

A RESPONSE TO R. A. WEBB'S COMMENTS

DEWEY J. MOORE¹

Indiana State University

Webb (1972) accepts as a plausible hypothesis that autistic children have (or had) high intellectual potential. He also agrees that normal human social development has its basis in instinctive processes and is preceded by considerable perceptual organization. He further agrees that such perceptual organization is unlikely to occur in utero. Having agreed with all the major tenets of the theory, Webb proceeds to argue its conclusions by citing truisms from normal development as his evidence. He conveniently ignores the fact that early infantile autism is not a normal state.

Webb anticipates that if the theory is valid there should be degrees of autism. He notes correctly that such variability is not observed. The IQ limit of 170+ was not chosen capriciously or arbitrarily. As hypothesized in the original paper, a projected intellectual level of such magnitude is necessary for the critical six months to pass in utero. Less potential permits adequate opportunity for imprinting and/or perceptual organization after birth. There are probably time limits for imprinting in humans as in lower animals. After expiration of these limits, the developmental opportunity is lost. Imprinting seems to be an all-or-nothing-at-all phenomenon. In autism, it is hypothesized to be a nothing-at-all phenomenon. Without such basic perceptual or-

ganization, the human socialization process fails to proceed normally. Hence, Webb is correct in his observation that early infantile autism does not exist in degrees, but he is correct for reasons other than those he states.

Finally, Webb rejects the theory's rejection of Rimland's (1964) early brain damage hypothesis on grounds that the brain is more susceptible to damage when it is developing rapidly. His restatement of this commonly accepted truism lends nothing to the discussion. Significantly he failed to comment on the fact that the negative effects of sensory deprivation are relatively greater for genetically brighter organisms.

The theory demands flexibility of thought, rigidity of logic, and a renunciation of many time-worn and erroneous beliefs regarding autism. Hopefully, experimenters and theoreticians will scrutinize past, present, and future knowledge regarding the status of autistics in the light of new hypotheses. One avenue of exploration might attempt to determine how biochemical differences that appear to exist in the reticular activating systems of autistics approximate those found in normals who have undergone sensory deprivation.

REFERENCES

- RIMLAND, B. *Infantile autism*. New York: Appleton-Century-Crofts, 1964.
WEBB, R. A. A comment on D. J. Moore and D. A. Skiel's "Toward a theory of early infantile autism." *Psychological Review*, 1972, 79, 278-279.

¹ Requests for reprints should be sent to Dewey J. Moore, Department of Special Education, Indiana State University, Terre Haute, Indiana 47809.

PSYCHOLOGICAL REVIEW

VOL. 79, No. 4

JULY 1972

ELIMINATION BY ASPECTS:

A THEORY OF CHOICE¹

AMOS TVERSKY²

Hebrew University of Jerusalem

Most probabilistic analyses of choice are based on the assumption of simple scalability which is an ordinal formulation of the principle of independence from irrelevant alternatives. This assumption, however, is shown to be inadequate on both theoretical and experimental grounds. To resolve this problem, a more general theory of choice based on a covert elimination process is developed. In this theory, each alternative is viewed as a set of aspects. At each stage in the process, an aspect is selected (with probability proportional to its weight), and all the alternatives that do not include the selected aspect are eliminated. The process continues until all alternatives but one are eliminated. It is shown (a) that this model is expressible purely in terms of the choice alternatives without any reference to specific aspects, (b) that it can be tested using observable choice probabilities, and (c) that it generalizes the choice models of R. D. Luce and of F. Restle. Empirical support from a study of psychophysical and preferential judgments is presented. The strategic implications of the present development are sketched, and the logic of elimination by aspects is discussed from both psychological and decision-theoretical viewpoints.

When faced with a choice among several alternatives, people often experience uncertainty and exhibit inconsistency. That is, people are often not sure which alternative they should select, nor do they always make the same choice under seemingly

identical conditions. In order to account for the observed inconsistency and the reported uncertainty, choice behavior has been viewed as a probabilistic process.

Probabilistic theories of preference differ with respect to the nature of the mechanism that is assumed to govern choice. Some theories (e.g., Thurstone, 1927, 1959) attribute a random element to the determination of subjective value, while others (e.g., Luce, 1959) attribute a random element to the decision rule. Most theoretical work on probabilistic preferences has been based on the notion of independence among alternatives. This notion, however, is incompatible with some observed patterns of preferences which exhibit systematic dependencies among alternatives.

This paper develops a probabilistic theory of choice, based on a covert elimination process, which accounts for observed

¹ The research was supported, in part, by National Science Foundation Grant GB-6782. Much of the work reported in this paper was accomplished while the author was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, California, during 1970-1971. I wish to thank the Center for the generous hospitality. I am grateful to David H. Krantz for many invaluable discussions throughout the years, to Maya Bar-Hillel for her assistance in both theoretical and experimental phases of the investigation, and to Edward N. Pugh for his help in the analysis of the data. I have also benefited from discussions with Clyde H. Coombs, Robyn M. Dawes, R. Duncan Luce, Jacob Marschak, J. E. Russo, and Paul Slovic.

² Requests for reprints should be sent to Amos Tversky, who is now at the Oregon Research Institute, P. O. Box 3196, Eugene, Oregon 97403.

dependencies among alternatives. The first section analyzes the independence assumption; the second section formulates a theory of choice and discusses its consequences; some experimental tests of the theory are reported in the third section; and its psychological implications are explored in the fourth and final section.

We begin by introducing some notation. Let $T = \{x, y, z, \dots\}$ be a finite set, interpreted as the total set of alternatives under consideration. We use A, B, C, \dots , to denote specific nonempty subsets of T , and A_i, B_j, C_k, \dots , to denote variables ranging over nonempty subsets of T . Thus, $\{A_i | A_i \supseteq B\}$ is the set of all subsets of T which includes B . The number of elements in A is denoted by a . Proper and non-proper set inclusion are denoted, respectively, by \supset and \supseteq . The empty set is denoted by ϕ . The probability of choosing an alternative x from an offered set $A \subseteq T$ is denoted $P(x, A)$. Naturally, we assume $P(x, A) \geq 0$, $\sum_{x \in A} P(x, A) = 1$ for any A , and $P(x, A) = 0$ for any $x \notin A$. For brevity, we write $P(x, y)$ for $P(x, \{x, y\})$, $P(x, y, z)$ for $P(x, \{x, y, z\})$, etc. A real-valued, nonnegative function in one argument is called a *scale*. Choice probability is typically estimated by relative frequency in repeated choices. It should be kept in mind, however, that other empirical interpretations of choice probability, such as confidence judgments (which are applicable to unique choice situations), might also be adopted.

Perhaps the most general formulation of the notion of independence from irrelevant alternatives is the assumption that the alternatives can be scaled so that each choice probability is expressible as a monotone function of the scale values of the respective alternatives. This assumption, called *simple scalability*, was first investigated by Krantz (1964, Appendix A). Formally, simple scalability holds if and only if there exists a scale u defined on the alternatives of T and functions F_n in n arguments, $2 \leq n \leq t$, such that for any $A = \{x, \dots, z\} \subseteq T$,

$$P(x, A) = F_n[u(x), \dots, u(z)], \quad [1]$$

where each F_n is strictly increasing in the first argument and strictly decreasing in the remaining $n - 1$ arguments provided $P(x, A) \neq 0, 1$. This assumption underlies most theoretical work in the field. The theory of Luce (1959), for example, is a special case of this assumption where

$$P(x, A) = F_a[u(x), \dots, u(z)] \\ = \frac{u(x)}{\sum_{y \in A} u(y)}, \quad [2]$$

Despite its generality, simple scalability (Equation 1) has strong testable consequences. In particular, it implies that for all $x, y \in A$,

$$P(x, y) \geq 1/2 \text{ iff } P(x, A) \geq P(y, A), \\ \text{provided } P(y, A) \neq 0. \quad [3]$$

Equation 3 asserts that the ordering of x and y , by choice probability, is independent of the offered set.³ Thus, if x is preferred to y in one context (e.g., $P(x, y) \geq 1/2$), then x is preferred to y in any context. Furthermore, if $P(x, y) = 1/2$ then $P(x, A) = P(y, A)$ for any A which contains both x and y . Thus, if an individual is indifferent between x and y , then he should choose them with equal probability from any set which contains them.

This assumption, however, is not valid in general, as suggested by several counterexamples and demonstrated in many experiments (see Becker, DeGroot, & Marschak, 1963b; Chipman, 1960; Coombs, 1958; Krantz, 1967; Tversky & Russo, 1969). To motivate the present development, let us examine the arguments against simple scalability starting with an example proposed by Debreu (1960).

