Subproblem analysis of discrimination shift learning*

DOUGLAS L. MEDIN
The Rockefeller University, New York, New York 10021

Although attention models assume that the two subproblems of a discrimination shift are treated as a single problem, learning of a nonreversal shift can be broken down into performance on the subproblem that changes reward conditions and the one that does not. A general analysis of performance on these subproblems, appropriate to a broad range of attention models, is developed in this paper. This analysis leads to a rejection of attention models for some S populations, but shows that attention models may have been rejected prematurely for other populations of Ss.

Since the work of the Kendlers a decade ago (Kendler & Kendler, 1962), investigators have compared the performance of various S populations on reversal and nonreversal shifts in order to assess the applicability of a two-stage model (usually described in terms of attention or mediation) for the behavior of Ss. The initial working assumption was that faster reversal than nonreversal shift performance implicated the presence of mediation or attention; slower reversal than nonreversal shift denied it. However, the latter implication does not hold, since the positive transfer on reversal shifts from attending to the relevant dimension may be offset by the negative transfer from specific cue-outcome relationships (MacKintosh, 1965).

More recently, Tighe, Glick, and Cole (1971) developed a subproblem analysis of nonreversal shift performance which may provide a more direct measure of the learning process operating during discrimination transfer problems. The particular variant of the reversal-nonreversal paradigm appropriate to this analysis is shown in Table 1. In the table, color is the relevant dimension while form is irrelevant in original learning. For the reversal shift, color continues to be the relevant dimension, and the S must learn to choose the other value (C_2) on that dimension. On the nonreversal shift, color is no longer relevant and form becomes the relevant dimension.

Tighe et al (1971) point out that this description implicitly assumes that Ss solve the original problem on the basis of dimensions and that the two different pairs of stimuli are treated as a single problem. One test of this assumption can be seen for the nonreversal shift problem. To execute a nonreversal shift, the reward conditions on one of the stimulus pairs changes, but for the other subproblem, the reward conditions are unchanged. By following the course of performance on the unchanged subproblem, one can obtain evidence as to whether Ss treat the two subproblems as a single discrimination problem. Tighe et al (1971) observed that performance on the unchanged subproblem was strikingly better than it was on the changed subproblem for 4-year-old children; however, there was little if any difference in performance on the two subproblems for 10-year-olds. On the basis of their analysis, the authors concluded that the 4-year-old Ss "tended as a group to learn the stimulus pairs as independent subproblems [p. 160]." Other experiments with various S populations (Graf & Tighe, 1971; Tighe & Tighe, 1971; Tighe & Frey, 1972) also appear to provide evidence for independent subproblem learning.

The term, independent subproblem learning, is open to at least two interpretations. It may be used loosely to refer to differences in performance on changed and unchanged subproblems, or it may refer more precisely to the absence of any mediating process to link the two subproblems. In this paper, we attempt to confine ourselves to this latter interpretation.

Before concluding that the subproblems were treated as independent, we must first provide evidence that attention models cannot handle the present data if they imply that the subproblems are treated as a single discrimination. Reference to the nonreversal shift shown in Table 1 will facilitate discussion.

From the point of view of attention theories, errors on the unchanged subproblem can occur when: (1) the S attends to the formerly relevant dimension, but the response strength to the formerly correct value is lowered to the extent that the S chooses the other value (C_2) on that dimension, (2) the S attends to the new relevant dimension but has not yet learned to choose the correct value (F_1) on this dimension, and (3) the S attends to some other, irrelevant, dimension and guesses incorrectly. To execute a nonreversal shift from color to form, the S must learn to attend to form rather than color and then to choose the specific value, F_1. At the start, the S will have some tendency to attend to the formerly relevant dimension (in this case, color) and to choose the value C_1. As long as this strategy persists, a S

Table 1
Reversal and Nonreversal Shift Paradigms

Related Stimuli				
Original Problem		Shift Problem		
		C_1F_1 (−)	C_2F_2 (+)	
	Reversal Shift	C_1F_2 (−)	C_2F_1 (+)	
C_1F_1 (+)	C_2F_2 (−)			
C_1F_2 (+)	C_2F_1 (−)			
		C_1F_1 (+)	C_2F_2 (−)	Unchanged Subproblem
	Nonreversal Shift	C_1F_2 (−)	C_2F_1 (+)	Changed Subproblem

*This research was supported by a grant (OEG 2 71 0532) from the Office of Education.

