74-29,699

GONELLA, Don Frederick, 1940-
DISCRIMINATION AND REVERSAL LEARNING OF RETARDATES: A TEST OF THE ZEAMAN AND HOUSE MODEL.

The City University of New York, Ph.D., 1974
Psychology, experimental

Xerox University Microfilms, Ann Arbor, Michigan 48106

DISCRIMINATION AND REVERSAL LEARNING OF RETARDATES:

A TEST OF THE ZEAMAN AND HOUSE MODEL

by

DON FREDERICK GONELLA[1]

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1974

[1] Now at Green Mountain College, Poultney, Vermont 05764.

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

August 5, 1974
August
date

Solomon Weinstock
Professor Solomon Weinstock
Chairman of Examining Committee

August 14, 1974
date

O. Jackson Cole
Deputy Executive Officer
Professor Florence Denmark
Executive Officer

Professor Wilfred A. Gibson

Professor Eric G. Heinemann

Professor Solomon Weinstock

Supervisory Committee

The City University of New York

Abstract

DISCRIMINATION AND REVERSAL LEARNING OF RETARDATES:
A TEST OF THE ZEAMAN AND HOUSE MODEL

by

Don Frederick Gonella

Advisor: Professor Solomon Weinstock

The Zeaman and House (1963) model for discrimination and reversal learning by retardates postulates 2 responses to be learned: an "observing response" in which the S must attend to the relevant stimulus-dimension, and an "instrumental response" in which the S must choose the correct cue within the relevant dimension. The initial probability of a correct observing response (P_o) is said to control the duration of chance level performance prior to actual solution. Once solution commences, improvement is rapid, but is thought to be a continuous function with a rate that is similar for most S groups. Computer generated data, based on the equations of the model, showed a plateau in the reversal function, the length of which depended on the number of stimulus-dimensions (n) and on the P_o at the outset of reversal. The objective of the present research was to check on these predictions, while incorporating safeguards against suspected methodological artifacts.

A modified WGTA was used with Ss ranging in MA from 4 to 10 yrs. and with a mean IQ of 54. A 2 x 2 factorial design was used with 15 Ss in each group. The treatment variables were stimulus dimensionality (2 vs. 4 dimensions) and pre-reversal criteria. After the criterion of either 8 or 11 successive correct responses was met, Ss underwent an additional 30 trials for the detection of "pseudo-learners." 11 Ss were disqualified as pseudo-learners before the sufficient number of true learners was obtained. Both pseudo-learners and true learners received reversal training.

The dimensionality variable had a significant effect on the reciprocal of trials-to-criterion in acquisition. Neither variable had a significant effect on reversal performance when the analysis included only true learners. Furthermore, no plateau was apparent in the forward reversal curve. However, if pseudo-learners were substituted for true learners of similar MA, both treatment variables had a significant effect upon plateau length in the directions predicted by the Zeaman and House model.

Further results pertain to the statistical property of stationarity. It was found that the probability of a correct response remained relatively stationary and near chance until the criterial run commenced. There was no evidence from backward curve data of any gradual rise in performance, either in acquisition or in reversal, indicating that learning occured in an all-or-none fashion.

It was suggested that data presented by Zeaman and his associates concerning the shape of the acquisition and reversal functions were artifactual. Problems in the use of backward curves and in the choice of criteria were also discussed.

Acknowledgements

Grateful acknowledgement is made to the members of the thesis committee, Prof. Wilfred A. Gibson, Prof. Eric G. Heinemann, and especially to Prof. Solomon Weinstock whose guidance and cooperation were an essential part of this thesis.

Acknowledgement is also made to my wife, J. Penny Gonella, for her patience, understanding, and encouragement throughout the course of these studies, and to my parents, Mr. and Mrs. Mario A. Gonella, for their assistance in the preparation of the graphs.

Special thanks are due to the administration and staff of the Brandon Training School, Brandon, Vermont, for their cooperation and assistance, without which this research could not have been done.

Table of Contents

List of Tables

List of Figures

Figure | Page

Introduction

Zeaman and House (1963) have proposed a multiple-process model of discrimination learning which generates interesting theoretical predictions and which has potential applications to mental retardation. It was initially offered as a theory of discrimination learning for moderately retarded children with only the suggestion that it might have greater generality.

Subsequently, Fisher and Zeaman (1973) merged the Zeaman and House attention theory with the retention theory of Atkinson and Shiffrin (1969). Although the newer "A-R" theory of retardate discrimination is more comprehensive, the functions for acquisition and reversal remain essentially the same as in Zeaman and House (1963). While the present research was conceived within the framework of the Zeaman and House model, it also pertains to the broader model.

The Model

Visual discrimination learning is assumed to require the learning of two distinct responses, with the probability of each being altered on every reinforced trial. First, the subject (S) must learn to attend to the relevant stimulus-dimension, such as color, size, or shape, or to make an "observing response." Then the S must make the correct choice of cues within the dimension, such as red or green, circle or square, etc., or to learn the correct "instrumental response."

The equations of the model reflect the probabilities that the S is attending to the various stimulus-dimensions and to the cues within those dimensions. In Zeaman and House (1963) there are two rate parameters which correspond to the two types of habit being acquired. Fisher and Zeaman (1973) added two more rate parameters for the extinction of the respective habits. A complete specification of the model will not be provided here. However it will be necessary to define some of the constructs and their relationships. Let

P_c = the probability of an overt correct response,

P_{ot} = the probability of S observing the relevant stimulus-dimension on Trial t,

P_{it} = the probability of the correct instrumental response on Trial t, given that S observed the relevant dimension,

θ_{oa} = the acquisition rate parameter for the observing response,

θ_{oe} = the extinction rate parameter for the observing response,

θ_{ia} = the acquisition rate parameter for the instrumental response,

θ_{ie} = the extinction rate parameter for the instrumental response.

The equations pertaining to the observing response are

$$P_{o(t+1)} = P_{ot} + \theta_{oa}(1 - P_{ot})$$

if the S chooses the positive stimulus while observing the relevant dimension, and

$$P_{o(t+1)} = P_{ot} - \theta_{oe}(1 - P_{ot})$$

if the S chooses the negative stimulus while observing the relevant dimension. In the original formulation, the probability of an overt correct response was

$$P_c = P_o P_i + \tfrac{1}{2}(1 - P_o).$$

However in the subsequent revision by Fisher and Zeaman this probability is expressed as the ratio of positive covert responses to the total number of covert responses.

Zeaman and House (1963) marshal evidence to suggest that the rate parameters (θ values) have little or no relationship to mental age (MA). The fact that Ss of different MAs require different numbers of trials to reach a criterion is attributed to differences in the length of an initial plateau of chance-level responding. They assert that the ultimate rise to criterion seems to occur within just the last 40 trials and shows a similar slope for all groups. The length of the initial plateau is a function of the initial probability of observing the relevant stimulus-dimension. In other words, retarded Ss suffer from an attentional deficiency; they have a low probability of observing the relevant dimension and relatively high probabilities of observing one or more other dimensions. If this model is accurate, it would seem to hold important implications both for the training of retardates and for the customary conception of the nature of intelligence.

Backward Curves

Zeaman and House advocate the use of the backward learning curve (Hayes, 1953) for revealing the true pattern of individual progress shortly before the criterion is met. In the construction of such curves, the sequence of responses of each individual S is juxtaposed so that all the Ss reach the criterion on the same trial, rather than start on the same trial. Successive percentages for the group curve are then computed working backward from the criterion, rather than forward from Trial 1. Points on the curve that are nearest the criterion will be based on most or all of the Ss, while points further removed from the criterion are based on fewer Ss, as the data of the more rapid learners run out. The median for trials-to-criterion is used to provide a convenient cut-off for such curves.

Although Hayes refers specifically to having the Ss start their criterial runs on a common trial, it can be deduced from his further comments that the Ss should actually have their last error on a common trial. This seemingly minor distinction can affect the shape of the resulting curve. If an S masters his task late in a series of trials, he may not be able to meet the criterion in that test session. His criterial run would then occur early in the next session, often starting with the first trial. In such cases, the several trials before the criterial run would show a high level of performance. If the Ss' data are lined up according to criterial runs, rather than the last error of each, there will be a bias toward a gradually rising curve, whether justified or not. Even when the last error method is used, such a bias may be operating but to a lesser extent. Backward curves presented herein have, with one exception, used the last error as the point of reference, although it sometimes occured several trials before the accepted criterial run.

The fact that the last trial before the criterial run must necessarily be an error renders that trial useless for plotting the curve. Hayes suggests that neither the last trial nor the criterial trials be used in curve construction. He suggests the use of post-criterial trials to determine the final percentages. He then would connect the pre-criterial and post-criterial data points, a procedure which in essence interpolates the points in between, and, more importantly, makes the implicit assumption of a gradual or continuous function. Recognizing that backward curves tend to rise more steeply than the usual, negatively accelerated forward curve, Hayes states that they do not prevent a gradual learning function when it is justified. He fails to mention that the interpolation between pre-criterial and post-criterial trials does prevent a discontinuous function. With one exception, which will be very apparent, the backward curves presented herein do not use post-criterial trials, but terminate with the second trial before the criterial run.