Suppose you are offered a choice among the following three records: a suite by Debussy, denoted D , and two different recordings of the same Beethoven symphony, denoted B_1 and B_2 . Assume that the two Beethoven recordings are of equal quality,

and that you are undecided between adding a Debussy or a Beethoven to your record collection. Hence, $P(B_1; B_2) = P(D; B_1) = P(D; B_2) = 1/2$. It follows readily from Equation 3 that $P(D; B_1, B_2) = 1/3$. This conclusion, however, is unacceptable on intuitive grounds because the basic conflict between Debussy and Beethoven is not likely to be affected by the addition of another Beethoven recording. Instead, it is suggested that in choosing among the three records, B_1 and B_2 are treated as one alternative to be compared with D . Consequently, one would expect that $P(D; B_1, B_2)$ will be close to one-half, while $P(B_1; B_2, D) = P(B_2; B_1, D)$ will be close to one-fourth, contrary to simple scalability (Equation 1). Empirical support for Debreu's hypothesis was presented by Becker et al. (1963b) in a study of choice among gambles. Although Debreu's example was offered as a criticism of Luce's model (Equation 2), it applies to any model based on simple scalability.

Previous efforts to resolve this problem (e.g., Estes, 1960) attempted to redefine the alternatives so that B_1 and B_2 are no longer viewed as different alternatives. Although this idea has some appeal, it does not provide a satisfactory account of our problem. First, B_1 and B_2 are not only physically distinct, but they can also be perfectly discriminable. Hence, there is no independent basis for treating them as indistinguishable. Second, the process of redefining choice alternatives itself requires an adequate theoretical analysis. Finally, data show that the principle of independence from irrelevant alternatives is violated in a manner that cannot be readily accounted for by grouping choice alternatives. More specifically, it appears that the addition of an alternative to an offered set "hurts" alternatives that are similar to the added alternative more than those that are dissimilar to it. Such an effect (of which Debreu's example is a special case) requires a more drastic revision of the principles underlying our models of choice.

The following example provides another illustration of the inadequacy of simple

scalability. Suppose each of two travel agencies, denoted 1 and 2, offers tours of Europe (E) and of the Far East (F). Let $T = \{E_1, F_1, E_2, F_2\}$ where letters denote the destination of the tours, and the subscripts denote the respective agencies. Let us assume, for simplicity, that the decision maker is equally attracted by Europe and by the Far East, and that he has no reason to prefer one travel agency over the other. Consequently, all binary choice probabilities equal one-half, and the probability of choosing each tour from the total set equals one-fourth. It follows from Equation 3, in this case, that all trinary probabilities must equal one-third. However, an examination of the problem suggests that in fact none of the trinary probabilities equals one-third; instead, some of them equal one-half while the others equal one-fourth.

Consider, for example, the set $\{E_1, F_1, F_2\}$. Since the distinction between the agencies is treated as irrelevant, the problem reduces to the choice between a tour of Europe and a tour of the Far East. If the latter is chosen, then either one of the agencies can be selected. Consequently, $P(E_1; F_1, F_2) = 1/2$, and $P(F_1; F_2, E_1) = P(F_2; F_1, E_1) = 1/4$. An identical argument applies to all other triples. Besides violating simple scalability, this example demonstrates that the same set of binary (or quaternary) probabilities can give rise to different trinary probabilities and hence the latter cannot be determined by the former. Put differently, this example shows that the probabilities of choosing alternatives from a given set, A , cannot be computed, in general, from the probabilities of choosing these alternatives from the subsets and the supersets of A . This observation imposes a high lower bound on the complexity of any adequate theory of choice.

A minor modification of an example due to L. J. Savage (see Luce & Suppes, 1965, pp. 334-335), which is based on binary comparisons only, illustrates yet another difficulty encountered by simple scalability. Imagine an individual who has to choose between a trip to Paris and a trip to Rome. Suppose he is indifferent between the two

³ Simple scalability is, in fact, equivalent (see Tversky, 1972) to the following order independence assumption. For $x, y \in A - B$, and $z \in B$, $P(x, A) \geq P(y, A)$ iff $P(z, B \cup \{x\}) \leq P(z, B \cup \{y\})$ provided the terms on the two sides of either inequality are not both 0 or 1.

trips so that $P(\text{Paris}; \text{Rome}) = 1/2$. When the individual is offered a new alternative which consists of the trip to Paris plus a \$1 bonus, denoted Paris +, he will undoubtedly prefer it over the original trip to Paris with certainty so that $P(\text{Paris} +; \text{Paris}) = 1$. It follows from Equation 3, then, that $P(\text{Paris} +; \text{Rome}) = 1$, which is counterintuitive. For if our individual cannot decide between Paris and Rome, it is unlikely that a relatively small bonus would resolve the conflict completely and change the choice probability from 1/2 to 1. Rather, we expect $P(\text{Paris} +; \text{Rome})$ to be closer to 1/2 than to 1. Experimental data (e.g., Tversky & Russo, 1969) support this intuition. Choice probabilities, therefore, reflect not only the utilities of the alternatives in question, but also the difficulty of comparing them. Thus, an extreme choice probability (i.e., close to 0 or 1) can result from either a large discrepancy in value or from an easy comparison, as in the case of the added bonus. The comparability of the alternatives, however, cannot be captured by their scale values, and hence simple scalability must be rejected. The above examples demonstrate that the substitution of one alternative for another, which is equivalent to it in some contexts, does not necessarily preserve choice probability in any context. The substitution affects the comparability among the alternatives, which in turn influences choice probability.

An alternative approach to the development of probabilistic theories of choice treats the utility of each alternative as a random variable rather than a constant. Specifically, it is assumed that there exists a random vector $U = (U_1, \dots, U_n)$ on $T = \{x, \dots, z\}$ (i.e., for any $y \in T$, U_y is a random variable) such that

$$P(x, A) = P(U_x \geq U_y \text{ for all } y \in A). \quad [4]$$

Models of this type are called random utility models. The only random utility models which have been seriously investigated assume that the random variables are independent. However, an extension of the last example (see Luce & Raiffa, 1957, p. 375) is shown to violate

any independent random utility model. To demonstrate, consider the trips to Paris and Rome with and without the added bonus. The expected binary choice probabilities in this case are $P(\text{Paris} +; \text{Paris}) = 1$, $P(\text{Rome} +; \text{Rome}) = 1$ but $P(\text{Paris} +; \text{Rome}) < 1$ and $P(\text{Rome} +; \text{Paris}) < 1$.

Assuming an independent random utility model, the first two equations above imply that there is no overlap between the distributions representing Paris and Paris +, nor is there an overlap between the distributions representing Rome and Rome +. The last two inequalities above imply that there must be some overlap between the distributions representing Rome and Paris +, as well as between the distributions representing Paris and Rome +. It is easy to verify that these conclusions are mutually inconsistent, and hence the above choice probabilities are incompatible with any independent random utility model. The representation of choice alternatives by independent random variables, therefore, appears too restrictive in general since, like simple scalability, it is incompatible with some eminently reasonable patterns of preference. In discussing the difficulties encountered by probabilistic theories of choice, Luce and Suppes (1965) wrote:

It appears that such criticisms, although usually directed toward specific models, are really much more sweeping objections to all our current preference theories. They suggest that we cannot hope to be completely successful in dealing with preferences until we include some mathematical structure over the set of outcomes that are simply substitutable for one another and those that are special cases of others. Such functional and logical relations among the outcomes seem to have a sharp control over the preference probabilities, and they cannot long be ignored [p. 337].

THEORY

The present development describes choice as a covert sequential elimination process. Suppose that each alternative consists of a set of aspects of characteristics,⁴ and that

⁴The representation of choice alternatives as collections of measurable aspects was developed by Restle (1961) who formulated a binary choice model based on this representation. As will be shown later, the present theory reduces to Restle's in the

at every stage of the process, an aspect is selected (from those included in the available alternatives) with probability that is proportional to its weight. The selection of an aspect eliminates all the alternatives that do not include the selected aspect, and the process continues until a single alternative remains. If a selected aspect is included in all the available alternatives, no alternative is eliminated and a new aspect is selected. Consequently, aspects that are common to all the alternatives under consideration do not affect choice probabilities. Since the present theory describes choice as an elimination process governed by successive selection of aspects, it is called the elimination-by-aspects (EBA) model.

In contemplating the purchase of a new car, for example, the first aspect selected may be automatic transmission: this will eliminate all cars that do not have this feature. Given the remaining alternatives, another aspect, say a \$3000 price limit, is selected and all cars whose price exceeds this limit are excluded. The process continues until all cars but one are eliminated. This decision rule is closely related to the lexicographic model (see Coombs, 1964; Fishburn, 1968), where an ordering of the relevant attributes is specified a priori. One chooses, then, the alternative that is best relative to the first attribute; if some alternatives are equivalent with respect to the first attribute, one chooses from them the alternative that is best relative to the second attribute, and so on. The present model differs from the lexicographic model in that here no fixed prior ordering of aspects (or attributes) is assumed, and the choice process is inherently probabilistic.

More formally, consider a mapping that associates with each $x \in T$ a nonempty set $x' = \{\alpha, \beta, \dots\}$ of elements which are interpreted as the aspects of x . An alternative x is said to include an aspect α whenever $\alpha \in x'$. The aspects could represent values

two-alternative case. A related representation of choice alternatives was developed by Lancaster (1966) who assumed that economic goods possess, or give rise to, multiple characteristics (or aspects) in fixed proportion, and that these characteristics determine the consumer's choice. Lancaster's theory, however, is nonprobabilistic.

