

Substitutability and Similarity in Binary Choices¹

AMOS TVERSKY

Hebrew University of Jerusalem, Jerusalem, Israel

AND

J. EDWARD RUSSO²

University of Michigan, Ann Arbor, Michigan 48104

The assumption that a binary choice probability is expressible as a monotone function of the scale values of the two alternatives is investigated. Four different conditions are shown to be equivalent forms of the same substitutability, or independence, principle which underlies most probabilistic theories of choice behavior. In a study of judgments of relative size, the independence principle is contrasted with the hypothesis that interstimulus similarity facilitates discrimination. The data reject the independence principle while supporting the similarity hypothesis.

Most probabilistic theories of choice behavior are based on a fundamental principle that has appeared in several different forms. The assumptions of simple scalability, strong stochastic transitivity, substitutability, and independence are different versions of the same basic principle. In the first part of the paper, these four assumptions are shown to be logically equivalent. In the second part, this principle is contrasted with an alternative hypothesis in an experimental study involving judgments of relative size.

To introduce the various conditions, let S be a set of alternatives or stimuli, denoted x, y, \dots , and let $P(x, y)$ be the probability that x is chosen over y . More specifically, we assume that $P(x, y) + P(y, x) = 1$ and that $P(x, x) = \frac{1}{2}$, for all x, y in S . Furthermore, it is assumed that all choice probabilities are neither 0 nor 1, that is, all preferences or discriminations are imperfect. These probabilities are usually estimated by the relative frequencies observed in binary choice experiments.

A set of binary choice probabilities satisfies *simple scalability* if there are real-valued functions F and u such that for all x, y in S

$$P(x, y) = F[u(x), u(y)], \quad (1)$$

¹ This work was supported in part by Public Health Service Grant MH-04236 and by National Science Foundation Grant GM-6782 to the University of Michigan.

² We thank David H. Krantz and H. William Morrison for their many helpful suggestions and criticism based on an earlier version of this paper.

where F is strictly increasing in its first argument and strictly decreasing in the second. This property, introduced by Krantz (1964), states that the effect of each stimulus, x , can be summarized by a single scale value, $u(x)$. Two alternatives are thus equivalent if, and only if, they have the same scale value. (Krantz's original formulation is slightly weaker as F must only be one-to-one in each argument. The two formulations, however, are equivalent if F is continuous in both arguments.)

Equation 1 is probably the most general formulation of independence between alternatives. The more elaborate choice models, such as Thurstone's (1927, case V) and Luce's (1959), require the stronger assumption that

$$P(x, y) = F[u(x) - u(y)].$$

The difference between the two assumptions is that in the former F is a function in two variables, $u(x)$ and $u(y)$, whereas in the latter it is a function of their difference. A detailed analysis of the relationships among the various probabilistic choice models can be found in Luce and Suppes (1965).

Despite its generality, simple scalability has several testable consequences. In particular, it implies that if $P(x, y)$ and $P(y, z)$ exceed one half then $P(x, z)$ exceeds both of them. This property, called *strong stochastic transitivity* (SST), is a probabilistic version of transitivity. Stated formally,

$$P(x, y) \geq \frac{1}{2} \quad \text{and} \quad P(y, z) \geq \frac{1}{2} \quad \text{imply} \quad P(x, z) \geq \max[P(x, y), P(y, z)], \quad (2)$$

where strict inequality in both hypotheses entails strict inequality in the conclusion. (The present formulation of SST is slightly stronger than the usual one, as the requirement of strict inequality is typically omitted.)

To derive (2) from (1), suppose $P(x, y) \geq \frac{1}{2} = P(y, y)$, hence $F[u(x), u(y)] \geq F[u(y), u(y)]$ and $u(x) \geq u(y)$ since F is increasing in its first argument. Consequently, $F[u(x), u(z)] \geq F[u(y), u(z)]$ or, $P(x, z) \geq P(y, z)$. Similarly, $P(y, z) \geq \frac{1}{2} = P(z, z)$ implies $u(y) \geq u(z)$ and $P(x, z) \geq P(x, y)$, since F is decreasing in its second argument. Finally, it is easy to verify that strict inequality in both hypotheses yields strict inequality in the conclusion, which completes the proof.

