3 Psychological research and theories on preferential choice

Jerome R. Busemeyer and Jörg Rieskamp¹

Understanding human preferential choice behavior is challenging because humans change their preferences across time and contexts. This chapter summarizes the basic behavioral findings from research on human preferential choice and reviews the psychological theories that have been proposed to account for the puzzling findings. The main theme that we attempt to convey to the reader is that a coherent view of an individual's underlying beliefs and values can only be recovered by carefully modeling the dynamic nature of the choice process through which these beliefs and values operate to produce observed behavior.

When examining people's choice behavior it becomes apparent that it varies substantially. For instance, Hey (2001) conducted a study in which 53 people repeatedly choose between pairs of simple gambles in five different sessions. In every session the same set of 100 pairs was presented. When assuming stable and deterministic preferences all people should have made identical choices in every session. However, it turned out that no single person always made the same choices across all five sessions. Instead, on average participants changed their preferences for 10 percent of pairs between two consecutive sessions; for on average 23 percent of the pairs they did not make identical choices in all five sessions. The seminal work by Mosteller and Nogee (1951) discovered early on that people's choice behavior varies and has a probabilistic character. More surprisingly, its consequences for theory-building are still not fully acknowledged. Axiomatic approaches to human choice behavior imply deterministic behavior and do not contain an error theory that could explain people's inconsistencies. The undeniable variability in people's behavior is often implicitly acknowledged by assuming that people's inconsistencies can be explained by unsystematic errors or 'white noise'. This implies that people's behavior varies around the theories' deterministic predictions. However, when people's inconsistencies are systematic such forms of 'tremble error' theories (Loomes et al., 2002) are not sufficient to explain human behavior. Moreover, without an explicit error theory it appears almost impossible to separate unsystematic from systematic inconsistencies and to unravel the mechanisms that underlie the systematic inconsistencies.

In this chapter we describe the two standard approaches to explaining the probabilistic character of choice behavior represented by fixed and random utility models. Second, we summarize the key empirical findings violating essential principles of utility theories. Next we present psychological models of choice behavior that have been suggested to explain some of the observed behavioral regularities. We start with static choice models followed by dynamic models that describe how people's preferences evolve over time. Then we summarize some additional and new directions, and finally we conclude by briefly comparing the different approaches and models against each other.

1 CLASSIC PROBABILISTIC UTILITY MODELS

Before delving into the behavioral research on preferential choice, it is useful to first spell out two classic utility theory approaches: fixed versus random utility models. These theories served as a primary guide for past research on preferential choice by implying specific choice principles that were tested empirically. These empirical tests revealed the inadequacies of the theories in describing people's behavior and let to the development of more descriptive and cognitively driven models. In the following we will describe and define these models using the notation and terms employed earlier in Rieskamp et al. (2006).

Assume that there exists a complete set of choice options $X = \{A_1, \ldots, A_n\}$ under consideration. The person may be presented with a subset $Y = \{B_1, \ldots, B_m\} \subseteq X, m \le n$ of this complete set. The probability that an individual chooses option B_i from the set Y is denoted $p(B_i|Y)$, with the constraint $p(B_i|Y) \ge 0$ and $\sum_{i \in Y} p(B_i|Y) = 1.0$. Note that psychological models normally assume that an individual's behavior is probabilistic, and so the theory needs to be defined with the probabilities at an individual (pooled across replications within an individual) rather than an aggregate level (pooled across individuals).

1.1 Fixed Utility Approach

According to the fixed utility model, a real value $u(A_i)$ can be assigned to each option $A_i \in X$ that remains fixed across choice sets Y. When presented with the subset Y, the probability of choosing option B_i equals:

$$p(B_i|Y) = f(u(B_i), u(B_1), \dots, u(B_{i-1}), u(B_{i+1}), \dots, u(B_m))$$
(3.1)

where the function f is a strictly increasing function of the first coordinate, $u(B_i)$, and a strictly decreasing function of each of the remaining coordinates (and the remaining can be permuted without changing the predictions). For example, according to a Luce (1959) ratio of strength model:

$$p(B_i|Y) = exp(u(B_i)) / \sum_{i \in Y} exp(u(B_i)).$$
(3.2)

A key idea of this class of models is that choice is inherently probabilistic and fundamentally unpredictable. Even with fixed utilities, a person's choice on each occasion remains probabilistic. The main simplifying assumption of this model is that the utilities assigned to the choice options do not depend on the choice set Y. In other words, utilities are context independent, also called simple scalable (Luce and Suppes, 1965). This context independence property produces a shortcoming of this class of models. Of course this property can be relaxed, but how to do this in a coherent and parsimonious manner is quite challenging.

1.2 Random Utility Approach

According to random utility models, on any choice occasion, the person samples an evaluation point w from a sample space W that determines the *n* real valued utilities $U(\omega)$

Psychological research and theories on preferential choice 51

= $[U_1(\omega), \ldots, U_n(\omega)]$. This sampling process produces *n* random variables U_i , $i = 1, \ldots, n$ that determine an *n* dimensional random utility distribution function denoted by F_X = $\Pr[U_1 \le u_1, \ldots, U_n \le u_n]$; when presented with a choice set *Y*, the person chooses the option that has the maximum randomly sampled value, max $\{U(B_1), \ldots, U(B_m)\}$. The choice probabilities for the set *Y* is based on the marginal *m* dimensional distribution F_Y , which is obtained by integrating F_X over the values of the options in *X* that are not presented in *Y*. The probability that option B_i is chosen from the presented set *Y* equals the probability that the randomly sampled value for the random variable U_i is the maximum:

$$p(B_{i}|Y) = \Pr[U_{i} = \max\{U_{i}, U_{i} \in Y\}].$$
(3.3)

For example, the Thurstone (1959) model assumes that the distribution function for the random utility vector $U = [U_1, \ldots, U_n]$ is multivariate normal.² The key idea behind this class of models is that once the evaluation point w is selected by the person, then choice becomes deterministic. Behavior is only probabilistic because we do not know the point w used to evaluate the options; we only know the probabilities of sampling these points. The main assumption of this class of models is that the same *n* dimensional distribution function F_X is used for all choice sets *Y*. In other words, the distribution function is context independence turns out to produce shortcomings for explaining certain choice phenomena with this class of models. This property can be relaxed, but how to do this in a coherent and parsimonious manner presents a major challenge.

1.3 Comparison of the Two Utility Approaches

The ideas motivating these two approaches are fundamentally different: Fixed utility models imply that choice is fundamentally indeterministic. In contrast, random utility models imply that choice is completely deterministic. Despite these conceptual differences, it is often difficult to empirically discriminate these ideas. For example, the Luce (1959) model can be mathematically derived from a random utility model with an identical and independently extreme value distributed error (Yellot, 1977). More general extreme value random utility models form the basis of many economic choice models (McFadden, 1981), and they are also very popular among marketing researchers (Louviere et al., 2000). In sum, fixed and random utility models represent the predominant approach (in particular, in economics) of predicting human choice behavior. Both approaches should be evaluated empirically by the major behavior findings.

2 BASIC CHOICE BEHAVIOR FINDINGS

This section reviews the basic empirical findings from preferential choice that have accumulated across the last 50 or so years from behavioral economics, consumer research and psychology. As mentioned earlier, much of this research was targeted at basic properties implied by the two classic probabilistic utility approaches described above. The research examines whether human choice behavior obeys basic properties such as transitivity, independence, regularity, and stationarity.

2.1 Transitivity

One of the most basic properties of choice to examine empirically is transitivity. If a person prefers Beethoven to Mozart and Mozart to Chopin, then transitivity implies that the person will also prefer Beethoven to Chopin. It appears unreasonable to violate this principle repeatedly.

In general, the transitivity property is considered to be one of the main axioms of rational choice. Formally, transitivity is defined by a mathematical relation called a 'preference' denoted \geq so that $B_i \geq B_j$ means that option B_i is preferred or indifferent to option B_j . Transitive preferences must satisfy $B_i \geq B_j \land B_j \geq B_k \rightarrow B_i \geq B_k$ for all i,j,k in X. This is required by deterministic utility models in order to postulate a real valued utility function for ordering preference, $u: X \rightarrow Reals$.

Transitivity is often justified as an axiom of choice behavior by arguing that violations of this axiom permit a person to be turned into a money pump. If a person prefers B to C then the person should be willing to pay money to exchange C for B; likewise if the person prefers A to B then the person should be willing to pay money to exchange Bfor A; finally if the person also prefers C to A then the person should be willing to pay to exchange A for C, thus returning to their original position, but after losing money on three exchanges. However, it is unlikely that these intransitive preferences could be exploited by building a money pump (compare, Chu and Chu, 1990). Instead, people will presumably notice their intransitive cycles at some point, making the money pump a 'bogeyman' that only demonstrates in principle the irrationality of intransitive choices but one that would never be observed (Lopes, 1996).