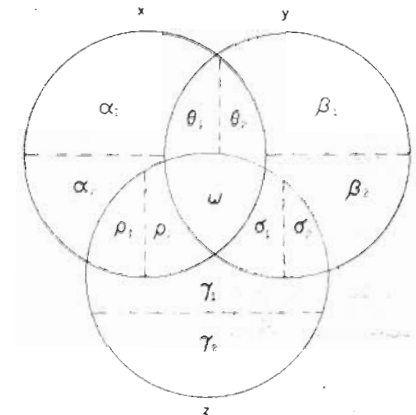


FIG. 1. A graphical representation of aspects in the three-alternative case.

along some fixed quantitative or qualitative dimensions (e.g., price, quality, comfort), or they could be arbitrary features of the alternatives that do not fit into any simple dimensional structure. The characterization of alternatives in terms of aspects is not necessarily unique. Furthermore, we generally do not know what aspects are considered by an individual in any particular choice problem. Nevertheless, as is demonstrated later, this knowledge is not required in order to apply the present model, and its descriptive validity can be determined independently of any particular characterization of the alternatives.

To clarify the formalization of the model, let us first examine a simple example. Consider a three-alternative set $T = \{x, y, z\}$, where the collections of aspects associated with the respective alternatives are

$$x' = \{\alpha_1, \alpha_2, \theta_1, \theta_2, \rho_1, \rho_2, \omega\},$$

$$y' = \{\beta_1, \beta_2, \theta_1, \theta_2, \sigma_1, \sigma_2, \omega\},$$

and

$$z' = \{\gamma_1, \gamma_2, \rho_1, \rho_2, \sigma_1, \sigma_2, \omega\}.$$

A graphical representation of the structure of the alternatives and their aspects is presented in Figure 1. It is readily seen that α_i , β_i , and γ_i ($i = 1, 2$) are, respectively, the unique aspects of x , y , and z ; that θ_i , σ_i , and ρ_i are, respectively, the

aspects shared by x and y , by y and z , and by x and z ; and that ω is shared by all three alternatives. Since the selection of ω does not eliminate any alternative, it can be discarded from further considerations. Let u be a scale which assigns to each aspect a positive number representing its utility or value, and let K be the sum of the scale values of all the aspects under consideration, that is, $K = \sum u(\alpha)$ where the summation ranges over all the aspects except ω . Using these notations we now compute $P(x, T)$.

Note first that x can be chosen directly from T if either α_1 or α_2 is selected in the first stage (in which case both y and z are eliminated). This occurs with probability

$$P(x; y) = \frac{u(\alpha_1) + u(\alpha_2) + u(\beta_1) + u(\beta_2)}{u(\alpha_1) + u(\alpha_2) + u(\beta_1) + u(\beta_2) + u(\beta_3) + u(\beta_4) + u(\beta_5) + u(\beta_6)}, \text{ etc.}$$

More generally, let T be any finite set of alternatives. For any $A \in T$ let $A' = \{\alpha | \alpha \in x' \text{ for some } x \in A\}$, and $A'' = \{\alpha | \alpha \in x' \text{ for all } x \in A\}$. Thus, A' is the set of aspects that belongs to at least one alternative in A , and A'' is the set of aspects that belongs to all the alternatives in A . In particular, T' is the set of all aspects under consideration, while T'' is the set of aspects shared by all the alternatives under study. Given any aspect $\alpha \in T'$, let A_α denote those alternatives of A which include α , that is, $A_\alpha = \{x | x \in A \text{ and } \alpha \in x\}$.

The elimination-by-aspects model asserts that there exists a positive scale u defined on the aspects (or more specifically on $T' - T''$) such that for all $x \in A \subseteq T$

$$P(x, A) = \frac{\sum_{\alpha \in A' - A''} u(\alpha) P(x, A_\alpha)}{\sum_{\beta \in A' - A''} u(\beta)} \quad [6]$$

provided the denominator does not vanish. Note that the summations in the numerator and the denominator of Equation 6 range, respectively, over all aspects of x and A except those that are shared by all elements of A . Hence, the denominator of Equation 6 vanishes only if all elements of A share the same aspects, in which case it is assumed that $P(x, A) = 1/u$.

$[u(\alpha_1) + u(\alpha_2)]/K$. Alternatively, x can be chosen via $\{x, y\}$ if either θ_1 or θ_2 is selected in the first stage (in which case z is eliminated), and then x is chosen over y . This occurs with probability $[u(\theta_1) + u(\theta_2)] \times P(x; y)/K$. Finally, x can be chosen via $\{x, z\}$ if either ρ_1 or ρ_2 is selected in the first stage (in which case y is eliminated), and then x is chosen over z . This occurs with probability $[u(\rho_1) + u(\rho_2)]P(x; z)/K$. Since the above paths leading to the choice of x from T are all disjoint,

$$P(x, T) = (1/K)[u(\alpha_1) + u(\alpha_2) + [u(\theta_1) + u(\theta_2)]P(x; y) + [u(\rho_1) + u(\rho_2)]P(x; z)] \quad [5]$$

where

Equation 6 is a recursive formula. It expresses the probability of choosing x from A as a weighted sum of the probabilities of choosing x from the various subsets of A (i.e., A_α for $\alpha \in x'$), where the weights (i.e., $u(\alpha)$; $\sum u(\beta)$) correspond to the probabilities of selecting the respective aspects of x .

Consider a special case of the elimination-by-aspects model where all pairs of alternatives share the same aspects, that is, $x' \cap y' = z' \cap w'$ for all $x, y, z, w \in T$. Since aspects that are common to all the alternatives of T do not affect the choice process, the alternatives can be treated as (pairwise) disjoint, that is, $x' \cap y' = \emptyset$ for all $x, y \in T$. In this case, Equation 6 reduces to

$$P(x, A) = \frac{\sum_{\alpha \in x'} u(\alpha)}{\sum_{\beta \in A'} u(\beta)}$$

since $\alpha \in x'$ implies $A_\alpha = \{x\}$, and $P(x, \{x\}) = 1$. Letting

$$u(x) = \sum_{\alpha \in x'} u(\alpha)$$

yields

$$P(x, A) = \frac{u(x)}{\sum_{y \in A} u(y)}$$

Hence, in the present theory, Luce's model

(Equation 2) holds whenever the alternatives can be regarded as composed of disjoint aspects.

Next, examine another special case of the model where only binary choice probabilities are considered. In this case, we obtain

$$P(x; y) = \frac{\sum_{\alpha \in x' - y'} u(\alpha)}{\sum_{\alpha \in x' - y'} u(\alpha) + \sum_{\alpha \in y' - x'} u(\alpha)} = \frac{u(x' - y')}{u(x' - y') + u(y' - x')}, \quad [7]$$

where $x' - y' = \{\alpha | \alpha \in x' \text{ and } \alpha \notin y'\}$ is the set of aspects that belongs to x but not to y ; $y' - x' = \{\beta | \beta \in y' \text{ and } \beta \notin x'\}$ is the set of aspects that belongs to y but not to x ; and $u(x' - y') = \sum_{\alpha \in x' - y'} u(\alpha)$. Equation 7 coincides with Restle's (1961) model. The EBA model, therefore, generalizes the choice models of Luce and of Restle.

The elimination-by-aspects model has been formulated above in terms of a scale u defined over the set of relevant aspects. It appears that the application of the model presupposes prior characterization of the alternatives in terms of their aspects. However, it turns out that this is not necessary because the EBA model can be formulated purely in terms of the alterna-

tives, or more specifically, in terms of the subsets of T .

To illustrate the basic idea, consider the example presented in Figure 1. There we assume that α_i, β_i, \dots ($i = 1, 2$) are all distinct aspects. According to the elimination-by-aspects model, however, there is no need to distinguish between aspects that lead to the same outcome. For example, the selection of either α_1 or α_2 eliminates both y and z ; the selection of either θ_1 or θ_2 eliminates z ; and the selection of either ρ_1 or ρ_2 eliminates y . From the standpoint of the elimination-by-aspects model, therefore, there is no need to differentiate between α_1 and α_2 , between θ_1 and θ_2 , or between ρ_1 and ρ_2 . Thus we can group all the aspects that belong to x alone, all the aspects that belong to x and y but not to z , etc. Let $\{x\}$ denote the aspects that belong to x alone (i.e., α_1 and α_2), $\{x, y\}$ the aspects that belong only to x and y (i.e., θ_1 and θ_2), $\{x, z\}$ the aspects that belong only to x and z (i.e., ρ_1 and ρ_2), etc.⁵ The representation of the grouped aspects in the three-alternative case is displayed in Figure 2.

