JOURNAL OF MATHEMATICAL PSYCHOLOGY 15, 215–233 (1977)

The Choice Axiom after Twenty Years*

R. DUNCAN LUCE

Department of Psychology and Social Relations, Harvard University, Cambridge, Massachusetts 02138

This survey is divided into three major sections. The first concerns mathematical results about the choice axiom and the choice models that devolve from it. For example, its relationship to Thurstonian theory is satisfyingly understood; much is known about how choice and ranking probabilities may relate, although little of this knowledge seems empirically useful; and there are certain interesting statistical facts. The second section describes attempts that have been made to test and apply these models. The testing has been done mostly, though not exclusively, by psychologists; the applications have been mostly in economics and sociology. Although it is clear from many experiments that the conditions under which the choice axiom holds are surely delicate, the need for simple, rational underpinnings in complex theories, as in economics and sociology, leads one to accept assumptions that are at best approximate. And the third section concerns alternative, more general theories which, in spirit, are much like the choice axiom.

Perhaps I had best admit at the outset that, as a commentator on this scene, I am qualified no better than many others and rather less well than some who have been working in this area recently, which I have not been. My pursuits have led me along other, somewhat related routes. On the one hand, I have contributed to some of the recent, purely algebraic aspects of fundamental measurement (for a survey of some of this material, see Krantz, Luce, Suppes, & Tversky, 1971). And on the other hand, I have worked in the highly probabilistic area of psychophysical theory; but the empirical materials have led me away from axiomatic structures, such as the choice axiom, to more structural, neural models which are not readily axiomatized at the present time. After some attempts to apply choice models to psychophysical phenomena (discussed below in its proper place), I was led to conclude that it is not a very promising approach to these data, and so I have not been actively studying any aspect of the choice axiom in over 12 years. With that understood, let us begin.

MATHEMATICAL AND STATISTICAL RESULTS

The Choice Axiom

Anyone reading this article is probably already familiar with the choice axiom (axiom I of Luce, 1959), so I confine myself to a brief reminder. One part says that if a choice set S

* A slightly abridged version of this article will appear in a memorial volume for P. F. Lazarsfeld edited by J. S. Coleman, R. K. Merton, and P. Rossi. Drs. D. M. Green, T. Indow, A. A. J. Marley, P. Suppes, A. Tversky, and J. I. Yellott, Jr. commented on an earlier draft of this paper. In several cases, errors of fact or interpretation were pointed out as well as various omissions and obscurities; my thanks to each for his thoughtful remarks.

ISSN 0022-2496

contains two elements, a and b, such that a is never chosen over b when the choice is restricted to just a and b; then a can be deleted from S without affecting any of the choice probabilities. The other, more substantive part, concerns choice situations where all the deletions of the first part have been carried out. Suppose R is a subset of S; then the choice probabilities for the choice set R are assumed to be identical to the choice probabilities for the choice set S conditional on R having been chosen, i.e., for a in R

$$P_R(a) = P_S(a \mid R).$$

Several easy consequences of this are that the ratio $P_R(a)/P_R(b)$ is independent of R (the constant-ratio rule), that the pairwise probabilities satisfy

$$P(a, b)\, P(b, c)\, P(c, a) = P(a, c)\, P(c, b)\, P(b, a)$$

(the product rule), and that there exists a numerical ratio scale v over T such that for all a in R,

$$P_R(a) = v(a) \Big/ \sum_{b \text{ in } R} v(b).$$

We say $v(a)$ is a response strength associated with response a, and the axiom implies response probability is proportional to response strength. This representation is closely related to the logit analysis used in statistics. For two-element choice sets, it was proposed earlier by Bradley and Terry (1952), and so some authors (e.g., Suppes & Zinnes, 1963) refer to these as BTL models. Others, especially economists, refer to it as the strict utility model. As Fantino and Navarick (1974) note, it recurs repeatedly in people's thinking about choice probabilities.

The Relation to Thurstone's Theory

It was apparent from the outset that in many situations the numbers predicted by the choice axiom differ but little from those predicted by the special case of Thurstone's discriminable dispersion model (called the random utility model by economists and probit analysis by statisticians) in which the random variables are independent, are normally distributed, and have the same variance. Indeed, the paired comparison situation leads to the logistic distribution for the choice model and the normal ogive for Thurstone's, and these are well known to be very similar except in the tails. One study (Burke & Zinnes, 1965), aimed at empirical tests, concerned ways to try to maximize the difference between the two models. Large differences in the prediction of $P(a, c)$ in terms of $P(a, b)$ and $P(b, c)$ arise only when one of these probabilities is very near 0 and the other is very near 1, but it is not practical to estimate such extreme values with sufficient accuracy to make the test workable.

The first major step in understanding the exact relationship between these two models, due to Block and Marschak (1960), is the theorem that for any system of probabilities satisfying the choice axiom, there is an independent Thurstone model defined on the whole real line that yields these probabilities. Their proof was very indirect and later

E. Holman and A. A. J. Marley (reported in Luce & Suppes, 1965) gave a direct proof. In particular, an independent Thurstone model with random variables of the form

$$P(X \leqslant t) = f(t)^{v(x)}, \qquad f(-\infty) = 0, f(\infty) = 1, -\infty < t < \infty,$$

yields the choice model. If one restricts this to random variables that differ only in their means—a shift family—then one obtains the double exponential[1] of the form

$$P(X \leqslant t) = e^{-e^{-\alpha t+\beta}}.$$

A question not then answered, and only recently resolved by McFadden (1974) and independently and under appreciably less restrictive conditions[2] by Yellott (1977), is whether this distribution is unique or not. That is to say, while we can reproduce any set of probabilities satisfying the choice axiom by using appropriate double exponential distributions, can we do the same thing with some different distribution? The answer is "No" provided the Thurstone model is restricted to a shift family and the choice set has at least three elements. This is by no means obvious, nor is it easy to prove. The case of just binary choices is more complex; Bradley (1965) discussed it.

In the course of investigating this problem, Yellott has brought to our attention two related facts. The first, which was also pointed out earlier in the context of paired comparison models by Thompson and Singh (1967) (see also Beaver & Rao, 1972), is that the double exponential distribution arose very many years ago in statistics (Fisher & Tippett, 1928; Gumbel, 1958) as an answer to the question: given n independent, identically distributed random variables, what is the asymptotic ($n \rightarrow \infty$) distribution of the maximum? This is similar to the question answered in the central limit theorem, but with "maximum" substituted for "average" (or "sum"). There are various possibilities, one of which is the double exponential which arises when the underlying distribution has an upper exponential tail.