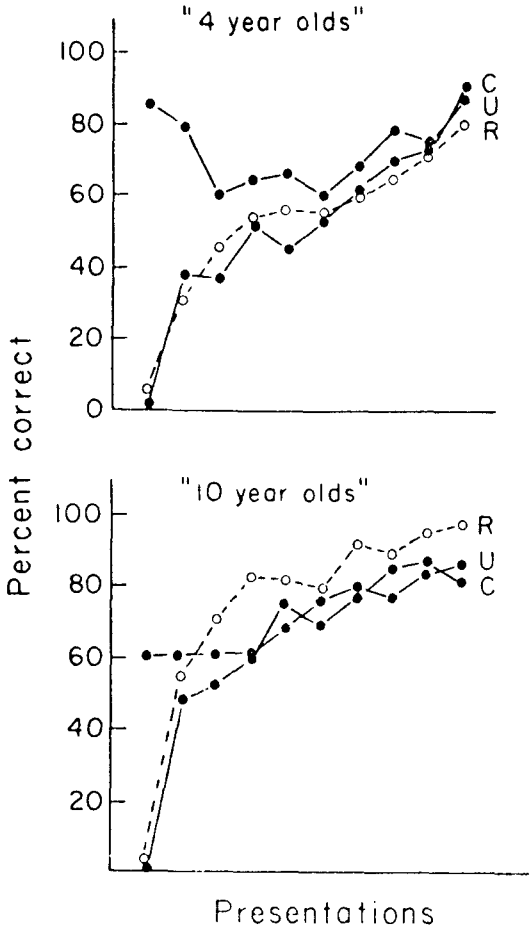

Fig. 1. Performance on the changed (C) and unchanged (U) subproblems of a nonreversal shift along with reversal shift (R) performance for 4-year-old and 10-year-old subjects.

will perform perfectly on the unchanged subproblem and produce errors on the changed subproblem. Therefore, attention models make the qualitatively correct prediction that performance will be better on the unchanged subproblem than on the changed subproblem. The particular rates at which the dimension-observing and cue-selection processes operated will determine the exact predictions of two process models concerning performance on changed and unchanged subproblems. Consider the data shown in Fig. 1. The top panel shows performance on the unchanged problem to be consistently better than on the changed problem; the bottom panel shows very little difference. The 4- and 10-year-old curves in Fig. 1 correspond roughly to data of 4- and 10-year-olds described by Tighe et al. Yet both sets of curves were generated by a dimensional attention model which assumes that the subproblems are responded to as a single problem (Zeaman & House, 1963). No attempt was made to fit the Tighe et al data precisely beyond describing the gross features of the curves.

One could argue that the differences between 4- and 10-year-olds reflect differences in amount of mediation, but this conclusion must be made more precise. According to the Zeaman-House model, mediation occurs on every trial, so it may not be meaningful to talk in terms of amount of mediation. However, since mediation is reflected in the rate change parameter for observing responses, in one sense "amount of mediation" could refer to: (1) how large this parameter is in an absolute sence, or (2) how large it is relative to the rate change parameter for specific responses. This distinction may have important implications, but it is outside the scope of the present paper to consider them. For the simulated data of Fig. 1, the dimensional rate change parameter was not varied. In the simulations, the rate change parameter for specific responses was set to be much smaller (.15 vs .40) for the 4-year-olds than for the 10-year-olds, but the rate change parameter for dimensional observing responses was the same for each group. These assumptions are consistent with a recent experiment by Dickerson, Novik, and Gould (1972), which appears to yield direct evidence that young children are relatively slow in changing instrumental choice responses. Performance on an unchanged subproblem relative to a changed subproblem does not necessarily reveal the presence or absence of independent subproblem learning.