The scale value of a collection of aspects is defined as the sum of the scale value of its members, that is, $U(\{x\}) = u(\alpha_1) + u(\alpha_2)$, $U(\{x, y\}) = u(\theta_1) + u(\theta_2)$, etc. For simplicity of notation we write $U(\{x\})$ for $U(\{\alpha_1\})$, $U(\{x, y\})$ for $U(\{\theta_1\})$, etc. Thus, Equation 5 is expressible as

$$P(x, T) = \frac{U(\{x\}) + U(\{x, y\})P(x; y) + U(\{x, z\})P(x; z)}{U(\{x\}) + U(\{y\}) + U(\{z\}) + U(\{x, y\}) + U(\{x, z\}) + U(\{y, z\})} \quad [8]$$

where

$$P(x; y) = \frac{U(\{x\}) + U(\{x, z\})}{U(\{x\}) + U(\{y\}) + U(\{x, z\}) + U(\{y, z\})},$$

etc.

The essential difference between Equations 5 and 8 lies in the domain of the scales: in Equation 5, u is defined over individual aspects, whereas in Equation 8 U is defined over collections of aspects which are associated, respectively, with the subsets of

⁵ In this paper, the superfix i is used exclusively to denote collections of aspects. It should not be confused with a common use of this symbol to denote set complement.

T . The method by which Equation 5 is translated into Equation 8 can be applied in general.

Each proper subset A of T is associated with the set \bar{A} of all aspects that are included in all the alternatives of A and are not included in any of the alternatives that do not belong to A . That is, $\bar{A} = \{\alpha \in T' | \alpha \in x' \text{ for all } x \in A \text{ and } \alpha \notin y' \text{ for any } y \notin A\}$. The scale U is defined by $U(\bar{A}) = \sum_{\alpha \in \bar{A}} u(\alpha)$. It is shown in the appendix that the elimination-by-aspects model, defined in Equation 6, holds if and only if there exists a scale U defined on $\{\bar{A} | A \subset T\}$ such

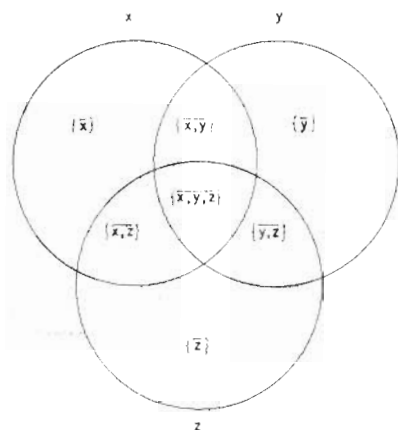


FIG. 2. A graphical representation of the grouped aspects in the three-alternative case.

that for all $x \in A \subseteq T$

$$P(x, A) = \frac{\sum_{B_i \in \mathcal{A}} U(\tilde{B}_i) P(x, A \cap B_i)}{\sum_{A_i \in \mathcal{A}} U(\tilde{A}_i)} \quad [9]$$

where $\mathcal{A} = \{A_i : A_i \cap A \neq \emptyset, \emptyset\}$, provided the denominator does not vanish. (According to the present theory, the denominator can vanish only if $P(x, A) = 1/a$.) The significance of this result lies in showing how the elimination-by-aspects model can be formulated in terms of the subsets of T without reference to specific aspects. Note that for $A \subset T$, $U(\tilde{A})$ is not a measure of the value of the alternatives of A ; rather it is a measure of all the evaluative aspects that are shared by all the alternatives of A and by them only. Thus, $U(\tilde{A})$ can be viewed as a measure of the unique advantage of the alternatives of A . The reader is invited to verify that in the three-alternative case, Equation 9 reduces to Equation 8.

Before discussing the consequences of the EBA model, let us examine how it resolves the counterexamples described in the previous section. First, consider Debreu's record selection problem where $T = \{D, B_1, B_2\}$. Naturally, the two Beethoven recordings have much more in common with each other than either of

them has with the Debussy record. Assume, for simplicity, that any aspect shared by D and one of the B records is also shared by the other B record, hence D can be treated as (aspectwise) disjoint of both B_1 and B_2 . Suppose $U(\tilde{B}_1) = U(\tilde{B}_2) = a$, $U(\tilde{B}_1, \tilde{B}_2) = b$, and $U(\tilde{D}) = a + b$. A graphical illustration of this representation is shown in Figure 3.

It follows readily, under these assumptions, that all the binary choice probabilities are equal, since

$$P(B_1; B_2) = \frac{a}{2a} = \frac{1}{2} = \frac{a+b}{2(a+b)} \\ = P(D; B_1) = P(D; B_2),$$

yet the trinary choice probabilities are unequal, since

$$P(D; B_1, B_2) = \frac{a+b}{3a+2b} > \frac{a+b(a/2a)}{3a+2b} \\ = P(B_1; B_2, D) = P(B_2; B_1, D).$$

In fact, as a (or a/b) approaches 0, the left-hand side approaches $1/2$ while the right-hand side approaches $1/4$. Hence, according to the elimination-by-aspects model, all three records can be pairwise equivalent, and yet the probability of choosing D from the entire set can be as high as $1/2$ whenever B_1 and B_2 include the same aspects.

Second, consider Savage's problem of choosing between trips, and let $T = \{P, R, P+, R+\}$, where P and R denote, respectively, trips to Paris and Rome, while $+$ denotes a small monetary bonus. Here it is natural to suppose that $P+$ includes Paris (in the sense that all aspects of the

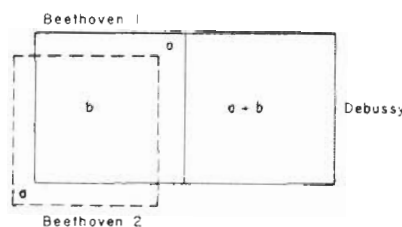


FIG. 3. A graphical illustration of the analysis of the record selection problem.

latter trip are included in the former). On the other hand, $P+$ does not include Rome because each of these trips has some aspects that are not shared by the other. Similarly, $R+$ includes Rome but not Paris. The relations among the four alternatives are illustrated in Figure 4.

Letting $U(P+) = U(R+) = a$, and $U(P, P+) = U(R, R+) = b$, yields

$$P(P; R) = \frac{b}{2b} = \frac{1}{2} = \frac{a+b}{2(a+b)} \\ = P(P+; R+),$$

$$P(P+; P) = P(R+; R) = \frac{a}{a} = 1, \quad \text{and}$$

$$P(P+; R) = P(R+; P) = \frac{a+b}{a+2b},$$

which can take any value between $1/2$ and 1 , depending on the relative weight of the bonus. Thus, the above pattern of binary choice probabilities, which violates simple scalability (Equation 1) and any independent random utility model (Equation 4), arises naturally in the present model. Essentially the same solution to this problem (which involves only binary probabilities) has been proposed by Restle (1961).

The reader is invited to show how the elimination-by-aspects model can accommodate the example described earlier of choice among tours of Europe or the Far East with each of two travel agencies.

Consequences

In the following discussion we assume that the elimination-by-aspects model is valid, and list some of its testable consequences. The derivations of these properties are presented in Tversky (1972).

Regularity: For all $x \in A \subseteq B$,

$$P(x, A) \geq P(x, B). \quad [10]$$

Regularity asserts that the probability of choosing an alternative from a given set cannot be increased by enlarging the offered set. This is probably the weakest form of noninteraction among alternatives. Although regularity seems innocuous, it is

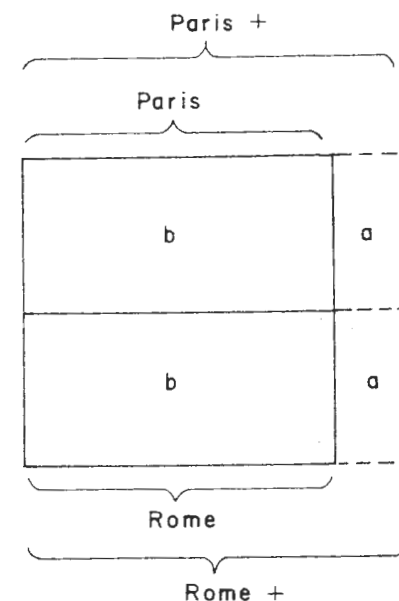


FIG. 4. A graphical illustration of the analysis of the choice between trips.

worth noting that the replacement of \geq by $>$ in Equation 10 violates the expected preference pattern in the record selection problem.

The following consequence of the elimination-by-aspects model involves binary probabilities only. Since it generalizes the algebraic notion of transitivity, it is called *moderate stochastic transitivity*.

Moderate stochastic transitivity:

$$P(x; y) \geq 1/2 \text{ and } P(y; z) \geq 1/2 \text{ imply} \\ P(x; z) \geq \min[P(x; y), P(y; z)]. \quad [11]$$

If we replace \min by \max in the conclusion of Equation 11, we obtain a stronger condition called *strong stochastic transitivity*. This latter property (which is not a consequence of the present model) is essentially equivalent to simple scalability in the binary case. If we replace the conclusion of Equation 11 by $P(x; z) \geq 1/2$, we obtain a weaker condition called *weak stochastic transitivity*, which is a con-

sequence of the existence of an ordinal utility scale satisfying $u(x) \geq u(y)$ iff $P(x; y) \geq 1/2$.

The next consequence of the EBA model has not been investigated previously to the best of my knowledge. It relates binary and trinary choice probabilities by a property called the *multiplicative inequality*.