In a rather cursory description of backward curve construction, Zeaman and House (1963) use an example that is greatly at variance with the Hayes (1953) procedure. They divide their Ss into subgroups according to the number of days needed to meet the criterion. Then a forward learning curve is constructed for each subgroup. Last, the separate forward curves are shifted to a common end-point before they are averaged together. Although the resulting curve may have some unique advantages, it is certainly not the backward curve described by Hayes. Because it is fundamentally a composite of forward curves, it inevitably yields a continuous function, albeit steeper than the usual forward curve. It is also more gradual than a true backward curve.

Whether all of the so-called backward curves of Zeaman and House are really composites of forward curves is not clear. If not, there remain the questions of whether they used post-criterial trials for interpolation and whether they lined up Ss according to their last errors or their criterial runs. In any event, the curves that they present are always continuous. This is in accord with their model which may be characterized as a two-stage growth model.

One may note that Trabasso (1963) collected data to compare a single-stage growth model (Bourne & Restle, 1959) with a two-stage all-or-none model (Restle, 1962). He used a concept-identification task with a successive procedure. Backward learning curves were constructed for both real and statistical Ss. The observed data fit the all-or-none prediction; that is, the probability of a correct response remained near the chance level until the last error was made. Unlike Trabasso, Zeaman and House have generally employed simultaneous discrimination. One of the objectives of this research was to examine the shape of the backward learning curve under the condition of simultaneous discrimination.

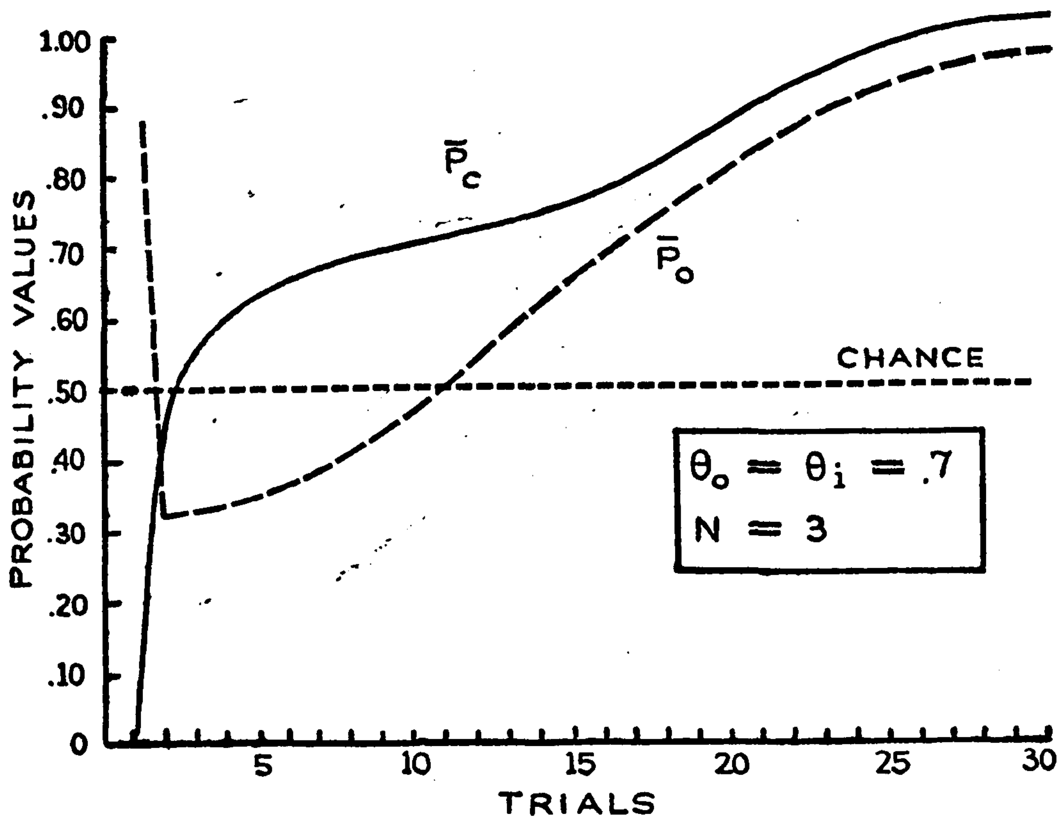
The Mid-Reversal Plateau

One of the more interesting predictions of the Zeaman and House theory is that under certain conditions one may expect to see a mid-reversal plateau in the forward reversal curve. That is, in the reversal function there sometimes will occur an inflection from negative to positive acceleration followed by another inflection back to negative acceleration. Zeaman and House suggest that the actual occurence of the plateau in data should have probative value for their theory, since they know of no other quantitative theory that predicts it.

The cause of the plateau, according to Zeaman and House, is the extinction of the relevant observing response in the early trials of reversal. This explanation was arrived at through inspection of expected operator approximations of mean probabilities, both of a relevant observing response ($\overline{P}_o$) and of an overt correct choice ($\overline{P}_c$) on successive trials. Whereas the expected $\overline{P}_o$ is high at the outset of reversal, it immediately makes a sharp descent and then gradually regains its former magnitude. The drop in $\overline{P}_o$ coincides with a plateau in the expected operator function for $\overline{P}_c$. These operator functions are depicted in Figure 1. It should be noted that since these are mean probabilities which yield the expected operator functions, they may not truly reflect patterns occuring within individuals.

Zeaman and House do point out that the average curve of a group does not necessarily have the same form as those of the individual Ss; see Sidman (1952) and Estes (1960). It is for this reason that they favor the use of backward rather than forward curves for depicting acquisition data.

Figure 1. Expected operator approximations of theoretical reversal learning, where $\overline{P}_c$ is the expected mean probability of a correct response and $\overline{P}_o$ is the expected mean probability of the S observing the relevant stimulus-dimension. (Adapted from Zeaman & House, 1963, p. 181.)


1.00
.90
.80
.70
.60
.50
.40
.30
.20
.10
0
PROBABILITY VALUES
$\bar{P}_c$
$\bar{P}_o$
CHANCE
$\theta_o = \theta_i = .7$
N = 3
5
10
15
20
25
30
TRIALS

Nevertheless, their reversal data is presented in the form of forward curves. Perhaps this was done to maintain comparability with the expected operator functions. However, if it is at all plausible that acquisition might be an all-or-none affair (Trabasso, 1963), then it would seem plausible that reversal might also be so. Consequently, another goal of this research was to examine the shape of the reversal function in backward, as well as forward, curves.

The exact extent of the mid-reversal plateau is supposed within the theory to depend upon four variables: the number of stimulus-dimensions present (n), the probability of the S observing the relevant dimension at the outset of reversal (P_o), and the rate parameters for the acquisition of the observing response (θ_{oa}) and the instrumental response (θ_{1a}). Methods are suggested for manipulating P_o. The rate parameters, on the other hand, are said to be structural features of the organism "reflecting fixed capacities possibly relatable to fixed components of intelligence." (Fisher & Zeaman, 1973, p. 173)

Based on "statistical children" which they generated randomly in accord with the model, Zeaman and House suggest that the plateau phenomenon may only occur for a limited range of parameter combinations. Furthermore, it is supposed to be strongly under the influence of the rate parameters, for which no means of manipulation is suggested.

The theoretical possibility presents itself that the plateau may never be observed among real Ss. They, however, do present data obtained from trainable retardates with MAs from two to six years which presumably do show the effect. Therefore it seemed reasonable to suppose that the problem of manipulating these parameters need not be of great concern in comparable research.

One method for varying P_o is to use different pre-reversal criteria for different groups of Ss. The more stringent the criterion, the closer should P_o be to one. Zeaman and House used six successive correct responses in a series of 25 trials as an example of an easy criterion, and ten as an example of a more stringent criterion. The easy criterion is supposed to accentuate the mid-reversal plateau.

Grant (1947) shows that the probability is greater than .15 that a run of six correct will occur somewhere in a series of 25 trials, strictly by chance, with the probability increasing to 1.00 as the S is brought back for successive days of training. Thus when the easy criterion is used it is quite possible that some "pseudo-learners" could be taken into reversal training, and, depending upon their number, could significantly affect the shape of the reversal function. They would perform at a chance level in reversal until such time as they again achieved the criterion.

Also, according to the theory, the greater the number of stimulus-dimensions (n) in the problem, the greater should be the plateau effect in reversal. However, the value of n is also an index of task difficulty. Given that equivalent criteria are used, a lower proportion of the S population should be able to learn a more difficult task than should be able to learn an easier task. Consequently, a higher proportion of the S population would ultimately become either pseudo-learners or non-learners.

In view of the above arguments, it seemed that the occurence of a mid-reversal plateau could be a methodological artifact having little or no relevance to actual learning processes. Consequently, procedures were employed in this study for the detection of pseudo-learners prior to reversal. Specifically, the criteria were raised slightly, and, more importantly, the Ss underwent an additional day of testing after they met their criteria, prior to reversal.