On the other hand, attention theories would not be consistent with an outcome showing no errors on the unchanged subproblem. Graf and Tighe (1971) have observed errorless performance on unchanged problems in turtles, and this contradicts attention models which imply that the subproblems are treated as a single discrimination. However, the special case of errorless performance on the unchanged subproblem would likely tend to be rare, at least for human Ss.

In the remainder of the paper, a quite general test for subproblem independence is developed. First, we shall try to demonstrate that the analysis is sufficiently general to cover all currently proposed attention models, then develop the analysis, and finally apply it to some extant data. The test for nonindependence is quite weak in that it is not sensitive to many particular aspects of models. But this is a virtue, since, when nonindependence is indicated, the implications are very strong.

SCOPE OF THE ANALYSIS

In the simplest versions of attention theory (e.g., Sutherland, 1964; Zeaman & House, 1963), it is assumed that the S attends to and learns about values on only a single dimension on a particular trial. However, it is possible to construct models which assume that choices are generated by values on a single dimension but provide that a S may learn about more than one cue on a trial (e.g., Lovejoy, 1968). These models may be designated as having single-cue control and multiple-cue learning. The present analysis will apply directly to both models.

A little elaboration will be required to extend the analysis to multiple-cue control models allowing for multiple-cue learning (e.g., Sutherland & MacKintosh,

1971; Fisher & Zeaman, 1972). Sutherland and MacKintosh assume that performance is based on the values of the dimension of highest strength and other dimensions whose strengths are within some constant amount of the strongest dimension. The next paragraphs show that many multiple-cue control models can be represented as single-cue control models and suggest an alternative formulation of Sutherland and MacKintosh's assumption which will make multiple-cue control models amenable to the subproblem analysis.

Let D_i equal the probability that Dimension i alone is activated on a trial and $D_{i,k}$ equal the probability that both Dimensions i and k are activated. Further, let a_{ij} be the probability of Response j, given than Dimension i controls responding. Assuming, for the moment, just two dimensions ($D_1 + D_2 + D_{1,2} = 1$), the probability of Response j(P_j) is

$$P_j = D_1 a_{1j} + D_2 a_{2j} + D_{1,2}[R a_{1j} + (1 - R)a_{2j}], \quad (1)$$

where R is a probability which governs the combination rule when both dimensions are simultaneously activated. When R = 1/2, one obtains the averaging rule adopted by Sutherland and MacKintosh (1971). One can generate a maximizing rule by letting R = 0 or 1, depending upon whether a_{1j} is less than or greater than a_{2j}. Since the bracketed term in Eq. 1 is bounded by the $a_{1j} - a_{2j}$ interval, multiple-cue control may be represented by the following single-cue control equation:

$$P_j = D'_1 a_{1j} + (1 - D'_1)a_{2j}, \quad (2)$$

where D'_1 represents the probability that Dimension 1 effectively controls responding on a particular trial.

It is easy to show that Eq. 2 is equivalent to Eq. 1 whenever $D'_1 = D_1 + D_{1,2}R$. This result is restricted to multiple-cue control models for which Eq. 1 is applicable, but to my knowledge there are no multiple-cue control attention models whose assumptions are inconsistent with Eq. 1. In other words, one can represent multiple-cue control in a single-cue control framework.

With Eq. 2 in hand, one might represent multiple-cue control models by the relative strength of dimensions ($0 \leqslant D_i < 1$) raised to a power as:

$$D'_1 = \frac{{D_1}^n}{\sum_{i=1} {D_i}^n}, \quad (3)$$

where the power, n, takes the place of the constant in the Sutherland and MacKintosh formulation. When n = 0, all dimensions have equal weight, regardless of their strength; when n = 1, an averaging rule for strengths is implied; and for larger n, the strongest dimension will approach complete control of responding.

TEST FOR SUBPROBLEM INDEPENDENCE AND SOME DATA

The following analysis applies to attention models falling into the framework defined in the following minimal assumptions: (1) For dimensions about which a S learns, reward increases and nonreward decreases the probability that the dimension will effectively control responding; (2) the probability of responding to a particular value along a dimension (given that the dimension effectively controls responding) is increased by reward and decreased by nonreward; and (3) the two subproblems of discriminations (such as those described earlier) are treated as a single discrimination.