Multiplicative inequality:

$$P(x; y, z) \geq P(x; y)P(x; z). \quad [12]$$

The multiplicative inequality asserts that the probability of choosing x from $\{x, y, z\}$ is at least as large as the probability of choosing x from both $\{x, y\}$ and $\{x, z\}$ in two independent choices. It is conjectured that the elimination-by-aspects model implies a much stronger form of the multiplicative inequality, namely, $P(x, A \cup B) \geq P(x, A)P(x, B)$ for all $A, B \subseteq T$.

Equations 10 and 12 can be combined to yield

$$\min[P(x; y), P(x; z)] \geq P(x; y, z) \geq P(x; y)P(x; z). \quad [13]$$

Thus, trinary choice probabilities are bounded from above by regularity, and from below by the multiplicative inequality. A geometric representation of Equation 13

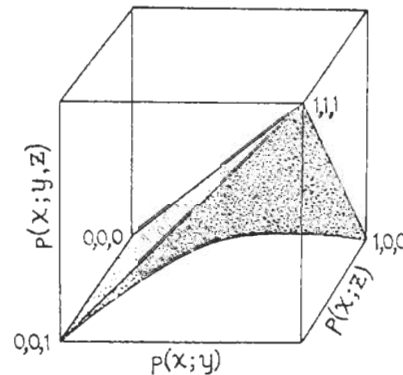


FIG. 5. A geometric representation of the admissible values (shaded region) of the trinary probability $P(x; y, z)$ given the binary probabilities $P(x; y)$ and $P(x; z)$, under Equation 13.

which displays the admissible range of $P(x; y, z)$ given the values of $P(x; y)$ and $P(x; z)$ is given in Figure 5. It shows that the trinary probability must lie between the lower and upper surfaces generated, respectively, by the multiplicative inequality (Equation 12) and regularity (Equation 10).

The significance of the above consequences stems from the fact that they provide measurement-free tests of the elimination-by-aspects model, that is, tests which do not require estimation of parameters.

For a given set of alternatives T , the elimination-by-aspects model has $2^t - 3$ free parameters, or U values (the number of proper nonempty subsets of T minus an arbitrary unit of measurement), while the number of independent data points of the form $P(x, A)$, $x \in A \subseteq T$, is

$$\sum_{n=2}^t (n-1) \binom{t}{n} = (t-2)2^{t-1} + 1.$$

Hence, there are always at least as many data points as parameters in the present model; the former exceeds the latter whenever $t > 3$. In general, therefore, the scale values are uniquely determined by the choice probabilities except in some particular situations, for example, when $P(x, A) = 1/a$ for all $x \in A \subseteq T$.

Even in the case where $t = 3$, in which the number of parameters (five) equals the number of data points, the choice probabilities are severely constrained. The volume of the subspace generated by the present model is less than 1/2% of the volume of the entire parameter space which is a five-dimensional unit hypercube. The probability that a point sampled at random, from a uniform distribution over the parameter space, satisfies the present model, therefore, is less than .005 in this case.

Additional consequences and further developments of the elimination-by-aspects model are presented in Tversky (1972). They include a generalization of the present model, an extension to ranking, and a proof that the EBA model is a random utility model, though not an independent one.



FIG. 6. Typical stimulus slides from each of the three tasks.

TESTS

In contrast to the many theoretical studies of probabilistic models of preference (see, e.g., Becker et al., 1963a; Luce & Suppes, 1965; Marschak, 1960; Morrison, 1963), there have been relatively few empirical studies in which these models were tested. Moreover, much of the available data are limited to binary choices, and most studies report and analyze only group data (see, e.g., Rummelhart & Greeno, 1971). Unfortunately, group data usually do not permit adequate testing of theories of individual choice behavior because, in general, the compatibility of such data with the theory is neither a necessary nor a sufficient condition for its validity. (For an instructive illustration of this point, see Luce, 1959, p. 8.) The scarcity of appropriate data in an area of considerable theoretical interest is undoubtedly due to the difficulties involved in obtaining adequate estimates of choice probabilities for an individual subject, particularly outside the domain of psychophysics.

Two consequences of the present model were tested in previous studies. In an experiment involving choice among gambles, Becker et al. (1963b) showed that although simple scalability (Equation 1) is systematically violated, the regularity condition (Equation 10) is generally satisfied. Similarly, although strong stochastic transitivity was violated in several studies (e.g., Coombs, 1958; Krantz, 1967; Tversky & Russo, 1969), moderate stochastic transitivity was usually supported. (For some specified conditions under which moderate stochastic transitivity, as well as weak stochastic transitivity, is violated, see Tversky, 1969.) The fact that simple

scalability and strong stochastic transitivity are often violated while regularity and moderate stochastic transitivity are typically satisfied provides some support, albeit nonspecific, for the present theory. The following experimental work was designed to obtain a more direct test of the EBA model.

Method

To test the model, three different tasks were selected. The stimuli in Task A were random dot patterns, in a square frame, varying in size (of square) and density (of dots). Subjects were presented with pairs and triples of frames and instructed to choose, in each case, the frame which contained the largest number of dots. The stimuli in Task B were profiles of college applicants with different intelligence (I) and motivation (M) scores. The scores were expressed in percentiles (relative to the population of college applicants), and displayed as bar graphs. Subjects were presented with pairs and triples of such profiles and asked to select, in each case, the applicant they considered the most promising. The stimuli in Task C were two-outcome gambles of the form (p, x) , in which one wins $\$x$ with probability p and nothing otherwise. Each gamble was displayed as a pie diagram, where the probabilities of winning and not winning were represented, respectively, by the black and white sectors of the pie. Subjects were presented with pairs and triples of gambles and were asked to choose the gamble they would prefer to play. (At the end of the study, each subject actually played for money five of the gambles chosen by him in the course of the study. The gambles were played by spinning an arrow on a wheel of fortune and the subjects won the indicated amount if the arrow landed on the black sector of the wheel.) Examples of the three types of stimuli are shown in Figure 6.

The same eight subjects participated in all three tasks. They were students in a Jerusalem high school, ages 16-18. Subjects were run in a single group. The stimuli were projected on slides and each subject indicated his choices by checking an appropriate box on his response sheet. The study consisted of 12 one-hour sessions, three times a week,

for four weeks. The first two sessions were practice sessions in which the problems and the procedure were introduced and the subjects familiarized themselves with the stimuli of the task.

Each experimental session included all three tasks, and the ordering of the tasks was randomized across sessions. Within each task, subjects were presented with various pairs and triples formed from a basic set of $4 \times 4 = 16$ two-dimensional stimuli. One set of three stimuli of each type was isolated and replicated more than other sets. The entire triple was replicated 30 times (three per session) while each of the pairs within this triple was replicated 20 times (two per session). The following discussion is concerned with the analysis of these triples. Each triple was constructed so that no alternative dominates another one with respect to both dimensions, and so that two of the elements, called x and y , are very similar to each other, while the third element, z , is relatively dissimilar to each of them.⁶

The subjects were paid a flat fee for the completion of all the sessions. In addition, each subject received a bonus proportional to the number of correct numerosity judgments made by him, and was allowed to play, for money, five gambles selected randomly from those chosen by him during the study.

Results

The analysis of the results begins by testing the constant-ratio rule which is essentially equivalent to Luce's (1959) model. According to this rule,

$$\frac{P(x;y)}{P(y;x)} = \frac{P(x,A)}{P(y,A)} \quad x, y \in A. \quad [14]$$

provided the denominators do not vanish. The constant-ratio rule is a strong version of the principle of independence from irrelevant alternatives. It requires that the ratio of $P(x,A)$ and $P(y,A)$ (not merely their order as required by simple scalability) be independent of the offered set A .

⁶The following stimuli were employed in the study. Task A: $x = (13 \times 13, 4/5)$, $y = (14 \times 14, 3/4)$, and $z = (28 \times 28, 1/5)$ where the first component of each stimulus is the size of the underlying matrix used to generate the pattern, and the second component is the proportion of cells of the matrix that contain dots. Task B: $x = (78, 25)$, $y = (75, 35)$, and $z = (60, 90)$ where the first and second components of each pair denote, respectively, intelligence and motivation scores of the applicants. Task C: $x = (1/5, 4.00)$, $y = (1/3, 3.50)$, and $z = (2/3, 1.00)$, where the first and second component of each pair are, respectively, the probability of winning and the amount to be won in each of the gambles in Israeli pounds.

Let $T = \{x, y, z\}$, and define

$$P_y(x; z) = \frac{P(x, T)}{P(x, T) + P(z, T)},$$

$$P_z(y; z) = \frac{P(y, T)}{P(y, T) + P(z, T)}.$$

Hence, by the constant-ratio rule,

$$P(x; z) = P_y(x; z)$$

and

$$P(y; z) = P_z(y; z). \quad [15]$$

Put differently, the binary probability $P(x; z)$ should equal $P_y(x; z)$, computed from the trinary probabilities, since under Equation 14 the presence of y is "irrelevant" to the choice between x and z .