In the present study an attempt was made to test the plateau predictions of the Zeaman and House model by varying the pre-reversal criteria and by varying the size of n. Increasing n should lead to a larger plateau, as should the weakening of the criterion.

Method

A two-by-two factorial design was used, with the experimental variables being the pre-reversal criterion and the number of stimulus-dimensions present. Fifteen Ss were assigned to each of the four groups.

Subjects

All of the Ss were classified as retardates and were either students or residents of the Brandon Training School in Brandon, Vermont. They were selected primarily on the basis of their MAs, which ranged from 49 to 120 months. In all cases, the Peabody Picture Vocabulary Test was used for the determination of MA and IQ.

Previous research (Zeaman & House, 1963) indicated that there might be a considerable number of non-learners in the sample. Since the four experimental treatments were not of equal difficulty, it was anticipated that strict random sampling would lead to groups of unequal competence. Therefore, a system of stratified random sampling was used. The S pool was divided into four strata according to MA. Each of the four treatment groups received an equal number of Ss from each of the four strata, with non-learners being replaced.

The 60 Ss who were successful had a mean MA of 82.9 months, a mean IQ of 54.3, and a mean CA of 213.3 months.

The four groups did not differ significantly in any of these variables.

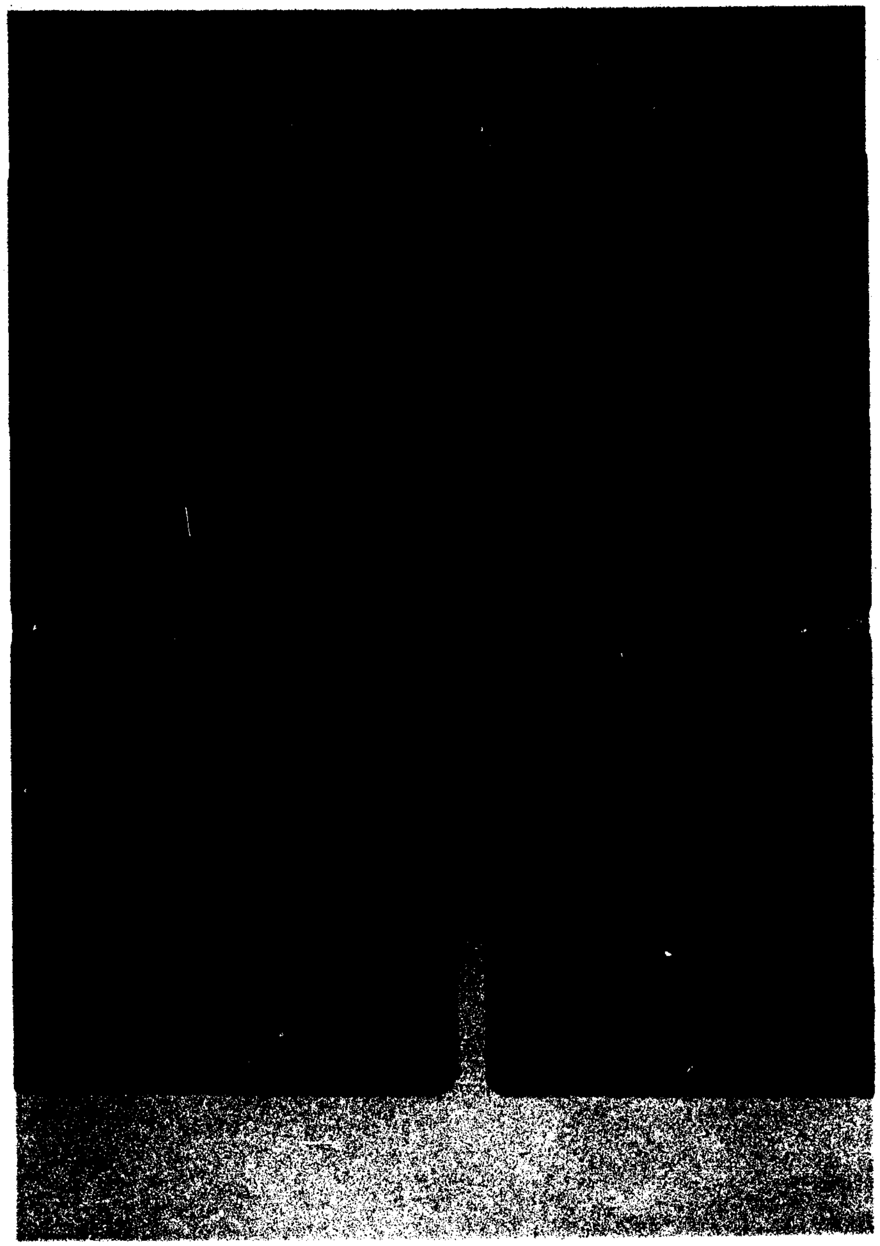
An additional 19 Ss were classified as either non-learners, pseudo-learners, or uncooperative. They had a mean MA of 65.1 months, a mean IQ of 44.1, and a mean CA of 224.4 months.

Apparatus

A modified version of the Wisconsin General Test Apparatus was constructed, similar to the one depicted in Zeaman and House (1963). Photographs of the apparatus may be seen in Figure 2. The apparatus, which consisted essentially of a sliding tray for the presentation of stimuli and a one-way window, framed with lights, between the S and the experimenter (E), was placed on a card table. The tray was approximately 30 inches wide and 12 inches deep. Centered on the tray, 12 inches apart, were two food cups having diameters of 2½ inches. The food cups were partially filled with plaster of Paris to create a concave surface which was lined with white felt to reduce auditory cues. Before a trial began the tray was prepared behind the screen. Then it was pushed toward the S to start the trial.

The stimulus objects which were presented differed from one another in one or three dimensions, depending on the desired difficulty of the task. The dimensions were shape, color, and height, with shape being the only dimension that

Figure 2. Photographs of the apparatus that was used, showing front and back views, with tray in and out, and the stimulus objects.



was relevant to a correct solution. Zeaman and his associates regard position as an additional dimension. The two shapes were circle and square; the two colors were red and green. The heights were 3/4 of an inch and 1 1/2 inches. All the stimulus objects were mounted on 1/4 inch Masonite bases, cut into 4 inch squares and painted white.

In order to have all combinations of stimulus characteristics, eight stimulus objects were necessary. These were arranged into four pairs, with the exact pairing dependent on the dimensionality of the task. Table 1 describes the pairings which were used. The order of their presentation was by random permutations of four.

Procedure

Shape was chosen as the relevant stimulus dimension for this study because previous research has indicated its high saliency for retarded Ss. In fact, next to position, shape is perhaps the easiest discrimination for them to learn. (Zeaman & House, 1963)

Because position was an irrelevant dimension in the task, the placement of the positive stimulus was randomly determined using permutations of 10 digits. In other words, in every 10 trials the positive stimulus was on the left 5 times and on the right 5 times.

On his first day the S was seated in front of the apparatus and the following instructions were read to him.

Table 1

Descriptions of the Stimulus Objects

When two dimensions were present, the following pairs were used.

Pair A	Tall, red square	vs.	Tall, red circle
Pair B	Short, red square	vs.	Short, red circle
Pair C	Short, green square	vs.	Short, green circle
Pair D	Tall, green square	vs.	Tall, green circle

When four dimensions were present, the following pairs were used.

Pair A	Tall, red square	vs.	Short, green circle
Pair B	Short, red square	vs.	Tall, green circle
Pair C	Short, green square	vs.	Tall, red circle
Pair D	Tall, green square	vs.	Short, red circle

"We are going to play a candy game. When this tray comes out there may be a piece of candy in one of these two holes. But you won't know which hole has the candy because they will be covered up by blocks like these. If you look closely at the blocks you can solve the problem and find the candy on your first choice each time. When you have learned what the clue is, and make the right choices all the time, then I'll buy you a free Coke. Sometimes there won't be any candy in the tray, but most of the time there will be. You can either eat it now or save it in the dish, whichever you prefer."

The E then sat behind the screen and the trials began. The S received 30 trials a day except on the day that he happened to reach the criterion, at which point the testing was stopped and the S was treated to a Coke. The criterion was either 8 or 11 successive correct choices, depending upon the group to which the S was assigned. If the S did not reach his criterion in 15 days of testing, he was judged a non-learner.

To begin any trial, the E placed an M & M candy in the appropriate cup, covered the cups with the appropriate stimulus objects, and then pushed the tray forward. The S would pick up one of the objects. If he chose correctly, he found the candy and the E said, "Good." If the S chose incorrectly, the E said, "No," and then instructed the S to look in the other cup and to take the candy if there was any. The tray was then withdrawn. The interval between the starts of successive trials was about 15 seconds, with most Ss finishing in less than 10 minutes.