Strong stochastic transitivity, in turn, implies the following *substitutability* condition for all x, y, z in S .

$$P(x, z) \geq P(y, z) \quad \text{if and only if} \quad P(x, y) \geq \frac{1}{2}. \quad (3)$$

This property may also be stated as a conjunction of two implications: (i) $P(x, z) > P(y, z)$ implies $P(x, y) > \frac{1}{2}$, and (ii) $P(x, z) = P(y, z)$ implies $P(x, y) = \frac{1}{2}$. The equivalence of the two forms is readily established.

To derive (3), assume (i) is false; hence $P(x, z) > P(y, z)$ but $P(x, y) \geq \frac{1}{2}$. There are two cases to be considered, $P(x, z) \geq \frac{1}{2}$ and $\frac{1}{2} > P(x, z)$. First, suppose $P(x, z) \geq \frac{1}{2}$; hence by SST $P(y, z) \geq P(x, z)$, contrary to our hypothesis. Second,

suppose $\frac{1}{2} > P(x, z)$; hence by hypothesis $\frac{1}{2} > P(y, z)$ or $P(z, y) > \frac{1}{2}$, but since $P(y, x) \geq \frac{1}{2}$ it follows from SST that $P(z, x) \geq P(z, y)$, contrary to our hypothesis that $P(x, z) > P(y, z)$.

Next, assume (ii) is false; hence $P(x, z) = P(y, z)$ but $P(x, y) \neq \frac{1}{2}$, say $P(x, y) > \frac{1}{2}$. There are three cases to be considered, $P(x, z) > \frac{1}{2}$, $\frac{1}{2} > P(x, z)$, and $P(x, z) = \frac{1}{2}$. First, suppose $P(x, z) = P(y, z) > \frac{1}{2}$; hence by SST, $P(x, z) > P(y, z)$, a contradiction. Second, suppose $\frac{1}{2} > P(x, z)$ or $P(z, x) > \frac{1}{2}$; hence by SST $P(z, y) > P(z, x)$ or $P(x, z) > P(y, z)$, a contradiction. Finally, suppose $P(x, z) = \frac{1}{2} = P(z, y)$; hence by SST $P(x, y) = \frac{1}{2}$ as required. This completes the derivation of the substitutability condition. Essentially the same result was obtained by Block and Marschak (1960, Theorem 4.1).

A set of binary choice probabilities satisfies the *independence* condition if for any x, y, z, w in S ,

$$P(x, z) \geq P(y, z) \quad \text{if and only if} \quad P(x, w) \geq P(y, w). \quad (4)$$

Thus, if two stimuli (x, y) are ordered according to their choice probabilities relative to some fixed standard then, under Eq. 4, the ordering is independent of the particular standard. Essentially the same property plays an important role in the theory of conjoint measurement (see Tversky, 1967). To derive independence from substitutability, suppose $P(x, z) \geq P(y, z)$; hence by applying (3) twice, $P(x, y) \geq \frac{1}{2}$ and $P(x, w) \geq P(y, w)$ as required.

The independence condition, in turn, implies simple scalability. To demonstrate, choose a fixed element z and define a real-valued function u on S by $u(x) = P(x, z)$. Next, define another real-valued function F by the equation $F[u(x), u(y)] = P(x, y)$. To show that F is well defined, suppose $u(x) = u(x')$ and $u(y) = u(y')$. Then $P(x, z) = P(x', z)$ so $P(x, y) = P(x', y)$ by (4). Also, $P(y, z) = P(y', z)$, so $P(y, x') = P(y', x')$ by (4). Hence, $P(x, y) = P(x', y')$, i.e., $F[u(x), u(y)] = F[u(x'), u(y')]$. Reversing these steps shows that F is one-to-one in each component. Finally, to show that F is strictly increasing in the first argument, suppose $u(x') > u(x)$; hence, by construction together with independence, $P(x', y) > P(x, y)$ and $F[u(x'), u(y)] > F[u(x), u(y)]$ as required. An analogous argument applied to the second component shows that F is strictly decreasing in the second component, which completes the derivation of simple scalability. Essentially the same result was established by Krantz (1964, Theorem 4).

