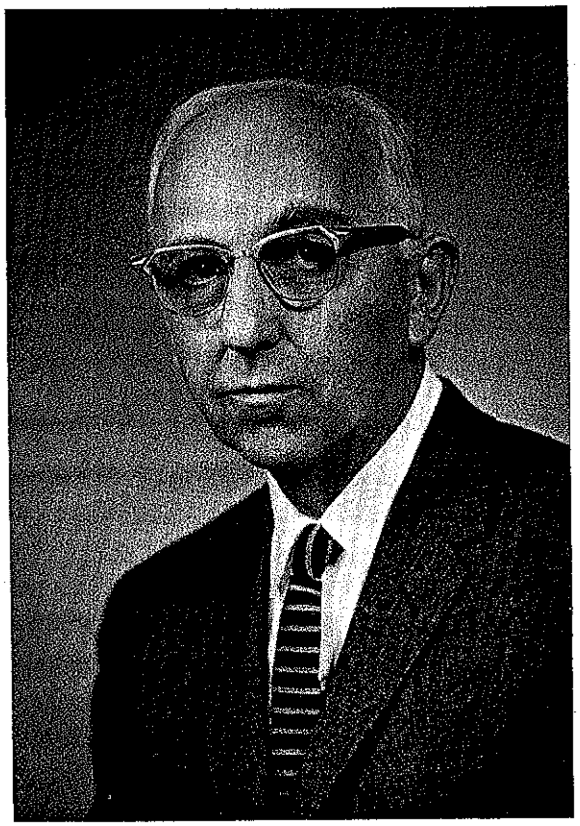

Harry Helson

Adaptation-Level Theory

A SYMPOSIUM

Edited by M. H. APPLEY

GRADUATE SCHOOL
UNIVERSITY OF MASSACHUSETTS
AMHERST, MASSACHUSETTS

1971

ACADEMIC PRESS New York and London

ADAPTATION-LEVEL CODING OF STIMULI AND SERIAL POSITION EFFECTS[1]

Gordon H. Bower

Traditionally, AL theory had its origins and main data base in describing people's judgments of unidimensional stimuli. Accepting the notion that a potential stimulus becomes transformed into an effective stimulus by virtue of an *encoding process,* AL theory provides a hypothetical (and quantitative) description of this encoding—that the functional stimulus is the deviation of the physical stimulus from an internal norm, the AL. This characterization of the functional stimulus has many implications and applications which Helson has systematically pursued and reviewed in his book, *Adaptation-Level Theory* (Helson, 1964).

I wish to apply the theory to a phenomenon which occurs universally whenever subjects learn identifying responses to multiple stimuli that differ from one another along one dimension or in one attribute. The universal phenomenon is the "serial position effect" (SPE). Except for a brief allusion in his book to Murdock's (1960) theory of stimulus distinctiveness, Helson does not discuss SPEs in relation to AL theory, and I hope to show more firmly the relation between the phenomenon and the theory. The plan of this contribution is as follows: First, a few experiments will be reviewed to indicate the ubiquity of the SPE; second, considering theories for the SPE, I shall describe Murdock's (1960) theory, describe a new experiment testing it, then criticize that theory; third, an alternative theory of the SPE will be briefly proposed that is more compatible with AL concepts; and fourth, the major failing of these theories will be brought out in experiments implying that human adults also map (encode) stimuli along various sensory dimensions onto a common conceptual domain composed of abstract elements arrayed in a linear order.

[1] The research reported in this article was supported by Grant MH-13950, from the National Institute of Mental Health.

Serial-Position Effects in Associative Learning

It is well to begin by citing a few examples of the phenomenon that commands our attention. It arises when subjects learn to associate "identifying responses" to stimuli which are values of a unidimensional physical variable. In the jargon of verbal learning, this is paired-associate learning. The stimuli may be on either substitutive or intensity dimensions. Examples would be tones varying in pitch, loudness, or duration, gray squares varying in size, brightness, or spatial position, lines varying in length, and so on.

The experiments involve selecting several values along one of these dimensions, assigning some unrelated identifying responses to the stimuli, then training human subjects on a paired-associate (PA) procedure until they learn the stimulus–response (S–R) assignments. The procedures differ from the psychophysical method of absolute judgments in that the PA responses are unordered with respect to the stimuli, whereas the responses in absolute-judgment experiments are always ordered numerically or verbally (for example, from "very low" to "very high") and assigned to the stimuli with a perfect correlation. Also in the PA procedure, there is an experimentally defined "correct response" to each stimulus and the subject receives feedback regarding this after each guess. These procedural aspects are missing in the absolute-judgment experiments.

Experiments of this sort have been done many times and quite a few have been published. In every one of the published experiments there is a large SPE in learning. The subjects make the fewest errors on the stimulus values at the end of the range, whereas they make the most errors on stimuli in the middle of the range. A bowed curve is obtained when some measure of performance (for example, correct responses or errors) is plotted against the value of the stimulus on the physical dimension.[2]

The basic phenomena to be explained are illustrated in Figs. 1 and 2. Figure 1 shows the relative SP error curve from an experiment by Ebenholtz (1963). The stimuli were 10 locations of a 1-inch red patch; the 10 cells were arrayed in a vertical column 10.3 inches high. The responses were 10 pronounceable nonsense syllables assigned randomly to the spatial locations. Training was done by the method of anticipation to a criterion of one perfect trial. The spatial stimuli were presented each trial in a different random temporal order. The chief

[2] Many of the curves to be shown will be "percentage-normalized" curves, the errors or correct responses at each position being expressed as a percentage of the total errors or total correct responses made, summing over all positions. This description of SP curves cancels out the absolute numbers of errors or corrects, and inspects only the relative numbers at the several positions. This procedure follows in the tradition of McCrary and Hunter (1953), who found that relative SP error curves for rote serial learning were remarkably constant even though total errors varied considerably with manipulations of meaningfulness, pacing rate, intertrial interval, and similar variables.

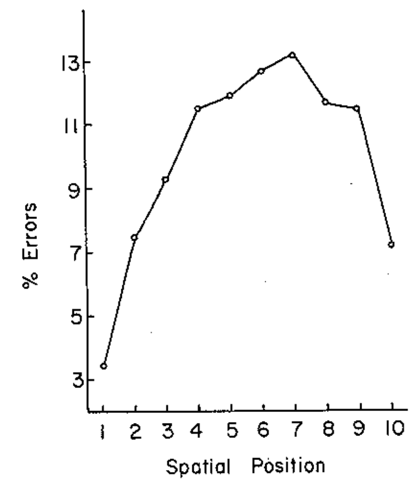

Fig. 1. Percentage of total errors contributed by each positional stimulus during paired-associate learning. Data are conditions D–T and I combined, as reported by Ebenholtz (1963).

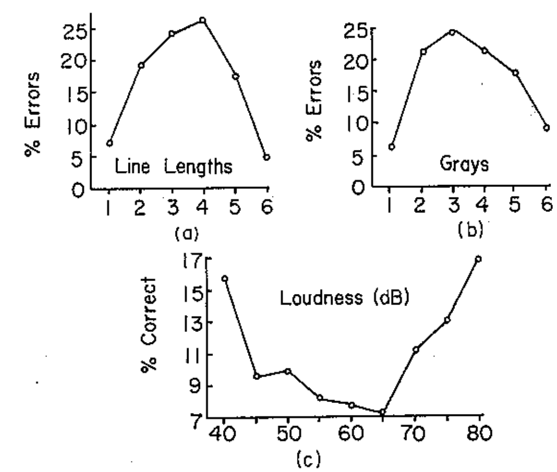

Fig. 2. Relative percentage error curves from the author's experiments using as stimuli (*a*) lines varying in length, and (*b*) gray squares varying redundantly in size and brightness. Part (*c*), from Murdock (1960), is relative percentage correct curves for tones varying in loudness.

thing to notice about Fig. 1 is the bowed shape, the fact that fewest errors occurred on the extreme stimuli. Jensen (1962) reported similar SP curves.