The second fact is a new and very pretty result which, had Thurstone known it, might have led him to postulate the double exponential rather than the normal distribution. Yellott asked and answered the following question. Suppose choices are determined by a Thurstone model involving random variables that are independent and a shift family and that the choice probabilities exhibit the following plausible invariance property: let each element of a choice set be replicated k times—think of choices among books or phonograph records with several identical copies of each—the probability of choosing any particular element is independent of the value of k. Yellott has shown that under this invariance assumption the distributions in the Thurstone model must be double

[1] There is a good deal of terminological confusion. Sometimes the LaPlace distribution is called the double exponential, whereas $e^{-e^{-x}}$ is often called the log–Weibull distribution because the distribution e^{-x^k}, $x \geqslant 0$, is known as the Weibull distribution. Both Holman and Marley and McFadden stated the result in Weibull form, but in order to have the variables run from $-\infty$ to ∞, Yellott uses the double exponential form.

[2] McFadden's proof is more restrictive in two respects. He assumed that the distribution of the random variable is translation complete and that there are infinitely many objects of choice; neither is required in Yellott's proof.

exponentials, and hence the choice axiom must hold. In many contexts, Yellott's condition is so compelling that his theorem means Thurstone's model with a shift family of independent random variables and the choice axiom stand or fall together.

It should be noted that Yellott's invariance condition is closely related to Debreu's (1960) counterexample to the choice axiom: Suppose a person is indifferent between a and b and a' and b, $P(a, b) = P(a', b) = \frac{1}{2}$ and that a and a' are very similar so $P(a, a') = \frac{1}{2}$. Because of the similarity of a and a', Debreu contends $P_A(a) = P(a, b) = \frac{1}{2}$, where $A = \{a, a', b\}$. However, according to the choice axiom

$$P(a', b) = P(a, b) \quad \text{iff} \quad P_A(a) = P_A(a'),$$
$$P(a', a) = P(a', b) \quad \text{iff} \quad P_A(b) = P_A(a),$$

whence $P_A = \frac{1}{3} < P(a, b)$. The key difference between Yellott and Debreu is that the former uses equal numbers of replications whereas the latter does not.

Limit Properties

Let L_n denote a probability distribution on the first n integers, $n \geqslant 2$. Steele (1974) called a sequence of such distributions, $n = 2, 3, \ldots$, a *Luce process* provided they exhibit the constant-ratio rule, i.e., for each $m < n$ and for $i, j \leqslant m$,

$$P_n(i)/P_n(j) = P_m(i)/P_m(j).$$

Next, let us define for each L_n a measure P_n on the unit interval by placing atoms at the points k/n, $1 \leqslant k \leqslant n$, with mass $P_n(k)$. Using the results of Karamata (1933) on regularly varying functions, Steele showed that if the sequence L_n is a Luce process and if the associate measures P_n converge in distribution to P, then for some α, $0 \leqslant \alpha \leqslant \infty$, $P[0, x] = x^{\alpha}$. So far as I know, this result has yet to be put to use.

Discard and Acceptance Mechanisms

Luce (1960) suggested that choices may be effected by systematically discarding elements from a choice set. If so and if we denote the probability of discarding b from A by $Q_A{}^*(b)$, then a natural inductive hypothesis is

$$P_A(a) = \sum_{b \text{ in } A-\{a\}} Q_A{}^*(b)\, P_{A-\{b\}}(a).$$

There it was shown that if the choice axiom also holds, then

$$Q_A{}^*(b) = [1 - P_A(b)]/[|A| - 1],$$

where $|A|$ is the number of elements in A. So in this very special case the choice probabilities can arise from a discard mechanism.

Marley (1965) studied the problem more generally. We say the choice probabilities satisfy *regularity* if, for a in $A \subseteq B$, $P_A(a) \geqslant P_B(a)$. This is a very weak property which almost everyone believes to be true of choice probabilities. However, as Corbin and Marley (1974) have pointed out, even it may not always be true. For example, a guest in deference to his host selects the second most expensive item on a menu; this choice of course is altered by adding an entree more expensive than any on the current menu. Marley showed that regularity is sufficient, but not necessary, for the discard equation to hold, that the discard probabilities are unique when the inequality in regularity is strict, and he gave an expression for them.

Following Marley (1968), let P^* denote aversion probabilities, i.e., $P_A{}^*(a)$ is the probability that a is selected from A as the least desired element. One can define an acceptance condition that parallels the discard one by

$$P_A{}^*(a) = \sum_{b \text{ in } A-\{a\}} Q_A(b)\, P^*_{A-\{b\}}(a).$$

These two conditions, discard and acceptance, are called strong if $Q^* = P^*$ and $Q = P$.

Another plausible relation between the two kinds of probabilities is Marley's *concordant condition*

$$P_A(a)\, P^*_{A-\{a\}}(b) = P_A{}^*(b)\, P_{A-\{b\}}(a).$$

He showed that the concordant condition implies that both the strong discard and the strong acceptance conditions hold, but that neither set of choice probabilities can satisfy the choice axiom.

Rankings

Georgescu-Roegen (1958, 1969) and Luce (1959) independently discussed how choice and ranking probabilities might interrelate, and Block and Marschak (1960) and Marschak (1960) extended some of these results considerably.

Let us denote the choice probabilities by $P_A(a)$ and the probability of the rank order $a_1 > a_2 > \cdots > a_n$ by $p(a_1, ..., a_n)$. Let U be the universal, finite set under consideration. A major difference among assumptions is whether the rankings are just of U or of all of its subsets.

Georgescu-Roegen (1958) discussed the following two interrelations.

A. Suppose ranking probabilities exist for all subsets of U, then for any set $\{a_1, ..., a_n\} \subseteq U$,

$$p(a_2, a_3, ..., a_n) = p(a_1, a_2, a_3, ..., a_n) + p(a_2, a_1, a_3, ..., a_n) + \cdots + p(a_2, a_3, ..., a_n, a_1).$$

B. Suppose the ranking probabilities exist for all subsets of U, then for $A_n = \{a_1, ..., a_n\} \subseteq U$,

$$P_{A_n}(a_1) = \sum_{\rho} p(a_1, a_{\rho(2)}, ..., a_{\rho(n)}),$$

where the sum extends over all permutations of $a_2, ..., a_n$.

Note that for $A_2 = \{a_1, a_2\}$, B yields

$$P(a_1, a_2) = p(a_1, a_2),$$

which together with A applied inductively yields that $P(a_1, a_2)$ can be expressed as the sum of the probabilities of all rankings of U for which a_1 is ahead of a_2. This property was suggested by Luce (1959) and generalized by Block and Marschak (1960) to the following. Let $R(a, A)$ denote all of the rank orderings of U in which a in A ranks ahead of all elements in $A - \{a\}$.