Using transitivity of implications, the previous discussion is summarized by the following result.

THEOREM. *The following conditions are equivalent:*

- (i) *simple scalability,*
- (ii) *strong stochastic transitivity,*

- (iii) *substitutability, and*
- (iv) *independence.*

All four properties, therefore, capture the same principle that pairwise choice probabilities can be expressed as a monotone function of some underlying scale values in such a way that if two alternatives are equivalent in one context, they are substitutable for each other in any context.

Although this principle has dominated much of the theoretical work in the field, research exists indicating that choice probabilities are affected by comparability factors which cannot be accounted for by any model based on simple scalability. In particular, Coombs (1958) presented subjects with gray color patches varying in brightness and asked them to select that patch closest to their ideal image of gray. As predicted by Coombs' unfolding theory, SST was violated for some specified triples of stimuli that lay on both sides of the subject's ideal point. More recently, Krantz (1967) has demonstrated serious violations of simple scalability in judgments of similarity between pairs of monochromatic colors. Krantz showed that the probability of choosing one pair of stimuli as more similar than another pair is affected by comparability factors between the pairs, over and above the similarity between the elements of each pair. These results indicate that choice probabilities reflect not only the scale values of the alternatives, but also the degree of difficulty of the comparison. Consequently, substitutability is violated as alternatives may be substitutable in some contexts but not in others.

The similarity between stimuli has long been considered a determinant of the degree of comparability between them. In fact, it has been hypothesized that for a fixed difference between the psychological scale values, the more similar the stimuli, the easier the comparison or the discrimination between them.

The present study investigated the simple scalability principle and the above similarity hypothesis in judgments of relative size. Geometric figures, varying in size and shape independently, were used as stimuli. The subjects were asked to judge which of two figures presented to them had a bigger area. Each stimulus was paired with each of two standards that were maximally dissimilar from each other with respect to shape. If simple scalability holds, then the orders of the choice probabilities obtained under the two standards should coincide. This is precisely the independence condition of Eq. 4. If, on the other hand, shape similarity facilitates judgments of relative size, then the independence condition must be violated by some specified pairs of stimuli. The two opposing predictions are contrasted in the following experiment.

METHOD

Stimuli. The stimuli consisted of two sets of geometric figures: rectangles and lenses, where a lens was a figure formed by the intersection of two circles of equal radii. Each set contained

20 variable stimuli varying in size and shape, and two standard stimuli of the same size but with different shapes. The 20 stimuli in each set formed a factorial design with four size levels and five shape levels. The size levels were determined by the ratio of the stimulus area to the area of the standards. The ratios of .91, .94, 1.06, and 1.09 were used in both sets. The shape levels were determined by the ratio of width to length in the rectangle set and by the ratio of the minor axis to the major axis in the lens set. The ratios used in the rectangles were 2/3, 1/2, 1/3, 1/4, and 1/5, while the ratios used in the lenses were .70, .55, .40, .30, and .20. The shape levels of both sets were denoted by the letters *a* through *e*. In both sets the standards had the two extreme shape levels. One standard (*s_a*) was closest to a square (or a circle), while the other standard (*s_e*) had the most elongated shape. A schematic illustration of the stimulus sets including sets of rectangles and lenses with the same area and all five different shape levels is presented in Fig. 1.

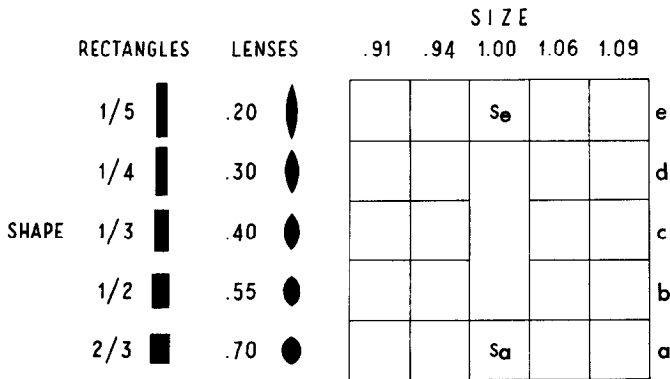


FIG. 1. An illustration of the stimulus sets.