The concept of transitivity is difficult to apply to probabilistic choice behavior because it is difficult to define the preference relation \geq when choices are inconsistent. One way is to define $B_i \geq B_j \ll p(B_i \{B_i, B_j\}) \geq .50$ (Luce, 2000). This immediately leads to a definition of weak stochastic transitivity:

$$p(B_{i}|\{B_{i},B_{j}\}) \ge .50, p(B_{j}|\{B_{j},B_{k}\}) \ge .50 \to p(B_{i}|\{B_{i},B_{k}\}) \ge .50$$
(3.4)

for all *i*, *j*, *k* in *X*.

The fixed utility class of models must satisfy weak stochastic transitivity, and moreover, this class must also satisfy a stronger version called strong stochastic transitivity:

$$p(B_{i}|\{B_{i}, B_{j}\}) \ge .50, p(B_{j}|\{B_{j}, B_{k}\}) \ge .50 \rightarrow p(B_{i}|\{B_{i}, B_{k}\}) = \max\{p(B_{i}|\{B_{i}, B_{i}\}), p(B_{i}|\{B_{i}, B_{k}\})\}$$
(3.5)

for all i,j,k in X. This follows from the fact that $p(B_i|\{B_i,B_j\}) \ge .50 \rightarrow u(B_i) \ge u(B_j)$ and $p(B_j|\{B_j,B_k\}) \ge .50 \rightarrow u(B_j) \ge u(B_k)$ and together this implies $u(B_i) \ge u(B_k)$, so that $p(B_i|\{B_i,B_k\}) \ge p(B_j|\{B_j,B_k\})$ and $p(B_i|\{B_i,B_k\}) \ge p(B_j|\{B_i,B_j\})$.

The random utility class of models does not need to satisfy weak stochastic transitivity. The decision-maker can be transitive from each point of view w, but averaging across these different preference orders produced by different points of view can violate weak stochastic transitivity (Regenwetter et al., 2011). For instance, suppose that across different occasions, a person experiences the following three transitive preference orders: $(B_1 > B_2 > B_3)$, $(B_2 > B_3 > B_1)$, $(B_3 > B_1 > B_2)$ where A > B indicates strict preference. If

we assume that these preferences occur equally often, then we observe that $p(B_1|\{B_1,B_2\}) = 2/3$, $p(B_2|\{B_2,B_3\}) = 2/3$, but $p(B_1|\{B_1,B_3\}) = 1/3$, violating weak stochastic transitivity, which is called the 'Condorcet paradox'.

Although random utility models do not need to satisfy weak stochastic transitivity they need to satisfy another transitivity property called the triangular inequality for binary choices (Becker et al., 1963), if we assume no indifference, $p(U_i = U_i) = 0$:

$$p(B_{i}|\{B_{i}, B_{i}\}) + p(B_{i}|\{B_{i}, B_{k}\}) - p(B_{i}|\{B_{i}, B_{k}\}) \le 1.$$
(3.6)

This can be shown as follows. For brevity, define p(xyz) as the probability for the strict transitive order x > y > z. Then $p(x|\{x,y\})+p(y|\{y,z\}) - p(x|\{x,z\}) \le 1 \rightarrow p(z|\{x,z\}) \le p(z|\{y,z\}) + p(y|\{x,y\}) \rightarrow p(zxy) + p(zyx) + p(yzx) \le p(zyx) + p(zxy) + p(xzy) + p(yzz) + p(yzx) +$

Given the importance of the transitivity property, one would expect a firm resolution with regard to its empirical status. Beginning with May (1954), a long series of investigations has appeared reporting violations of transitivity (see review by Rieskamp et al, 2006). Many of these studies were designed to replicate the well-known experiment by Tversky (1969), who reported violations of weak stochastic transitivity. Recently, however, the results regarding transitivity have been called into question because of inadequate methods for statistically testing this property (Iverson and Falmagne, 1985; Regenwetter et al., 2011). In response, several new statistical methods have been developed for testing transitivity, and these new results send a more mixed message. Tsai and Böckenholt (2006) used a random utility mixture model to compare a parametric formulation of an intransitive model versus a more constrained transitive model, and they found that chi-square difference tests rejected the constrained transitive model in favor of the more general intransitive model. Regenwetter et al. (2011) developed nonparametric statistical tests of the triangular inequality and found that most participants did not produce statistically significant violations. Birnbaum and Schmidt (2008) developed statistical models for choice patterns based on true preferences plus response errors, and they found that most participants produced patterns consistent with a model of transitive true preferences plus response error. In sum, in spite of the large body of research it appears unclear whether people violate weak stochastic transitivity. Recent evidence using adequate statistic tests shows that these violations appear to be the exception rather than the rule.

More systematic, reliable, and robust evidence has been found for violations of strong stochastic transitivity (Mellers and Biagini, 1994; Rumelhart and Greeno, 1971; Tversky and Russo, 1969). The psychological reason for the violations of strong stochastic transitivity is the following: when a person is faced with choices between multidimensional options, the context produced by a pair of options is used to single out some dimensions for making comparisons while ignoring other dimensions. For example, one pair of options A and B might differ largely in a tradeoff among the quality features but seem similar in price – for this pair the person may tend to ignore the price difference and focus more on the quality features. In another example comparing B and C again the price difference appears negligible after giving the quality dimension more attention. However, when comparing A and C the difference in price might become substantial, so that the person may tend to refocus more heavily on price for the comparison. Thus the

dimensions used to make the comparison change across choice pairs and this change in the basis for comparison results in violations of strong stochastic transitivity.

2.2 Independence

The next most basic property of choice to be examined empirically is independence, which is concerned with invariance of a preference relation between two options with changes in the choice set that contains these options. More formally suppose $Y \subseteq X$ and also $Z \subseteq X$; then independence states (Tversky, 1972b)

$$p(B_{j}|\{B_{j}, B_{j}\} \cup Y) \ge p(B_{j}|\{B_{j}, B_{j}\} \cup Y) \leftrightarrow p(B_{j}|\{B_{j}, B_{j}\} \cup Z) \ge p(B_{j}|\{B_{j}, B_{j}\} \cup Z).$$
(3.7)

This property is also required by fixed utility theories: $p(B_j|\{B_j, B_j\} \cup Y) \ge p(B_j|\{B_j, B_j\} \cup Y) \ge p(B_j|\{B_i, B_j\} \cup Z) \ge p(B_j|\{B_i, B_j\} \cup Z)$. However, this property is not required by the general random utility model (although the standard multinomial logit random utility model with statistically independent utilities satisfies this property). The independence property directly reflects the context independence property underlying the fixed utility model. There are different ways as to how this principle can be systematically violated in empirical studies. One very prominent way is the so-called 'similarity effect' and another is the more recently documented 'compromise effect'.

The 'similarity effect' was initially examined by Tversky (1972a). The essential idea is to examine the preferences for options, say A versus B, with or without the context of another option A^* . Options A and B are designed to be qualitatively different or dissimilar – for example option A could be high in quality and price whereas option B is low in quality and price. Option A^* is designed to be very similar and competitive with option A, for example it could also be high in quality and price but just a bit worse in quality but therefore a bit cheaper. It has been empirically found that in binary choican can be that A^* preferred to B, $p(A|\{A,B\}) \ge p(B|\{A,B\})$, but when option A^* is included into the choice set then A^* competes with A but not so much with B, so that $p(A|\{A,B,C\})$ $\le p(B|\{A,B,C\})$. These findings were later extended by Tversky and Sattath (1979). In the transportation literature the similarity effect is often explained by the red-bus blue-bus example, where the decision-maker chooses between different transportation options, such as a car and a red bus. When another (blue) bus is added to the choice set that only differs in color with the red bus it harms the choice share of the red bus but not of the car (Train, 2003).

The 'compromise effect' was initially examined by Simonson (1989). Again, the essential idea is to examine the preferences for options, say *B* versus *C* with or without the context of another option *A*. In this case, option *C* is designed to be a compromise between the two other options *A* and *B*, and the latter are designed to have extreme tradeoffs on two different dimensions. For example, option *A* may be very high in quality and price whereas option *B* is very low in quality and price, and option *C* is moderate in quality and price. In the context of a choice between *B* and *C*, option *C* does not appear as any kind of compromise – option *B* is simply lower than *C* in quality and price; but if another extreme option *A* is included in the choice set, then option *C* emerges as a compromise that lies between the two extremes of *A* and *B*. The main finding is that when comparing options *B* and *C* one finds $p(B|\{B,C\}) \ge .50$; but when option *A* is

included into the choice set then the compromise becomes favored so that $p(C|\{A, B, C\}) \ge p(B|\{A, B, C\})$. This work has been replicated and extended in many subsequent studies (see review by Kivetz et al., 2004).