During the initial learning of the discrimination, only 20 of the 30 trials had candy. The 10 unrewarded trials were randomly scattered among the last 25 trials. In the session on the day after an S had reached his criterion, 15 trials were rewarded and 15 were not. This was a test session, intended to reveal to the E any pseudo-learners without providing much extra training for the true learners. The purpose of the 10 unrewarded trials during each session of the initial training was to prepare the Ss for this test. To be judged a "true learner" then, an S not only had to reach his criterion, but also had to choose correctly in at least 26 of the 30 trials on the test day.

Regardless of an S's performance in the test session, he was still continued into reversal training in the next session. However, only those Ss who had achieved 26 out of 30 correct in the test session were included in the aforementioned 60 Ss.

Reversal training was conducted in the same manner as the initial training with two exceptions. First, of course, is the fact that the positive and the negative stimuli were reversed. In this case, the original positive stimulus was the square and the negative was the circle. The other change in procedure was that during reversal training all trials held reinforcement. The S received 30 M & M candies a day, unless he achieved the criterion of 11 successive correct choices, at which time he received another Coke. If a S was not able to reach the criterion in 11 days of testing (330 reinforced trials) testing was terminated. Happily, this only occured among pseudo-learners.

Results

The first analyses to be presented are based upon the data of the 60 Ss judged to be true learners. Subsequently the data of the disqualified Ss will also be considered.

Acquisition

An analysis of variance performed on the trials-to-criterion measure revealed no significant differences between the groups although the group medians differed greatly. This analysis is summarized in Table 2. Table 3 shows the means and medians of the four groups. Inspection of the frequency distributions showed considerable positive skewness in all four groups. Therefore, another analysis of variance was performed using the reciprocal of the trials-to-criterion score. As is shown in Table 4, the treatment variable of dimensionality did produce significant differences.

Effect of nonreinforced trials. Since including 10 nonreinforced trials in each day's testing was a departure from the procedure used by Zeaman and House, contingency tables were created based on whether correct or incorrect responses followed reinforced or nonreinforced trials. Separate analyses were done for the group of 30 Ss exposed to two stimulus-dimensions and for the group exposed to four stimulus-dimensions.

Table 2

Analysis of Variance of

Trials-to-Criterion Measure in Acquisition

Source	SS	df	MS	F
Criteria (C)	686.8	1	686.8	.15
Dimensions (D)	13290.8	1	13290.8	2.88
C x D	2680.0	1	2680.0	.58
Within Groups	258758.2	56	4620.7	
Total	275415.9	59		

Table 3

Means and Medians of Trials-to-Criterion Measure

by Treatment Groups

Group	Means	Medians
Two Dimensions, Easy Criterion (2D-8C)	45.1	26
Two Dimensions, Hard Criterion (2D-11C)	65.4	41
Four Dimensions, Easy Criterion (4D-8C)	88.4	92
Four Dimensions, Hard Criterion (4D-11C)	82.6	79

Table 4

Analysis of Variance of Reciprocal of Trials-to-Criterion Measure in Acquisition

Source	SS	df	MS	F
Criteria (C)	.00299	1	.00299	1.50
Dimensions (D)	.01309	1	.01309	6.56*
C x D	.00002	1	.00002	.01
Within Groups	.11176	56	.00200	
Total	.12787	59		

* $p < .05$

For the two-dimension group, the proportion of correct responses following reinforced trials in no case differed significantly from the proportion following nonreinforced trials. Separate chi-square tests were done using all pre-solution trials combined, the last 50 trials, the last 20 trials, and Vincentized data by halves. Also, it made no difference if the responses that followed reinforced trials were subdivided according to whether they followed correct or incorrect (corrected) responses.

The same types of contingency tables used for the two-dimension group were also created from the data of the four-dimension group. Six of the resulting chi-squares were insignificant and four were significant at the 5% level.

Two of the significant chi-squares were computed from tables in which the responses following reinforced trials were subdivided. One of these tables was for all pre-solution trials; the other was for the data of the first half. In both, the proportion correct after reinforced errors (correction trials) was greater than that after reinforced correct responses. It should be noted that the chi-squares from the analogous tables in which responses following reinforced trials were not subdivided were not significant. In both cases the proportion correct following nonreinforced trials was not significantly different from that for reinforced trials.

The other two significant chi-squares were both from the data of the last 20 trials. In these cases, the proportion correct after nonreinforced trials was greater than that after reinforced trials. Since it seems unlikely that learning could have taken place during the nonreinforced trials it may indicate a motivational effect.

The significant chi-squares may be due to alpha error. In any event, the nonreinforced trial procedure did not retard learning. Furthermore, the effect was small and we shall assume in the subsequent analyses that the nonreinforced trials acted as "probe" or "blank" trials.

Effect of MA on trials-to-criterion. It should be noted that the mean MA of the *S*s in this study was higher than that of the *S*s typically used by Zeaman and House (1963). They used *S*s with MAs of six years or less, whereas the present *S*s ranged from four to ten years. One quarter of the *S*s in the present study did have MAs of less than six years.

Using the top 4 and the bottom 4 *S*s from each group of 15, a three-dimensional analysis of variance was done. As expected, MA did have a significant effect on the reciprocal of trials-to-criterion. The other two factors were not significant in this analysis, but that may be attributed to the small number of *S*s in each cell, and consequently to the reduced power of the test. This analysis is summarized in Table 5.

Because of the difference in mean MA between the present Ss and the Zeaman and House Ss, wherever MA might have influenced the results an effort was made to examine this possibility and to apply appropriate tests.

Backward curves. Before presenting the backward curves of the four treatment groups and various combinations thereof, something should be said about the possible bias of such curves that was mentioned earlier.

Nine Ss from various groups obviously achieved solution too late in a test session to meet their criteria. Their criterial runs began with Trial 1 of the next session, although at least the last response of each on the previous day was correct. If the backward curve for these Ss were to use the criterial run, rather than the last error, as its point of reference, the curve would show a gradual rise to criterion.

Another 11 Ss, who also had a correct last response preceding their criterial day, began their criterial runs within the first 10 trials of a session. It is possible that some of these Ss also had achieved solution on the previous day, but had undergone some forgetting. If this were the case, their backward curve would also show a somewhat gradual rise to criterion.

Table 5

Analysis of Variance of Reciprocal of

Trials-to-Criterion Measure in Acquisition

with MA as a Third Factor

Source	SS	df	MS	F
Mental Age (M)	.00849	1	.00849	4.75*
Criteria (C)	.00181	1	.00181	1.01
Dimensions (D)	.00559	1	.00559	3.13
M x C	.00002	1	.00002	.01
M x D	.00611	1	.00611	3.42
C x D	.00159	1	.00159	.89
M x C x D	.00059	1	.00059	.33
Within Groups	.04291	24	.00179	
Total	.06711	31		

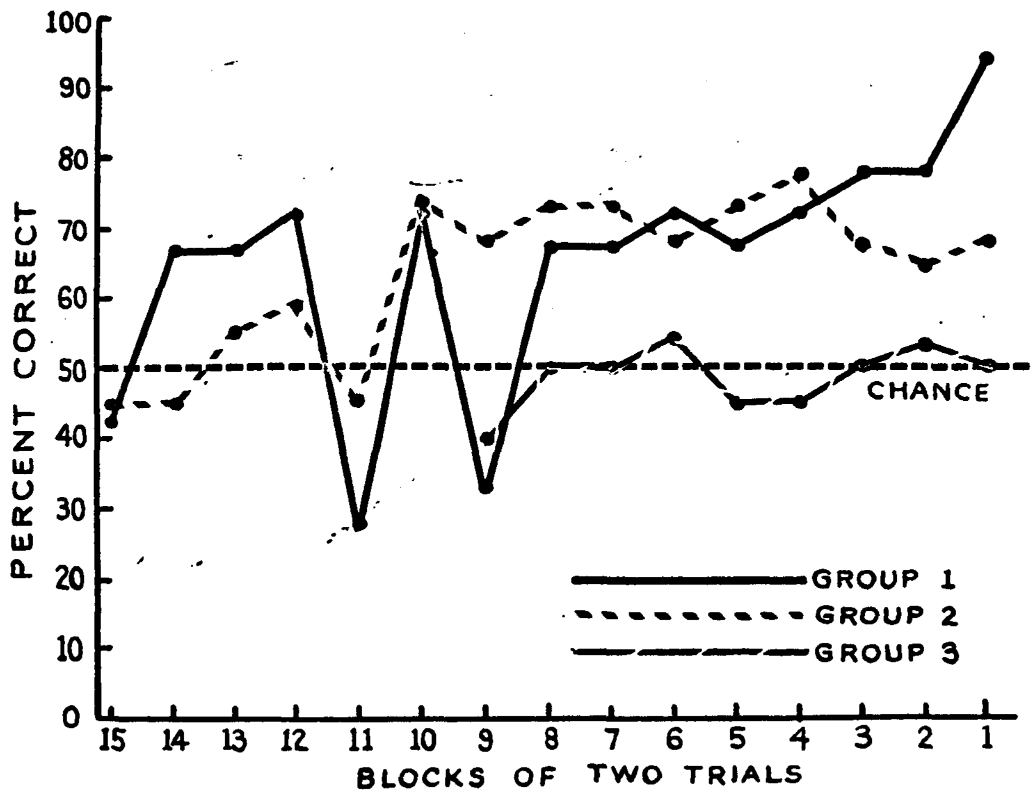
* $p < .05$

In Figure 3 are shown the backward curves of three groups. The top curve is from the 9 Ss who started their criterial runs on the first trial of a session. The middlemost curve is from the 11 Ss that started their runs within the first 10 trials of a session. The bottom curve is from the 15 Ss that met their criteria in their very first test session, and therefore could not have achieved a prior solution. All three of these curves use the criterial run as their point of reference, although for the bottom two curves the last errors would also coincide. The effect of late solution on the shape of the backward curve is apparent.