In the present study, the alternatives were designed so that x and y are much more similar to each other than either of them is to z . Hence, the similarity hypothesis that is incorporated into the elimination-by-aspects model predicts that the addition of alternative y to the set $\{x, z\}$ will reduce $P(x, T)$ proportionally more than $P(z, T)$. That is, the similar alternative, x , will lose relatively more than the dissimilar alternative, z , by the addition of y . Likewise, y is expected to lose relatively more than z by the introduction of x . Contrary to the constant-ratio rule, therefore, the similarity hypothesis implies

$$P(x; z) > P_y(x; z) \quad [16]$$

and

$$P(y; z) > P_z(y; z).$$

To test the constant-ratio rule, the observed (binary) relative frequencies $\hat{P}(x, z)$ and $\hat{P}(y, z)$ were compared, respectively, with $\hat{P}_y(x; z)$ and $\hat{P}_z(y; z)$ computed from the trinary relative frequencies, separately for each one of the subjects. The observed and the computed values for all subjects are shown in Table 1 for each of the three tasks.

It seems that the constant-ratio model (Equation 14) holds in the psychophysical task (A), and that it fails in the two preference tasks (B and C) in the manner predicted by the similarity hypothesis (Equation 16). Out of 16 individual comparisons in each task (two per subject), Equation 16 was satisfied in 13 and 15 cases, respec-

TABLE 1
OBSERVED AND PREDICTED PROPORTIONS (UNDER THE CONSTANT-RATIO MODEL) FOR EACH TASK

Subject	Task A (dots)				Task B (applicants)				Task C (gambles)			
	$\hat{P}(x; z)$	$\hat{P}_y(x; z)$	$\hat{P}(y; z)$	$\hat{P}_z(y; z)$	$\hat{P}(x; z)$	$\hat{P}_y(x; z)$	$\hat{P}(y; z)$	$\hat{P}_z(y; z)$	$\hat{P}(x; z)$	$\hat{P}_y(x; z)$	$\hat{P}(y; z)$	$\hat{P}_z(y; z)$
1	.50	.43	.45	.43	.65	.44	.30	.26	.35	.12	.50	.46
2	.60	.27	.35	.33	.55	.37	.75	.58	.60	.53	.70	.68
3	.25	.38	.40	.41	.55	.38	.60	.41	.25	.26	.50	.29
4	.70	.75	.30	.67	.40	.46	.40	.32	.60	.43	.70	.35
5	.65	.52	.35	.39	.65	.45	.55	.40	.20	.16	.50	.41
6	.40	.39	.45	.52	.35	.20	.40	.38	.65	.54	.60	.44
7	.85	.26	.45	.44	.75	.77	.35	.40	.55	.42	.65	.50
8	.15	.14	.45	.57	.55	.52	.40	.29	.55	.35	.70	.43
Overall proportion	.425	.405	.400	.466	.556	.463	.469	.388	.469	.354	.606	.466
p	ns		ns		< .05		< .10		< .01		< .01	

tively, in Tasks B and C ($p < .05$ in each case⁷), and only in 7 cases in Task A. Essentially the same result was found in additional analyses.

The relatively small number of observations does not permit an adequate test of individual comparisons. Hence, the observed and the computed choice frequencies were pooled over subjects. The results of a chi-square test of Equation 15 against Equation 16, based on these data, are shown in the last row of Table 1 for each comparison in each of the tasks. The same pattern emerges from the analysis of the pooled data: the observed proportions are significantly higher than the computed ones in Tasks B and C, but not in Task A.

Since the constant-ratio model is not acceptable, in general, the simplest version of the elimination-by-aspects model, which is compatible with the similarity hypothesis, was selected next. Recall that the test stimuli were designed so that x and y are very similar to each other while z is relatively dissimilar to either of them (see Footnote 6). Thus, we assume that neither x nor y share with z any aspect that they do not share with each other. Consequently, aside from the aspects shared by all three stimuli, z can be regarded as (aspectwise) disjoint from both x and y . That is, we assume that, to a reason-

able degree of approximation, $U(x, z) = U(y, z) = 0$. This assumption reduces the number of free parameters (from five to three) at the cost of some loss in generality.

Let $U(x) = a$, $U(y) = b$, $U(z) = c$, and $U(x, y) = d$ (see Figure 7). Under this special case of the model, there exist non-negative a , b , c , and d such that

$$P(x; y) = \frac{a}{a+b}, \quad P(y; z) = \frac{b+d}{b+d+c},$$

$$P(x; z) = \frac{a+d}{a+d+c}, \quad [17]$$

$$P(x; y, z) = \frac{a+d}{a+b+c+d},$$

and

$$P(z; x, y) = \frac{c}{a+b+c+d}.$$

For three alternatives, there are five independent data points (three binary and two trinary). In the absence of any restrictions on the parameters, the likelihood function of the data is maximized by using the observed relative frequencies as estimates of the parameters, in which case the dimensionality of the parameter space, denoted $d(\Omega)$, equals five. In the above version (Equation 17) of the elimination-by-aspects model, we can set c , say, arbitrarily, whence the observed proportions are all expressible in terms of three param-

⁷This significance level should be interpreted with caution because of the potential dependency between the observations of each subject.

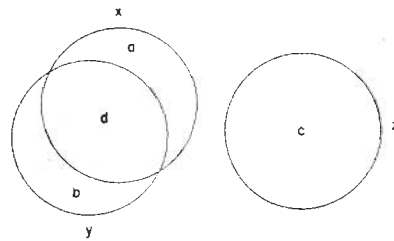


FIG. 7. A graphical illustration of the tested version of the EBA model (Equation 17).

ters (a , b , and d), and the dimensionality of the restricted parameter space, denoted $d(\omega)$, equals three. Let λ be the likelihood ratio $L(\omega)/L(\Omega)$, where L denotes the maximum value of the likelihood function under the respective model. If Equation 17 holds, then the statistic $-2\ln\lambda$ has an approximate chi-square distribution with $d(\Omega) - d(\omega) = 2$ degrees of freedom.

Chandler's (1969) STEPIT program was employed to obtain maximum likelihood estimates of the parameters under Equation 17 with $c = 1$. The values of the test statistics are reported in Table 2, along with the estimates of d , for each subject in all tasks.

Table 2 exhibits a very good correspondence between the observed proportions and the tested version (Equation 17) of the EBA model: only 2 out of 24 tests permit rejecting the model at the conservative .1 level. It should perhaps be noted that

a correspondence between observed choice probabilities and the elimination-by-aspects model does not necessarily imply that the subjects are actually following a strategy of elimination by aspects. They might, in fact, employ a different strategy that is well approximated by the elimination-by-aspects model. The study of the actual strategies employed by subjects in choice experiments may perhaps be advanced by investigating choice probabilities in conjunction with other data such as reaction time, eye movements, or verbal protocols.

The relation between the predictions of the constant-ratio model (Equation 15) and the similarity hypothesis (Equation 16) can be further investigated using the obtained estimates of the parameter d , reported in Table 2. It is easy to verify that the constant-ratio model is compatible with Equation 17 if and only if $d = 0$, while the similarity hypothesis implies $d > 0$. Hence, if the former holds, the estimates of d should be close to 0, whereas if the latter holds, the estimates should be substantially positive. (The magnitude of d should be interpreted in the light of the facts that all parameters are nonnegative and $c = 1$, see Equation 17 and Figure 7.) Inspection of Table 2 reveals that the majority of the d estimates in Task A are zero, while the majority of the d estimates in Tasks B and C are substantially positive. This agrees with the results of previous analyses (summarized in Table 1) according to which the constant-ratio model is satisfied in Task A, but not in Tasks B and C.

Taken together, the experimental findings suggest the hypothesis that the constant-ratio model is valid for choice among *unitary* alternatives (e.g., dots, colors, sounds) that are usually evaluated as wholes, but not for *composite* alternatives (e.g., gambles, applicants) that tend to be evaluated in terms of their attributes or components. This hypothesis is closely related to a suggestion made by Luce (1959):

If we call a decision that is not subdivided into simpler decisions an elementary choice, then possibly we can hope to find Axiom 1 [Luce's choice axiom] directly confirmed for elementary choices but probably not for more complex ones [p. 133].

Research on multidimensional scaling based on similarity, or proximity, data (e.g., Shepard, 1964a; Torgerson, 1965) has also shown that judgments of unitary and composite stimuli (sometimes referred to as analyzable and unanalyzable) are governed by different rules. Much additional research, however, is required in order to assess the validity and the generality of the proposed hypothesis.

Finally, the distinction between unitary and composite stimuli is logically independent of whether the inconsistency reflected in choice probabilities is attributable to imperfect discrimination or to a conflict among incompatible criteria. (For a discussion of this last distinction, see Block & Marschak, 1960.) Although choice experiments in psychophysics typically involve imperfect discrimination with unitary stimuli while preference experiments are usually concerned with conflict among composite alternatives, the other two combinations also exist.

DISCUSSION

Strategic Implications

A major feature of the elimination-by-aspects model is that the probability of selecting an alternative depends not only on its overall value, but also on its relations to the other available alternatives. This gives rise to study of strategic factors in the design and the presentation of choice alternatives. Specifically, the present model provides a method for investigating questions concerning optimal design or location of alternatives in order to maximize (or minimize) choice probability under specified constraints. The following examples are intended to illustrate the scope and the nature of such a study.