Figure 2 shows similar bowed-shape curves this time for the stimulus dimensions of length of a line, of brightness, and size of a gray patch (from our experiments), and of the loudness of a 1000-Hz tone (from Murdock, 1960). The Murdock data are relative correct SP curves (inverted from the relative error SP curves of the other data).

Serial-position curves such as those illustrated above have been found several times for these dimensions as well as for heaviness, pitch of a tone, temporal durations, and areas of a square or a circle. So the claim is that an SPE results any time a subject learns identifying responses to specific stimuli arranged along a one-dimensional continuum.

Murdock's Theory on Stimulus Distinctiveness

Murdock (1960) collated several SP curves like those above, and offered a theory to predict the shape of these curves. He noted the similarity of his theory to AL theory; Helson subsequently (1964) reviewed Murdock's paper with approval, considering it to be a good application of AL concepts. A critical reexamination of Murdock's theory seems warranted.

Murdock's concern was to devise a measure of the relative distinctiveness of a stimulus within a given set of stimuli, and to relate this measure to the relative accuracy of performance when the subjects were giving absolute judgments or learning PA responses to these stimuli. First, using the Weber–Fechner law, he assumed that the effective stimulus would be the logarithm of the physical stimulus. Second, the distinctiveness of a given stimulus within a set was defined to be proportional to the sum of the distances between it and each of the other stimuli in the set. Table I illustrates the method of computation for a set of five

TABLE I

Computation of PD Measures for Five Logarithmically Spaced Stimuli[a]

Stimuli	Sum of distances $(d)^b$	% distinctiveness (PD)
1	1+2+3+4 = 10	10/40 = .250
2	1+1+2+3 = 7	7/40 = .175
3	1+1+2+2 = 6	6/40 = .150
4	1+1+2+3 = 7	7/40 = .175
5	1+2+3+4 = 10	10/40 = .250

[a] According to Murdock's (1960) hypothesis.
[b] Sum of all distances = 40.

logarithmically spaced stimuli, denoted simply as 1, 2, 3, 4, and 5. The distances from each stimulus to every other stimulus in the set are indicated in the second column. The distinctiveness of each stimulus relative to that of all other stimuli is shown in the third column. Murdock assumed that the percentage-distinctiveness (PD) measure would be predictive of the *relative* degrees of accuracy in performance to these stimuli. This amounts to assuming that whatever the overall level of correct performance to the stimuli (provided it is not perfect nor at chance), the ratio of correct-response probability for stimulus i to that for stimulus j should be the same as the ratio of the two distinctiveness measures. Therefore, the relative percentage correct curve should be identical to the theoretical PD curve; in other words, relative SP curves (for correct responses—but, by implication, for errors too) should be predictable from simply knowing the stimuli involved in the learning task. Murdock compared the predicted PD numbers with several observed SP gradients and found neither significant nor systematic discrepancies. He thus concluded that this PD measure gave an adequate account of relative SP curves. By implication, the frame-of-reference or AL concepts used in Murdock's hypothesis received further confirmation.

Anchor Stimuli in Learning

Inherent in AL theory and Murdock's hypothesis are many implications about learning with unidimensional stimuli. One set of implications, which my student, Peter Arnold, and I have tested, concerns the effect of anchor stimuli on the SP curve obtained during PA learning. The experimental procedure and theoretical predictions may be illustrated by reference to Table II. The stimuli 1–9 are nine horizontal line lengths logarithmically spaced from 1 inch long up

TABLE II

Schematic Procedure for PA Learning to Middle Stimuli 3, 4, 5, 6, 7 with or without Anchors at the Low or High End[a]

	Log stimulus								
Condition	1	2	3	4	5	6	7	8	9
No anchors	p	p	R_1	R_2	R_3	R_4	R_5	p	p
Low anchors	+	+	R_1	R_2	R_3	R_4	R_5	p	p
High anchors	p	p	R_1	R_2	R_3	R_4	R_5	+	+
Both anchors	+	+	R_1	R_2	R_3	R_4	R_5	+	+

[a] The $R_1, \ldots, R_5$ are nonsense-syllable response terms; + denotes that the stimulus was presented; and p denotes that the stimulus was replaced by a neutral picture.

to 9.81 inches long (Weber fraction of 1/3). These were shown one to a flash card; the subject looked at each and either responded or did not respond. All subjects learned to pair the five middle stimuli S_3–S_7 with five unrelated nonsense syllables–JUL, KAW, TES, NEY, and DIZ. The subjects were trained by the conventional anticipation method at a 3:3-sec. rate. During presentation cycles of these five items, four other stimuli were also presented, mixed in randomly with the critical stimuli. These four extra stimuli required no response–in fact, a checkmark on the card informed the subject of this–that he was merely to look at the stimulus.

Four experimental conditions were formed differing in the four extra stimuli. For the no-anchor group, the extra stimuli were colored pictures (advertisements) taken from *Life* magazine. These were intended to be so unlike the line stimuli that they would not be pooled into the AL for judging the critical lines, and the pictures should not influence the learning of responses to the lines. For the low-anchor condition, two of the pictures were replaced by cards displaying the shortest two lines (1.00 and 1.333 inches, respectively). These were shown printed on cards but with a checkmark, so the subject knew that no response was required to such stimuli. For the high-anchor group, the two longest lines (8 and 9) replaced the same two pictures. For the both-anchors group, all four pictures were replaced by the two shortest lines (1,2) and by the two longest lines (8,9). Each subject was tested ten trials, or to a criterion of three consecutive errorless cycles on the five critical stimuli, whichever occurred first.

We shall focus upon the SP curve for stimuli S_3–S_7 for the four experimental groups. What are the expectations of Murdock's hypothesis regarding these SP curves? The theoretical calculations are summarized in Table III and the percentage-distinctiveness measures are shown in graph form (Fig. 3). In following through the calculations of the PD measures, one should be reminded that responses are being collected only to stimuli S_3–S_7, and therefore the PD measures for these stimuli must sum to unity.

In Fig. 3a the SP curves predicted for the no-anchor and both-anchors conditions are compared. The AL, midpoint, and minimum of the symmetric SP curves are predicted to be the same in these two conditions, but the flatness of the two gradients should differ. In particular, the presence of low- and high-anchor stimuli should reduce the former distinctiveness of the end points of the critical stimuli, S_3 and S_7, and therefore performance to these end stimuli should not be much better than that to the middle stimulus. Incidentally, intuition would also suggest that the both-anchors condition should have poorer absolute performance than the no-anchors condition, as well as a flatter relative gradient. In Fig. 3b the relative SP curves predicted for the low- and high-anchor conditions are compared. Performance on end stimuli near the anchors should be depressed relative to the midpoint of the series, and the PD of the stimulus at the opposite end should be enhanced.

TABLE III

Theoretical Calculations of PD Measures for Four Experimental Conditions[a]

		Stimulus								
Condition	Measure	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
No anchors	d	–	–	10	7	6	7	10	–	–
	PD	–	–	.250	.175	.150	.175	.250	–	–
Low anchors	d	21	16	13	12	13	16	21	–	–
	PD	–	–	.173	.160	.173	.214	.280	–	–
High anchors	d	–	–	21	16	13	12	13	16	21
	PD	–	–	.280	.214	.173	.160	.173	–	–
Both anchors	d	36	29	24	21	20	21	24	29	36
	PD	–	–	.218	.192	.182	.192	.218	–	–

[a] Line d gives the summed distance measures from that stimulus to every presented stimulus, and the PD measure is d divided by the sum of d measures for stimuli S_3–S_7.

