparisons**

For any pair of options $r, s$, the value $V_{rs}$ that enters into the context dependent drift rate of the MLBA is

$$V_{rs} = \frac{|\Delta I_{rs}|}{|\Delta D_{rs}| + 1} + \alpha \cdot \frac{\Delta D_{rs}}{|\Delta I_{rs}| + 1}$$

The first term in $V_{rs}$ measures the magnitude of the absolute difference on the indifference dimension relative to the absolute difference on the dominance dimension; absolute values are used because the sign of the indifference comparison does not matter. The second term is the parallel relative comparison of the dominance dimension to the indifference dimension; in this case, the sign of the difference on the dominance dimension matters - a dominated option will have a negative comparison relative to a dominating one, and vice-versa; and the sign of this comparison is crucial for explaining context effects. The value $V_{rs}$ can be negative; for this reason, a constant $I_0$ is added in calculating the context dependent drift rates, below. Also, $V_{rs}$ is not symmetric because, as noted above, $\Delta D_{rs}$ is nonnegative if $r$ dominates option $s$ and nonpositive if $s$ dominates option $r$. Larger $\alpha > 0$ gives more importance to the difference along the dominance dimension than to the difference along the indifference dimension.

**Context dependent drift rates**

The final stage gives the context dependent drift rate for each $r \in X$:

$$d_X(r) = I_0 + \sum_{\{r,s\} \in X \atop r \neq s} V_{rs}.$$  

The constant $I_0 \geq 0$ helps avoid non-termination in the LBA model.

Trueblood et al. (2013a) show that this context dependent LBA model can produce the three context effects, described earlier, with numerous sets of parameter values. The similarity effect is the consequence of the indifference comparison, the attraction effect is a consequence of the dominance comparison, and the compromise effect is the consequence of the subjective value function. They also demonstrate that the model can handle various properties of the context effects when participants are under time pressure.
Note that the above positive results regarding the prediction of context effects by the MLBA model were obtained by adding a context dependent “front end” to a context independent LBA model. This suggests taking a fresh look at various choice-only models that can handle some or all such context effects in choice, and extending them to response time by, say, adding a context independent LBA model that is driven by the “front end” parameters of the choice model. We consider possible such (choice) models in 1.5.6.

The 2N-ary choice tree model for N-alternative preferential choice
The 2\(N\)-ary choice tree model for \(N\)-alternative preferential choice (Wollschläger and Diederich, 2012) builds on earlier information sampling models, such as multi-alternative decision field theory (MADFT, Roe et al., 2001) and the leaky competing accumulator (LCA) model (Usher and McClelland, 2001, 2004), in ways that resonate with the multiattribute linear ballistic accumulator (MLBA) model and the parallel best-worst LBA model. In contrast to the MLBA model, the 2\(N\)-ary choice tree model specifies details of the temporal accumulation of information, rather than having a front-end that incorporates the overall effect of such accumulated information.

In contrast to MADFT and LCA, and in agreement with the general form of the parallel best-worst LBA model of Section 1.5.4, the 2\(N\)-ary choice tree model assigns two counters to each of the \(N\) alternatives in the current choice set - one samples information in favor of the respective alternative, the other samples information against it. The difference of these two quantities for each alternative at any given time describes the preference state for that alternative at that time. Sampling time is discretized, with the length \(h\) of the sampling interval chosen arbitrarily with a shorter time interval leading to more precision in the calculation of expected choice probabilities and choice response times. At each time point, the model selects one (positive or negative) counter for one alternative, and increases its level by one; 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. Also, updating counters at discrete points in time creates a discrete structure of possible combinations of counter states which can be interpreted as a graph or, more precisely, as a (2\(N\)-ary) tree. There is a stopping rule that depends on the preference states associated with the alternatives; each preference state is the difference between the cumulative positive and negative count for the relevant alternative. The most general stopping rule considered by Wollschläger and Diederich (2012) has
two separate thresholds for the difference count for each alternative, with one positive, which leads to choosing an option if its difference count reaches that threshold first, and the other negative, which leads to eliminating an option from further consideration if its difference count reaches that threshold first; in the latter case, later stages of the process are completed without counts for the eliminated option.