First, consider a problem of binary comparisons. Suppose y and z are given, and we search for x such that $P(x; y)$ is maximized under the constraints that z has no aspects in common with any other alternative, and that $P(y; z)$ and $P(x; z)$ are fixed. By the former constraint, z can be viewed as a standard of comparison. Hence, the latter constraint can be interpreted as meaning that the overall values of y and

of x (evaluated relative to z) are held fixed. Thus, only the position of x relative to y can be varied to maximize $P(x; y)$. Under these conditions, the present model implies that if $P(x; z) > P(y; z)$, x' should include as much of y' as possible. If, on the other hand, $P(x; z) < P(y; z)$, x' should include as little of y' as possible. The degree of overlap between x' and y' can be regarded as an index of the difficulty of comparing them. If x' includes y' , the comparison is trivial, and $P(x; y)$ is maximal. If x' and y' are disjoint, the comparison is much more difficult, and $P(x; y)$ is less extreme.

In the light of this interpretation, the above result asserts that it is in the best interest of the favored alternative to make the comparison as easy as possible, while it is in the best interest of the nonfavored alternative to make the comparison as difficult as possible. This certainly makes sense: any increase in the difficulty of comparing the alternatives adds "error" to the judgment process and makes $P(x; y)$ closer to 1/2. According to this logic, drastically different policies are prescribed depending on whether x is the favored or the nonfavored alternative. Advertising campaigns based on slogans such as "All aspirins are the same—why pay more?" and "This car is completely different from any other car in its class," illustrate, respectively, the policies recommended to the favored and the nonfavored alternatives. Note that these policies could be employed in the design of products as well as in their advertisements.

Second, let $T = \{x, y, \dots, z\}$ and suppose that all pairwise choice probabilities are fixed and that we wish to select a set $A \subseteq T$ which includes both x and y so that the ratio $P(x; A)/P(y; A)$ is maximized. According to the elimination-by-aspects model, the above ratio is maximized when A consists of alternatives (which are not dominated by y) that "cover" as much of y as possible without "covering" much of x . If x and y are products in some market A , for example, then the present model predicts that the relative advantage of x over y is maximized when the other available products are as similar to y and dissimilar to x as possible. The example of choice

TABLE 2
VALUES OF THE TEST STATISTIC AND THE ESTIMATED VALUES OF d FOR EACH SUBJECT IN EACH OF THE TASKS

Subject	Task A (dots)		Task B (applicants)		Task C (gambles)	
	χ^2	d	χ^2	d	χ^2	d
1	.133	.29	2.179	.14	.040	.46
2	3.025	.89	1.634	.92	.001	.58
3	.849	0	.159	1.18	2.022	.14
4	5.551*	0	6.864*	.51	1.053	1.56
5	.951	0	.428	1.23	.887	0
6	.401	0	.405	.42	.157	1.18
7	3.740	0	.083	0	.304	1.00
8	4.112	0	.038	.37	1.241	1.44

Note.— $df = 2$.
* $p = .1$.

among records discussed in the introduction and the similarity effect demonstrated in Table 1 illustrate the point. Note that this maximization problem cannot be investigated in Luce's model (Equation 2), for example, since by the constant-ratio rule $P(x:A)/P(y:A) = P(x;y)/P(y;x)$, $x, y \in A$, and hence is independent of A . According to the EBA model, in contrast, the above ratio can, in principle, be arbitrarily large, provided $P(x;y) \neq 0$.

Thus, if the present theory is valid, one can take advantage of the so-called "irrelevant alternatives" to influence choice probabilities. This result is based on the idea that the introduction of an additional alternative "hurts" similar alternatives more than dissimilar ones. This is a familiar notion in the context of group choice. The present development suggests that it is an important determinant of individual choice behavior as well. In practice, problems such as the design of a product or a political campaign involve many specific constraints concerning the nature of the product or the candidate. To the extent that these constraints can be translated into the present framework, the elimination-by-aspects model can be used (or abused) to determine the optimal design, or location, of choice alternatives.

Psychological Interpretation

The EBA model accounts for choice in terms of a covert elimination process based on sequential selection of aspects. Any such sequence of aspects can be regarded as a particular state of mind which leads to a unique choice. In light of this interpretation, the choice mechanism at any given moment in time is entirely deterministic; the probabilities merely reflect the fact that at different moments in time different states of mind (leading to different choices) may prevail. According to the present theory, choice probability is an increasing function of the values of the relevant aspects. Indeed, the elimination-by-aspects model is compensatory in nature despite the fact that at any given instant in time, the choice is assumed to follow a conjunctive (or a lexicographic) strategy. Thus, the present model is compensatory

"globally" with respect to choice probability but not "locally" with regard to any particular state of mind.

In the proposed model, aspects are interpreted as desirable features; the selection of any particular aspect leads to elimination of all alternatives that do not contain the selected aspect. Following the present development, one can formulate a dual model where aspects are interpreted as disadvantages, or regrets, associated with the alternatives. According to such a model, the selection of a particular aspect leads to the elimination of all alternatives that contain the selected aspect. This model is also based on the notion of elimination by aspects, except that here an alternative is chosen if and only if none of its aspects is selected, whereas in the model developed in this paper an alternative is selected if and only if it includes all the selected aspects. The former model may be more appropriate when the defining features of the alternatives are naturally viewed as undesirable. In choosing among various insurance policies, for example, it may be more natural to apply the strategy of elimination by aspects to the various risks and premiums, treated as disadvantages or regrets, than to interpret them as relative advantages with respect to some reference points.⁵

Although the present model has been introduced and discussed in terms of aspects, we have shown that it requires no specific assumptions concerning the structure of these aspects. In the course of the investigation, however, assumptions concerning the structure and/or the relative weights of aspects were sometimes introduced. In discussing the Paris-Rome problem, for example, we assumed that Paris + (i.e., a trip to Paris plus an added bonus) includes Paris in the sense that all aspects of the latter are included in the former. Similarly, in analyzing Debreu's example, we assumed that the two recordings B_1 and B_2 of the Beethoven

⁵ George Miller remarked that people seem to be better at finding what is wrong with an alternative than what is good about it. This certainly is true of some people, who might then find the "negative" version of the model less objectionable or more compatible with their way of thinking.

symphony are very similar to each other, whereas the suite by Debussy is relatively dissimilar to either of them. Essentially the same assumption was employed in the analysis of the experimental data. In all these instances, specific assumptions about the structure or the relative weights of aspects were added to the model on the basis of some prior analysis of the alternatives. The addition of such assumptions strengthens the predictions of the model and tightens its empirical interpretation. These assumptions, however, must be carefully examined because the inadequacy of an added assumption can erroneously be interpreted as a failure of the model.

To illustrate this point, consider the following example of choice between articles of clothing. Let J denote a jacket, S a pair of matching slacks, and C a coat. Suppose that the coat is more valuable than the jacket, so $P(C;J) > 1/2$. But since the slacks and the jacket are well matched, $P(JS;CS) > 1/2$, where JS and CS denote the options consisting of the combined respective articles. Both JS and CS share the same article, S ; hence one might be tempted to interpret S as a collection of aspects shared by the two alternatives. According to the elimination-by-aspects model, such aspects could be deleted without affecting the choice process. Consequently, under the proposed interpretation of S , $P(JS;CS) = P(J;C)$ contrary to the assumptions. Further reflection, however, reveals that the interpretation of S as a collection of aspects common to both options is inappropriate. The fact that the jacket and the slacks form an attractive outfit implies that this alternative has some gestaltlike properties, or that the option JS includes some aspects that are not included in either J or S alone. Hence, the fact that the option JS includes both J and S as components does not, by itself, justify the conclusion that the aspects of JS can be partitioned into those associated with J and S alone.

Rational Choice and the Logic of Elimination by Aspects

The following television commercial serves to introduce the problem. "There are more

than two dozen companies in the San Francisco area which offer training in computer programming." The announcer puts some two dozen eggs and one walnut on the table to represent the alternatives, and continues: "Let us examine the facts. How many of these schools have on-line computer facilities for training?" The announcer removes several eggs. "How many of these schools have placement services that would help find you a job?" The announcer removes some more eggs. "How many of these schools are approved for veterans' benefits?" This continues until the walnut alone remains. The announcer cracks the nutshell, which reveals the name of the company and concludes: "This is all you need to know in a nutshell."

This commercial illustrates the logic of elimination by aspects; it also suggests that this logic has some normative appeal as a method of choosing among many complex alternatives. The appeal of this logic stems primarily from the fact that it is easy to state, defend, and apply. In choosing among many complex alternatives such as new cars or job offers, one typically faces an overwhelming amount of relevant information. Optimal policies for choosing among such alternatives usually require involved computations based on the weights assigned to the various relevant factors, or on the compensation rates associated with the critical variables. Since man's intuitive computational facilities are quite limited (Shepard, 1964b; Slovic & Lichtenstein, 1971), the above method is difficult to apply.