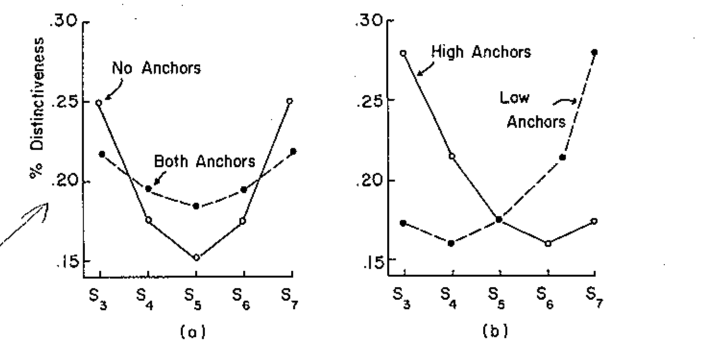

Fig. 3. Percentage-distinctiveness curves for groups with high or low anchors (b) or both high and low anchors or neither (a) graphed from values in Table III.

Only if judged [illegible]

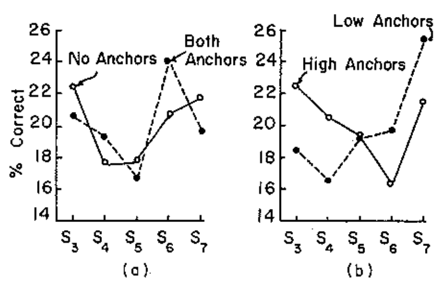

Fig. 4. Observed relative percentage correct curves for the four experimental conditions. The results are to be compared to theoretical predictions shown in Fig. 3: (a) the no-anchors and both-anchors groups; (b) the low- and high-anchors groups.

Those are the predictions. The actual results of the experiment are shown in Fig. 4, which gives the relative percentage correct curves, after pooling data from two different replications of this experiment (a total of approximately 20 *S*s per condition). The no-anchors and both-anchors conditions on the left side do not compare too favorably to the theoretical expectations in Fig. 3. The both-anchors condition was supposed to yield a flatter SP curve; instead it produced an irregular SP curve, with the point at stimulus S_6 simply out of line according to any sensible interpretation. The end points on the both-anchors SP curve are lower than for the no-anchors curve, as expected, but the middle stimulus is also lower. The prediction that both-anchors *S*s would make more errors than the no-anchors *S*s was confirmed in both replications of the experiment. So one is left without any clear theoretical decision in this case, presumably because the effects are small relative to the sampling variability of the mean errors at each point.

In Fig. 4b SP curves for the high- and low-anchor groups are given, and these compare favorably with the expectations in Fig. 3. With the low-anchors group, the minimum point of correct responding is shifted down toward the anchors and the gradient is raised at the upper end; with the high anchors group, the minimum is shifted to the right (in other words, to longer line lengths).

To summarize, then, the anchor experiment has given only equivocal support to Murdock's theory. The results with asymmetrical anchors came out as expected, whereas those with both anchors were not entirely as predicted.

Critique of Murdock's Theory

Although Murdock's theory delivers some accurate predictions, I think it can be shown to be wrong. First, it is a simple matter to concoct "*Gedanken* experiments" in which the theory's predictions must be false; second, as will be shown later, most of the theory's predictions of SP curves do not differentiate it from a large class of simple, alternative theories. That means, of course, that the SP data that have been fit by Murdock's theory are not particularly critical or discriminating.

First, a *Gedanken* experiment for which the theory's prediction about percentage correct responses must surely be false. Consider an absolute-judgment experiment or PA learning experiment arranged as follows: Beginning with three widely spaced stimuli, say, tones of 5, 45, and 85 dB. above the threshold, we then add very many new tones that differ by less than a just noticeable difference (JND) from the low and high tones. For example, by adding two low and two high tones, the new set for discrimination might be 3.9, 4.0, 4.1, 45, 84.9, 85.0, and 85.1 dB. Murdock's PD measure implies that the end stimuli, 3.9 and 85.1 dB., will be the most distinctive and will have the most correct responses, whereas the middle stimulus, 45 dB., will be the least distinctive and will have the fewest correct responses. One need not run this experiment to know that the prediction must be wrong; in such an experiment, the middle stimulus would surely have the most correct responses, because it stands out from the clumps of confusing tones at the high and low ends of the range. Along the same vein, Murdock's hypothesis implies that the relative percentage correct curve for three equally log-spaced stimuli will be the same, independent of their physical spacing. Thus, relative performance in discriminating tones of 5, 45, and 85 dB. is predicted to be the same as that in discriminating 44.9, 45.0, and 45.1 dB. But clearly, by manipulating the spacing one can create sets of three tones for which absolute judgments are either random (PDs = .33), intermediate (PDs = .38, .25, .38), or perfect (PDs = .33) in accuracy. The point is that Murdock's relativistic hypothesis does not take proper account of the spacing between the physical stimuli, nor does it address the question of how indistinguishable stimuli can be predicted by the theory to differ in relative discriminability.

A further point is that Murdock's theory provides no clear way to incorporate the effect on the SP curve of wide variations in differential frequencies of presenting the series stimuli. For example, if, in the no-anchors condition, the longer stimuli (S_6 and S_7) had been presented three times as frequently as the shorter stimuli (S_3-S_5) the AL should shift upwards and have an effect on the SPE similar to that observed in the high-anchors condition. Such relative frequency effects are seen in the shapes of generalization gradients (see, for example, Capehart, Tempone, & Hebert, 1969) and would doubtless occur in relative SP curves, yet it is not clear how Murdock's hypothesis can be addressed to such matters.

Perhaps a more serious criticism of Murdock's approach is that it proposes an *ad hoc* descriptive measure of PD, but it provides no mechanistic explanation whatsoever for why that measure ought to be correlated with performance. It tells us nothing about how learning occurs, what kinds of errors are likely and how they are overcome, how responses are eventually differentiated to nondistinctive stimuli, and so on. It gives no account of generalization errors, and there can be no doubt about such errors. For example, in Fig. 5 (for our line-length experiment) a pooled gradient of remote intrusion errors, expressing the relative frequency with which an error was the response actually appropriate to a stimulus one, two, three, or four steps away is given. The precise shape of the gradient is not as important as the fact that it decreases monotonically. Such intrusion-error gradients are the most elementary manifestations of stimulus

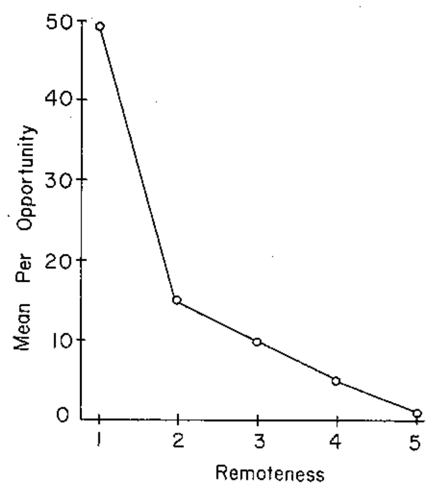

Fig. 5. Gradient of remote intrusion errors in the line-length experiment. Data are pooled over serial positions, with correction for differential logical opportunity for errors of different degrees of remoteness.

generalization, of giving responses R_3 or R_5 to S_4 because S_4 is "close to" S_3 and S_5.

One would like to have some "mechanism" theory of the learning and generalization process. In the next section it will be shown that the most elementary learning-and-generalization model derived from AL theory can deliver predictions of SP curves that are practically indistinguishable from those implied by Murdock's measure.