Subjects. One hundred and sixty-eight inmates of the Detroit House of Correction participated in the experiment. Seven subjects were deleted because they failed to perform the task according to instructions. Of the remaining subjects, 78 were presented with rectangles and 83 were presented with lenses.

Procedure. Each standard was paired with all 20 variable stimuli from the same set, yielding a total of 40 pair comparisons between rectangles and 40 pair comparisons between lenses. The subjects were asked to judge which member of a pair had the larger area. The stimuli were projected on a screen for a period of about 10 sec, under normal viewing conditions. The experimental session consisted of a practice period and three replications of a complete stimulus set (40 pair comparisons). The presentation order was randomized. The session, including the practice period, lasted about 2 hr, and the subjects were run in groups of approximately 45 each. In order to motivate the subjects, payments were given on the basis of the proportion of correct responses. Payments consisted of packs of cigarettes, which are used as currency in the prison; the average subject received about three packs.

RESULTS

The data consist of a set of 5×4 matrices in which the rows correspond to the shape levels, the columns correspond to the size levels, and the cell entries are the

relative frequencies of choosing the variable stimulus over the standard. Since each standard was compared with all variable stimuli, two such data matrices were obtained for each subject. The first test of the independence principle was based on the comparisons where the variable stimulus had one of the two extreme shape levels. Hence, only the bottom and the top rows of each data matrix were utilized in this analysis. These rows correspond to the stimuli with the least and the most elongated shapes, designated by a and e , respectively.

Let x_{ai} and x_{ei} , $i = 1, \dots, 4$, denote stimuli of the two extreme shapes (a and e) and the same size (i). If the independence principle is valid, then for any size level, i ,

$$P(x_{ai}, s_a) \geq P(x_{ei}, s_a) \quad \text{if and only if} \quad P(x_{ai}, s_e) \geq P(x_{ei}, s_e). \quad (5)$$

That is, the order of the choice probabilities is independent of the standard. If, however, the similarity hypothesis is valid and shape similarity facilitates the judgments, then the more similar the stimuli with respect to shape, the easier the size discrimination between them. Consequently, the comparison between x_{ai} and s_a is easier than that between x_{ei} and s_a , since s_a and x_{ai} have the same shape. Similarly, the comparison between x_{ei} and s_e is easier than that between x_{ai} and s_e since s_e and x_{ei} have the same shape. If both variable stimuli in Eq. 5 were of the same subjective area and if both standards were of the same subjective area, then the similarity hypothesis would imply that

$$P(x_{ai}, s_a) \geq P(x_{ei}, s_a) \quad \text{if and only if} \quad P(x_{ei}, s_e) \geq P(x_{ai}, s_e). \quad (6)$$

That is, opposite orders should be obtained under the two standards, in direct contradiction to the earlier prediction. Since in the present design, however, the stimuli in the pairs (s_a, s_e) and (x_{ai}, x_{ei}) are of equal objective rather than subjective area, Eq. 6 is not a necessary consequence of the similarity hypothesis. Nevertheless, one would expect Eq. 6 to be satisfied for some pairs of stimuli if the similarity hypothesis is true.

To compare the predictions of (5) and (6), the following measure (M) of the degree of correspondence between the orders obtained under the two standards was devised. To each size level ($i = 1, 2, 3, 4$), $+1$ was assigned if Eq. 5 was satisfied, and -1 was assigned if Eq. 6 was satisfied. The cases in which a tie occurred in at least one of the two orders do not provide an adequate basis for comparing the orders and were, therefore, discounted from the analysis. The value of M for a given individual is simply the sum of the $+1$'s and -1 's (over the four size levels) normalized by the number of untied comparisons. Thus, M ranges from $+1$ to -1 , where $+1$ is predicted by Eq. 5, -1 is predicted by Eq. 6, and 0 is expected on the basis of random choice. The value of M is essentially an average Kendall's tau where tau is based on two stimuli only.