B′. $$P_A(a) = \sum_{\rho \text{ in } R(a,A)} p(\rho).$$

Luce (1959) proposed that the ranking probabilities might arise by repeated selections of the best element. This can be formulated in either of two ways depending on which ranking probabilities are assumed to exist.

R. If the ranking probabilities exist for all subsets of U,

$$p(a_1, a_2, ..., a_n) = P_{A_n}(a_1)\, p(a_2, a_3, ..., a_n).$$

R′. If the ranking probabilities exist on U,

$$p(a_1, a_2, ..., a_n) = P_U(a_1)\, P_{U-\{a_1\}}(a_2) \cdots P(a_{n-1}, a_n).$$

The major results appear to be these:

(1) (Block & Marschak, 1960) A system of choice probabilities on the subsets of U form a random utility (Thurstone) model iff there exist ranking probabilities on U such that B′ holds.

(2) (Block & Marschak, 1960; Luce, 1959 for $n = 3$) A system of choice probabilities on the subsets of U satisfy the choice axiom iff there exists ranking probabilities on U such that B′ and R′ hold.

(3) (Georgescu-Roegen, 1969) Suppose a system of choice probabilities and ranking probabilities on the subsets of U satisfy R. Then the choice axiom is equivalent to A.

Georgescu-Roegen (1958) asserted that if there are systems of rankings and choice probabilities satisfying A and B, then it is not possible for the choice probabilities to be a function just of the binary ones. In particular, then, the choice axiom may not hold. This has been questioned by Chakrabarti (1969) and by me in correspondence and replied to by Georgescu-Roegen (1969). It is clear that some misunderstanding exists. By (3) above and by the simple fact, pointed out by Georgescu-Roegen (1969), that B follows from R, we see that R and the choice axiom imply A and B. But if he is correct, these imply that the choice axiom cannot hold. The difficulty lies in the proof in the 1958 paper.

A complication arises when we admit that rankings can be generated by choice mechanisms in at least two ways—either by repeatedly selecting the best alternative or by repeatedly selecting the worst. Luce and Suppes (1965) summarize the impossibility theorem that arises if one demands both that the probability of the ranking not depend on which process is sused and that the choice axiom holds. More insight is provided by Marley (1968) who showed that this ranking hypothesis is equivalent to assuming the concordant condition which, as we noted above, implies that the choice axiom does not hold.

Applications

As we see in the next section, psychologists have been at pains to test the adequacy of the choice axiom as a descriptive model, but there has been relatively little application of it as the foundation of some other theory. I made some attempts in Luce (1959, 1963) and Luce and Galanter (1963) to build a psychophysical theory on it, and Nakatani (1972) and Shipley (1965) extended those ideas, but realistically I do not think it has been a particularly successful program.

In sociology, there are two large texts on mathematical sociology. There is a brief mention of the choice axiom in the introduction to Coleman (1964), but he does not appear to employ it later; whereas Fararo (1973) devotes much of one of 25 chapters to expositing the idea. However, so far as I can see, nothing is really built on it.

In economics, there appear to be two sorts of applications. One is typified by the work of McFadden and his colleagues (for a survey, see McFadden, 1976) in which different possible models of choice are used to fit empirical data which arise in economic contexts of some complexity. This work is more in the spirit of data handling and reduction of statistics than economic theory. The other, which is more deeply theoretical, attempts to reconstruct the foundations of rational economic theory, basing it on probabilistic rather than algebraic assumptions. Work of this sort was initiated by Georgescu-Roegen (1936, 1958). The latter paper raised the key questions: "I. What axioms are logically necessary and experimentally justified, to relate the multiple choice probability to that of the binary choice? II. Granted these axioms, what is the reflection of the 'average' demand into the binary choice map?" Halldin (1974) undertook the investigation of the second question on the assumption that the choice axiom is a reasonable answer to the first question. As he pointed out, the choice axiom is a possible probabilistic formulation of "... a leading ... and powerful ... principle in the traditional economic theory of consumer behavior: by the definition of rationality, the choice made from any given budget set B depends exclusively on the ordering of the alternatives in B." He proceeds to develop on this foundation a generalization of the classical theory of demand. He summarizes his results as follows. "On the basis of Luce's theory then, we conclude that whereas the mode or most probable demand of the uncertain consumer will always coincide with his classical and optimal demand, his mean or expected demand will not, that is, his expected demand will not reveal preferences as demand reveals preferences in the classical case, nor can it be identified with the demand he would have if, as in classical theory, his preferences were perfectly certain."

Experimental Tests

By now, quite a variety of experiments have either been performed to test or been interpreted in terms of the choice axiom, and so it is helpful to group them into related categories. Even so, only the briefest mention of individual studies is possible. I have elected to group most of the animal studies together at the end.

Binary Data

Because the choice axiom implies the binary constraint

$$\frac{P(a, b)\, P(b, c)\, P(c, a)}{P(b, a)\, P(c, b)\, P(a, c)} = 1,$$

known as the product rule, binary (paired comparison) data provide an indirect test of the axiom. This property is equivalent to the binary choice model of Bradley and Terry (1952),

$$P(a, b) = \frac{v(a)}{v(a) + v(b)}.$$

In an early paper on taste preferences, Hopkins (1954) used Bradley's (1953) estimation scheme and a χ^2 evaluation and found the Bradley–Terry model satisfactory. Burke and Zinnes (1965) fitted both the binary Thurstone and the Bradley–Terry model to three published sets of data—Gulliksen and Tukey (1958) in which 200 subjects evaluated pairs of handwriting specimens, Guilford (1954) in which a single subject judged each pair of weights 200 times, and Thurstone (1959) in which 266 subjects judged the seriousness of crimes. They estimated two of the three probabilities, predicted the third, and used χ^2 as a measure of goodness of fit. Thurstone's data were indecisive and the other two favored the Thurstone model. As they pointed out, assigning all of the variability to just one of the three probabilities is not really very satisfactory. Hohle (1966) carried out another analysis of existing data: lifted weights (Guilford, 1931), preferences among vegetables (Guilford, 1954), samples of handwriting (Hevner, 1930), and numerousness (Hohle). His more sophisticated analysis involved maximum likelihood estimates for the scale values of both the Bradley–Terry and Thurstone models and evaluated the fit in terms of $-2 \ln \lambda$, where λ is the likelihood ratio, which is asymptotically distributed as χ^2. By this test, the Bradley–Terry model was favored, although in over half of the comparisons the discrepancies were significant. In the worst case, the average absolute deviation was 0.032. Skvoretz, Windell, and Fararo (1974) used as stimuli two sets of five occupations each, with one occupation in common. A group of 79 subjects were presented with all 36 pairs and asked to "choose that occupation which, *in general*, you accord a higher social standing." Using Bradley's procedures, these were fitted to the binary model and the appropriate χ^2 test yielded $p > .25$. So, for the average subject, the binary model fit very well. Later, under rankings, I discuss other data from this study.