A Stimulus-Generalization Model

This model uses simple notions of associative learning and stimulus generalization, but the effective stimulus is the discrepancy of the physical stimulus from the AL. The evidence for this latter view has been marshalled effectively in a paper by Capehart *et al.* (1969). Along with the usual weighted-log definition of the AL, we define the effective code, C_i, for physical stimulus S_i as the algebraic difference between log S_i and log AL, all of which is equivalent to the log of the ratio of S_i to the AL. The substantive assumptions are that learning and generalization of responses occurs in the standard "Spencian" manner but with respect to the C_i measures. That is, reinforcement of response R_i in the presence of specific stimulus S_i causes R_i to become associated maximally to C_i; but the response generalizes to other stimuli accordingly as their codes are close to C_i. For illustrative purposes here, let us assume that the stimulus-generalization gradient is linear when habit is plotted on the C scale, namely, against the logarithm of the physical stimulus. In particular, for stimuli spaced by steps of one log unit, the illustrative generalization gradient for response R_i to stimulus S_x will be

$$_xH_i = \begin{cases} {}_iH_i[1-{}^1\!/_3\,|C_i-C_x|] & \text{for} \quad |C_i-C_x| \leqslant 3 \\ 0 & \text{otherwise,} \end{cases}$$

where $_iH_i$ is the habit strength of R_i to coded stimulus C_i. Since $C_i = \log S_i - \log AL$, the absolute difference between the Cs is expressible simply as $\log S_i/S_x$.

To perform computations with this model, consider the habit profile existing after one training cycle through a list of five pairs. Suppose S_1-S_5 denote the stimuli as equally spaced on a log scale. If the stimuli are presented equally often, then the AL will be at S_3 and the coded stimuli will have values -2, -1, 0, $+1$, and $+2$ for S_1-S_5, respectively. The calculations of performance are given in Table IV, where sequential effects are ignored but it is assumed that training has brought the correct habit to a strength of three at each stimulus. The entries in Table IV are habit strengths of the column response to the row stimulus, and the (3, 2, 1) generalization gradient has been used here. Across any row of Table IV, say that for S_1, the strengths of the several response tendencies to that

TABLE IV

Computations of Habit Strengths for Generalization Model

Stimuli	Coded cues	Responses R_1	R_2	R_3	R_4	R_5	Correct/sum	Relative PD
S_1	−2	3	2	1	0	0	3/6	.240
S_2	−1	2	3	2	1	0	3/8	.180
S_3	0	1	2	3	2	1	3/9	.160
S_4	+1	0	1	2	3	2	3/8	.180
S_5	+2	0	0	1	2	3	3/6	.240

stimulus can be seen. It is presumably these remote response tendencies that are manifested in gradients of remote intrusion errors to each stimulus.

A "response axiom" is needed for deciding how the hypothetical *S* should respond. A simple one to use for the moment is Luce's (1959) choice rule which supposes that the probability of R_i to an eliciting stimulus is the ratio of the strength of R_i to that stimulus divided by the summed strengths of all responses to that stimulus. The probability of a correct response, calculated according to this ratio rule, is shown in the next-to-last column of Table IV. In order to compare this model with Murdock's, these correct response probabilities have to be normalized by dividng them by the sum of the correct-response probabilities to all stimuli (which is in fact the expected total number of correct responses if the items S_1-S_5 were to be tested now). This division is shown in the final column PD, and this is the theoretical measure which is to be compared to actual data as well as to the PD predictions from Murdock's hypothesis.

The first point to notice about the numbers in Table IV is that a symmetric SP curve is predicted (see the last column), with the fewest errors on the end stimuli and the most errors on the middle stimulus. Second, by reading across any row we obtain an impression of the source of errors. In particular, the probability of an error of *n* steps remoteness should decrease with *n*, in a regular gradient of remote intrusions, as shown in Fig. 5.

The next inquiry is whether and how well this generalization model can mimic the PD predictions of Murdock's earlier measure. The PD predictions for the generalization model (G) and for Murdock's measure (M) were calculated for sets of 4, 5, 6, or 7 equally log-spaced stimuli. These calculated values are displayed for comparison in Table V. The most remarkable feature of Table V is the closeness of the G and M measures in every case. The largest absolute discrepancy is around .015, which is too small a difference to detect, given the sampling error of the usual points in an SP curve. We may therefore conclude that the stimulus-generalization theory mimics Murdock's PD measures for stimuli equally spaced on a logarithmic scale. For that reason, it will fit all those

TABLE V

Comparison of Generalization (G) and Murdock (M) Theories on PD Measures for Training on Sets of 4, 5, 6, or 7 Stimuli Equally Spaced on a Logarithmic Scale

	Number of pairs: 4		5		6		7	
Stimulus	G	M	G	M	G	M	G	M
S_1	.285	.300	.240	.250	.207	.213	.182	.187
S_2	.214	.200	.180	.175	.155	.159	.136	.143
S_3	.214	.200	.160	.150	.138	.127	.121	.116
S_4	.285	.300	.180	.175	.138	.127	.121	.107
S_5			.240	.250	.155	.159	.121	.116
S_6					.207	.213	.136	.143
S_7							.182	.187

PD curves based on equally log-spaced stimuli which Murdock fit and offered as evidence for his theory.

The generalization theory also handles the effects of anchor stimuli if it is assumed that certain neutral or "inhibitory" responses come to be associated to the anchor stimuli, and that these inhibitory tendencies generalize to neighboring stimuli just as do active response tendencies. The effect of low anchors (S_1, S_2) on the response tendencies to the active stimuli S_3-S_7 is illustrated in Table VI, where R_1 and R_2 denote "inhibitory responses" in this context. These inhibitory responses may occur to stimuli S_3-S_7, probably causing omissions and thus lowering the probabilities of correct responses to

TABLE VI

Generalization Theory Applied to the Low-Anchor Condition[a]

Stimuli	R_1	R_2	R_3	R_4	R_5	R_6	R_7	Correct/sum	PD
S_1	3	2	1						
S_2	2	3	2	1					
S_3	1	2	3	2	1			3/9	.178
S_4		1	2	3	2	1		3/9	.178
S_5			1	2	3	2	1	3/9	.178
S_6				1	2	3	2	3/8	.200
S_7					1	2	3	3/6	.267

[a] Here, R_1 and R_2 denote "inhibitory responses," whereas R_3–R_7 are relevant experimental responses.

stimuli near the low end of the scale. The PD measure shown in the last column of Table VI may be compared to the PD numbers in Table V (for 5 stimuli with no anchors) and to Murdock's prediction for the low-anchor condition in Table III (also in Fig. 3b). The predictions for the high-anchors condition would be symmetrical to those for the low-anchors condition. The main point to note is that the expected PD gradient for the low-anchor condition for the generalization theory is virtually indistinguishable from that predicted by Murdock's hypothesis. So, insofar as Murdock's hypothesis fits the gradients from the anchor study, to that extent the generalization theory is also supported. It should be noted that the generalization theory could accommodate the anchor results in this instance only by assuming that "inhibitory" or neutral responses become conditioned to the anchor stimuli. It does not suffice to merely assume that anchors alter the AL in this experiment, since that factor consistently cancels out of the relative generalization gradients.