Additional evidence on this point is found by relating the final performance of Ss on the day before they met the criterion with the mean trial on which their criterial runs began in the next session. Table 6 summarizes this data. It can be seen that the more the correct responses at the end the previous session, the earlier the criterial run is likely to begin. That this is not merely a reflection of incremental learning will be shown by subsequent data on the statistical property of "stationarity" and also by inspection of the backward curves to be presented.

Figure 4 shows the backward acquisition curves of each of the four treatment groups. Figure 5 combines groups that were exposed to the same stimuli in order to achieve

Figure 3. Backward curves illustrating the bias introduced by Ss achieving solution late in a session. Group 1 Ss began their criterial runs on the first trial of a session. Group 2 Ss began their runs within the first 10 trials of a session, excluding the first, and also had responded correctly on the last trial of the previous session. Group 3 Ss achieved criterion on the first day of testing. For the curve of Group 1 only, the last errors of Ss do not coincide. Instead, the criterial runs coincide for all three curves.



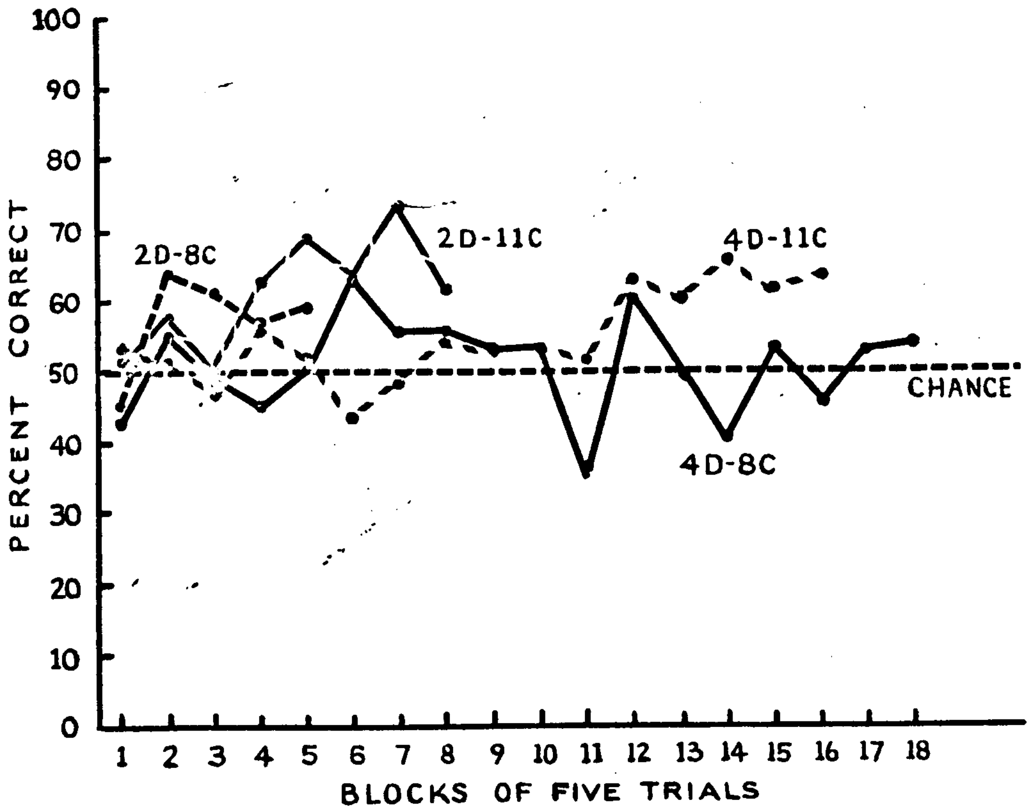
100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
CHANCE
GROUP 1
GROUP 2
GROUP 3
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
BLOCKS OF TWO TRIALS

Table 6

Data Showing the Relation of
Last Responses in Pre-criterial Session
with Trial on which Criterial Runs Began

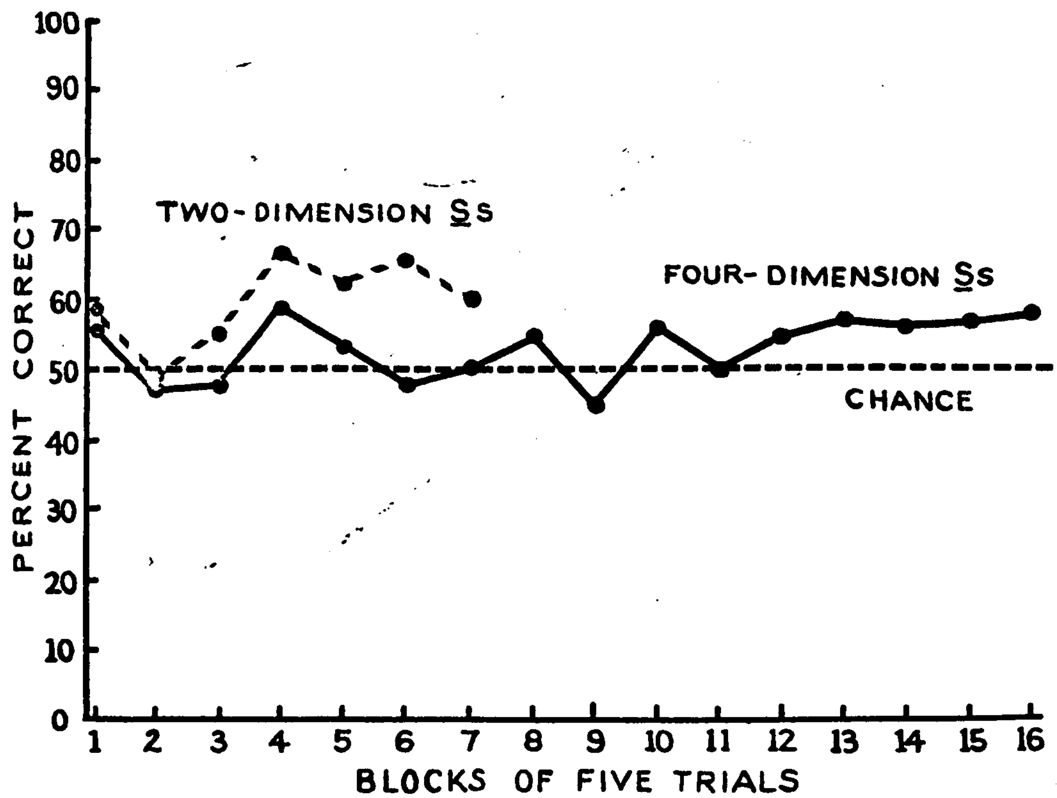
Ss	N	Mean Trial of Start of Run
Met criterion on first day of testing	15	12.73
Last response of previous day incorrect	17	10.35
Last response of previous day correct	7	13.14
Last 2 responses of previous day correct	8	11.00
Last 3 responses of previous day correct	2	4.50
Last 4 responses of previous day correct	1	1.00
Last 5 responses of previous day correct	2	3.50
Last 6 responses of previous day correct	2	4.00
Last 7 responses of previous day correct	4	3.50
Last 9 responses of previous day correct	1	1.00
Last 10 responses of previous day correct	1	1.00
Total	60	

Figure 4. Backward acquisition curves of the four treatment groups. The placement and length of each curve were determined by the median number of trials-to-criterion for that group. Each curve is based on the data of 15 Ss.



100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
2D-8C
2D-11C
4D-11C
CHANCE
4D-8C
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
BLOCKS OF FIVE TRIALS

Figure 5. Backward acquisition curves of groups combined according to stimulus-dimensionality. The placement and length of each curve were determined by the median number of trials-to-criterion for that group. Each curve is based on the data of 30 <u>S</u>s.



100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
TWO-DIMENSION Ss
FOUR-DIMENSION Ss
CHANCE
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
BLOCKS OF FIVE TRIALS

a more stable pair of curves. In Figure 6 the same groups are again combined in order to focus on just the last 30 trials before criterion.

Figure 7 presents backward curves based on just the highest and lowest Ss in MA. The high group had MAs between 92 and 120 months. The low group went from 49 to 74 months.