Several authors, probably motivated by Debreu's (1960) counter example to the choice axiom (actually, to simple scalability; see below), have emphasized the need to take into account similarity relations among the stimuli. Restle (1961) proposed a binary choice theory which has recently been generalized by Tversky (1972a, b) to arbitrary sets of alternatives. Rumelhart and Greeno (1971) designed a study in which certain subgroups of stimuli were similar, but different groups were not. Their data clearly disconfirmed the Bradley–Terry model and were consistent with the Restle model. Their procedure of estimation was criticized by Edgell, Geisler, and Zinnes (1973), but their conclusion was not altered by correcting the estimates.

Nonbinary Data: The Constant Ratio Rule

The constant ratio rule (CRR) (Clarke, 1957) version of the choice model has attracted considerable attention, especially as a way of dealing with confusion matrices. Early studies include Anderson (1959), Clarke (1957, 1959), Clarke and Anderson (1957), and Egan (1957a, b). These authors all concluded that the CRR is reasonably satisfactory. Later studies along much the same lines, including Atkinson, Bower, and Crothers (1965, pp. 146–150, analyzing data of W. K. Estes), Hodge (1967), Hodge and Pollack (1962), Pollack and Decker (1960), and Waganaar (1968), continued to interpret the data as reasonably favorable, although not always. Hodge (1967, p. 435) put it this way: "Taken as a whole, the predictions are good enough to make the CRR a useful tool in designing practical stimulus displays and response systems (cf., Engstrand & Moeller, 1962). As a model of choice behavior, however, the rule as presently formulated is somewhat unsatisfactory. The results of the present experiment and those of Hodge and Pollack demonstrate that the CRR has difficulty predicting the response proportions of single dimensional ensembles, especially when the 2 by 2 matrices involve variations in the spacing-range conditions and practice on the task."

But even the apparently favorable cases are cast into doubt by Morgan (1974) who worked out a likelihood ratio test for the CRR and reexamined some of the published data, which he found departs significantly from predicted values. His discussion of the difficulties in running a suitable experiment to test the CRR is illuminating.

Yellott and Curnow (1967) ran a study in which one of three possible target letters was flashed briefly at random in a 4×4 matrix along with 15 other nontarget letters. On some trials the subject had all three responses available and on the remainder one of the two incorrect responses was eliminated by the experimenter after the presentation and before the response. Assuming that choices were a mixture of guesses and observations, it was found that the choice axiom accounted well for the data of the four subjects. When an error was made, on trials for which all three responses were available, they required a second choice. These were not well predicted, but they argued that this is not a satisfactory test of the choice axiom because of sequential effects introduced by the selection of trials.

Tversky (1972b) conducted a psychophysical and two preference studies of the CRR. The psychophysical data exhibited no significant departure, but in the preference data the observed binary proportions were larger than those predicted from trinary data. This is plausible in the light of his model, which generalizes both Restle and the CRR (see below), in which stimulus similarity plays a role.

Nonbinary Data: Response Biases

In Luce (1959) and more fully in Luce (1963) a class of models was developed in which the scale values associated with responses are decomposed multiplicatively into two factors, a response bias b_j associated solely with the response and a stimulus factor, η_{ij} being the impact of stimulus i on response j. The choice axiom then yields the form

$$P(r_j \mid s_i) = \eta_{ij} b_j \Big/ \sum_k \eta_{ik} b_k .$$

In the latter paper I attempted, with some degree of success, to use this to account for various psychophysical data, but certain phenomena proved recalcitrant to this approach. Perhaps the most striking of these occurs in absolute identification, where the information transmitted for stimuli spread over a large range grows almost linearly with log n until about $n = 7$ at which point it becomes constant (Miller, 1956) or for 10 or more stimuli it grows linearly with a range up to a point (15–20 dB in auditory intensity) after which its rate is greatly reduced (Braida & Durlach, 1972). Indeed, until very recently no model has successfully accounted for these phenomena (Luce, Green, & Weber, 1976).

Shipley and Luce (1964) collected considerable data (weights) and attempted to fit them with this model with, at best, modest success. Shipley (1965) collected additional auditory data and found that the simple model would not do and she developed a number of elaborate modifications of it which, however, have not been pursued. Townsend (1971) fit the model to an alphabet confusion matrix with rather better results. Broadbent (1967) argued that it is a suitable model to explain the word-frequency effect in the perception of words, but his interpretation was sharply questioned by Catlin (1969) and Nakatani (1970) and defended by Treisman (1971).

Nakatani (1968, 1972) proposed a generalization of the response bias model which he calls confusion-choice recognition theory. He has had success in fitting it to word recognition data. For a summary, see Luce and Green (1974, p. 307).

β-Learning Model

The form of biased choices just described suggested (Luce, 1959) a possible mechanism for learning, not unrelated to the linear operator models then in vogue. The idea was that on trial n there is a set of response strengths $v_n(a)$, a in A, and if alternative a is selected on trial n then depending on the outcome it is changed to $\beta v_n(a) = v_{n+1}(a)$ on trial $n + 1$. A positive reinforcement corresponds to $\beta > 1$ and a negative one to $0 < \beta < 1$. In terms of the probabilities, these operators are nonlinear, but they exhibit instead of linearity the very powerful property of commutativity. This captures a learning process in which there is no forgetting or attenuation of the past.

Mathematical properties of these operators were investigated by Bush (1960), Kanal (1960, 1962a, b), Lamperti & Suppes (1960), and Luce (1959). In Luce (1964) whole families of commutative operators were investigated and, in general, were shown to behave much like the β model. Bush, Galanter, & Luce (1959) conducted some experimental tests of the model, using animals, but their conclusions were sharply questioned by Sternberg (1963, p. 96–97) who pointed out that optimal estimates of the parameters were considerably different from their nonoptimal ones.

Rankings

Although some nice theoretical results relate rankings to choices, few relevant data have been reported. Part of the reason for this is that once one goes beyond three alternatives, the number of observations required to get reasonably accurate estimates of ranking probabilities is nearly prohibitive. For example, with $n = 5$ there are 120 possible rankings, and so 100 observations per rank yields a total of 12,000 judgments.

I am aware of only two sets of data. One, collected by W.K. Estes, was analyzed by

Atkinson *et al.* (1965, pp. 150–155). The averages from 17 subjects seem to provide support for the ranking hypothesis R (see above).