This is all I wish to say comparing the generalization model to Murdock's PD hypothesis. I have emphasized the similarities of their predictions of SP curves for stimuli that are equally spaced on a logarithmic scale. The two theories could probably be differentiated in a number of ways. For example, the generalization theory explains the deleterious effects on performance of subliminal spacing of stimuli, which was a criticism of Murdock's hypothesis raised earlier. Also, the theories may differ in their predictions of relative accuracy of performance with unequally spaced stimuli, but it would appear that the relevant differentiating experiments have not been done. In any comparative tests of the competing theories, it must be kept in mind that the height and slope of the generalization gradient are parameters to be selected in the one theory. My selection of a height of 3 and slope of $1/3$ was purely fortuitous and casual, and no special significance is to be read into this choice of parameters.[3]

Conceptual Dimensions

Both Murdock's theory and the generalization theory outlined above are incomplete because they do not give adequate recognition to several "cognitive" aspects of learning responses to linear orders. First, some evidence will be reviewed showing that the SPE occurs with words located symbolically along a

[3] With the learning rules as stated (that is, linear increments in the advantage of the correct response), Luce's response rule will not predict eventual perfect learning of responses to a small set of stimuli. An alternative response rule that predicts nearly 100% correct responding (for a small stimulus set) is Hull's (1943), where the habit strengths of different responses to each stimulus are normal random variables, and the choice is to that response with the highest momentary strength. This rule is quite similar to Luce's axiom (see Suppes & Zinnes, 1963) except at the extreme values which are of concern in this footnote.

semantic or conceptual continuum. Second, other evidence will be reviewed showing transfer of arbitrary "response scales" both within and between sensory dimensions in such a manner as to force one to a belief in an abstract, conceptual linear ordering of cognitive elements which has multiple sensory inputs. These two points, along with an important counterargument to the first, will be taken up in turn.

Symbolically Ordered Stimuli

When stimuli vary in physical characteristics like brightness or heaviness or linear extent, there would appear to be a physiological effect of such stimuli which varies monotonically as the stimulus varies. That is, one can practically see, in the functioning of the peripheral nervous system, the grounds for assuming distinct physiological "dimensions" in these cases. But SPEs appear just as strongly with derived dimensions or with symbolic, "conceptual" dimensions where there is no possibility of an elementary correspondence to a physical stimulus.

One example of a "derived" dimension is elapsed time, a dimension for which there is no physical stimulus. When PA responses are learned to a series of time intervals used as separate stimuli, SPEs arise, as Murdock (1960) has shown. The most well-known SPE occurs, of course, when the subjects learn to produce a temporal series of items, by the method of rote serial anticipation. Of the multiple cues possible for the responses in rote serial learning, a plausible one is the time interval elapsed since the beginning of the list; if the time per item is t, then an elapsed-time interval of nt would be the cue for the $(n+1)$th response item. Murdock's data implies that the learning of associations to a randomly presented set of regular time intervals would itself produce an SPE.

Another plausible cue for the response in rote serial learning is the *number of events* elapsed since the start of the trial, so that something like a subjective count of items might be kept by the subject. Associating responses to variations in number of events *qua* stimuli would also probably produce our ubiquitous SP curve. With a constant rate of item presentations, elapsed time and number of events are confounded, but they could easily be unconfounded in special experiments. The point of these remarks is that elapsed-time intervals and number of events elapsed are cues implicit in the rote serial-learning procedure and that these acting alone produce an SPE in PA procedures; therefore, they may be considered as likely contributors to the total SPE observed in the rote serial procedure.

Serial-position effects with symbolic dimensions have been found for numerals and for conceptual (semantic) dimensions. Examples of the first type are experiments by Ebenholtz (1966), who used the first 8–10 integers as stimuli to be associated to 10 unrelated nouns as response terms (in his Experiment 1) or to 8 nonsense-syllable responses (in his Experiment 2), and Pollio and Draper (1966), who used the integers 1–5 as responses to be associated with nonsense-syllable stimuli. The SP curves are shown in Fig. 6. The Ebenholtz data (Fig. 6a) are relative error percentages with digit stimuli; the Pollio and Draper data (Fig. 6b) are mean trials to criterion for pairs with digit responses of one to five, but the relative percentage error curve should have a similar shape. It is clear that high and low numbers are learned the fastest whether they serve as stimuli or as responses for the other element of the pair. This is not simply a

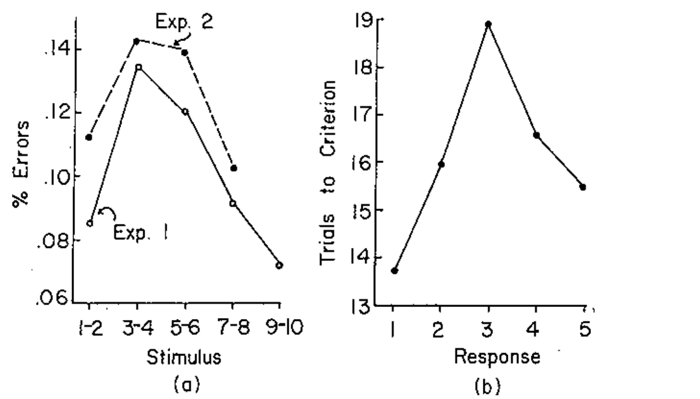

Fig. 6. Relative-error gradients to digits used (*a*) as stimuli (results taken from Ebenholtz, 1966); or (*b*) as responses in PA learning (results taken from Pollio & Draper, 1966).

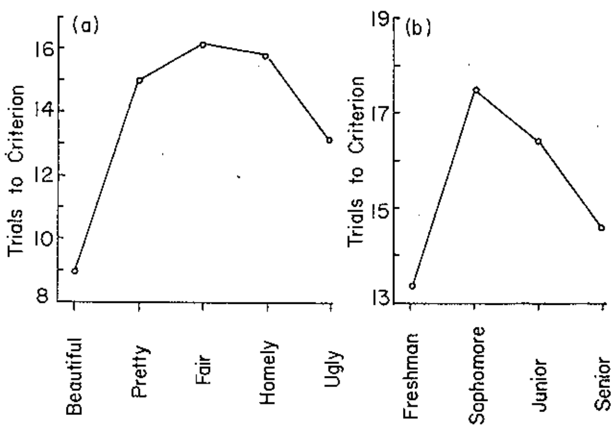

Fig. 7. Trials to criterion for pairs with the response terms arrayed on a semantic continuum: (*a*) data from Pollio and Deitchman (1964) (*b*) data from DeSoto and Bosley (1962).

function of the "meaningfulness" or association value of these digits (as scaled by Battig & Spera, 1962) since, for example, 3, 4, 5, and 6 have higher-rated association values than do 7 and 8, but the former are learned slower (as Fig. 6a shows).

Serial-position curves to words arrayed along an implicit semantic dimension have been studied by DeSoto and Bosley (1962) and by Pollio and Deitchman (1964). Examples of the obtained SP curves are shown in Fig. 7. Pollio and Deitchman (Fig. 7a) had their subjects learn to associate an adjective response (beautiful, pretty, fair, homely, ugly) to 15 girls' names, each adjective being assigned as a response to three names. DeSoto and Bosley (Fig. 7b) had their *S*s learn to associate a school class (freshman, sophomore, junior, senior) to 16 male names, four in each class. The significant fact in both cases is that response words at the extreme ends of the semantic scale are more easily learned. This suggests that the internal representation of the meanings of the words are ordered in proximity as though they were points along a one-dimensional continuum.

Induced Linear Orderings

In the experiments previously reviewed, it could be said that the stimuli existed in a linear order for the subject before he came into the experiment, this prior ordering revealing itself in SPEs when the person learns to associate any responses to stimuli lying along that dimension. However, a simple principle exists for creating linear order among any new set of stimuli which previously were unordered. To induce a linear ordering onto a new set of unrelated elements, one simply associates them one to one with a second set of elements that were previously linearly ordered, perhaps on a "primary sensory" basis. To be more explicit, the principle is this: If elements $a_1, a_2, \ldots, a_n$ are linearly ordered in some primitive sense, and if we take any unrelated elements whatever, $b_1, b_2, \ldots, b_n$ and associate the two series in pairs (a_i, b_i), then the set $b_1, b_2, \ldots, b_n$ will acquire a derivative linear ordering. The critical implication of this derivative linear ordering is that stimulus generalization among the b_is will now vary (in more or less degree) according to the proximities of the associated *a*s. That is, the proximity structure of the *a* set has been induced onto the *b* set, which were formerly unrelated elements.