2.3 Regularity

Another very basic property of choice to be examined empirically is regularity, which asserts that the addition of an option to a choice set should never increase the probability of selecting an option from the original set. More formally, suppose $Z \subset Y \subseteq X$ presented for choice. Then regularity states

$$p(B_i \mid B_i \leq Z) \ge p(B_i \mid B_i \leq Y).$$
(3.8)

This property holds for random utility theories since it is less likely that B_i has the highest utility in a larger set, Y, than in a smaller set, Z, because $p(B_i | Z) = \Pr[U(B_i) = \max\{U(B_j), B_j \in Z\}] \ge \Pr[U(B_i) = \max\{U(B_j), B_j \in Z\}] \times \Pr[U(B_i) = \max\{U(B_k), B_k \in Y-Z\}|U(B_i) = \max\{U(B_i), B_i \in Z\}] = p(B_i | Y).$

Violations of regularity were first found by Huber et al. (1982), which are called asymmetric dominance effects and/or attraction effects. The essential idea here is to examine the preferences for options, say A versus B with or without the context of another *deficient* option D. Like the 'similarity effect,' options A and B are designed to be qualitatively different or dissimilar – for example option A could be high on quality and price whereas option B is low in quality and price. Also like the 'similarity effect', option D is designed to be very similar to option A; but the key difference needed to produce an attraction effect is to make D deficient or defective as compared to option A, for example D is of slightly lower quality but the same price as A so that D is dominated by A. In this situation, D is rarely chosen, but including D to the choice set increases the probability of choosing option A so that $p(A|\{A,B\}) \leq p(A|\{A,B,D\})$. The same result can be obtained even when D is not necessarily dominated by A but just much less attractive as compared with A (Huber and Puto, 1983). These findings have been replicated in many studies (Wedell, 1991). (For review see Heath and Chatterjee, 1995.)

2.4 Stationarity

All choices take time, and the time taken to make a decision can change the choice that is finally made. For example, suppose you wake one morning to find two emails, each of which is an invitation to present a keynote speech at an attractive venue, but unfortunately on the same day. Choosing between these mutually exclusive offers may take substantial time to think through the advantages and disadvantages, and deadlines for making the decision could affect the final decision by preventing you from thinking through all the consequences.

If choice probabilities do not change as a function of deliberation time (excluding time to read the choices), then they are stationary. Define the probability of choosing an option A from a set $Y \subset X$ conditioned on deliberating for a period of time t as $p_t(A|Y)$. Stationarity states that $p_t(A|Y) = p(A|Y)$ for $t > t_0$ where t_0 is time necessary to read the choice options. Fixed and random utility models are static models that provide no

mechanisms for predicting the effects of decision time on choice probability. This oversimplification becomes a problem for these theories when stationarity is violated. On the one hand, one could argue that fixed and random utility models only apply for choices without time constraints, assuming that the probabilities converge to some asymptote $p_t(A|Y) \rightarrow p_{\infty}(A|Y)$. On the other hand, even when there is no explicit deadline, there is a cost for taking time to deliberate that puts time pressure on the decision-maker.

The effects of decision time on choice probability are now well established by several different lines of experimental research (Svenson and Maule, 1993). Consumer choices systematically reverse under time pressure (Svenson and Edland, 1987). Choices between uncertain actions systematically change as a function of deliberation time (Busemeyer, 1985) and even reverse under time pressure (Diederich, 2003). Compromise and attraction effects become even larger when decision-makers are encouraged to deliberate longer (Dhar et al., 2000; Pettibone, 2012). Decision-makers also tend to switch strategies for making decisions when pressed for time (Ben Zur and Bresnitz, 1981; Rieskamp and Hoffrage, 2008).

Time is an important factor in choice behavior because information about the choice options must be accumulated across time, and the type of information ultimately entering the choice process depends on how much time is allocated to making the decision (Wallsten and Barton, 1982). Under short time constraints, the decision-maker may have time only to focus on the most important dimension and ignore a number of other relevant aspects of the decision problem. When given more time to process all of the dimensions, the initial preference established by the first dimension can be overcome by competing information accumulated on many other relevant dimensions. Alternatively decision strategies could change from optimal to heuristic under time pressure (Payne et al., 1993).

Clearly, decision time plays a major role in emergency types of decisions, such as medical or military decisions (Janis and Mann, 1977). However, decision time also plays an important role in day-to day economic choices by consumers because of the well-known tradeoff between effort and accuracy when choosing strategies for decision-making (Payne et al., 1993). Time has a cost and consumers are often unwilling to spend a large amount of costly time to make decisions (Wright, 1972).

There is another way to think about stationarity, one that concerns changes in the choice probabilities caused by learning from experience across a sequence of repeated decisions. There is now a large experimental literature on experienced-based choice (compare, Hertwig and Erev, 2009), and a growing theoretical effort to develop integrated models of learning and decision-making (Erev and Barron, 2005; Gonzalez and Dutt, 2011). However, the topic of learning goes beyond the intended range of this chapter.

2.5 Conclusions from Behavioral Findings

The overview of some of the main findings on human choice behavior illustrates that people often do not adhere to simple principles such as strong stochastic transitivity, independence, regularity and stationarity. People's preferences and choice behavior change as a function of the choice context as well as the amount of time taken to process the information for making a decision. However, standard fixed and random utility ()

Psychological research and theories on preferential choice 57

models are insufficiently sensitive to context or time pressure to capture these systematic changes in preference behavior. For instance, fixed utility theories assign values to options that are independent of the choice set, and random utility theories do not have any mechanism to explain the impact of available time on decision-making. Of course, one can change utilities in an ad hoc way for every context to fit these effects post hoc. For instance, the utility function could be defined to include a context effect term (for example, include an extra utility term for a compromise option), and different random utilities could be assumed for different deliberation times (for example, change the means for each time period). But these ad hoc fixes do not provide a scientific explanation that allows for a priori predictions in new contexts and time constraints. Psychological choice models, which are described next, tend to be more complex than the basic fixed and random utility models. They are designed to provide mechanisms to account for either context effects, or time pressure effects, or both.

3 STATIC PSYCHOLOGICAL MODELS

3.1 Thurstone Model

One of the earliest choice theories was put forth by Thurstone (1927). The theory was extended by Thurstone (1959) and later by Bock and Jones (1968). See Böckenholt (2006) for a recent review of the developments of this theory. Thurstone's theory is a random utility type of model for choice that assumes a multivariate normal distribution for the random utility vector $U \sim Normal(\mu, \Sigma)$, where μ is the mean and Σ is the variance–covariance matrix for the vector of random utilities U. This model is closely related to the probit model used in economics and marketing (Hausman and Wise, 1978). Parsimonious representations of the variance–covariance matrix Σ are obtained by formulating specific covariance structure models (Takane, 1987).

Being a random utility model, the Thurstone theory must satisfy the triangle inequality principle as well as the principle of regularity presented earlier. Furthermore, because it assumes a multivariate normal distribution, it also must satisfy weak stochastic transitivity, but it can violate strong stochastic transitivity (Halff, 1976).

When used in conjunction with multidimensional scaling methods, Thurstone's theory provides an effective way to model similarity effects that produce violations of independence and strong stochastic transitivity (Carrol and DeSoute, 1991). In one version of this model (called the wandering ideal point (WIP) model) each choice object is represented as a fixed point lying within a multidimensional attribute space. Each person has an ideal point, which also lies within this same attribute space, but this ideal point randomly changes (wanders) across replications according to a multivariate normal distribution. On any trial, the choice option having the shortest distance to the sample ideal point is chosen. Multidimensional scaling based on similarity judgments can be used to identify the dimension of the attribute space as well as the coordinates of the choice objects. The similarity relations between the choice objects can then be used to account for violations of independence and strong stochastic transitivity.

Thurstone's theory does not provide any mechanisms for explaining attraction effects and compromise effects. In particular, being a random utility model, it must

satisfy regularity. Another deficiency is that Thurstone's theory is a static model, and so it cannot explain changes in choice probability as a function of deliberation time. However, below we will describe ways that the model can be reformulated as a diffusion model that allows the mean and covariance to change across time.

3.2 Elimination by Aspects Model

One of the earliest formalizations of heuristic choice is Tversky's (1972a) elimination by aspects theory (EBA). This theory was built on earlier ideas presented by Restle (1961) in which each option is described by a set of aspects (or properties) and choice is based on the selection of aspects. The basic idea is that the when presented with a choice set $Y \subseteq X$, the person samples an aspect with a probability equal to its weight or importance, and then eliminates any option that does not contain this aspect. This elimination process continues until only one option remains, which is then chosen. This heuristic process leads to the following formulation for choice probability. Define Z as the power set of X; that is, the set of all subsets of X. Define $Z_j \in Z$ as one of the nonempty subsets of Z, which is assigned a weight $u(Z_j) \ge 0$. Then the probability of choosing option $B_i \in Y \subseteq X$ equals

$$p(B_i \mid Y) = \sum u(Z_i) \cdot p(B_i \mid Z_i) / \sum u(Z_i), \qquad (3.9)$$

where each sum extends across all $Z_j \subseteq Y$ that are strictly contained in Y, and it is assumed that the denominator is greater than zero.

Tversky (1972b) proved that the EBA satisfies weak stochastic transitivity, and he also proved that EBA can be re-expressed as a random utility model, so that it also satisfies regularity and the triangle inequality. However, the EBA model does not satisfy strong stochastic transitivity, and it also violates independence.