Referring again to the nonreversal shift involving related stimuli in Table 1, one can characterize a trial by the probabilities, C, F, and 1 – C – F, respectively, with which the formerly relevant, the now relevant, or some other dimension effectively controlled responding. In addition, one can describe the probability that the previously correct cue or the now correct cue was chosen (symbolized by c_1 and f_1, respectively), given that attention was directed toward the formerly relevant or now relevant dimension, respectively. Using these symbols and the assumption that subjects have a chance probability of being correct when they attend to neither color nor form, we can write equations for proportion correct on unchanged (P_u) and changed (P_c) subproblems.

Let $P^x_{u,n}$ denote the probability of a correct response on the unchanged pair by Subject x on Trial n, and similarly for the other quantities. Then

$$P^x_{u,n} = C^x_n c^x_{1,n} + F^x_n f^x_{1,n} + (1 - C^x_n - F^x_n)/2 \quad (4)$$

and

$$P^x_{c,n} = C^x_n(1 - c^x_{1,n}) + F^x_n f^x_{1,n} + (1 - C^x_n - F^x_n)/2. \quad (5)$$

Averaging over the N subjects, letting

$$\bar{P}_{u,n} = \frac{1}{N}\sum_x P^x_{u,n}$$

$$\bar{P}_{u,n} = \overline{C_n c_{1,n}} + \overline{F_n f_{1,n}} + (1 - \bar{C}_n - \bar{F}_n)/2 \quad (6)$$

and

$$\bar{P}_{c,n} = \overline{C_n(1 - c_{1,n})} + \overline{F_n f_{1,n}} + (1 - \bar{C}_n - \bar{F}_n)/2. \quad (7)$$

Even though there are two equations and four unknowns, by some simple algebra, one can at least see some relationships between these four parameters. Subtracting Eq. 7 from Eq. 6, one obtains

$$\bar{P}_{u,n} - \bar{P}_{c,n} = \overline{2C_n c_{1,n}} - \bar{C}_n, \quad (8)$$

and adding these two equations

Table 2
Proportion Correct on Changed ($\bar{P}_c$) and Unchanged ($\bar{P}_u$) Subproblems of the Nonreversal Shift in Blocks of 10 Trials and the Resulting Constraints on the Values of C, F, c_1, f_1

Parameters	End of Original Training	Shift Trials in Blocks of 10 Trials			
		1	2	3	4
$\bar{P}_u$	$\cong$ 1.00	.88	.92	.98	.98
$\bar{P}_c$	$\cong$ 1.00	.15	.40	.55	.62
$\bar{C}$	$\cong$ 1.00	.73-.97	.52-.68	.43-.47	.36-.40
$\bar{c}_1$	$\cong$ 1.00	$\geqslant$ .86	$\geqslant$ .89	$\geqslant$ .95	$\geqslant$ .94
$\bar{F}$	$\cong$ .00	.02-.27	.32-.48	.53-.57	.60-.64
$\bar{f}_1$	$\cong$ .50	$\geqslant$ .55	$\geqslant$ .87	$\geqslant$ .90	$\geqslant$ .94

$$P_{u,n} + P_{c,n} - 1 = \overline{2Ff}_{1,n} - \bar{F}_n. \tag{9}$$

For a given set of data, C, F, c_1, and f_1 cannot be estimated precisely, but one can observe certain constraints on their values. For example, since $c_{1,n}$ can be at most unity, $\bar{C}_n \geqslant \bar{P}_{u,n} - \bar{P}_{c,n}$, and similarly $\bar{F}_n \geqslant \bar{P}_{u,n} + \bar{P}_{c,n} - 1$. To obtain constraints on $\bar{c}_{1,n}$ and $\bar{f}_{1,n}$, however, terms in the form of $\overline{C_n c_{1,n}}$ rather than $\bar{C}_n \bar{c}_{1,n}$ are required. If C and c_1 were uncorrelated, one could replace $\overline{C_n c_{1,n}}$ with $\bar{C}_n \bar{c}_{1,n}$. But this is implausible, since one normally would expect a positive correlation between C and c_1. Whenever subjects attend to color and choose value 1, both C and c_1 will change in the same direction. Using the formula for correlation, we can see that