A corollary of these proximity relations is that an SP curve will arise if we now try to use elements from the *b* set as stimuli for PA learning. This is obviously an important principle, because it provides a possible way to think about how symbolic or conceptual linear orderings might have arisen from a history of analogical association to perceptually primitive orderings.

An experiment illustrating this principle was done by Phillips (1958), when she had her subjects associate six Turkish words (effectively, nonsense) with six shades of gray varying from black to white. Thereafter, one of the Turkish words was paired with electric shock so that a GSR was conditioned to that word. Finally, the other Turkish words were presented in order to observe the amount of GSR generalizing to them. The findings were that the GSR generalization between two Turkish words was higher when the gray patches associated to those two words were closer together. That is, the proximity metric of the gray patches had been induced onto the previously unordered set of nonsense words.

An experiment on an induced SPE was done successfully by Ebenholtz (1966). His subjects first learned a rote serial list of ten unrelated nonsense

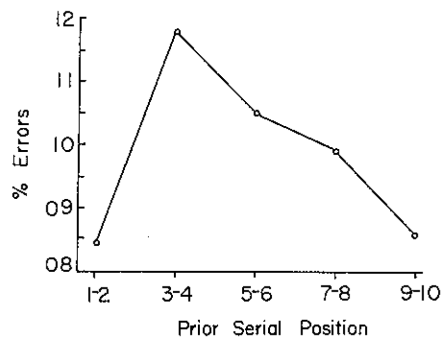

Fig. 8. Relative-error gradient during PA learning with syllable stimuli that formerly had occupied serial positions in a serial list learned before the PA task. (From Ebenholtz, S. M. Serial position effects of ordered stimulus dimensions in paired-associate learning. *Journal of Experimental Psychology*, 71, 1966, 132-137. Copyright 1966 by the American Psychological Association, and reproduced by permission.)

syllables. After the serial list was well learned, the ten syllables were then paired arbitrarily with ten unrelated nouns, and the person learned the list of ten syllable–noun pairs by the usual PA procedure. The inquiry centered upon the difficulty in learning a particular PA depending upon the SP that the stimulus syllable had occupied in the prior serial list. The results in this respect are shown in Fig. 8 in terms of the percentage error gradient, and again the SPE is evident. Syllables formerly at the initial and final ends of the serial list now serve as the more distinctive stimuli so that fewest PA errors are committed on them. This is an example of an SPE produced by an induced linear ordering: the SPs are the primary stimuli $a_1, a_2, \ldots, a_{10}$, while the nonsense syllables learned in serial order served as the secondary stimuli $b_1, b_2, \ldots, b_{10}$. The induced ordering of the b_i is then inferred by the SPE when they are later associated to the set of unrelated nouns $c_1, c_2, \ldots, c_{10}$.

One might suppose that most cases of symbolic dimensions (of numbers or word meanings) have the behavioral characteristics of linear orderings because

they have acquired their connotative meanings by association to concrete analogs of physical continua. For example, our vocabulary for locating events in egocentric time is thoroughly infused with spatial analogs—the future lies *ahead*, and one's past is *behind* one. As a second example, most people's subjective interpretation of numbers is as analog magnitudes—as heights, sizes, intensities, or distances along a line (the "real line"). These properties characterize the common, concrete exemplars used in teaching numerical concepts to children. We even use size adjectives in talking about "big" and "small" numbers, and speak of one number being "far" from another. Restle (1970) reported that the snap judgments which people make about numbers appear to be based on their initial fast interpretation in terms of analog magnitudes. Also, something like this analog correlation is relied upon in the psychophysical method of "magnitude estimation" in which the subject assigns a numeral representing the subjective intensity or quantity of some attribute of a stimulus.

One may recognize this principle of derivative linear orders, and its probable implication in the establishment of some conceptual dimensions, without believing this to be the complete story for all conceptual dimensions. Although the corresponding physical variables are clear for semantic continua like heavy to light, black to white, and cold to hot, it is not at all obvious what are the physical, analogical dimensions for semantic continua such as beautiful to ugly, love to hate, freedom to oppression, or many other contrastive oppositions in our language.

Perhaps the insufficiencies of the "derivative-ordering" principle will become apparent if we examine it more closely. The principle itself is a "low-level descriptive" law; it describes one procedure for inducing a given proximity structure upon a set of unrelated elements, namely, the association of elements of one set with elements of the other, in pairs (a_i, b_i). The principle states nothing about underlying mechanisms or representations which mediate these effects.

In this respect, there would appear to be two general classes of theories regarding the induced linear ordering; the two views may be illustrated by their differing interpretations of Phillips' experiment on Turkish words and gray squares. According to one interpretation, the Turkish word becomes directly conditioned to a "sensory response" or mental image corresponding to a fractional part of the perception of the associated gray patch. Staats (1968), for example, explicitly favors this alternative, with the presumption that such "sensory responses" are the intermediaries causing the Turkish words and gray squares to have the same proximity metric. An alternative interpretation is that the transfer of the metric to the Turkish word is mediated not by a mental image of the specific referent (the gray patch), but rather by a more abstract conceptual representative referring to the *relative position* of each stimulus in the range presented or, in Phillips' experiment, perhaps to just a verbal description or name for the physical stimulus. In the following I will develop this notion of an abstract linear order more fully and present some facts that I think are relevant.

A Cognitive Structure for Linear Orderings

The viewpoint to be espoused is that people either have innately or develop (learn) the general concept of a linear ordering of stimuli, with its primitive relations of comparison, progression, and betweenness with respect to a focused attribute. In early development, the concept is exemplified by physical attributes like brightness, size, distance, heaviness, and heat, but the concept becomes extended to include the more symbolic, semantic dimensions. In other words, the generalized internal representation of linear ordering is a cognitive structure or scaffold that is applied to the arrangement of a number of stimulus domains. These statements give explicit recognition to the communalities among various linear orderings of stimuli; it is these communalities which permit a type of interdimensional transfer which "mental images of specific referents" could never explain.

The hypothesis is that when response R_i is reinforced to stimulus S_i, it becomes associated not only to the coded version $C_i = \log(S_i/AL)$, but also to an implicit conceptual representative corresponding to C_i's relative position in the series, such as "first," "middle," or "last." The latter are relational concepts exemplified in a number of physical and semantic dimensions, and they are quite independent of the specific ranges or even the specific dimensions involved. That is, "position within a linearly ordered series" is a transsituational concept. It refers to a set of *intradimensional* invariants—the notions of "middle and end" stimuli always apply, regardless of the specific range of values involved. It also refers to a set of *cross-dimensional* invariants—the same notions apply regardless of the specific dimensions being discussed.

I will cite evidence of two kinds to support the view that responses become associated to these conceptual, relational stimulus codes. One kind concerns positive transfer of linear-ordered response scales between different ranges of values along one dimension; this is the traditional problem of "relational transposition" and it exemplifies my point about intradimensional invariance. The second kind of evidence involves transfer of linear orderings between two different dimensions, and this is addressed to my point about cross-dimensional invariance.

Transposition

Transposition studies examine transfer of relational responding to sets of stimuli *within* one and the same dimension. The usual transposition experiments

(see Reese, 1968) have been done with animals or children, typically with simultaneous presentation of two or three stimuli, the subject being rewarded for selecting a specific one out of the array. The following experiment[4] violates all these procedural strictures. We used adults learning PAs; in technical jargon, they were learning multiple "successive discriminations" with verbal responses. The change in procedure itself is a novelty in the literature on transposition.