Although some of the curves in Figures 4, 5, 6, and 7 do tend to hover more above than below the chance level, this may be due to the criterial artifact discussed above and depicted in Figure 3. Despite this problem, by and large these curves could hardly be used to support a gradual learning theory.

Stationarity. Three different approaches were used to determine whether or not the probability of a correct response remained stationary during the pre-solution trials. The results were mixed. When the data were Vincentized by halves, significant differences were found between the first and second halves. When backward learning curves were studied, no significant change in probability occured as the Ss neared solution. When tests were done on individual Ss, only 2 out of 60 Ss showed clearcut significance. Two more were possibly significant, but only if Tocher's modification of Fisher's exact test was used. (Siegel, 1956)

Referring first to the tests on individual Ss, the procedure was to take blocks of about 10 trials each.

Figure 6. Backward acquisition curves of groups combined according to stimulus-dimensionality. Each curve is based on the data of 30 Ss.

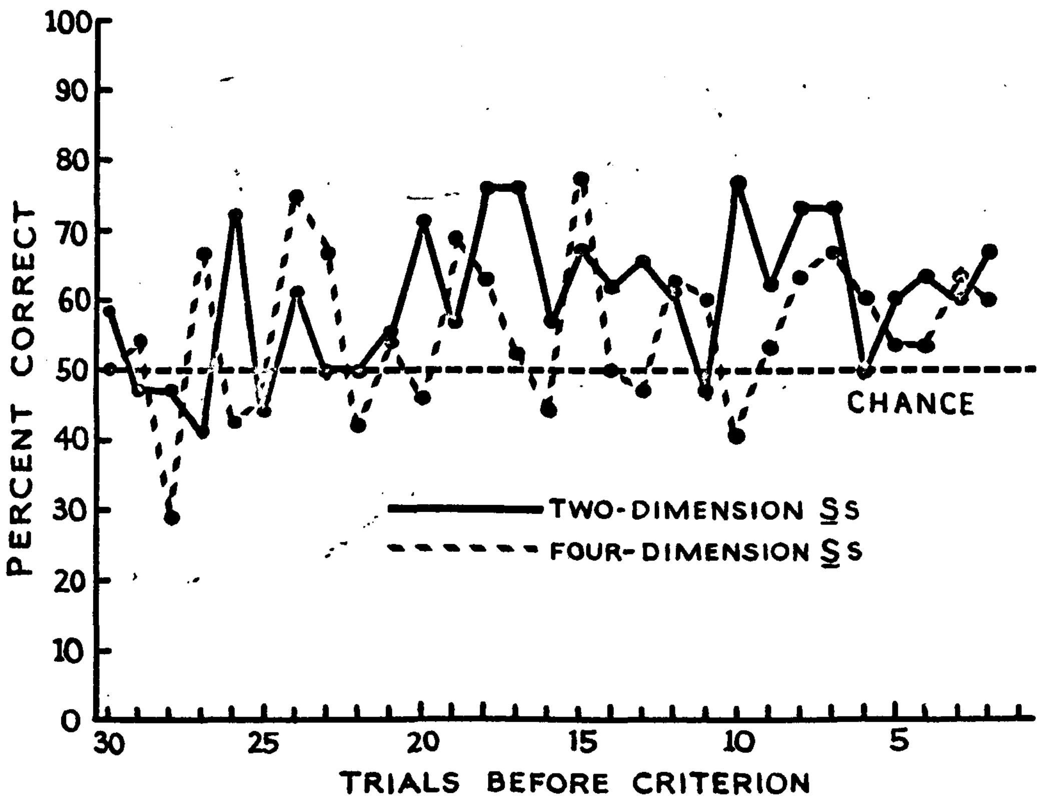
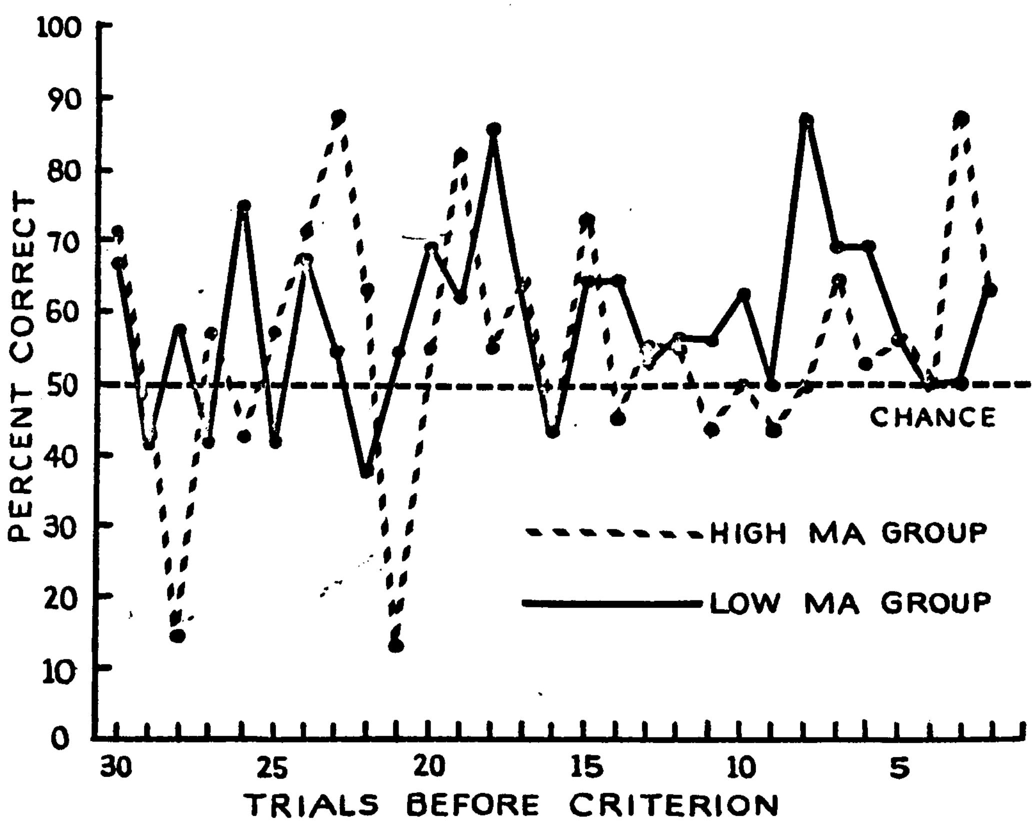

100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
CHANCE
TWO-DIMENSION Ss
FOUR-DIMENSION Ss
30
25
20
15
10
5
TRIALS BEFORE CRITERION

Figure 7. Backward acquisition curves of high and low MA groups. Each curve is based on the data of 16 Ss at its right end, and of fewer Ss as it moves toward the left.



PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
CHANCE
HIGH MA GROUP
LOW MA GROUP
30
25
20
15
10
5
TRIALS BEFORE CRITERION

Chi-square was used, except for cases where there was a fourfold table more suitable for Fisher's exact test. The alpha level was 5%, and 5% of 60 Ss is 3 Ss. Inasmuch as there were only from 2 to 4 Ss who showed significance, it cannot be said that the individual tests contradict the assumption of stationarity.

Stationarity tests based on the backward curve data were done in two different ways. First, chi-squares were computed for the trials between the medians and solutions, that is, the last 36 trials for the two-dimension groups and the last 81 trials for the four-dimension groups. No significant changes were found.

Since the data and theory of Zeaman and House (1963) indicated a rather steep incline shortly before the criterion was reached, it was thought possible that the analysis of too many trials at once might have masked this effect in the present data. Consequently, analyses of only the last 20 trials were done for each of the four groups separately and for all four combined. Again, none of the chi-squares were significant.

The results of the above tests, both from backward data and from individual data, run counter to the Zeaman and House assumption of gradual rather than all-or-none learning. Since the mean MA of the Ss in this study was higher than that of the Zeaman and House Ss, a separate analysis was performed for the 15 Ss in the present study who had MAs of less than six years. Chi-square tests based on the last 20 trials of just these 15 Ss again were not significant.

Three chi-squares were computed from the Vincentized data. One was for Ss exposed to two dimensions; one was for Ss exposed to four dimensions. The third was for all the Ss combined. All three were significant. The overall proportion correct in the first halves of the Ss' pre-solution trials was .492; for the second halves it was .556.

Although the results from the Vincentized data would seem to contradict the results from the backward data and the individual data, this may not actually be so. The backward data was focused primarily on what took place within the second halves of most Ss' data, and did not include much of the first half data. Also, for the Vincentized data, Ss who were slow to learn contributed much more heavily to the frequencies in the contingency tables than did the fast learners. It is conceivable that they were learning to connect the different stimulus-patterns to their responses in a paired-associates manner. Also, the criterial artifact discussed before could be responsible.

Considering the shape of the learning function proposed by Zeaman and House (1963), the tests based on backward data would seem much more pertinent to the questions at hand than would the tests of the Vincentized data.