The structures, though not the details, of the $2N$-ary choice tree model and of the MLBA model are rather similar, so we now summarize the former - they are: context free representations; local rescaling; contrast (positive and negative); weighted contrast (positive and negative); transition probabilities; noise; and leakage. Thus, the major differences are between the ballistic accumulators of the MLBA model and the transition probabilities of the $2N$-ary choice tree model, and the fact that the MLBA model has no decay, whereas the $2N$-ary choice tree model has decay. Each model well-fits the data presented by its authors.

To date, the $2N$-ary choice tree model has been applied to best choice. However, as in the parallel best-worst LBA model of Section 1.5.4, the choice tree model has a mechanism for accepting (rejecting) options. Thus, it is of interest to extend the $2N$-ary choice tree model to best-worst choice and compare its fits to the data from such choice with those of the parallel best-worst LBA model.

\section*{1.5.6 Context dependent models of choice}

As noted in Section 1.5.5, the predictions of context effects by the MLBA model are obtained by adding a context dependent “front end” to a context independent LBA model, which lead us to suggests taking a fresh look at various choice-only models that can handle some or all such context effects in choice, and extending them to response time by, say, adding a context independent LBA model that is driven by the “front end” parameters of the choice model.

Marley (1991) describes various context dependent choice models, the majority of them based on measures of binary advantage of one option relative to another. These include Rotondo’s (1986) binary advantage model, which is a generalization of Luce’s choice model that is not a random utility model; and a generalization of Rotondo’s model to a random advantage model - that is, where there exist (possibly dependent) random variables $A_X(z)$, $z \in X \subseteq \mathcal{A}$ of a particular form such that $B_X(x) = \Pr[A_X(x) > A_X(y), y \in X - \{x\}]$; clearly, this representation corresponds to the choice probabilities given by a context dependent horse race model (Section 1.5.3).
We now describe three other recent context dependent choice models.

**Berkowitsch, Scheibehenne, and Rieskamp**

Berkowitsch et al. (2013) develop a choice-only version of decision field theory (DFT, Roe et al., 2001) and compare its predictions to those of the multinomial logit and probit models; as with the original DFT (which predicts both choice and response time), this choice-only version of DFT can handle choice-based context effects. All the models under comparison accurately predict data from a relatively standard stated preference experiment (similar to those presented in Section 1.5.4), but the choice-only version of DFT is superior for designs where context effects are found; the latter is unsurprising, given that the tested multinomial logit and probit models are all random utility models (see Sections 1.5.4 and 1.5.5 for relevant commentary).

**Blavatskyy**

As discussed in Section 1.5.2, Blavatskyy (2012) developed a model for binary choice between risky gambles that has a natural extension to binary choice between multiattribute choice, and to choice between more than two such options. In this case, with

\[ z = (z_1, ..., z_m) \]

where \( z_i \) is the level of option \( z \) on attribute \( i \), we would want to derive (axiomatize) or assume the existence of utility functions \( u_i(z_i), i = 1, ..., m \), such for multiattribute options \( r, s \) have

\[ r \lor s = \max(u_1(r_1), u_1(s_1)), ..., \max(u_m(r_m), u_m(s_m)) \]
\[ r \land s = \min(u_1(r_1), u_1(s_1)), ..., \min(u_m(r_m), u_m(s_m)) \]

and a function \( \varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) and a real valued function \( u \) such that, say, for all \( r, s \)

\[ p(r, s) = \frac{\varphi[u(r) - u(r \land s)]}{\varphi[u(r) - u(r \land s)] + \varphi[u(s) - u(r \land s)]} \quad (1.65) \]

or, alternatively, for all \( r, s \)

\[ p(r, s) = \frac{\varphi[u(r \lor s) - u(s)]}{\varphi[u(r \lor s) - u(s)] + \varphi[u(r \lor s) - u(r)]}. \quad (1.66) \]

When \( u \) is additive, i.e.,

\[ u(z) = \sum_{i=1}^{m} u_i(z_i) \]
the representations in (1.65) and (1.66) give the same choice probabilities.