The design of this experiment can be seen by referring to Table VII. Paired-associate learning was carried out originally for all subjects (college students) with the six low-valued stimuli paired with unrelated nonsense-syllable responses. (In one experiment, the stimuli were nine horizontal lines of varying lengths; in the other, they were eight squares varying redundantly in brightness and size.) After 12 initial trials, the subject was told that he would be learning another set of pairs involving the same responses but a new range of stimuli, and he was exposed to one presentation cycle of the six longest stimuli, S_4–S_9, to familiarize him with the new range. (No response information was given at this time.) On the next cycle, the subject was permitted to respond if he wished, and he was informed of the new correct responses. Twelve transfer training trials then followed. Three conditions were formed that differed according to the new pairings between the old responses ($R_1 - R_6$) and the new stimuli (S_4–S_9).

TABLE VII

Stimulus-Response Assignments during Original Learning and Transfer for the Relational, Absolute, or Random Conditions[a]

	Stimulus								
Conditon:	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
Original learning:	R_1	R_2	R_3	R_4	R_5	R_6			
Relational				R_1	R_2	R_3	R_4	R_5	R_6
Absolute				R_4	R_5	R_6	R_2	R_1	R_3
Random				R_3	R_6	R_1	R_2	R_5	R_4

[a] Here, S_1-S_9 denote stimuli along a continuum; R_1-R_6 denote nonsense-syllable responses; entries denote S-R assignments; blanks denote nonpresented stimuli.

In the relational condition, the entire "response scale" was shifted up the stimulus scale, so that absolute stimuli S_4–S_6 had their responses altered. In the absolute condition, these three stimuli retained their former responses, while the three new values S_7–S_9 had responses R_1–R_3 assigned randomly to them. In the

[4] The experiment was conceived by David Tieman and the author, and was conducted at Stanford University by Lynn Carter and James Antognini.

random condition, responses were assigned to the six stimuli so that neither a relational nor an absolute basis for transfer was available.[5]

Comparison of the relational versus absolute groups in transfer is as follows: if *S*s have associated the responses R_1-R_6 to conceptual stimuli reflecting *relational* codes (like "first, middle, last"), then transfer to the relational response assignments should be relatively easy, whereas the absolute response assignments, which destroy the former relational order of the responses, should appear as difficult to learn in transfer as the random assignments. On the other hand, if stimuli are mainly reacted to on an *absolute* basis, the positive transfer should appear for the absolute assignments [which preserve three out of the six prior stimulus–response (S–R) assignments] but not for the relational assignments (which alter all specific S–R assignments). On this basis, one would predict that the relational and random groups should show similar poor transfer performance with the absolute condition showing best transfer.

TABLE VIII

Median Total Errors for Three Transfer Conditions Using Line-Length Stimuli or Squares Varying in Size and Brightness

Experiment	Relational	Absolute	Random
Line lengths	7.0	14.0	14.5
Gray squares	8.0	13.5	15.5

The results of the two replications of the experiment are shown in Table VIII in terms of the median total errors (summing over all six pairs) during the transfer trials for the three conditions. Both experiments show the same outcome pattern, namely, substantial positive transfer for learning the relational assignments, but virtually no transfer for the absolute assignments compared to the random control condition. The indicated differences are statistically significant at a suitable level in each experiment.

It appears as though the subjects were encoding the stimuli relationally—that is, relative to the end points and range in use—and associating syllables to these coded cues. The chief question to be raised with a strict relational interpretation of these data is why there were any errors at all in the relational condition. Various reasons might be given for the failure of perfect transfer, but most

[5] For convenience, I have described the design as starting everyone at the low end of the stimulus scale and shifting to the high end. In fact, only half the subjects were tested in this way, while the remainder transferred from the high to the low end of the scale. This counterbalancing in direction of shift is of no theoretical importance; the patterns of results were the same shifting in either direction, so this factor will not be mentioned further.

would require reference to specific or absolute characterizations of the stimuli.

The results in Table VIII exemplify *intra*dimensional transfer of a response scale. Now, that is the kind of transfer that is handled well by AL theory. In fact, the generalization theory outlined earlier, with the assumption that responses become associated to the code $C_i = \log(S_i/AL)$, explains the transposition results very neatly (see also James, 1953). During initial training with the set S_1-S_6, the AL is at 3.5 (recall, it is a log scale); therefore $C_1 = -2.5$, and this C value was formerly associated to R_1, which is currently the correct response. Transfer is less than perfect, according to this theory, because the AL does not shift immediately to the center of the new stimulus range; rather, it lags because of "residuals" from the training series. This hypothetical story implies particular details regarding the shift of error profiles for the different groups over the course of the transfer trials as the AL adjusts to the new stimulus series, but those details will not be pursued here.

I began this section by offering transposition as evidence for abstract conceptual elements arrayed in linear orders that mediate transfer. It does constitute such evidence; unfortunately, however, transposition of the type shown here is also explicable in terms of AL coding of stimuli. Sò the data are compatible with both hypotheses. I believe that both hypotheses are true, that both factors operate in a complementary fashion to promote transfer in this particular case, and that more ingenious experiments are required to unconfound them. A variety of intradimensional shift experiments might be tried. Discriminating ones would be those which study transfer when (*a*) the S–R assignments during transfer are simply the reverse ordering of those used in training, with or without a new stimulus set, or (*b*) the transfer stimulus set is an expansion (or contraction) of the stimulus range, centered around the training AL, with responses assigned according to their original relative scale position. In such cases, the abstract-mediation theory expects positive transfer, whereas a strict AL theory would predict little if any transfer. Until such experiments are done, the best evidence for the abstract-mediation theory comes from the studies on *cross*-dimensional transfer and intermodal interactions.

Cross-Dimensional Transfer of Response Scales

The most ingenious experiments of this type have been done by Ebenholtz. In one of these (Ebenholtz, 1965), students were first trained to associate eight nonsense syllables to eight lines varying in length from shortest to longest. They were then transferred to learning these same eight syllables as responses assigned to eight gray squares that varied in brightness. For one group of subjects, the ordering of the eight syllable responses assigned to the eight gray squares was the same as the ordering of responses that had been assigned to the short-to-long line lengths. That is, the response R_1 that had formerly been assigned to the shortest line was now assigned to the lightest gray; succeeding responses were also ordered, the response assigned to the darkest gray being that formerly learned to the longest line. The response assignments may be characterized by saying that the indices or ordinal positions of the two sets of stimuli assigned successively to the set of responses were perfectly correlated. The subjects learning these correlated response assignments were compared to others learning a random reshuffling of the eight syllable responses over the eight new gray-patch stimuli. Half of the subjects in the experiment were transferred from lines to grays, and half in the opposite direction.

The simple result was that the experimental subjects learned the correlated transfer pairs very much faster than the control subjects learned the random transfer pairs. This positive transfer would appear to be due to the learning of a "transdimensional" ordering of the responses—a response scale—which transcends specific sensory dimensions. Thus, by pairing the syllable MIB with the shortest line, it becomes partially associated to "least (or first) element in the series"; this is also the position occupied by the lightest gray patch, which is also paired with MIB, thus causing positive transfer. In this case the positive transfer cannot be accounted for by generalization of responses around specific dimensionalized stimuli. Nor does it seem helpful to assume that MIB comes to evoke a mental image of a specific line length, because that is not in the same mode as the gray squares to which transfer is being tested. In short, the results call for a set of *dimensionless* mediating elements such as "first, middle, last" in the series.[6]

Further research relevant to transdimensional transfer will be reviewed very briefly. In another study Ebenholtz (1963) showed high positive transfer, in either direction, between learning the *spatial positions* of nonsense syllables (top to bottom in a column) and learning their temporal position within a serial list (from beginning to end of the temporal list). Spatial and temporal locations are particularly susceptible to cross-talk or interaction with one another. In some of our experiments we have investigated these interactions by having the subjects learn to order items according to one stipulated dimension (say, their temporal order), while coincidentally the items also vary in a second dimension (their spatial position when presented).