In fact, the EBA model was originally designed to provide an elegant explanation of similarity effects on choice (Tversky, 1972a), but it cannot explain attraction and compromise effects. Mathematically, the EBA model is a Markov process (see Tversky, 1972b), and from this process one can derive choice response times (Busemeyer et al., 1988). However, the EBA model has not yet been used to explain changes in choice probability as a function of deliberation time.

3.3 Context Dependent Preference Model

After the discovery of attraction and compromise effects, Tversky and Simonson (1993) proposed the contextual dependent preference (CDP) model to account for these two new effects.³ This model assigns different values to an option depending on the other choice options in the choice set. The value of an option A presented in context of choice set Y is defined as

$$V(A|Y) = \sum_{k} \beta_{k} \cdot v(A_{k}) + \theta \sum_{Z \in Y} R(A, Z), \qquad (3.10)$$

where the index k refers to the k-th attribute. The first term on the right, $\Sigma_k \beta_k v(A_k)$, is a standard weighted sum of values for option A across the k attributes. This is a context

free contribution assigned to option A. In the second term, R(A,Z) is the relative advantage of option A over option Z, which is defined as

$$R(A,Z) = \frac{\sum_{k} a dv_k(A,Z)}{\sum_{k} a dv_k(A,Z) + \sum_{k} dis_k(A,Z)}$$
(3.11)

with $adv_k(A,Z) = v(A_k) - v(Z_k)$ if the advantage is positive (otherwise the advantage is zero), and $dis_k(A, Q) = \delta[(v(Z_k) - v(A_k))]$ if the disadvantage is positive (otherwise the disadvantage is zero). The function $\delta[x]$ is assumed to be a convex function that reflects loss aversion, and **q** is a constant that moderates the contextual effect. This contextual model provides a straightforward account of attraction and compromise effects based on the concept of loss aversion. Surprisingly, this same loss aversion principle prevents it from accounting for similarity effects (see Roe et al., 2001; Rieskamp et al., 2006, app. A). Furthermore, being a static model, it cannot explain the empirical fact that these effects grow larger with longer deliberation times. In fact, when fit to choices exhibiting both attraction and similarity effects, the model turns out to fit poorly and cannot fit similarity effects (Soltani et al., 2012).

3.4 Summary of Static Psychological Models

As mentioned earlier, the goal of these psychological choice models was to formulate psychological principles or mechanisms to capture context effects in a predictable manner. For example, the WIP Thurstone model uses spatial locations between options in the attribute space to derive predictions for similarity effects, and the EBA model derives predictions for similarity effects from the attribute sampling and elimination mechanism; the CDP model derives predictions for attraction and compromise effects using the principle of loss aversion. However, there are several serious limitations of these static models. First, there is no single model to account for all (similarity, compromise, attraction) effects simultaneously. Some models (WIP, EBA) account for similarity effects only, and others (CPA) account only for attraction and compromise effects. Second these models fail to describe how these context effects change across deliberation time. These deficiencies motivate the development of the dynamic models presented next.⁴

4 DYNAMIC PSYCHOLOGICAL MODELS

The static psychological models do not explicitly describe how people's preferences change as a function of deliberation time when there is a specific deadline. Nor do they describe the random time, T, chosen by the decision-maker to make a decision when there is no specific deadline. In contrast, dynamic models of choice behavior explicitly model how a deadline decision time affects people's choices, as well as the time T required to make a decision when there is no specific deadline.

4.1 'Horse Race' Choice Models

One of the first attempts to generalize the classic random utility family to include choice ponse time was a class called 'horse race' random utility models (Marley and Colonious, 1991). The underlying assumption is that when presented with a choice from set Y, each option is assigned a non-negative arrival time and option $B_i \subseteq Y$ is chosen if its 'arrival event' occurs first. It is assumed that the random arrival times for all options in the entire set X form a joint multivariate distribution, and the probability distribution for a subset $Y \subseteq X$ is obtained by marginalizing over the joint multivariate distribution of arrival times for the complete set X. The probability of choosing option B_i from set Y and this occurs after time t equals

$$p(B_i \wedge (T > t) \mid Y) = p([T(B_i) = \min\{T(B_i), B_i \in Y\}] \cap T(B_i) > t).$$
(3.12)

The 'horse race' choice model reduces to a random utility model when marginalizing over the time to make the decision; however, it extends the traditional random utility model by also providing a model of the distribution of choice time T.

Building on earlier ideas presented in Townsend and Ashby (1983), Otter et al. (2008) formulated an 'independent' version of the 'horse race' model. Otter et al. (2009) assumed that each choice option B_i in a set Y is associated with an independent counter $N_i(t)$ that counts up events favoring that option at time t. The first counter to reach a threshold wins the race and determines the choice. Events favoring an option occur at times distributed according to a Poisson process, and the rate assigned to each option depends on the attribute values of the choice option. The model was fit to choice and response time data from a survey involving 422 people, with each person providing stated choices for 18 different choice sets, and each choice set contained five television sets described by six attributes. The researchers found that by fitting both choice and response time (as compared to fitting only choice), they obtained a better marginal fit to the choice part of the data, which suggests that the model successfully extracts information contained in response times and that response times are informative about the cognitive processes underlying the observed choices.

The early versions of the 'horse race' choice model were random utility models, and so they inherited the triangle inequality and regularity properties. Therefore they could not explain violations of regularity such as the attraction effect. Nor were they used to account for similarity effects or the compromise effect. More recently, however, a linear ballistic accumulator (LBA) type of horse race model has been proposed that accounts for all three (similarity, attraction, compromise) context effects. The LBA model accomplishes this by introducing two new features into the horse race model – a front-end component that transforms choice stimuli into contextualized values and a back-end process that transforms these contextualized values into overt choices (Trueblood et al., 2014).

The independent Poisson 'horse race' model and the LBA models do account for changes in choice probability as a function of deliberation time. However, the Poisson model also predicts that the distribution of choice times becomes normal as the counter threshold increases (Otter et al., 2009). This prediction is problematic because, in fact, the choice response time distribution tends to become more positively skewed with longer mean deliberation times (Ratcliff and Smith, 2004). This problem is corrected by using the next class of dynamic models.

4.2 Sequential Sampling Choice Models

Sequential sampling models of decision-making were originally developed for Bayesian inference (DeGroot, 1970). Cognitive psychologists applied these models to a variety of cognitive tasks including sensory detection (Smith, 1995), perceptual discrimination (Laming, 1968; Link and Health, 1975; Usher and McClelland, 2001; Vickers, 1979), memory recognition (Brown and Heathcote, 2005; Ratcliff, 1978); categorization (Ashby, 2000; Nosofsky and Palmeri, 1997), probabilistic inference (Wallsten and Barton, 1982). Several sequential sampling models for preferential choice have also been proposed (Aschenbrenner et al., 1984; Bhatia, 2013; Fehr and Rangel, 2011; Glöckner and Betsch, 2008; Guo and Holyoak, 2002; Roe et al., 2001; Usher and McClelland, 2004).

The basic idea of many sequential sampling models is that when presented with a subset of *m* choice options $Y \subseteq X$ from a larger set *X*, a choice from his set takes some amount of deliberation time *T*. The decision process starts with an *m* state vector *P*(0), where each coordinate corresponds to the preference for one of the options in *Y*. This initial state *P*(0) reflects preferences for status quo options or biases from past experience. The initial state vector then evolves across time *t* by accumulating evaluations regarding the advantages and disadvantages of each option. After deliberating for a time *t*, the cumulative evaluations evolve to a new state vector *P*(*t*) according to the linear stochastic difference equation (technically an Ornstein-Uhlenbeck process – OU process),

$$dP(t+h) = P(t+h) - P(t) = -\Gamma \times P(t) \cdot h + \mu \cdot h + dB(t+h),$$
(3.13)

where dB(t+h) = B(t+h) - B(t) is a Brownian motion increment with mean zero and variance-covariance matrix $\Phi \cdot h$. The term, $\mu \cdot h + dB(t+h)$, is the new sample input into the process during the time step h, and $-\Gamma \cdot P(t) \cdot h$ is a feedback process that can be used to maintain preference stability. An important special case is the Wiener process, which is obtained by setting $\Gamma = 0$, but this allows preferences to grow without bound, and it also reduces the capability of accounting for context effects. The preference accumulation process continues until one of two stopping rules is satisfied – either a fixed or a variable stopping rule.