$$\overline{C_n c_{1,n}} - \bar{C}_n \bar{c}_{1,n} = (r_n) S_{C,n} S_{c1,n}, \tag{10}$$

where S refers to standard deviation and r_n is the correlation between $\bar{C}_n$ and $\bar{c}_{1,n}$. In order to proceed, an estimate of the correlation between $\bar{C}_n$ and $\bar{c}_{1,n}$, as well as an estimate of the standard deviations, is required.

Fortunately, some general results for correlation can be derived, and it will be assumed that the standard deviation can be as large as possible [$S_{\bar{C},n} \leqslant \bar{C}_n(1 - \bar{C}_n)$ and $S_{\bar{c}1,n} \leqslant \bar{c}_{1,n}(1 - \bar{c}_{1,n})$], since that will work against the eventual conclusions. The correlation between $\bar{C}_n$ and $\bar{c}_{1,n}$ can be partitioned into the value for the proportion of subjects whose responses were controlled by Dimension C on trial and the value for those subjects whose responses were controlled by some other dimension.

For the proportion $\bar{C}_{n-1}$ of subjects who are controlled by Dimension C on Trial n, the correlation will be

$$2c_{1,n-1} - 1.$$

This can be derived by writing out the possible changes in C and c_1 for the various outcomes and writing expressions for $S_{\bar{C},n}$ and $S_{\bar{c}1,n}$. One finds that the parameters for the amount of increase and decrease in C and c_1 all cancel, leaving a general result independent of rate change parameters. For the proportion $1 - \bar{C}_{n-1}$ of subjects not controlled by Dimension C on Trial n, the correlation will be zero. For all single-cue learning models, c_1 will not change, even though C will change due to indirect increases and decreases produced by the constraint that dimension control probabilities sum to unity. For multiple-cue learning models, c_1 will increase on a random half of the trials and decrease on the other half, independent of the trial outcome, while C will decrease or increase depending directly on the trial outcome, again producing a zero correlation. Therefore, by combining these two factors

$$r_n = \bar{C}_{n-1}(2\bar{c}_{1,n-1} - 1). \tag{11}$$

Using Eqs. 10 and 11 and assuming maximum standard deviations, one can substitute into Eq. 8 to obtain

$$\bar{P}_{u,n} - \bar{P}_{C,n} \leqslant 2[\bar{C}_n \bar{c}_{1,n} + \bar{C}_{n-1}(2\bar{c}_{1,n-1} - 1)\bar{C}_n(1 - \bar{C}_n)\bar{c}_{1,n}(1 - \bar{c}_{1,n})] - \bar{C}_n. \tag{12}$$

A similar expression may be written for F and f_1, and now one can observe parameter constraints for particular data. For a given set of data, these parameter constraints often prove to be severe enough to be informative. Data from an unpublished experiment using monkeys are shown in Table 2, along with the corresponding parameter constraints summed over blocks of 10 trials. First, note that the probability of attending to the previously relevant dimension decreases and the proportion of time that the currently relevant dimension is attended to shows a corresponding increase across trials. Attention models also lead one to expect that f_1 will increase, while c_1 will decrease across blocks of trials. The constrained value of f_1 increases across blocks as expected, but the value of c_1 shows no evidence of any decrease and remains near unity. That is, c_1 does not appear to change in spite of the fact that subjects have received nonrewards often enough to reduce drastically the probability of attending to color (the previously relevant dimension). Therefore, it seems that attention models cannot account for these data, since the parameter constraints would violate the assumption that nonrewards decrease the approach strengths to specific cues. One could speculate that nonreward only alters approach strengths by a tiny amount (not detectable in 40 trials), but this leads to the incorrect implication that a full reversal shift would be virtually impossible to accomplish. If nonreward had little or no effect on cue strengths, then Ss would be incapable of learning not to choose the formerly correct cue. Reversal shifts were mastered fairly efficiently in this experiment, the proportions correct being .34, .56, .56, and .62 in the first four 10-trial blocks of reversal shifts. In the case of the monkey shift performance, one can reject the assumption that the subproblems are treated as a single problem. Changes in attending and response probabilities

have not been specific beyond the constraint that they be monotonic, so these conclusions possess considerable generality.