In one of our designs, the subject learns a temporal series of nine letters, such as M Z L B X T Q K R, by the method of rote serial anticipation. But when each

[6] The coded stimuli $C_i = \log S_i/AL$ are dimensionless numbers (the dimensions cancel in the ratio), but will not fill the desired bill of particulars. First, the C_i have a particular directional orientation, whereas the results probably do not depend on whether, say, the shortest line is mapped onto the lightest gray or the darkest gray. Second, the C_i vary greatly according to the spacing of the two sets of stimuli, whereas the transfer of response scales would probably be evident for any discriminable spacing of the stimuli. However, these are empirical matters which cannot be prejudged with any certainty.

letter is presented visually, it occurs coincidentally in one of nine circles laid out in a row, so that that letter event has both temporal and spatial coordinates or index numbers. In one condition, the time and space indices are perfectly correlated—the first temporal element occurs in the spatial location farthest left, and succeeding elements in the time series march successively across the spatial array. In another condition, the two indices are uncorrelated, so that the first letter in the temporal series may appear in the fifth spatial location, the second in time may appear in the third location, and so on, in either constant or variable locations over trials. As the important result, relative to a "one-location" control group (the usual memory-drum slot), the temporal series is learned more rapidly when space is correlated with time, and more slowly when the two dimensions are uncorrelated. A similar difference between correlated and uncorrelated presentation occurs in the converse experiment, when the subjects are supposed to recall elements by their spatial location regardless of their temporal position in a presentation series. Many different conditions have been studied, checking out alternative explanations; a plausible explanation, however, is that of redundancy versus conflict of dual codes for the ordinal location of the elements. If time and space mapped onto totally distinct continua, and the person could attend selectively to either dimension according to how he is to order his recall, then there would be no confusion or interference in the uncorrelated presentation condition. However, to the extent that temporal position and spatial location map onto the same conceptual elements for "position in series," then learning the position in which an item is to be recalled is helped by time–space coordinates which coincide, but is hindered when they conflict. So, in this case, the cross-talk or interaction between two dimensions is explained by the idea that both project partially to a set of relative-position concepts which mediate the interaction.

Thus far I have reviewed evidence from learning and transfer experiments to support the idea of abstract concepts of relative position that are "aroused" by multiple sensory and semantic continua. But the transfer data may be indirect, and perhaps the most convincing evidence is the simplest, namely, cross-modality matching in psychophysical judgments. The subject can set a tone subjectively as loud as a light appears bright, or as loud as a weight feels heavy, and so on. Given any two linear continua, I suppose that the subjects could perform such cross-modality matchings reliably, especially if they are given a standard for each dimension ("call this loudness 10, and this brightness 10"). I am not concerned here with the form of the psychophysical relationship; for example, if Stevens (1957) is right, the logs of the matching physical variables will plot as straight lines. I am instead interested in our interpretation of how the performance can be done at all—how is it possible to compare the loudness of a tone to the brightness of a light? The only reasonable way anyone has imagined to do this is to transform each stimulus into a *dimensionless* scale, for example, a number expressing the ratio of the subjective magnitude of the current test stimulus to one's memory of the standard referent (the loudness called 10). This ratio is computed (or "immediately sensed") and the number assigned for both stimuli, and when the two numbers are equal, the person reports a "cross-modal match." So far as my argument goes, the important point is the conversion to a common "numerical magnitude" scale, since that qualifies as the abstract concept of linear order to which I have referred. The claim I am making for all such cases is essentially similar to that made by some linguists (see Shank, 1969), that an interlingual base of semantic and syntactic (universal) concepts exists into which all languages can be translated, and that it serves as the cognitive base for doing translations, say from English text into German or Swahili or Mandarin.

Verbal Mediation?

To reiterate, the transfer results require the notion of conceptual or cognitive elements that are linearly ordered, implicit elements or concepts that are transsituational or transdimensional. For human adults, the most likely candidates are words or "verbal responses." These could be verbal labels for stimuli along the continua, "shortest" to "medium" to "longest," or implicit, numerical category scales, "first, second, . . . , *n*th, . . . , last" in the series exposed to the subject. These are relational concepts, of course. According to this view, transposition within a dimension or between two different dimensions would be interpreted as being verbally mediated by implicit numerical labels that the subject applies to the training and testing stimuli. Were this "verbal mediation" theory the correct interpretation, then animals or preverbal children should not demonstrate transdimensional transfer of the sort Ebenholtz found with adult humans. On the other hand, it is possible to argue, somewhat as Piaget might, that linearly ordered stimulus domains exist in the world, that children acquire general cognitive mappings or representations of such linear orderings, and that perceptual notions of succession, progression, interposition, and ordering come into play long before the child learns the words required in counting or in describing linear order. In this interpretation, because linearly ordered stimulus domains exist in the world, the basic perceptual–conceptual scaffold is acquired preverbally; it is only later that numerical labels become attached to these perceptual–conceptual elements. I wish neither to elaborate upon nor to argue for either of these interpretations at the moment. But they do provide some of the larger context in which the present research may be placed.

References

Battig, W. F., & Spera, A. J. Rated association values of numbers from 0-100. *Journal of Verbal Learning and Verbal Behavior,* 1962, **1**, 200–202.

Capehart, J., Tempone, V. J., & Hebert, J. A theory of stimulus equivalence. *Psychological Review,* 1969, **76**, 405–418.

DeSoto, C. B., & Bosley, J. G. The cognitive structure of a social structure. *Journal of Abnormal and Social Psychology,* 1962, **64**, 303–307.

Ebenholtz, S. M. Position mediated transfer between serial learning and a spatial discrimination task. *Journal of Experimental Psychology,* 1963, **65**, 603–608.

Ebenholtz, S. M. Positional cues as mediators in discrimination learning. *Journal of Experimental Psychology,* 1965, **70**, 176–181.

Ebenholtz, S. M. Serial-position effects of ordered stimulus dimensions in paired-associate learning. *Journal of Experimental Psychology,* 1966, **71**, 132–137.

Helson, H. *Adaptation-level theory: An experimental and systematic approach to behavior.* New York: Harper, 1964.

Hull, C. L. *Principles of behavior.* New York: Appleton, 1943.

James, H. An application of Helson's theory of adaptation level to the problem of transposition. *Psychological Review,* 1953, **60**, 345–352.

Jensen, A. R. Temporal and spatial effects of serial position. *American Journal of Psychology,* 1962, **75**, 390–400.

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

McCrary, J. W., Jr., & Hunter, W. S. Serial position curves in verbal learning. *Science,* 1953, **117**, 131–134.

Murdock, B. B., Jr. The distinctiveness of stimuli. *Psychological Review,* 1960, **67**, 16–31.

Phillips, L. W. Mediated verbal similarity as a determinant of the generalization of a conditioned GSR. *Journal of Experimental Psychology,* 1958, **55**, 56–62.

Pollio, H. R., & Deitchman, R. The activational characteristics of a serial cognitive structure having oppositional end points. Mimeographed manuscript, University of Tennessee, 1964.

Pollio, H. R., & Draper, D. O. The effect of a serial structure on paired-associate learning. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 301–308.

Reese, H. W. *The perception of stimulus relations.* New York: Academic Press, 1968.

Restle, F. Speed of adding and comparing numbers. *Journal of Experimental Psychology,* 1970, **83**, 274–278.

Shank, R. C. A conceptual dependency representation for a computer-oriented semantics. Technical Report CS 130, Computer Science Department, Stanford University, 1969.

Staats, A. W. *Learning, language, and cognition.* New York: Holt, 1968.

Stevens, S. S. On the psychophysical law. *Psychological Review,* 1957, **64**, 153–181.

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