According to a fixed time stopping rule, the preference state evolves until some externally determined deadline time t occurs. In this case, the process evolves in an unconstrained manner until the deadline, t, at which point the option with the strongest preference state is chosen (that is, the option corresponding to the maximum coordinate of the P(T) vector). In the case of the fixed time stopping rule, the theory is equivalent to a dynamic version of a Thurstone model in which the state vector P(t) serves as an evolving vector of random utilities. However, unlike the static Thurstone model, the mean and the variance–covariance matrix of P(t) evolve across time, which accounts for changes in preferences as a function of deliberation time. Furthermore, if the feedback matrix is non-zero ($\Gamma \neq 0$), then both the mean and variance–covariance matrix of P(t) changes depending on the presented choice set Y. More specifically, if $\Gamma \neq 0$, then the joint distribution for a set of options Y cannot be derived by marginalization from a complete joint distribution defined over all possible options in X. Consequently, these theories do not have to satisfy

 (\bullet)

regularity.⁵ If the initial state P(0) starts out unbiased (i.e., P(0) = 0) then these theories satisfy both the stochastic transitivity and the triangle inequality. However, they generally violated by stochastic transitivity because of the changes in the variance–covariance matrix **F** across pairs of options.

According to the *variable* time stopping rule, preferences continue to evolve until the preference strength for one of the coordinates, corresponding to one of the options, exceeds a positive threshold. The first option exceeding the threshold is chosen, and the deliberation time equals the time it takes for an option to first cross a threshold. Figure 3.1 illustrates the process using the variable stopping time for three options $\{A, B, C\}$. In this figure, the horizontal axis represents time, and the vertical axis represents the preference state for each option; and in this case option A first crosses the threshold at a time equal to 2 seconds. The flat line near the top of the figure is called the threshold, and it is used to determine how strong the preference must be in order to make a decision. Increasing this threshold increases the average time to make a decision, which allows more evaluations of all of the options' advantages and disadvantages. Decreasing the



Note: The horizontal axis represents time, the vertical axis represents preference state for each option and the flat horizontal line represents the decision threshold that has to be crossed by one of the preference states so that a choice occurs. In this case, option A first crosses the threshold at a time equal to 2 seconds and is chosen by the decision-maker.

Figure 3.1 The preference state for each option/action changes over time by sampling information about the choice options

threshold decreases the average time to make a decision, which limits the ability to evaluate all the advantages and disadvantages of each option. The threshold is determined by the cost of sampling new information as compared to the gains and losses expected by the final decision after sampling more information. When the variable stopping time rule is used, sequential sampling models can be used to predict both the choice probabilities as well as the mean decision times. In fact, a strong test of the model is obtained by fitting the model parameters to the choice probabilities, and then using these same parameters to predict mean decision time.

Sequential sampling models provide very accurate accounts of the skewed shape of choice response time distributions (Ratcliff and Smith, 2004). In addition, they are designed to account for similarity, compromise and attraction effects using the same exact parameters for all three effects (see, Bhatia, 2013; Roe et al., 2001; Usher and McClelland, 2004). Furthermore, they make the strong a priori prediction that both compromise and attraction effects become larger with longer deliberation times, a prediction that has been confirmed by experimental evidence (Dhar et al., 2000; Pettibone, 2012).

As mentioned above, there are several versions of this general type of sequential sampling theory of preferential choice. One of the earliest was decision field theory (DFT) (represent the event of the eve hame parameters as the traditional Thurstone model with respect to the mean vector m and the covariance matrix Φ of the OU process. According to DFT, the mean μ and covariance matrix Φ are determined by the expectation and variance-covariance. respectively, of advantages or disadvantages for each option along a randomly sampled attribute (Roe et al., 2001). Different attributes are sampled across time by switching attention from one attribute to another during deliberation. Like the EBA model, attention-switching provides the mechanism used to account for similarity effects and violations of independence and strong stochastic transitivity. Also key to DFT, the coefficients in the feedback matrix Γ are det $\overline{\Omega}$ ned by the distance between options in a multi-attribute space of options (Hotaling et al., 2010). Three additional parameters are required to determine the feedback matrix G on the basis of the distances between the m options in the set Y (see Hotaling et al., 2010, for details). According to DFT, this feedback matrix is the key mechanism for predicting attraction and compromise effects in this theory. Importantly, DFT predicts all three effects (similarity, attraction and compromise), as well as other violations of independence, using the same fixed set of parameters. The feedback mechanism of DFT also explains why attraction and compromise effects become larger with longer deliberations. Decision field theory has been shown to predict choices between gambles better than popular models of decision-making under risk such as cumulative prospect theory (Rieskamp, 2008), and it also fits choices between consumer products better than another recently introduced cognitive model called the proportional difference choice model (Gonzalez-Vallejo, 2002; Scheibehenne et al., 2009). Finally it provides accurate cross validation predictions for mean decision time based on parameters derived from choice probabilities (Diederich and Busemeyer, 2003; Dror et al, 1999).

The leaky competing accumulator (LCA) model is another important version that shares many assumptions with DFT, but differs on some key mechanisms (Usher and McClelland, 2004).⁶ This model also assumes that people switch their attention during

the decision process to the different attributes and accumulate advantages or disadvantages for each option across time. This attention-switching mechanism allows LCA, similarly as with DFT, to explain violations of independence (such as the similarity effect) and violations of strong stochastic transitivity. The LCA model is also at the explain the attraction and compromise effects, but it uses a different mechanism and DFT. The LCA model does not assume distant dependence in the feedback matrix **G** (all off diagonal elements are the same). Instead it incorporates the same loss aversion principle as used in the context dependent preference model (Tversky and Simonson, 1993) to predict the compromise and attraction effects. The accumulation mechanism of LCA also enables the model to explain how attraction and compromise effects increase with longer deliberations. Decision field theory and the leaky competing accumulator model remain competitive by partly using different mechanisms to account for main behavior findings, but they are still quite similar and share many assumptions in contrast to more traditional random utility theories (Tsetsos et al., 2010).

Krajbich et al. (2010) developed a version of sequential sampling called the attention modulated drift-diffusion model (AM-DDM). Unlike DFT and LCA, it does not assume attention-switching across attributes, and instead it assumes attention shifting across choice alternatives. The AM-DDM assumes that each option accumulates random utilities according to a Wiener process ($\Gamma = 0$). The mean drift rate μ is simply determined by a traditional weighted sum of the attribute values, and $\Phi = \sigma^2 \mathbf{I}$, where σ^2 is the variance (a scalar). However, the mean drift rate is not constant during deliberation, but it changes depending on the option to which the decision-maker is attending. Using eye-movement recordings, it is possible to track the location of the gaze of the decision-maker at each moment during the decision process. The gaze is observed to shift from one option to another when they are spatially separated in a display. The mean drift rate of the Wiener process is assumed to change depending on the gaze of the decision-maker. The option currently being viewed is given more weight. If the option under gaze is advantageous then its advantage is enhanced during the gaze. If the option under gaze is disadvantageous, then this disadvantage is enhanced during the gaze. This model provided accurate quantitative predictions for complex relationships between eye fixation patterns and choices across time, as well as several fixation-driven decision biases. Unlike DFT, it does not make use of the feedback matrix ($\Gamma = 0$), and unlike LCA, it does not make use of loss aversion. Consequently, it has not been used to predict similarity, attraction, and compromise effects or other context effects and violations of independence.

The newest version of this class of sequential sampling models is the associative accumulation model (AAM) (Bhatia, 2013). Once again this model shares many assumptions with decision field theory such as sequential accumulation of advantages and disadvantages for options based on attention-switching across attributes to account for similarity effects. However, unlike decision field theory, the AAM does not make use of the feedback matrix Γ to account for attraction and compromise effects ($\Gamma = 0$). Instead it assumes that the average attention weight given to an attribute is mainly determined by the total absolute values assigned to the attribute summed across alternatives. Attributes present in extreme quantities in some alternatives, attributes present in many alternatives, and attributes present in especially salient alternatives, are more accessible relative to their competitors. This mechanism can account for not only compromise and attrac-

۲

 (\bullet)

tion effects, but also several other context effects such as alignment effects (Markman and Medin, 1995) and reference point effects (Tversky and Kahneman, 1991).

4.3 Summary of Dynamic Models

The dynamic psychological models are more complex than the static psychological models but they substantially expand their explanatory power. First, they are capable of predicting both choice probability and the distribution of time taken to make a choice. They are also capable of predicting the effects on time limits or time pressure on choice probabilities. Third, the dynamics provide mechanisms for simultaneously explaining all three types of context effects (similarity, attraction and compromise) as well as many other violations of independence and strong stochastic transitivity. None of the previously developed static models were found to be capable of predicting all these effects. This does not mean that it is not possible to discover a new static model that can do the job, but such a model remains to be developed.

Dynamic models, and sequential sampling models in particular, make strong predictions about the sampling process used to make decisions. One of the basic assumptions (except for the AM-DDM) is that information sampling is not affected by the person's current preference (formally, the input, $\mu \cdot h + dB(t+h)$, does not depend on the current state of preference P(t)). However recent studies using information search (Willemsen et al., 2011) and eye movements (Fiedler and Glöckner, 2012) indicates that the sampling process becomes biased to sample attributes that support the currently favored alternative. This finding suggests that the current preference state may feedback and modify the input into the accumulation process (see, for example, Glöckner and Betsch, 2008; Guo and Holyoak, 2002).