Response patterns. Inasmuch as each of the four pairs of stimuli was presented in two different arrangements, there were eight stimulus-patterns that any given S was exposed to. It seemed reasonable to suspect that some of the Ss might have learned in an all-or-none, paired-associates manner, with some stimulus-patterns being learned earlier and some later in the course of training. To check on this possibility, a 2 x 8 contingency table was created for each of the 60 Ss. The frequencies of correct and incorrect responses were tabulated for each of the stimulus-patterns, excluding the first presentation of each. Twelve chi-squares were significant. In other words, 12 Ss responded correctly to some stimulus-patterns more often than they did to others.

Upon examining the 12 contingency tables that showed significance, it became clear that many of them reflected a position preference on the part of the S. That is, left-handed patterns might be answered correctly with greater frequency than right-handed patterns, or vice-versa. Subsequent tests showed that 11 of the 12 did in fact have a significant position bias. This coincides with the E's recollection of some Ss who would repeatedly reach for the same side of the tray, for what seemed like hundreds of trials.

In order to determine which of these 12 tables were significant for reasons other than position preferences, the 2 x 8 tables were divided into two 2 x 4 tables for each S. The right-handed patterns were in one table, the left in the other. If the chi-squares from both of the smaller tables were not significant, it might be concluded that the significance of the larger table was due solely to a position bias. Eight of the 12 met this criterion of nonsignificance. For each of the other 4 Ss, only one of the 2 x 4 tables was significant. Three of these did in fact demonstrate position preferences, but this could not be hastily accepted as the entire cause of the significance of their 2 x 8 tables. Consequently, these four were singled out for further study.

None of the four Ss in question was among the four Ss who showed significant departures from stationarity. Stationarity tests on these four as a group, whether on the last 90 or the last 20 trials, were not significant. Individual tests on just the last 90 trials were also insignificant.

An additional type of stationarity test was devised, wherein the response sequences for each of the eight stimulus-patterns were tabulated separately. These response sequences could then be adjusted to have a common starting point or a common ending point. Both methods were used. None of the four Ss showed any departure from stationarity with these tests either.

Using the same response sequences, runs tests were done for each of the eight patterns for each of the four Ss. Of 32 tests, only 2 were significant. They happened to be from the same S, but the pattern of the responses was not very elucidating. This S did show a position preference.

Returning to the problem of why the four 2 x 8 tables were significant, certainly in three cases position preferences were a contributing factor. There was no evidence of paired-associates learning. Considering that there were 60 such tables, once again the likelihood of alpha error presents itself.

Table 7 shows the frequencies of the various test results according to treatments, and also indicates the extent of position preferences in relation to these test results.

Position preferences. The difference between the proportion of right and left responses for each S was tested using a z-test. Seventeen Ss, 11 from the four-dimension condition, showed a significant bias toward one side or the other.

The data of these 17 Ss were Vincentized by tenths, and the frequencies of each S's more frequent or preferred response were tabulated. Figure 8 shows the typical absorption into the position biases, and also the breaking

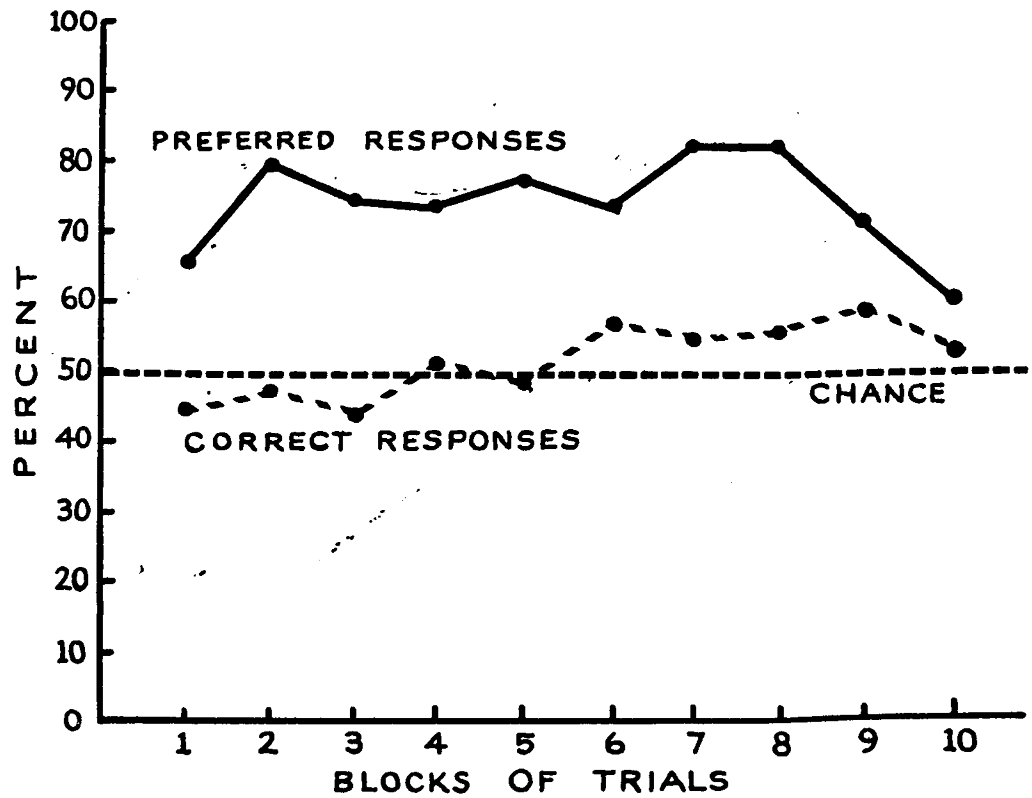
Table 7

Frequencies of Response Pattern Results according to Treatment

Results of Tests	Two Dimensions	Four Dimensions	Mean Percent Preferred Response
2x8 S*, both 2x4s NS**	3	5	76.0
2x8 S, one 2x4 S	1	3	69.5
2x8 NS, both 2x4s NS	10	12	56.5
2x8 NS, one 2x4 S	1	0	58.0
Insufficient Data	15	10	60.7
Total Subjects	30	30	

* S = Significant, ** NS = Not Significant

Figure 8. Vincentized data of the 17 Ss who showed significant position preferences.


PREFERRED RESPONSES
CORRECT RESPONSES
CHANCE
PERCENT
BLOCKS OF TRIALS
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10

out of these biases just prior to solution. The absorption process is illustrated more vividly by Figure 9 which uses just the first 15 trials of each S.

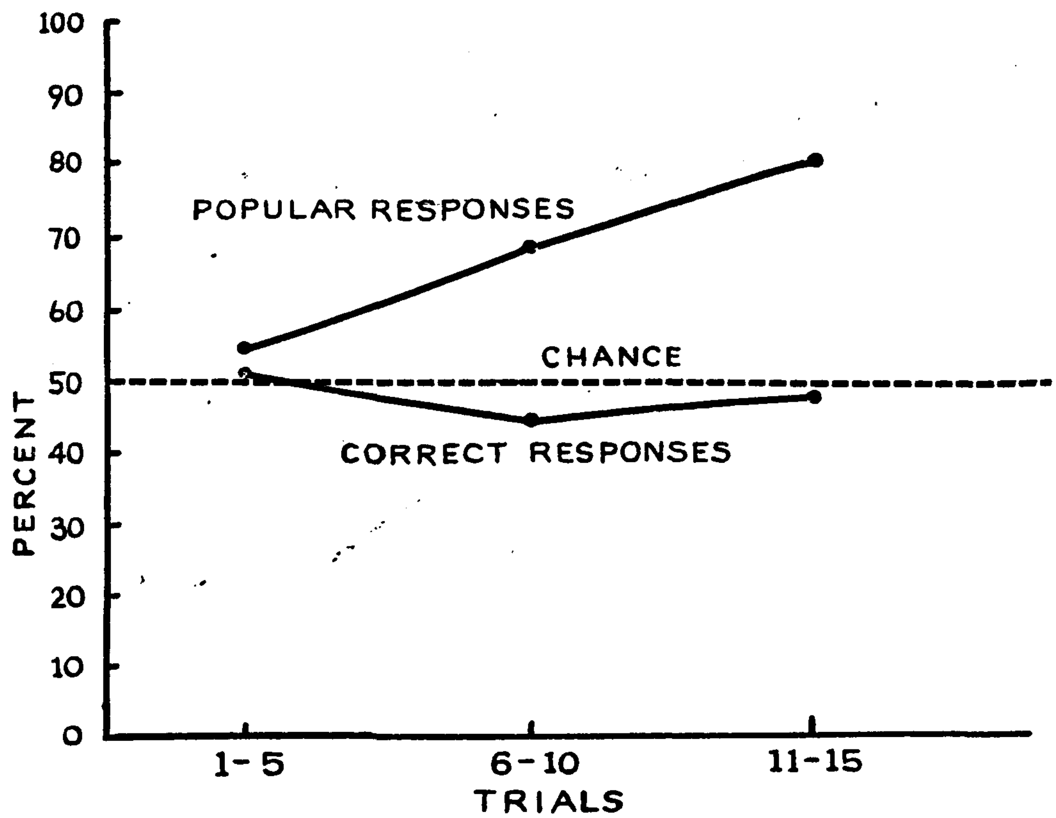
Reversal

Despite the differences in task difficulty which were evidenced in acquisition, there were no significant differences in the rate of reversal for the four groups. Table 8 summarizes this analysis. Neither were there significant differences when the reciprocal of trials-to-criterion was used, nor when high and low MA groups were compared in a three-way analysis of variance using reciprocals. Means and medians for trials-to-criterion in reversal are shown in Table 9. Although the differences were not significant, it may be noted that the direction of the differences was opposite to what is predicted by the "overlearning reversal effect" (Lovejoy, 1966) and by the theory of Zeaman and House (1963).