Blavatskyy (2009, 2012) extends his model (and, consequently, the above model) of binary probabilistic choice to choices among two or more alternatives with the following algorithm: i. Select an alternative at random from the choice set $X$; call it $x$. ii. Select another alternative at random from the choice set $X - \{x\}$; call it $y$. iii. Choose between $x$ and $y$ with $x$ (resp., $y$) chosen with probability $p(x, y)$ (resp., $p(y, x)$); iv. Set the winner of step iii. as the (new) alternative in step i.; v. Repeat these steps *ad infinitum*. Blavatskyy shows that the resultant set of objects involved in the representation of the choice probabilities for a set $X$ are the *arborescences*\(^{21}\) with the root of each arborescence a member of $X$ and the each directed edge $(r, s)$ of the arborescence having weight $p(r, s)$; and that the probability of choosing $x$ from $X$ is proportional to the sum of the products of these weights for each arborescence with root $x$. Blavatskyy (2012) summarizes several properties of this model, including the fact that, with appropriately selected binary choice probabilities, it predicts the asymmetric dominance effect and the attraction effect; in these predictions, he does not use a specific form for the choice probabilities, such as (1.65) or (1.66), above. Marley (1965) develops a related representation for discard (rejection, worst) probabilities motivated in a particular way from sets of preference (acceptance, best) choice probabilities where the latter satisfy regularity.

**Chorus**

Using the notation above for the representation of multiattribute options, and constants $\beta_i \geq 0$, Chorus (2013) assumes that: for $x \in X \subseteq A$,

$$B_X(x) = \frac{e^{u_X(x)}}{\sum_{z \in X} e^{u_X(z)}},$$

where, for $z \in X$

$$u_X(z) = -\sum_{s \in X - z} \sum_{i=1}^{m} \ln \left[ 1 + \exp(\beta_i [u_i(s_i) - u_i(s_i)]) \right].$$

Chorus (2013) shows that this referent dependent model explains various context effects. Using our earlier reasoning, the quantities $e^{u_X(x)}$ could be entered as the drift rates in an LBA model. Other recent work includes the related approach of Huang (2012) based on ideal and reference points, and Rooderkerk et al. (2011) who present a choice model with a context-free

---

\(^{21}\) An arborescence is a directed graph in which, for a given vertex $u$ (the root) and any other vertex $v$, there is exactly one directed path from $u$ to $v$. Equivalently, an arborescence is a directed, rooted tree in which all edges point away from the root.
partworth random utility component and a context-dependent random component. Bleichrodt (2007) develops an additive reference-dependent utility where the reference point varies across decisions and is one of the available options is each choice set, and Bleichrodt et al. (2009) axiomatize a representation for uncertain gambles with multiattribute outcomes, obtaining a version of prospect theory with referents on each attribute or option and with an additive representation of the outcomes.

1.6 Discussion, Open Problems, and Future Work

In the conclusion of their review chapter on preferential choice, Busemeyer and Rieskamp (2013) present an excellent summary of the similarities and differences between various approaches to the study of choice. The primary theme of that summary is that discrete choice models are primarily applied to the choices of large samples of people with relatively few data per person, with the primary goal of obtaining efficient estimates of economically relevant parameters, whereas process models account for choice, decision time, confidence, and brain activation, often with extensive data per person, with the primary goal of understanding the behavior of individuals. They agree that this is an over-simplification, and the results summarized or cited in our chapter illustrate ongoing multiple interactions between applied and basic research on choice, preference and utility. A significant part of the reason why standard accrual-halting models (Townsend et al., 2012) have not been used in applied areas is the “relatively high difficulty of estimating some sequential sampling models of decision making” (Berkowitsch et al., 2013). This difficulty is being alleviated by, for example, Berkowitsch et al.’s (2013) derivation of a tractable closed-form for the choice probabilities given by an asymptotic form of decision field theory (DFT) and the various linear ballistic accumulator (LBA) models for (best, worst, best-worst) choice and response times.