5 ADDITIONAL AND NEW DIRECTIONS

This section briefly reviews some additional important ideas for choice modeling coming from cognitive psychology and neuroscience.

5.1 Decision by Sampling

mand a covariance matrix **F** for the Brownian motion process. Another cognitive theory, called decision by sampling, provides more detailed mechanisms for the sequential evaluations entering the accumulation process (Stewart and Simpson, 2008; Stewart et al., 2006). Stewart et al. (2006) postulate that the subjective values of an attribute value are derived on-the-fly from comparisons with samples of other attribute values drawn from long-term memory. Furthermore, this long-term memory reflects the values experienced from real-world distributions. Stewart et al. (2006) showed that the shapes of utility functions, decision weighting functions and delay discounting functions estimated in decision research can be explained by the real-world distributions of gains, losses, probabilities, and delays that people experience.

۲

5.2 Quantum Choice Theory

Here we wish to return to one of the most fundamental issues concerning the probabilistic nature of choice. Why is choice probabilistic? One idea is that choice is fundamentally deterministic – each choice is determined by a real preference order or preference state existing at the moment immediately before a choice is made, but we (the theorist) do not have full information about this precise state. Another idea is that choice is fundamentally indeterministic – there is no real preference order or preference state existing at the moment immediately before a choice is made. Both the random utility and the sequential sampling class of models are deterministic in the sense that choice is based on a specific state that exists at the moment before the choice is made. Although the EBA model was originally cast as a probabilistic process, it can also be interpreted as a random utility model. Is there any empirical reason for questioning the deterministic underpinning of choice? Recently, interest has been growing in the use of principles based on quantum probability theory, which posits that choice is essentially indeterministic in nature (Lambert-Mogiliansky et al., 2009). The reason for considering a quantum approach is that preferences sometimes violate the law of total probability that lies at the heart of classic probabilistic models built upon deterministic mechanisms (Khrennikov, 2010; Busemeyer and Bruza, 2012; Pothos and Busemeyer, 2009; Yukalov and Sornette, 2011).

5.3 Decision Neuroscience

More recently, sequential sampling models of decision-making have attracted the interest of decision neuroscientists (Fehr and Rangel, 2011). Neuroscientists have been examining the neural basis for decision-making in the brains of Macaque monkeys using single cell recording techniques (Gold and Shadlen, 2001, 2002; Platt, 2002; Schall, 2003). Both the choice and the decision time of the monkeys were accurately predicted, trial by trial, based on where and when the neural activation crossed a threshold bound. Moreover, sequential sampling models have been fit to choice/response time data and then used to make a priori predictions for electrophysiological activation (Smith and Ratcliff, 2004). Although this research was based on single or multiple cell recording of saccadic eye movements from monkeys, converging evidence has been reported using cognitive tasks with humans. Researchers have recorded electrophysiological potentials from humans during a categorization task, called the lateralized readiness potential (Gratton et al., 1988). These potentials were recorded from the scalp above the premotor cortex that signals preparation for left- or right-hand movements to a cue. They found that choice and response times were determined by accumulation of the lateralized readiness potential to a threshold criterion. More direct evidence for accumulation to threshold as the basis for perceptual decision-making has been obtained from functional magnetic resonance imaging (Liu and Pleskac, 2011). Another imaging study by Gluth et al. (2012) explored the predictive accuracy of sequential sampling models for preferential choices. In their experiment the participants received sequential information about a stock company that could be bought or rejected. The results showed that a time-variant sequential sampling model using a decreasing rather than a fixed decision threshold was best in predicting the decision time and choice behavior. Furthermore the option's value as assumed by the sequential sampling model correlated with the brain activity in the

 (\bullet)

Psychological research and theories on preferential choice 67

ventromedial prefrontal cortex, the right and left orbitofrontal cortex, and the ventral striatum. These results indicate that the brain accumulates samples of information by forming an updated value representation in dopaminoceptive areas including the ventromedial prefrontal cortex and the ventral striatum. The conclusion from this research is the basic idea that decisions in the brain are based on the accumulation of noisy activation until a threshold is reached, which forms the basis for the sequential sampling models in cognitive science.

6 GENERAL CONCLUSIONS

In this chapter we have reviewed some of the main behavioral findings in behavioral economics, consumer research, and psychology. The list of findings provides a basis for evaluating the different approaches to predict and explain human choice behavior. On the one hand, we have shown that standard utility approaches, such as fixed and random utility models, have difficulties in explaining why people violate principles, such as strong stochastic transitivity, independence or regularity. On the other hand, we have shown that more cognitive models of decision-making that also aim to describe how preferential choices evolve dynamically over time can provide explanations for these findings. These models have been the result of the long and rich history in choice modeling in psychology. This work has led to more sophisticated models that have also increased the complexity of the models. The increase in theoretical complexity has been driven by two sources – the first is striving to explain the highly context-sensitive nature of human preferential choice behavior; the second is the attempt to account for more details of choice behavior including decision time, eve movements, information search and choice confidence (Pleskac and Busemeyer, 2010; Ratcliff and Starns, 2009; Van Zandt, 2000). Beyond that, decision neuroscientists have even begun to explain the neurological underpinnings of decision-making using electrophysiological recordings and/or functional magnetic resonance imaging during choice.

The psychological history of choice modeling began over 50 years ago with Luce's choice model (Luce, 1959), which was essentially a context-independent theory. Next, there were developments of Thurstone's (1959) theory to account for the effects of similarity between alternatives on choice. Aspect-based theories by Restle (1961) and Tversky (1972a) were also devised to account for similarity effects. However, these earlier models turned out to be inadequate to explain more complicated findings such as compromise and attraction effects. Dynamic models such as the sequential sampling models (DFT, LBA, LCA and AAM) are the first to be able to account for the full range of context effects in a coherent manner and predict how these effects change as a function of deliberation time. These models also add explanatory power by accounting for the strong effects of deliberation time on choice probability, and they gain empirical testability by making new predictions regarding the relations between choice probability and choice response time. More recent models, such as the decision by sampling model, provide a new capability for deriving the evaluations entering the sequential sampling accumulation process on the basis of long-term memory retrievals, where the memory is built from experience with real-world distributions. When developing sophisticated models, it is of course also important to examine to what extent the increased complexity

of the developed models can be justified by the increased goodness-of-fit in predicting human behavior. Only when the qualitative and quantitative increase in predictive accuracy of a model is large enough, the increased complexity can be justified.

Although the goals of psychologists and economist with respect to choice modeling overlap to some extent, there are also some clear differences. On the one hand, economists are often primarily interested in applying simple choice models to large samples of people with relatively few data per person, and the primary goal is to obtain efficient estimates of economically relevant parameters. For this goal, robust choice models are required that provide statistically efficient and computationally practical parameter estimates for further economic analysis. On the other hand, psychologists are primarily interested in explaining the cognitive processes that provide predictions across a wide range of puzzling findings using models that can simultaneously account for various measurements of choice including choice selection, decision time, confidence and brain activation. For this goal, more complex models are required that provide explanatory coherence and the power to predict new findings. But these are only two extreme positions, and more often the researcher is interested in both of these goals. For example, a marketing researcher may want to predict market share of a new product to be introduced to the market. The market share will depend on the context of the market; that is, the other products available. Given the huge costs of introducing a new product to the market (for example, product development and marketing) the precision with which the future share of the product can be predicted becomes essential. In this situation, using a complex choice model that requires higher methodological effort to be estimated can pay off in terms of obtaining better predictions of people's behavior.

NOTES

- 1. This work was supported by NIDA grant 5R01DA030551 to the first author and by a SNSF research grant 100014_130149 to the second author.
- 2. The Thurstone model has always been a theory of choice. Initially it was applied to psychophysics (Thurstone, 1927), but later (Thurstone, 1959) it was applied to values and preferences (Bock and Jones, 1968).
- 3. For some reason, similarity effects are ignored in this article.
- 4. Recently, Soltani et al. (2012) propose a range renormalization model to account for both attraction and similarity effects. This is also a static choice model. However, this model has not been shown to predict compromise effects and, furthermore, it is a static model that fails to describe how the effects change with deliberation time.
- 5. Tony Marley (personal communication, 9 November 2011) pointed out to the first author of this chapter that if $\Gamma = \mathbf{0}$ then this model becomes a special case of the 'horse race' model described in Marley and Colonious (1991) and satisfies regularity.
- 6. Strictly speaking, the LCA is not an OU process because they impose a lower threshold on the preferences so that they reflect off a bound at zero. This bound is used to make the model more similar to neural activations that must always be positive.

REFERENCES

Aschenbrenner, K.M., D. Albert and F. Schmalhofer (1984), 'Stochastic choice heuristics', *Acta Psychologica*, **56** (1–3), 153–66.

Ashby, F.G. (2000), 'A stochastic version of general recognition theory', *Journal of Mathematical Psychology*, **44** (2), 310–29.