The distributions of the M values are presented in Table 1. The obtained distributions were positively skewed in both stimulus sets. The average M values were $-.38$ for rectangles and $-.24$ for lenses, both of which were significantly ($p < .05$) negative according to the significance test for Kendall's tau.

TABLE 1
THE FREQUENCY DISTRIBUTIONS OF THE INDIVIDUALS' M VALUES

M	Rectangles	Lenses	Total
+1	12	16	28
$+\frac{1}{3}$	3	4	7
0	9	11	20
$-\frac{1}{3}$	3	2	5
-1	36	32	68

The overall relative frequencies of choosing the variable stimuli over the standards, totaled for all subjects, are given in Table 2. The group data provide strong support for the similarity hypothesis and strong evidence against the independence principle, which is violated in all cases. In every column of Table 2, the cell entries are ordered oppositely under the two standards in complete agreement with Eq. 6.

TABLE 2
THE OVERALL FREQUENCY OF CHOOSING THE VARIABLE
STIMULUS AS LARGER THAN THE STANDARD

Standard	Size level: Shape	1	2	3	4	
s_a Rectangles	e	93	106	154	180	
	a	23	41	205	208	
	s_e	e	16	26	198	210
		a	58	73	117	125
s_a Lenses	e	129	153	202	203	
	a	40	48	222	232	
	s_e	e	37	49	211	215
		a	44	56	100	102

The only implication of the similarity hypothesis that is independent of perceived area values is that violations of independence should be due to shape similarity. That is, if the discrimination between s_a and x is better than that between s_a and y whereas the discrimination between s_e and x is worse than that between s_e and y , then s_a should be more similar to x than to y , and s_e should be more similar to y than to x . Letting $Q(x, y)$ denote the proportion of correct area judgments between x and y , it is readily seen that any strict violation of independence is expressible in the form

$$Q(x, s_a) > Q(y, s_a) \quad \text{and} \quad Q(y, s_e) > Q(x, s_e) \quad (7)$$

To test the similarity hypothesis, two sets of pairs of variable stimuli, where the elements in each pair have the same area, were employed in this analysis. Set I contains all such pairs where one element has shape a and the other element has shape e , while Set II contains all such pairs where one element has shape b and the other element has shape d . In accord with the natural partial order of shape similarity between stimuli, all pairs from Sets I and II were classified as follows: a pair of variable stimuli (x, y) satisfying (7) is compatible with the similarity hypothesis if and only if x is less elongated than y . That is, if either (x, y) belongs to Set I and x has shape a while y has shape e , or if (x, y) belongs to Set II and x has shape b while y has shape d . The proportions of pairs of rectangles and lenses, from Sets I and II, which satisfy (7) and are compatible with the similarity hypothesis are given in Table 3. If all violations

TABLE 3
THE PROPORTIONS OF PAIRS OF STIMULI WHICH
VIOLATE THE INDEPENDENCE PRINCIPLE AND ARE
COMPATIBLE WITH THE SIMILARITY HYPOTHESIS

	Rectangles	Lenses	Total
Set I	80/83	62/72	142/155
Set II	41/60	28/42	69/102
Total	121/143	90/114	211/257

of independence were due to indiscriminability or random error, then only one half of the violations should be compatible with the similarity hypothesis. However, all the entries of Table 3 are significantly greater than one half. (The proportion of pairs of lenses from Set II is significant at the .05 level, whereas all other proportions are significant at the .01 level.) The results show that most violations of independence are attributable to similarity, and that this effect is stronger in Set I than in Set II where the similarity differences are less extreme. The results of Set II show that the similarity

hypothesis is supported even when the variable stimuli do not have the same shape as the standards.