This same basic analysis was applied to the experiments of the Tighes and their associates cited earlier in this paper by estimating trial-by-trial performance from published graphs. In addition, data from the Dickerson et al (1972) experiments were made available for the independence test through the courtesy of those researchers. In every case involving animal subjects, this independence assumption could be rejected, but in no case could this assumption be rejected for data obtained with children.

IMPLICATIONS

The proposed analysis is fairly weak in that it is not a very sensitive test for independent responding to subproblems, but it does make available strong inferences when it leads to the rejection of the idea that the subproblems are treated as a single discrimination.

The major restriction on concluding that subproblem independence has been observed is the possibility, raised by House and Zeaman (1963), that subjects may attend to compounds during training and transfer. For example, the compound of color and form might be treated as an additional dimension. The assumption that subjects respond to compounds has not been formally incorporated into attention models. The subproblem analysis could serve to indicate when a compounding explanation must be evoked.

In summary, the subproblem analysis developed by Tighe and his associates represents an advance in understanding reversal and nonreversal shift performance in relation to learning strategies. However, unless one restricts oneself to cases where no errors are made on unchanged subproblems, some further theoretical analysis is required to aid interpretation. This paper has provided a general analysis that allows one to test the appropriateness of a large class of attention models for reversal-nonreversal shift comparisons. The test is insensitive to particular assumptions of specific models, yet leads to a rejection of the class of attention models for a number of subject populations. Equally interesting, the analysis casts some doubts on earlier statements concerning the developmental increases in mediation unless these conjectures are restricted to changes in a relative, rather than an absolute, sense.

REFERENCES

Dickerson, D. J., Novik, N., & Gould, S. A. Acquisition and extinction rates as determinants of age changes in discrimination shift behavior. Journal of Experimental Psychology, 1972, 95, 116-122.

Fisher, M. A., & Zeaman, D. An attention-retention theory of retardate discrimination learning. In N. R. Ellis (Ed.), *International review of research in mental retardation*. Vol. 6. New York: Academic Press, 1972.

Graf, V., & Tighe, T. Subproblem analysis of discrimination shift learning in the turtle. Psychonomic Science, 1971, 25, 257-259.

House, B. J., & Zeaman, D. Miniature experiments in retardates discrimination learning. In L. P. Lipsitt and C. C. Spiker (Eds.), *Advances in child development and behavior*. Vol. I. New York: Academic Press, 1963.

Kendler, H. H., & Kendler, T. S. Vertical and horizontal processes in problem solving. Psychological Review, 1962, 69, 1-16.

Lovejoy, E. *Attention in discrimination learning*. San Francisco: Holden-Day, 1968.

MacKintosh, N. J. Selective attention in animal discrimination learning. Psychological Bulletin, 1965, 64, 124-150.

Sutherland, N. S. The learning of discriminations by animals. Endeavor, 1964, 23, 69-78.

Sutherland, N. S., & MacKintosh, N. J. *Mechanisms of animal discrimination learning*. New York: Academic Press, 1971.

Tighe, T. J. Subproblem analysis of discrimination shift learning in the pigeon. Psychonomic Science, 1972, 29, 139-141.

Tighe, T., & Frey, K. Subproblem analysis of discrimination shift learning in the rat. Psychonomic Science, 1972, 28, 129-133.

Tighe, T. J., Glick, J., & Cole, M. Subproblem analysis of discrimination shift learning. Psychonomic Science, 1971, 24, 159-160.

Tighe, T., & Tighe, L. S. Stimulus control in children's learning. In *Proceedings of the Sixth Minnesota Symposium on Child Psychology*. Minneapolis: University of Minnesota Press, 1971.

Zeaman, D., & House, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963. Pp. 159-223.

(Received for publication January 31, 1973;
revision received March 6, 1973.)