The other study is in Skvoretz, Windell, and Fararo (1974) who had 155 subjects rank order the two sets of occupations that had been used in the paired comparisons study with another group of subjects. They then employed assumption B to estimate the pairwise choice probabilities, fit the binary choice model to get scale values, used these to generate choice probabilities, and from these and the ranking postulate R predicted the ranking probabilities. A χ^2 test strongly rejected this model. They concluded that the ranking postulate is in error; however, it is really the combination of B, the choice axiom, and R that are rejected. In particular, they do not report for these data how well the binary choice model works.

Simple Scalability

Although the evidence so far reviewed makes one suspicious that the choice axiom is not terribly accurate empirically, to me the most compelling data against it are the several studies that were explicitly designed to undermine the whole broad class of models satisfying simple scalability. Luce and Suppes (1965) defined the *constant utility models* to be those for which there is a real-valued function u over the alternatives and real function F_n on n real arguments such that

$$P_A(a_1) = F_n[u(a_1); u(a_2),..., u(a_n)],$$

where $A = \{a_1, a_2, ..., a_n\}$. If, in addition, we assume that F_n is strictly increasing in the first argument and strictly decreasing in the remaining $n - 1$, then we say *simple scalability* holds. Tversky (1972a) showed this to be equivalent to the following simple observable property called *order independence*: for a, b in $A - B$ and c in B,

$$P_A(a) \geqslant P_A(b) \quad \text{iff} \quad P_{B\cup\{b\}}(c) \geqslant P_{B\cup\{a\}}(c),$$

provided none of the probabilities is 0. This generalizes earlier results on binary simple scalability (Krantz, 1967; Tversky & Russo, 1969) which had been shown to be equivalent to each of the following well-known properties:

Strict Stochastic Transitivity. $P(a, b) \geqslant \frac{1}{2}$ and $P(b, c) \geqslant \frac{1}{2}$ imply $P(a, c) \geqslant \max[P(a, b), P(b, c)]$, and $>$ in either hypothesis implies $>$ in the conclusion.

Substitutability. $P(a, c) \geqslant P(b, c)$ iff $P(a, b) \geqslant \frac{1}{2}$.

Independence. $P(a, c) \geqslant P(b, c)$ iff $P(a, d) \geqslant P(b, d)$.

To begin with, against simple scalability is Debreu's (1960) example. It was raised against the choice axiom, but the argument actually only uses order independence (see p. 7). Almost everyone, including me, agrees that Debreu is empirically correct, and so simple scalability must be wrong.

Empirically, we also have several unambiguous studies. Coombs (1958) showed that preferences among shades of gray include triples of alternatives, which he could identify on the basis of his unfolding ideas, that violate strict stochastic transitivity and hence simple scalability (the full impact of his result was not fully realized by most of us at the time). Chipman (1960) found in a study of subjective probability that he could reject strict stochastic transitivity for 6 of 10 subjects. Krantz (1967) showed that substitutability, and hence simple scalability, failed for judgments about monochromatic colors. Whether or not one alternative can be substituted for another was found to be context dependent. And Tversky and Russo (1969) exhibited failures of independence, and hence simple scalability, in judgments of the relative size of simple geometric figures that varied in two dimensions. "It was found that the similarity between stimuli facilitate the discrimination between them. But since the similarity between two stimuli can be varied without changing their scale values, simple scalability, and hence independence, must be violated" (p. 11).

These results are deeply disturbing because their variety of domains make it difficult to see in what domain simple scalability, let alone the choice axiom, may hold. And at the same time, all of the data reported are in large part consistent with simple scalability, and so it is difficult to abandon the idea completely. As we see, Tversky was led from these problems to his choice-by-elimination model, but even that may not be without problems.

Expected Random Utility (Thurstone) Model

Becker, DeGroot, and Marschak (1963b), drawing upon a theoretical analysis in Becker, DeGroot, and Marschak (1963a), designed a study of choices among gambles which had the following property. No expected random utility model predicts that one of the alternatives (which was an average of the other two) in each choice set would ever be chosen. There were 25 such sets and 60 of the 62 subjects selected the forbidden alternative at least once in 25 times. This study casts doubt upon just about any way of generalizing the expected utility property in a probabilistic context.

Animal Data

In his text, Greeno (1968) analyzed several sets of data. Food preferences of rats (Young, 1947) gave good agreement with the product rule. However, the proportions of time spent in alternative activities (Allison, 1964) did not satisfy the product rule very well. And finally, monkey data on times devoted to activities allowed good predictions.

By far the most systematic program of animal research that bears on these matters is that of R. J. Herrnstein and his collaborators. He has concerned himself with choice situations defined in an operant framework in terms of different keys associated with different reinforcement schedules. The behavior observed is key presses and reinforcements received. The major generalization from these data is that strength of responding is proportional to its relative rate of reinforcement. Put more formally, if there are n alternative responses, r_i denotes the number of reinforcements delivered per ith response, r_0 is a constant, and R_i is the number of occurrences of response i, then

$$R_i = kr_i \Big/ \sum_{j=0}^{n} r_j .$$

If P_i denotes the proportion of i responses, then

$$P_i = R_i \Big/ \sum_{j=1}^{n} R_j$$

$$= \left(kr_i \Big/ \sum_{j=0}^{n} r_j\right) \Big/ \left(\sum_{j=1}^{n} kr_j \Big/ \sum_{h=0}^{n} r_h\right)$$

$$= r_i \Big/ \sum_{j=1}^{n} r_j$$

which, of course, is the choice axiom representation, but with the scale value having an explicit empirical interpretation.

For a general discussion of these ideas, summaries of the data, and more detailed references, see Fantino and Navarick (1974), Herrnstein (1970, 1974), and de Villiers and Herrnstein (1976). Herrnstein and Loveland (1976) have studied the product rule during extinction trials and found that the choice axiom prediction is not nearly as extreme as the data.

Generalizations

As we have seen, there is much evidence that choices are not generally governed by axioms as simple as the choice axiom. Indeed, the difficulty pervades the whole broad class of models exhibiting simple scalability. That being so, various other approaches have been taken.

There is of course the general class of Thurstone models, which economists call random utility models. These are the ones with a random variable U_a associated with each response alternative and the assumption that

$$P_A(a) = \Pr(U_a \geqslant U_b \text{ for all } b \text{ in } A).$$

Mostly, only independent random utility models have been used; these are ones in which the random variables are assumed to be independent of each other. As Tversky (1972b) pointed out, a familiar example (Luce & Raiffa, 1957, p. 375) makes clear that this model will not suffice. Consider nearly equivalent trips to Paris and to Rome, and let the added symbol $+$ denote some minor added benefit, such as a slight reduction in price. It is evident that for $a =$ either Paris or Rome, $P(a+, a) = 1$, but that it is quite possible for both $P(\text{Paris} +, \text{Rome}) < 1$ and $P(\text{Rome} +, \text{Paris}) < 1$. Assuming an independent Thurstone model, the first two equalities imply no overlap between the distributions of a and $a+$ for $a =$ Paris and $a =$ Rome, whereas the last two inequalities imply an overlap between Paris + and Rome and between Rome + and Paris. These statements are mutually inconsistent for independent random variables.