The final analysis was also based on the frequencies of correct area judgments. From the original 5×4 frequency matrices of each subject, a 5×1 column vector was computed whose entries were the number of correct choices for each shape level, summed over areas. Two such vectors were obtained for each subject, one under each standard, and the rank order correlation (Kendall's tau) between the two vectors was computed. The independence principle predicts a perfect positive correlation between the two vectors. A zero correlation is expected under the assumption of random choice, and a perfect negative correlation is expected under the similarity hypothesis provided stimuli with equal objective area are equal in subjective area. Since the stimuli were not equated in subjective area, however, one would expect a negative but not a perfect correlation. The distributions of the tau values from each stimulus set are presented in Table 4. The obtained distributions were positively

TABLE 4
THE FREQUENCY DISTRIBUTION OF THE INDIVIDUALS' TAU-VALUES

Tau	Rectangles	Lenses	Total
+0.76-+1.00	1	0	1
+0.51-+0.75	5	5	10
+0.26-+0.50	7	5	12
+0.01-+0.25	9	11	20
0	8	8	16
-0.01--0.25	15	6	21
-0.26--0.50	17	17	34
-0.51--0.75	11	17	28
-0.76--1.00	5	14	19

skewed, as 102 subjects had negative values as compared with 43 subjects with positive values. The average tau was $-.15$ for the rectangles and $-.26$ for the lenses, both of which were significantly negative ($p < .01$) according to a test for the significance of Kendall's tau.

The relative frequencies (totaled over subjects) of correct choices for each shape level under the two standards are shown in Fig. 2 for the rectangles and Fig. 3 for the lenses.

The orderings of these values, under the two standards, appear inversely related to each other and monotonically related to similarity, in accordance with the similarity hypothesis and in contradiction to the independence principle.

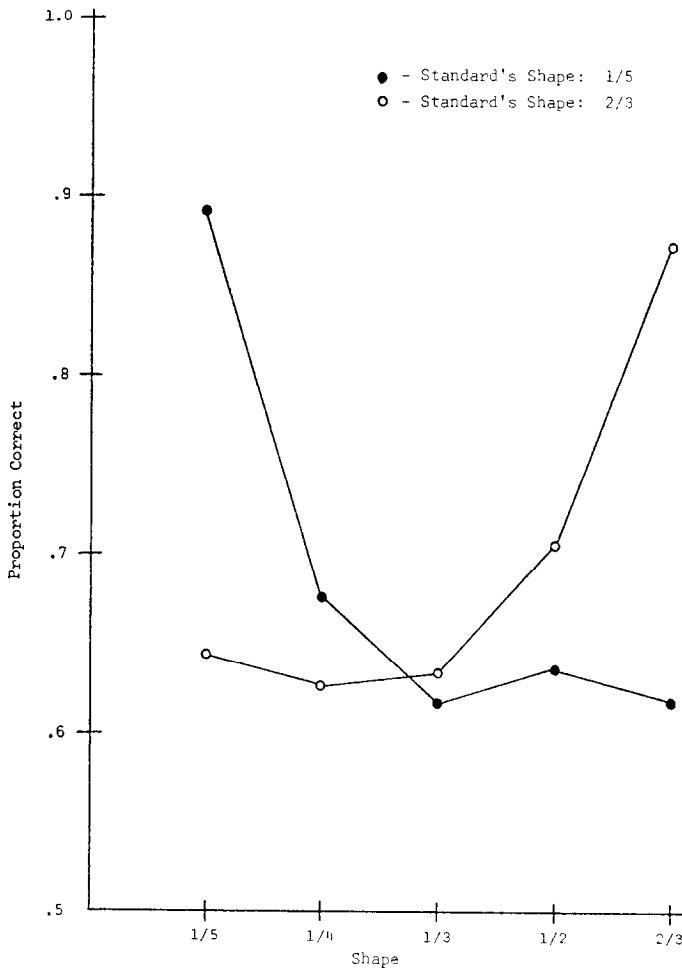


FIG. 2. Proportion of correct choice between rectangles ($N = 936$) summed over subjects and area values.

DISCUSSION

Two incompatible principles of choice were compared in the present study. The data provide evidence against the independence principle and for the similarity hypothesis. It was found that the similarity between stimuli facilitates 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.