Moreover, it seems that people are reluctant to accept the principle that (even very important) decisions should depend on computations based on subjective estimates of likelihoods or values in which the decision maker himself has only limited confidence. When faced with an important decision, people appear to search for an analysis of the situation and a compelling principle of choice which will resolve the decision problem by offering a clear-cut choice without relying on estimation of relative weights, or on numerical computations. (Altogether people seem to have more confidence in the rationality of

their decisions than in the validity of their intuitive estimates, and the fact that the former depends on the latter is often met with a mixture of resistance and unhappiness.)

The strategy of elimination by aspects (illustrated by the above commercial) provides an example of such a principle: It is relatively easy to apply, it involves no numerical computations, and it is easy to explain and justify in terms of a priority ordering defined on the aspects. Inasmuch as people look for a decision rule that not only looks sensible, but which also seems easy to defend to oneself as well as to others, the principle of elimination by aspects appears attractive. Its uncritical application, however, may lead to very poor decisions. For virtually any available alternative, no matter how inadequate it might be, one can devise a sequence of selected aspects or, equivalently, describe a particular state of mind that leads to the choice of that alternative.

Indeed, the purpose of advertisement is to induce a state of mind in the decision maker which will result in the purchase of the advertised product. This is typically accomplished by increasing the salience and the availability of the desired state of mind. Being influenced by such factors, people are often lured into adopting a state of mind which, upon further reflection, appears atypical or inadequate. Shepard (1964b) tells of a person who is induced to purchase the *Encyclopedia Britannica* by imagining how he would read it in his free time and impress his friends with his newly acquired knowledge. Only after failing to consult the *Encyclopedia Britannica* for a long period of time does the person realize how inappropriate the state of mind was that had led him to purchase those many dusty volumes.

From a normative standpoint, the major flaw in the principle of elimination by aspects lies in its failure to ensure that the alternatives retained are, in fact, superior to those which are eliminated.

In the problem addressed by the above commercial, for instance, the existence of placement services that would help

the trainee to find a job is certainly a desirable aspect of the advertised program. Its use as a criterion for elimination, however, may lead to the rejection of programs whose overall quality exceeds that of the advertised one despite the fact that they do not offer placement services.

In general, therefore, the strategy of elimination by aspects cannot be defended as a rational procedure of choice. On the other hand, there may be many contexts in which it provides a good approximation to much more complicated compensatory models and could thus serve as a useful simplification procedure. The conditions under which the approximation is adequate, and the manner in which this principle could be utilized to facilitate and improve decision making, are subjects for future investigations.

REFERENCES

- 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. Probabilities of choices among very similar objects. *Behavioral Science*, 1963, 8, 306-311. (b)
 BLOCK, H. D., & MARSCHAK, J. Random orderings and stochastic theories of responses. In J. Olkin, S. Ghurye, W. Hoeldding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics*. Stanford: Stanford University Press, 1960.
 CHANDLER, J. P. STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science*, 1969, 14, 81-82.
 CHURMAN, J. S. Stochastic choice and subjective probability. In D. Willner (Ed.), *Decisions, values, and groups*. Vol. 1. New York: Pergamon Press, 1960.
 COOMBS, C. H. On the use of inconsistency of preferences in psychological measurement. *Journal of Experimental Psychology*, 1958, 55, 1-7.
 COOMBS, C. H. *A theory of data*. New York: Wiley, 1964.
 DEBREU, G. Review of R. D. Luce, *Individual choice behavior: A theoretical analysis*. *American Economic Review*, 1960, 50, 186-188.
 ESTES, W. K. A random-walk model for choice behavior. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences, 1959*. Stanford: Stanford University Press, 1960.
 FISHER, P. C. Utility theory. *Management Science*, 1968, 13, 435-453.
 KRATZ, D. H. *The scaling of small and large color differences*. (Doctoral dissertation, University of Pennsylvania) Ann Arbor, Mich.: University Microfilms, 1964. No. 65-5777.
 KRATZ, D. H. Rational distance function for

- multidimensional scaling. *Journal of Mathematical Psychology*, 1967, 4, 226-245.
 LANCASTER, K. J. A new approach to consumer theory. *Journal of Political Economy*, 1966, 74, 132-157.
 LUCE, R. D. *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959.
 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, III*. New York: Wiley, 1965.
 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.
 MORRISON, H. W. Testable conditions for triads of paired comparison choices. *Psychometrika*, 1963, 28, 369-390.
 RUSTEN, 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 Rostle choice models. *Journal of Mathematical Psychology*, 1971, 8, 370-381.
 SHEPARD, R. N. Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, 1964, 1, 54-87. (a)
 SHEPARD, R. N. On the subjectively optimum selection among multiattribute alternatives. In M. W. Shelly & G. L. Bryan (Eds.), *Human judgments and optimality*. New York: Wiley, 1964. (b)
 SLOVIC, P., & LICHTENSTEIN, S. C. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 1971, 6, 649-744.
 THURSTONE, L. L. A law of comparative judgment. *Psychological Review*, 1927, 34, 273-286.
 THURSTONE, L. L. *The measurement of values*. Chicago: University of Chicago Press, 1959.
 TORGERSON, W. S. Multidimensional scaling of similarity. *Psychometrika*, 1965, 30, 379-393.
 TVERSKY, A. Intransitivity of preferences. *Psychological Review*, 1969, 76, 31-48.
 TVERSKY, A. Choice by elimination. *Journal of Mathematical Psychology*, 1972, in press.
 TVERSKY, A., & KESSE, J. E. Similarity and substitutability in binary choices. *Journal of Mathematical Psychology*, 1969, 6, 1-12.

APPENDIX

This appendix establishes the equivalence of the two formulations (Equations 6 and 9) of the elimination-by-aspects model. Let T be a finite set of alternatives. For each $x \in T$, let x' denote the set of aspects associated with x . For any $A \subseteq T$, define $A' = \{a | a \in x' \text{ for}$

some $x \in A\}$, $A'' = \{a | a \in x' \text{ for all } x \in A\}$, and $\bar{A} = \{a | a \in x' \text{ for all } x \in A \text{ \& } a \notin y' \text{ for any } y \in A\}$. We wish to show that there exists a positive scale U on $T' - T''$ satisfying Equation 6 if and only if there exists a scale U on $\{\bar{A} | A \in T\}$ satisfying Equation 9.

It follows at once from the above definitions that $\{\bar{A} | A \in T\}$ forms a partition of $T' - T''$, since any $a \in T' - T''$ belongs to exactly one \bar{A} . Suppose Equation 6 holds. For any $A \subseteq T$, define $U(\bar{A}) = \sum_{a \in \bar{A}} u(a)$. By the

positivity of u , U is nonnegative and $U(\bar{A}) = 0$ iff $\bar{A} = \emptyset$. Note that if $a, b \in \bar{B}$ then for all $A \subseteq T$, $A \cap a = A \cap b = A \cap B$. Furthermore, since $\{B_i | x \in B_i\}$ forms a partition of x' , the numerator in Equation 6 can be expressed as

$$\begin{aligned} \sum_{a \in x' - A''} P(x, A \cap a) u(a) \\ &= \sum_{x \in B_i, \bar{A} \neq A} \sum_{a \in \bar{B}_i} P(x, A \cap a) u(a) \\ &= \sum_{x \in B_i, \bar{A} \neq A} P(x, A \cap B_i) \sum_{a \in \bar{B}_i} u(a) \\ &= \sum_{B_i \neq A} P(x, A \cap B_i) U(\bar{B}_i). \end{aligned}$$

(The condition $x \in B_i$ under the summation sign is deleted because for any $x \in B_i$, $P(x, A \cap B_i) = 0$. Similarly, since $\{\bar{B} | \bar{B} \in A' - A''\} = \{\bar{B} | \bar{B} \in \bar{A}_i \text{ for some } A_i \text{ such that } A_i \not\subseteq A \text{ and } A_i \cap A \neq \emptyset\}$, the denominator in Equation 6 can be expressed as

$$\sum_{\bar{B} \in A' - A''} u(\bar{B}) = \sum_{\bar{A}_i \in \bar{A}} U(\bar{A}_i)$$

where $\bar{A} = \{A_i | A_i \cap A \neq A, \emptyset\}$.

Thus, Equation 6 reduces to Equation 9, since

$$\begin{aligned} \frac{\sum_{a \in x' - A''} u(a) P(x, A \cap a)}{\sum_{\bar{B} \in A' - A''} u(\bar{B})} &= \frac{\sum_{B_i \neq A} U(\bar{B}_i) P(x, A \cap B_i)}{\sum_{\bar{A}_i \in \bar{A}} U(\bar{A}_i)} \end{aligned}$$

Conversely, suppose Equation 9 holds. That is, there exists a scale U such that $P(x, A)$ is given by the right-hand side of the above equation. For any $x \in T$, let $x' = \{A_i | x \in A_i\}$, and $u = U$. Hence Equation 9 reduces to Equation 6. Finally, if either of the above denominators vanishes, so does the other.

(Received February 7, 1972)