Nonetheless, there is currently no clear formulation of exactly what kind of (contextual) complexity is required in accrual-halting models, nor have such models been extensively used for prediction of, say, revealed choices (e.g., purchases) from stated choices (e.g., surveys) or of new phenomena. The (context-dependent) accrual-halting models that we have presented in this chapter involve numerous complex processes, such as context free representations; local rescaling; contrast (positive and negative); weighted contrast (positive and negative); transition probabilities; noise; and leakage. Further empirical study along the lines of Teoderescu and Usher (2013), with follow-
up theoretical work, is needed to clarify the conditions under which some or all these, and possibly other, structures are needed.

We chose not to present context-dependent versions of the material on choice probabilities in Section 1.2 and later related sections. We have developed the quite-natural notation for such extensions, but substantive results are required for such notation to be useful - for instance, can one formulate a polytope for best choices that satisfies one, or more, of the standard choice-based context effects? A second major limitation of that material on choice probabilities is just that - it is for choice probabilities, only, not response times. At this time, we have no concrete ideas of how to join polytopes for choice with representations of response time, but it sounds like a fascinating topic.

One open problem that we are somewhat optimistic about being able to solve (at least for small numbers of options) is the characterization of the linear ordering polytope of best-worst choice probabilities.

Finally, there have been significant theoretical, empirical, and statistical advances in the past 10 years in the study of properties of binary choice probabilities such as the triangle inequality and related properties. However, there has been very little work testing the properties of binary choice probabilities between, say, gambles, where the deterministic representation of the gamble is assumed to satisfy certain behavioral properties (accounting indifferences) in the sense of Section 1.4.3; Regenwetter et al.’s (2013) work on cumulative prospect theory (see Section 1.5.1) is indirectly related to this issue, though it evaluates (probabilistic) representations, rather than (probabilistic) behavioral properties.
Acknowledgments

We thank E. Bokhari, C. Davis-Stober, J.-P. Doignon, A. Popova, and J. Trueblood for helpful comments on earlier partial drafts or portions of the text. Numerous other colleagues were extremely helpful with suggestions on the content and structure of the chapter. We thank Ying Guo for computing facet-defining inequalities for best-worst choice using PORTA. This research has been supported by Natural Science and Engineering Research Council Discovery Grant 8124-98 to the University of Victoria for Marley. The work was carried out, in part, whilst Marley was a Research Professor (part-time) at the Centre for the Study of Choice, University of Technology Sydney. National Science Foundation grants SES # 08-20009, SES # 10-62045, CCF # 1216016 and an Arnold O. Beckman Research Award from the University of Illinois at Urbana-Champaign each supported Regenwetter as Principal Investigator. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of colleagues, funding agencies, or employers.
References


Figure 1.1 A sample stimulus from a ternary paired comparison study using binary lotteries. Fig. 4 of the Online Supplement to Regenwetter and Davis-Stober (2012). Reproduced with permission from the American Psychological Association.
Figure 1.2 A sample question from a study on preferences for micro-generation of electricity using solar panels. Fig. 1 of Marley and Islam (2012). Reproduced with permission from Elsevier.
Figure 1.3 A hypothetical choice between two treatment options described by their success (survival) rate, life expectancy, and health outcome.

Figure 1.4 1-compound invariance. The indifference marked by a heavy line is implied by the other three indifferences. Fig. 3.2 of Bleichrodt et al. (2013). Reproduced with permission from Elsevier.
Racing LBA Accumulators

Start points vary randomly from trial to trial.

Drift rates vary randomly from trial to trial.

Figure 1.5 Illustrative example of the decision processes of the linear ballistic accumulator (LBA). Fig. 1 of Hawkins et al. (2013a). Reproduced with permission from John Wiley & Sons [To copy editor: In Press, will get permission when available online].