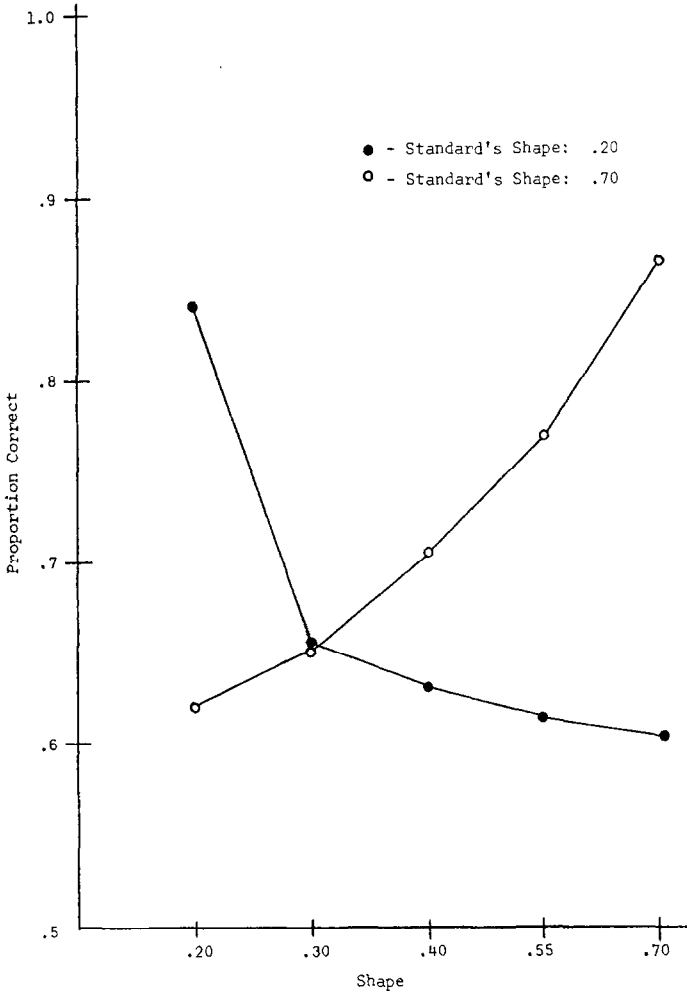


FIG. 3. Proportion of correct choices between lenses ($N = 996$) summed over subjects and area values.

Although these findings hold for both types of stimuli (rectangles and lenses), their applicability to other stimuli and to different types of judgments are left to be explored. If, as available data indicate, simple scalability is violated in many contexts, then both theoretical and applied research on choice behavior should be fundamentally reevaluated.

Is this the end of simple scalability? Not necessarily. It should be recalled that in the present study, the stimuli were paired so as to maximize the similarity effect.

Simple scalability may still hold for more homogeneous sets of pair comparisons. Moreover, all violations of simple scalability have been obtained in studies of pair comparisons. This, however, is not the only empirical procedure for estimating binary choice probabilities. Alternatively, a single-stimulus method may be employed to obtain replicated magnitude estimates for each stimulus. The $P(x, y)$ may then be defined as the probability that a value assigned to x exceeds a value assigned to y . In this method each stimulus is presented alone so that comparability factors cannot operate. Simple scalability may very well be satisfied by choice probabilities estimated in this fashion.

Finally, to the extent that the similarity hypothesis is applicable to the decisions of consumers or voters, it suggests the intriguing possibility of influencing choice probabilities between products or candidates by manipulating the similarity between them.

REFERENCES

- BLOCK, H. D. AND MARSCHAK, J. Random orderings and stochastic theories of responses. In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, and H. Mann (Eds.), *Contributions to probability and statistics*. Stanford: Stanford Univer. Press, 1960.
- COOMBS, C. H. On the use of inconsistency of preferences in psychological measurement. *Journal of Experimental Psychology*, 1958, **55**, 1-7.
- KRANTZ, D. H. The scaling of small and large color differences. Unpublished doctoral dissertation, University of Pennsylvania, 1964.
- KRANTZ, D. H. Rational distance functions for multidimensional scaling. *Journal of Mathematical Psychology*, 1967, **4**, 226-245.
- LUCE, R. D. *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959.
- LUCE, R. D. AND SUPPES, P. Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 3. New York: Wiley, 1965. Pp. 249-410.
- THURSTONE, L. L. A law of comparative judgment. *Psychological Review*, 1927, **34**, 273-286.
- TVERSKY, A. Additivity, utility, and subjective probability. *Journal of Mathematical Psychology*, 1967, **4**, 175-201.

RECEIVED: October 20, 1967