- Becker, G.M., M.H. DeGroot and J. Marschak (1963), 'Stochastic models of choice behavior', *Behavioral Science*, **8** (1), 41–55.
- Ben Zur, H. and S.J. Bresnitz (1981), 'The effect of time pressure on risky choice behavior', *Acta Psychologica*, **47** (2), 89–104.
- Bhatia, S. (2013), 'Associations and the accumulation of preference', *Psychological Review*, **120** (3), 522-43.
- Birnbaum, M.H. and U. Schmidt (2008), 'An experimental investigation of violations of transitivity in choice under uncertainty', *Journal of Risk and Uncertainty*, **37** (1), 77–91.
- Bock, R.D. and L.V. Jones (1968), The Measurement and Prediction of Judgment and Choice. Oxford: Holden-Day.
- Böckenholt, Ú. (2006), 'Thurstonian-based analyses: past, present, and future utilities', *Psychometrika*, **71** (4), 615–29.
- Brown, S. and A. Heathcote (2005), 'A ballistic model of choice response time', *Psychological Review*, **112** (1), 117–28.
- Busemeyer, J.R. (1985), 'Decision making under uncertainty: a comparison of simple scalability, fixed-sample, and sequential-sampling models', *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 11 (3), 538–64.
- Busemeyer, J.R. and J.T. Townsend (1993), 'Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment', *Psychological Review*, **100** (3), 432–59.
- Busemeyer, J.R., B.H. Forsyth and G. Nozawa (1988), 'The use of choice response time to distinguish between elaborated versions of the elimination by aspects and suppression of aspects choice models', *Journal of Mathematical Psychology*, **32** (3), 341–49.
- Carrol, J.D. and G. DeSoute (1991), 'Toward a new paradigm for the study of multiattribute choice behavior', *American Psychologist*, **46** (4), 342–52.
- Chu, Y.P. and R.L. Chu (1990), 'The subsidence of preference reversals in simplified and marketlike experimental settings – a note', American Economic Review, 80 (4), 902–11.
- DeGroot, M.H. (1970), Optimal Statisical Decisions, New York: McGraw-Hill.
- Dhar, R., S.M. Nowlis and S.J. Sherman (2000), 'Trying hard or hardly trying: an analysis of context effects in choice', *Journal of Consumer Psychology*, 9 (4), 189–200.
- Diederich, A. (2003), 'MDFT account of decision making under time pressure', *Psychonomic Bulletin & Review*, **10** (1), 157–66.
- Diederich, A. and J.R. Busemeyer (2003), 'Conflict and the stochastic-dominance principle of decisionmaking', *Psychological Science*, **10** (4), 353–59.
- Dror, I.E., J.R. Busemeyer and B. Basola (1999), 'Decision making under time pressure: an independent test of sequential sampling models', *Memory & Cognition*, 27 (4), 713–25.
- Erev, I. and G. Barron (2005), 'On adaptation, maximization, and reinforcement learning among cognitive strategies', *Psychological Review*, **112** (4), 912–31.
- Fehr, E. and A. Rangel (2011), 'Neuroeconomic foundations of economic choice recent advances', Journal of Economic Perspectives, 25 (4), 3–30.
- Fiedler, S. and A. Glöckner (2012), 'The dynamics of decision making in risky choice: an eye-tracking analysis', *Frontiers in Psychology*, **3**, 335, doi: 10.3389/fpsyg.2012.00335.
- Glöckner, A. and T. Betsch (2008), 'Modeling option and strategy choices with connectionist networks: towards an integrative model of automatic and deliberate decision making', *Judgment and Decision Making*, 3 (3), 215–28.
- Gluth, S., J. Rieskamp and C. Büchel, C. (2012), 'Deciding when to decide: time-variant sequential sampling models explain the emergence of value-based decisions in the human brain', *Journal of Neuroscience*, **32** (31), 10686–98.
- Gold, J.I. and M.N. Shadlen (2007), 'The neural basis of decision making', Annual Review of Neuroscience, 30 (July), 535–74.
- Gold, J.I. and M.N. Shadlen (2002), 'Banburismus and the brain decoding the relationship between sensory stimuli, decisions, and reward', *Neuron*, **36** (October), 299–308.
- Gold, J.I. and M.N. Shadlen (2001), 'Neural computations that underlie decisions about sensory stimuli', *Trends in Cognitive Sciences*, **5** (1), 10–16.
- Gonzalez, C. and V. Dutt (2011), 'Instance based learning: integrating decisions from experience in sampling and repeated choice experiments', *Psychological Review*, **118** (4), 523–51.
- Gonzalez-Vallejo, C. (2002), 'Making trade-offs: a probabilistic and context-sensitive model of choice behavior', *Pschological Review*, **109** (1), 137–54.
- Gratton, G., M.G.H. Coles, E.J. SirevaagC.W. Eriksen and E. Donchin (1988), 'Pre- and poststimulus activation of response channels: a psychophysiological analysis', *Journal of Experimental Psychology. Human Perception and Performance*, 14 (3), 331–44.
- Guo, F.Y. and K.J. Holyoak (2002), 'Understanding similarity in choice behavior: a connectionist model',

 (\bullet)

in W.D.Gray and C.D.Schunn (eds), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, Mahwah, NJ: Erlbaum, pp. 393–8.

Halff, H.M. (1976), 'Choice theories for differentially comparable alternatives', *Journal of Mathematical Psychology*, **14** (3), 244–6.

Hausman, J.A. and D.A. Wise (1978), 'A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous references', *Econometrica*, **46** (2), 403–26.

Heath, T.B. and S. Chatterjee (1995), 'Asymmetric decoy effects on lower quality versus higher quality brands: meta-analytic and experimental evidence', *Journal of Consumer Research*, 22 (3), 268–84.

Hertwig, R. and I. Erev (2009), 'The description-experience gap in risky choice', *Trends in Cognitive Science*, **13** (12), 517–23.

Hey, J.D. (2001), 'Does repetition improve consistency?', Experimental Economics, 4 (1), 5-54.

Hotaling, J.M., J.R. Busemeyer and J. Li (2010), 'Theoretical developments in decision field theory: a comment on K. Tsetsos, N. Chater and M. Usher', *Psychological Review*, 117 (4), 1294–8.

Huber, J. and C. Puto (1983), 'Market boundaries and product choice: illustrating attraction and substitution effects', *Journal of Consumer Research*, **10** (1), 31–44.

Huber, J., J.W. Payne and C. Pluto (1982), 'Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypthesis', *Journal of Consumer Research*, 9 (1), 90–98.

Iverson, G.J. and J.-C.Falmagne. (1985), 'Statistical issues in measurement', *Mathematical Social Sciences*, 10 (2), 131–53.

Janis, I.L. and L. Mann (1977), Decision Making: A Psychological Analysis of Choice, Conflict, and Decision Making, New York: Free Press.

Khrennikov, A.Y. (2010), Ubiquitous Quantum Structure: From Psychology to Finance: Berlin: Springer.

Kivetz, R., O. Netzer and V. Srinivasan (2004), 'Alternative models for capturing the compromise effect', Journal of Marketing Research, 41 (3), 237–57.

Krajbich, I., C. Armel and A. Rangel (2010), 'Visual fixations and the computation and comparison of value in simple choice', *Nature Neuroscience*, **13** (10), 1292–8.

Lambert-Mogiliansky, A., A. Zamir and H. Zwirn (2009), 'Type indeterminacy: a model of the KT (Kahneman-Tversky) man', *Journal of Mathematical Psychology*, **53** (5), 349-61.

Laming, D.R. (1968), Information Theory of Choice-Reaction Times. New York: Academic Press.

Link, S.W. and R.A. Health (1975), 'A sequential theory of psychological discrimination', *Psychometrika*, **40** (1), 77–111.

Liu, T. and T.J. Pleskac (2011), 'Neural correlates of evidence accumulation in a perceptual decision task', *Journal of Neurophysiology*, **106** (5), 2383–98.

Loomes, G., P.G. Moffat and R. Sugden (2002), 'A microeconomic test of alternative stochastic theories of risky choice', *Journal of Risk and Uncertainty*, 24 (2), 103–30.

Lopes, L.L. (1996), 'When time is of the essence: averaging, aspiration, and the short run', *Organizational Behavior & Human Decision Processes*, **65** (3), 179–89.

Louviere, J.J., D.A. Hensher and J.D. Swait (2000), *Stated Choice Models: An Analysis and Applications*, Cambridge: Cambridge University Press.

Luce, R.D. (1959), Individual Choice Behavior, New York: Wiley.

Luce, R.D. (2000), Utility of Gains And Losses: Measurement-Theoretical and Experimental Approaches. Mahwah, NJ: Lawrence Erlbaum.

Luce, R.D. and P. Suppes (1965), 'Preference, utility, and subjective probability', in R.D. Luce, R.B. Bush and E. Gallanter (eds), *Handbook of Mathematical Psychology*, vol. 3, New York: Wiley, pp.249–410.

Markman, A. and D.L. Me 1995), 'Similarity and alignment in choice', Organizational Behavior and Human Decision Processes, (1), 117-30.