So the only remaining real possibility that retains the intuitive idea of a utility indicator is some form of a nonindependent Thurstone model. Although Restle (1961) and Eisler

(1964) have suggested binary generalizations, it was not until Tversky (1972a, b) (see also Corbin & Marley, 1974) that a full generalization was proposed. Tversky took as his point of departure an equation which in spirit is much like the discard condition stated above but which admits that more than one alternative may be discarded at a time. If $B \subseteq A$, let $Q_A(B)$ denote the probability that the alternatives A–B are discarded, then he proposed the Markovian condition

$$P_A(a) = \sum_{\substack{B \\ B \subsetneq A}} Q_A(B)\, P_B(a)$$

as describing the elimination process. By itself, this imposes no restrictions on the choice probabilities. However, he suggested that the transition probabilities A may satisfy the following condition: Suppose $B, C \subseteq A \subseteq T$, then

$$Q_A(B)/Q_A(C) = \sum_{\substack{D \\ D \cap A = B}} Q_T(D) \Big/ \sum_{\substack{D \\ D \cap A = C}} Q_T(D),$$

provided that both denominators are positive. No rationalization was offered for this condition. From it and the Markovian condition on the elimination process, he showed that there is a function U over the subsets of T such that

$$P_A(a) = \sum_{\substack{B \\ B \cap A \neq A}} U(B)\, P_{A \cap B}(a) \Big/ \sum_{\substack{B \\ B \cap A \neq \phi}} U(B).$$

An alternative argument for this recursive form is given in Tversky (1972b), where he identified each alternative by the set of aspects (or components) that characterizes it. He suggested that people proceed by selecting an aspect in proportion to its measure and then eliminate all alternatives not possessing it. This process is continued until exactly one alternative remains, which is their choice. Such a process can be translated into an equation of the above form.

This model reduces to the choice model in the special circumstance where the pairs of alternatives are aspect disjoint. Furthermore, for binary alternatives it is equivalent to Restle's (1961) model. But perhaps its nicest feature is the number of desirable consequences that can be derived from it. Tversky (1972a) showed that it implies regularity (see the section on Discard and Acceptance Mechanisms), moderate stochastic transitivity: if $P(a, b) \geqslant \frac{1}{2}$ and $P(b, c) \geqslant \frac{1}{2}$, then $P(a, c) \geqslant \min[P(a, b), P(b, c)]$; and $P_{\{a,b,c\}}(a) \geqslant P(a, b)\, P(a, c)$. The latter multiplicative inequality was generalized in Sattath and Tversky (1976) to: $P_{A \cup B}(a) \geqslant P_A(a)\, P_B(a)$. Thus, it and regularity provide lower and upper bounds on $P_{A \cup B}(a)$ in terms of $P_A(a)$ and $P_B(a)$. It is recalled that regularity is a condition that nearly everyone believes to be true of most choices. And whereas strong stochastic transitivity (max for min above), which is equivalent to binary simple scalability, has been rejected in data, moderate stochastic transitivity has not been.

Tversky also showed how to construct ranking probabilities so that property B of the section on ranking holds. Since by Block and Marschak's theorem, this is equivalent to the choice probabilities satisfying a random utility model, this means Tversky's model has this property (of course, it does not involve independent random variables).

To the best of my knowledge, the only data so far reported that conceivably might reject this elimination-by-aspects model are the gambling ones of Becker, DeGroot, and Marschak (1963b) that reject every expected random utility model; however, the trouble may well be with the expected utility property. Corbin and Marley's (1974) example against regularity raises doubts about any random utility model and so EBA, but it has not yet been explored empirically. Indow (1975) in a survey article on the choice axiom and Tversky's model questions whether it or any mathematical model can capture all that is involved. "... It may not be quite realistic as a [substantive] model for actual choice behavior of human beings. Choice is in essence *conflict* and will not proceed in such a straightforward way defined in the eliminating process. Such cases occur rather frequently in which no alternative is easily eliminated from a certain stage of decision making on and they create conflict or attractiveness of aspects changes their balance from moment to moment, which also leads to conflict. Certainly, to take all these features into consideration will make mathematical formulation untractable. Then, computer simulation will be a possible way of approach." Others, however, are unwilling yet to give up the quest for a general structural model or at least for several classes of such models to cover different domains.

Conclusion

After somewhat less than 20 years, where does the choice axiom stand? As a descriptive tool, it is surely imperfect; sometimes it works well, other times not very well. As Debreu and Restle made clear and as has been repeatedly demonstrated experimentally, it fails to describe choice behavior when the stimulus set is structured in such a way that several alternatives are treated as substantially the same. It probably also fails whenever the experimental subjects have the belief—which I fear may often be the case—that they should employ the response alternatives about equally often. This tends to introduce some form of response bias which can well differ from one experimental run to another. Keep in mind that once we enter the path of strict rejection of models on the basis of statistically significant differences, little remains. To the best of my knowledge, the only property of general choice probabilities that has never been empirically disconfirmed is regularity—a decrease in the choice set does not decrease the probability of choosing any of the remaining alternatives—but even that looks suspect in a Gedanken experiment. Despite these empirical difficulties, there remains a tendency to invoke the choice axiom in many behavioral models—often implicitly. This is partly because it is so simple and the resulting computations are so easy.

Perhaps the greatest strength of the choice axiom, and one reason it continues to be used, is as a canon of probabilistic rationality. It is a natural probabilistic formulation of K. J. Arrow's famed principle of the independence of irrelevant alternatives, and as such it is a possible underpinning for rational, probabilistic theories of social behavior. Thus, in the development of economic theory based on the assumption of probabilistic individual choice behavior, it can play a role analogous to the algebraic rationality postulates of the traditional theory.

However long the choice axiom may prove useful, at least during the 1960s and 1970s it contributed to the interplay of ideas about choices that arose in economics, psychology, and statistics.

References

Allison, J. Strength of preference for food, magnitude of food reward, and performance in instrumental conditioning. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 217–223.

Anderson, C. D. The constant-ratio rule as a predictor of confusions among various stimuli of brief exposure duration. Technical Note, AFCRC-TN-58-60, Hearing and Communication Laboratory, Indiana University, 1959.

Atkinson, R. C., Bower, G. H., & Crothers, E. J. *An introduction to mathematical learning theory.* New York: Wiley, 1965.

Beaver, R. J., & Rao, P. V. The use of limit theorems in paired and triple comparison model building. *Journal of Mathematical Psychology*, 1972, **9**, 92–103.