Marley, A.A.J. and H. Colonious (1991), 'The "horse race" random utility model and its competing risks interpretation', *Journal of Mathematical Psychology*, 36 (1), 1–20.

May, K. (1954), 'Intransitivity, utility, and the aggregation of preference patterns', Econometrica, 22 (1), 1-13.

McFadden, D. (1981), 'Econometric models of probabilistic choice', in C.F. Manski and D. McFadden (eds), Structural Analysis of Discrete Data with Econometric Applications, Cambridge, MA: MIT Press, pp. 198–272.

Mellers, B.A. and K. Biagini (1994), 'Similarity and choice', Psychological Review, 101 (3), 505-18.

Mosteller, F. and P. Nogee (1951), 'An experimental measurement of utility', *Journal of Political Economy*, **59** (5), 371–404.

Nosofsky, R.M. and Palmeri, T.J. (1997), 'An exemplar-based random walk model of speeded classification', *Psychological Review*, **104** (2), 266–300.

Otter, T., G.M. Allenby and T. van Zandt (2008), 'An integrated model of discrete choice and response time', Journal of Marketing Research, 45 (5), 593–607.

Payne, J.W., J.R. Bettman and E.J. Johnson (1993), *The Adaptive Decision Maker*, New York: Cambridge University Press.

- Pettibone, J. (2012), 'Testing the effect of time pressure on asymmetric dominance and compromise in choice', *Judgment and Decision Making*, **7** (4), 513–23.
- Platt, M.L. (2002), 'Neural correlates of decisions', Current Opinion in Neurobiology, 12 (2), 141-8.
- Pleskac, T.J. and J.R. Busemeyer (2010), 'Two-stage dynamic signal detection: a theory of choice, decision time, and confidence', *Psychological Review*, **117** (3), 864–901.
- Pothos, E.M. and J.R. Busemeyer (2009), 'A quantum probability model explanation for violations of "rational" decision making', *Proceedings of the Royal Society*, *B.*, **276** (1665), 2171–8.
- Ratcliff, R. (1978), 'A theory of memory retrieval', Psychological Review, 85 (2), 59-108.
- Ratcliff, R. and P. Smith (2004), 'A comparison of sequential sampling models for two-choice reaction time', *Psychological Review*, **111** (2), 333–67.
- Ratcliff, R. and J. Starns (2009), 'Modeling confidence and response time in recognition memory', *Pschological Review*, **116** (1), 59–83.
- Regenwetter, M., J. Dana and C.P. Davis-Stober (2011), 'Transitivity of preferences', *Psychological Review*, **118** (1), 42–6.
- Restle, F. (1961), Psychology of Judgment and Choice, New York: Wiley.
- Rieskamp, J. (2008), 'The probabilistic nature of preferential choice', Journal of Experimental Psychology: Learning, Memory, & Cognition, 34 (6), 1446–65.
- Rieskamp, J. and U. Hoffrage (2008), 'Inferences under time pressure: how opportunity costs affect strategy selection', Acta Psychologica, 127 (2), 258–76.
- Rieskamp, J., J.R. Busemeyer and B.A. Mellers (2006), 'Extending the bounds of rationality: evidence and theories of preferential choice', *Journal of Economic Literature*, 44 (3), 631–61.
- Roe, R.M., J.R. Busemeyer and J.T. Townsend (2001), 'Multialternative decision field theory: a dynamic connectionist model of decision making', *Psychological Review*, **108** (2), 370–92.
- Rumelhart, D.E. and J.G. Greeno (1971), 'Similarity between stimuli: an experimental test of the Luce and Restle choice models', *Journal of Mathematical Psychology*, **8** (3), 370–81.
- Schall, J.D. (2003), 'Neural correlates of decision processes: neural and mental chronometry', Current Opinion in Neurobiology, 13 (2), 182–6.
- Scheibehenne, B., J. Rieskamp and C. Gonzalez-Vallejo (2009), 'Cognitive models of choice: comparing decision field theory to the proportional difference model', *Cognitive Science*, 33 (5), 911–39.
- Simonson, I. (1989), 'Choice based on reasons: the case of attraction and compromise effects', Journal of Consumer Research, 16 (2), 158–74.
- Smith, P.L. (1995), 'Psychophysically principled models of visual simple reaction time', *Psychological Review*, 102 (3), 567–93.
- Smith, P.L. and R. Ratcliff (2004), 'The psychology and neurobiology of simple decisions', Trends in Neurosciences, 27 (3), 161–8.
- Soltani, A., B. De Martino and C. Camerer (2012), 'A range-normalization model of context dependent choice: a new model and evidence', *PLoS Computational Biology*, 8 (7), e1002607.
- Stewart, N. and K. Simpson (2008), 'A decision-by-sampling account of decision under risk', in N. Chater and M. Oaksford (eds), *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*, Oxford: Oxford University Press, pp. 261–76.
- Stewart, N., N. Chater and G.D.A. Brown (2006), 'Decision by sampling', Cognitive Pschology, 53 (1), 1-26.
- Svenson, O. and A. Edland (1987), 'Changes of preference under time pressure: choices and judgements', Scandanavian Journal of Psychology, 28 (4), 322–30.
- Svenson, O. and A.J. Maule (1993), *Time Pressures and Stress in Judgment and Decision Making*, New York: Plenum.
- Takane, Y. (1987), 'Analysis of covariance structures and binary choice data', *Cognition and Communication*, **20**, 45–62.
- Thurstone, L.L. (1927), 'Law of comparative judgment', Psychological Review, 34 (4), 273-6.
- Thurstone, L.L. (1959), The Measurement of Values, Chicago, IL: University of Chicago Press.
- Townsend, J.T. and F.G. Ashby (1983), Stochastic Modeling of Elementary Psychological Processes, Cambridge: Cambridge University Press.
- Train, K.E. (2003), Discrete Choice Methods with Simulation, Cambridge: Cambridge University Press.
- Tsai, R.C. and U. Böckenholt (2006), 'Modelling intransitive preferences: a random-effects approach', *Journal of Mathematical Psychology*, **50** (1), 1–14.
- Trueblood, J.S., S.D. Brown and A. Heathcote, A. (2014), 'The multi-attribute linear ballistic accumulator model of context effects in multi-alternative choice', *Psychological Review*, forthcoming.
- Tsetsos, K., M. Usher and N. Chater, N. (2010), 'Preference reversal in multialternative choice', *Psychological Review*, **117** (4), 1275–93.
- Tversky, A. (1969), 'Intransitivity of preferences', Psychological Review, 76 (1), 31-48.
- Tversky, A. (1972a), 'Elimination by aspects: a theory of choice', *Psychological Review*, 79 (4), 281–99.
- Tversky, A. (1972b), 'Choice by elimination', Journal of Mathematical Psychology, 9 (4), 341-67.

- Tversky, A. and D. Kahneman. (1991), 'Loss aversion in riskless choice: a reference-dependent model', *Quarterly Journal of Economics*, **106** (4), 1039–61.
- Tversky, A and J.E. Russo (1969), 'Similarity and substitutability in binary choices', *Journal of Mathematical Psychology*, 6 (1), 1–12.

Tversky, A and S. Sattath (1979), 'Preference trees', Psychological Review, 86 (6), 542-72.

Tversky, A. and I. Simonson (1993), 'Context dependent preferences', Management Science, 39 (10), 1179-89.

Usher, M. and J.L. McClelland (2001), 'The time course of perceptual choice: the leaky, competing accumulator model', *Psychological Review*, **108** (3), 550–92.

Usher, M. and J.L. McClelland (2004), 'Loss aversion and inhibition in dynamic models of multialternative choice', *Psychological Review*, **111** (3), 757–69.

Van Zandt, T. (2000), 'ROC curves and confidence judgments in recognition memory', Journal of Experimental Psychology: Learning, Memory, & Cognition, 26 (3), 582–600.

Vickers, D. (1979), Decision Processes in Visual Perception, New York: Academic Press.

- Wallsten, T.S. and C. Barton (1982), 'Processing probabilistic multidimensional information for decisions', Journal of Experimental Psychology: Learning, Memory, and Cognition, 8 (5), 361-84.
- Wedell, D.H. (1991), 'Distinguishing among models of contextually induced preference reversals', Journal of Experimental Psychology: Learning, Memory, & Cognition, 17 (4), 767–78.
- Willemson, M.C., U. Böckenholt and E.J. Johnson (2011) 'Choice by value coding and value construction: processes of loss aversion', *Journal of Experimental Psychology: General*, 140 (3), 303–24.
- Wright, P. (1972), 'The harassed decision maker: time pressures, distractions, and the use of evidence', *Journal* of Applied Psychology, **59** (5), 555–61.

Yellot, J.I. (1977), 'The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution', *Journal of Mathematical Psychology*, 15 (2), 109–44.

Yukalov, V.I. and D. Sornette (2011), 'Decision theory with prospect interference and entanglement', *Theory* and Decision, **70** (3) 283–328.

 (\bullet)