Becker, G. M., DeGroot, M. H., & Marschak, J. Stochastic models of choice behavior. *Behavioral Science*, 1963, **8**, 41–55. (a)

Becker, G. M., DeGroot, M. H., & Marschak, J. An experimental study of some stochastic models for wagers. *Behavioral Science*, 1963, **8**, 199–202. (b)

Block, H. D., & Marschak, J. Random orderings and stochastic theories of responses. In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics.* Stanford, Calif.: Stanford University Press, 1960. Pp. 97–132.

Bradley, R. A. Some statistical methods in taste testing and evaluation. *Biometrics*, 1953, **9**, 22–38.

Bradley, R. A. Another interpretation of a model for paired comparisons. *Psychometrika*, 1965, **30**, 315–318.

Bradley, R. A., & Terry, M. E. Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, 1952, **39**, 324–345.

Braida, L. D., & Durlach, N. I. Intensity perception II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America*, 1972, **51**, 483–502.

Broadbent, D. E. Word-frequency effect and response bias. *Psychological Review*, 1967, **74**, 1–15.

Burke, C. J., & Zinnes, J. L. A paired comparison of pair comparisons. *Journal of Mathematical Psychology*, 1965, **2**, 53–76.

Bush, R. R. Some properties of Luce's beta model for learning. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences*, 1959. Stanford, Calif.: Stanford University Press, 1960. Pp. 254–264.

Bush, R. R., Galanter, E., & Luce, R. D. Tests of the "beta model." In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford, Calif.: Stanford University Press, 1959. Pp. 382–399.

Catlin, J. On the word-frequency effect. *Psychological Review*, 1969, **76**, 504–506.

Chakrabarti, S. K. A note on the relation between binary and multiple choice probabilities. *Econometrica*, 1969, **37**, 726–727.

Chipman, J. S. Stochastic choice and subjective probability. In D. Willner (Ed.), *Decisions, values and groups*, Vol. 1. New York: Pergamon, 1960. Pp. 70–95.

Clarke, F. R. Constant-ratio rule for confusion matrices in speech communication. *Journal of the Acoustical Society of America*, 1957, **29**, 715–720.

Clarke, F. R. Proportion of correct responses as a function of the number of stimulus-response alternatives. *Journal of the Acoustical Society of America*, 1959, **31**, 835.

Clarke, F. R., & Anderson, C. D. Further test of the constant-ratio rule in speech communication. *Journal of the Acoustical Society of America*, 1957, **29**, 1318–1320.

Coleman, J. S. *Introduction to mathematical sociology*. Glencoe, Ill.: Free Press, 1964.
Coombs, C. H. On the use of inconsistency of preferences in psychological measurement. *Journal of Experimental Psychology*, 1958, **55**, 1–7.
Corbin, R., & Marley, A. A. J. Random utility models with equality: An apparent, but not actual, generalization of random utility models. *Journal of Mathematical Psychology*, 1974, **11**, 274–293.
Debreu, G. Review of R. D. Luce, Individual choice behavior: A theoretical analysis. *American Economic Review*, 1960, **50**, 186–188.
de Villiers, P. A., & Herrnstein, R. J. Toward a law of response strength. *Psychological Bulletin*, 1976, **83**, 1131–1153.
Edgell, S. E., Geisler, W. S. III, & Zinnes, J. L. A note on a paper by Rumelhart and Greeno. *Journal of Mathematical Psychology*, 1973, **10**, 86–90.
Egan, J. P. Message repetition, operating characteristics, and confusion matrices in speech communication. Technical Report 57-50, Hearing and Communication Laboratory, Indiana University, 1957. (a)
Egan, J. P. Monitoring task in speech communication. *Journal of the Acoustical Society of America*, 1957, **29**, 482–489. (b)
Eisler, H. A choice model for paired comparison data based on imperfectly nested sets. *Psychometrika*, 1964, **29**, 363–370.
Engstrand, R. D., & Moeller, G. The relative legibility of ten simple geometric figures. *American Psychologist*, 1962, **17**, 386.
Fantino, E., & Navarick, D. Recent developments in choice. In G. H. Bower (Ed.), *The psychology of learning and motivation*, Vol. 8. New York: Academic Press, 1974. Pp. 147–185.
Fararo, T. J. *Mathematical sociology*. New York: Wiley, 1973.
Fisher, R. A., & Tippett, L. H. C. Limiting forms of the frequency distribution of the largest or the smallest of a sample. *Proceedings of the Cambridge Philosophical Society*, 1928, **24**, 180–190.
Georgescu-Roegen, N. The pure theory of consumer's behavior. *Quarterly Journal of Economics*, 1936, **50**, 545–593.
Georgescu-Roegen, N. Threshold in choice and the theory of demand. *Econometrica*, 1958, **26**, 157–168.
Georgescu-Roegen, N. The relation between binary and multiple choices: some comments and further results. *Econometrica*, 1969, **37**, 728–730.
Greeno, J. G. *Elementary theoretical psychology*. Reading, Mass.: Addison–Wesley, 1968.
Guilford, J. P. Some empirical tests of the method of paired comparisons. *Journal of Genetic Psychology*, 1931, **5**, 64–77.
Guilford, J. P. *Psychometric methods* (2nd ed.). New York: McGraw–Hill, 1954.
Gulliksen, H., & Tukey, J. W. Reliability for the law of comparative judgment. *Psychometrika*, 1958, **23**, 95–110.
Gumbel, E. J. *Statistics of extremes*. New York: Columbia Press, 1958.
Halldin, C. The choice axiom, revealed preferences, and the theory of demand. *Theory and Decision*, 1974, **5**, 139–160.
Herrnstein, R. J. On the law of effect. *Journal of the Experimental Analysis of Behavior*, 1970, **13**, 243–266.
Herrnstein, R. J. Formal properties of the matching law. *Journal of the Experimental Analysis of Behavior*, 1974, **21**, 159–164.
Herrnstein, R. J., & Loveland, D. H. Matching in a network. *Journal of the Experimental Analysis of Behavior*, 1976, **26**, 143–153.
Hevner, K. An empirical study of three psychophysical methods. *Journal of Genetic Psychology*, 1930, **4**, 191–212.
Hodge, M. H. Some further tests of the constant-ratio rule. *Perception & Psychophysics*, 1967, **2**, 429–437.
Hodge, M. H., & Pollack, I. Confusion matrix analysis of single and multidimensional auditory displays. *Journal of Experimental Psychology*, 1962, **63**, 129–142.

HOHLE, R. H. An empirical evaluation and comparison of two models for discriminability scales. *Journal of Mathematical Psychology*, 1966, **3**, 174–183.

HOPKINS, J. W. Incomplete block rank analysis: Some taste test results. *Biometrics*, 1954, **10**, 397–399.

INDOW, T. On choice probability. *Behaviometrika*, 1975, **2**, 13–31.

KANAL, L. Analysis of some stochastic processes arising from a learning model. Ph.D. dissertation, University of Pennsylvania, 1960.

KANAL, L. A functional equation analysis of two learning models. *Psychometrika*, 1962, **27**, 89–104.(a)

KANAL, L. The asymptotic distribution for the two-absorbing-barrier beta model. *Psychometrika*, 1962, **27**, 105–109. (b)

KARAMATA, J. Sur un mode de croissance régulière: Théorèmes fondamentaux. *Bulletin de la Société Mathématique de France*, 1933, **61**, 55–62.

KRANTZ, D. H. Rational distance functions for multidimensional scaling. *Journal of Mathematical Psychology*, 1967, **4**, 226–245.

KRANTZ, D. H., LUCE, R. D., SUPPES, P., & TVERSKY, A. *Foundations of measurement*, Vol. 1. New York: Academic Press, 1971.

LAMPERTI, J., & SUPPES, P. Some asymptotic properties of Luce's beta learning model. *Psychometrika*, 1960, **25**, 233–241.

LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.

LUCE, R. D. Response latencies and probabilities. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences*, 1959. Stanford: Stanford University Press, 1960. Pp. 298–311.

LUCE, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 1. New York: Wiley, 1963. Pp. 103–189.

LUCE, R. D. Some one-parameter families of commutative learning operators. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford, Calif.: Stanford University Press, 1964. Pp. 380–398.

LUCE, R. D., & GALANTER, E. Discrimination. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 1. New York: Wiley, 1963. Pp. 193–243.

LUCE, R. D., & GREEN, D. M. Detection, discrimination, and recognition. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception*, Vol. II. New York: Academic Press, 1974. Pp. 299–342.

LUCE, R. D., GREEN, D. M., & WEBER, D. L. Attention bands in absolute identification. *Perception & Psychophysics*, 1976, **20**, 49–54.

LUCE, R. D., & RAIFFA, H. *Games and decisions*. New York: Wiley, 1957.

LUCE, R. D., & SUPPES, P. Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 3. New York: Wiley, 1965. Pp. 249–410.

MARLEY, A. A. J. The relation between the discard and regularity conditions for choice probabilities. *Journal of Mathematical Psychology*, 1965, **2**, 242–253.

MARLEY, A. A. J. Some probabilistic models of simple choice and ranking. *Journal of Mathematical Psychology*, 1968, **5**, 333–357.

MARSCHAK, J. Binary-choice constraints and random utility indicators. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences*, 1959. Stanford: Stanford University Press, 1960. Pp. 312–329.

MCFADDEN, D. Conditional logit analysis of quantitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics*. New York: Academic Press, 1974. Pp. 105–142.

MCFADDEN, D. Quantal choice analysis: a survey. *Annals of Economic and Social Measurement*, 1976, **5**, 363–390.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

MORGAN, B. J. T. On Luce's choice axiom. *Journal of Mathematical Psychology*, 1974, **11**, 107–123.

NAKATANI, L. H. *A confusion-choice stimulus recognition model applied to word recognition*. Ph.D. dissertation, University of California at Los Angeles, 1968.

NAKATANI, L. H. Comments on Broadbent's response bias model for stimulus recognition. *Psychological Review*, 1970, **77**, 574–576.

NAKATANI, L. H. Confusion-choice model for multidimensional psychophysics. *Journal of Mathematical Psychology*, 1972, **9**, 104–127.

POLLACK, I., & DECKER, L. Consonant confusions and the constant-ratio rule. *Language and Speech*, 1960, **3**, 1–6.

RESTLE, F. *Psychology of judgment and choice*. New York: Wiley, 1961.

RUMELHART, D. L., & GREENO, J. G. Similarity between stimuli: an experimental test of the Luce and Restle choice models. *Journal of Mathematical Psychology*, 1971, **8**, 370–381.

SATTATH, S., & TVERSKY, A. Unite and conquer: A multiplicative inequality for choice probabilities. *Econometrica*, 1976, **44**, 79–89.

SHIPLEY, E. F. Detection and recognition: experiments and choice models. *Journal of Mathematical Psychology*, 1965, **2**, 277–311.

SHIPLEY, E. F., & LUCE, R. D. Discrimination among two- and three-element sets of weights. In R. C. Atkinson (Ed.), *Studies in Mathematical Psychology*. Stanford, Calif.: Stanford University Press, 1964. Pp. 218–232.

SKVORETZ, J., WINDELL, P., & FARARO, T. J. Luce's axiom and occupational prestige: test of a measurement model. *Journal of Mathematical Sociology*, 1974, **3**, 147–162.

STEELE, J. M. Limit properties of Luce's choice theory. *Journal of Mathematical Psychology*, 1974, **11**, 124–131.

STERNBERG, S. Stochastic learning theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 2. New York: Wiley, 1963. Pp. 1–120.

SUPPES, P., & ZINNES, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 1. New York: Wiley, 1963. Pp. 1–76.

THOMPSON, W. A., JR., & SINGH, J. The use of limit theorems in paired comparison model building. *Psychometrika*, 1967, **32**, 255–264.

THURSTONE, L. L. *The measurement of values*. Chicago: University of Chicago Press, 1959.

TOWNSEND, J. J. Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, 1971, **9**, 40–50.

TREISMAN, M. On the word frequency effect: Comments on papers by J. Catlin and L. H. Nakatani. *Psychological Review*, 1971, **78**, 420–425.

TVERSKY, A. Choice by elimination. *Journal of Mathematical Psychology*, 1972, **9**, 341–367. (a)

TVERSKY, A. Elimination by aspects: a theory of choice. *Psychological Review*, 1972, **79**, 281–299. (b)

TVERSKY, A., & RUSSO, J. E. Substitutability and similarity in binary choices. *Journal of Mathematical Psychology*, 1969, **6**, 1–12.

WAGENAAR, W. A. Application of Luce's choice axiom to form-discrimination. *Nederlands Tijdschrift voor de Psychologie*, 1968, **23**, 96–108.

YELLOTT, J. I., JR. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 1977, **15**, 109–144.

YELLOTT, J. I., JR., & CURNOW, P. F. Second choices in a visual span of apprehension task. *Perception & Psychophysics*, 1967, **2**, 307–311.

YOUNG, P. T. Studies of food preference, appetite, and dietary habit: VII. Palatability in relation to learning and performance. *Journal of Comparative and Physiological Psychology*, 1947, **40**, 37–72.

RECEIVED: September 8, 1976