Reversal curves. One of the stated goals of this research was to examine the shape of the reversal function in both forward and backward curves. With forward curves this was a relatively simple matter. With backward curves, however, there were unforeseen difficulties.

Inasmuch as reversal is a rapid process for most Ss, points on the backward curve, going from right to left, soon become unreliable as the data of each succeeding S are

Figure 9. The first 15 trials of the 17 Ss who showed significant position preferences.



100
90
80
70
60
50
40
30
20
10
0
PERCENT
POPULAR RESPONSES
CHANCE
CORRECT RESPONSES
1-5
6-10
11-15
TRIALS

Table 8

Analysis of Variance of Trials-to-Criterion in Reversal

Source	SS	df	MS	F
Criteria (C)	1297.3	1	1297.3	1.87
Dimensions (D)	120.4	1	120.4	.17
C x D	.8	1	.8	.00
Within Groups	38780.4	56	692.5	
Total	40198.9	59		

Table 9

Means and Medians of Trials-to-Criterion Measure in Reversal by Treatment Groups

Group	Means	Medians
Two Dimensions, Easy Criterion (2D-8C)	15.07	6
Two Dimensions, Hard Criterion (2D-11C)	24.13	13
Four Dimensions, Easy Criterion (4D-8C)	12.00	3
Four Dimensions, Hard Criterion (4D-11C)	21.67	4
Overall	18.22	4.5

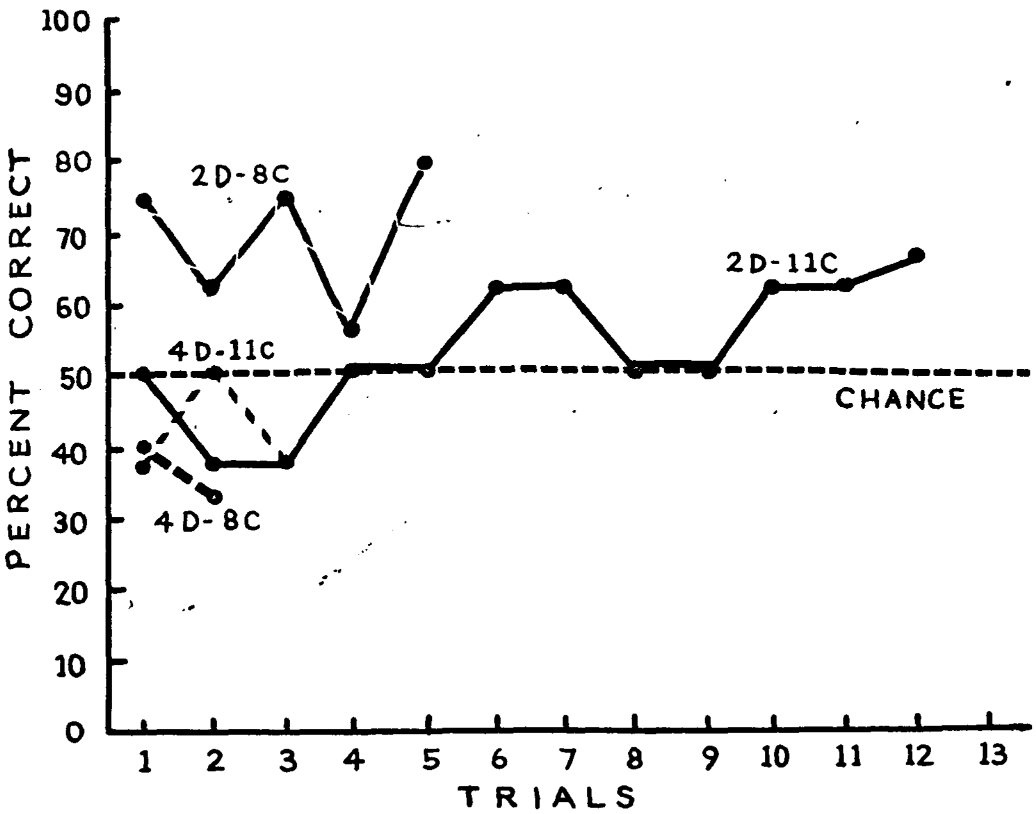
used up. Also, about a third of the Ss couldn't be used at all, as they made only one error prior to their criterial runs. For the same reasons, backward and forward curves are not comparable. In forward curves, once a S has met his criterion he is assumed to continue at the same level of performance, and the assumed data is used in curve construction. In backward curves, when one reaches the beginning of a S's data, that S drops out of the picture. The remaining (leftward) portion of the curve will be uninfluenced by that S.

Figure 10 shows the backward curves of the four treatment groups. The points in this figure are based on a mean of only 8.6 Ss. That these curves are unreliable is demonstrated by one curve having the typical S start reversal with a rate of 75% correct.

In Figure 11 the groups are combined according to treatment variables in an effort to obtain more stable backward curves. Only one of the four curves in the figure shows any hint of a gradual function.

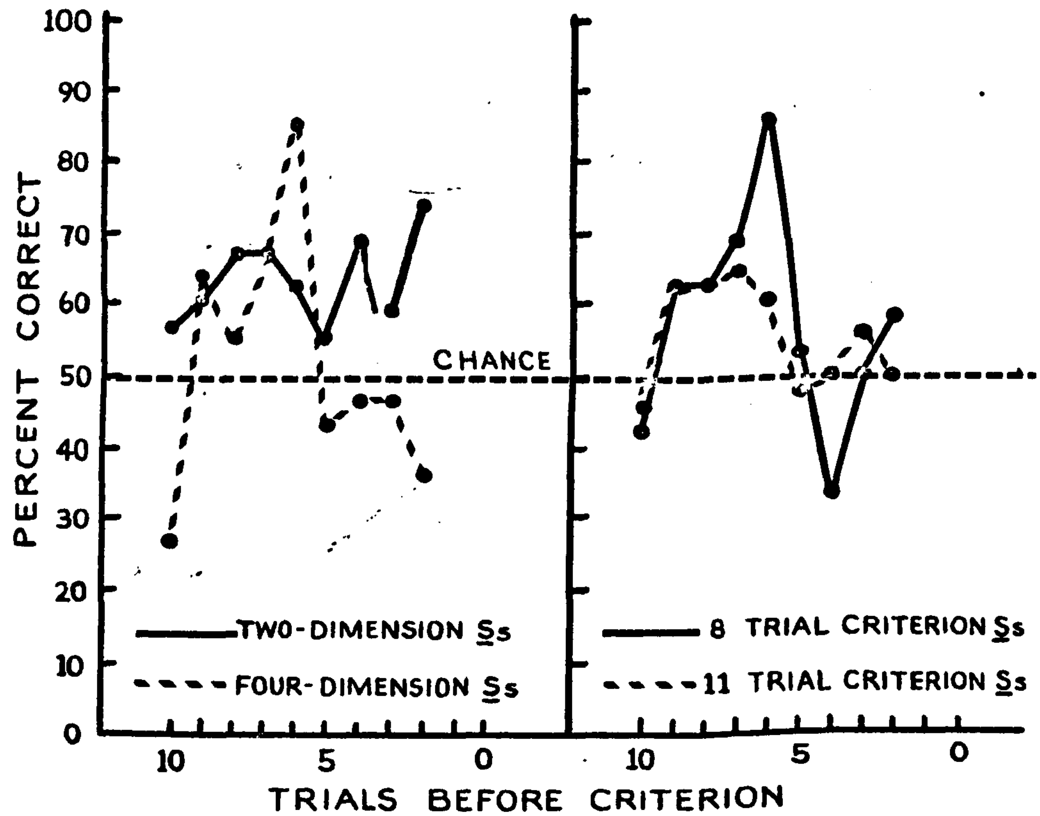
Figure 12 shows the forward reversal curves of the two groups that were supposed to have the greatest and least amounts of the mid-reversal plateau effect according to Zeaman and House (1963). Inspection of these curves reveals that the group which was supposed to have the least plateau had the most, and vice-versa. However, it is not established

Figure 10. Backward reversal curves of the four treatment groups. Placement and length of each curve were determined by the median number of trials-to-criterion for that group. Each curve is based on the data of from 9 to 13 Ss at its right end, and from 8 to 9 Ss at its left end.



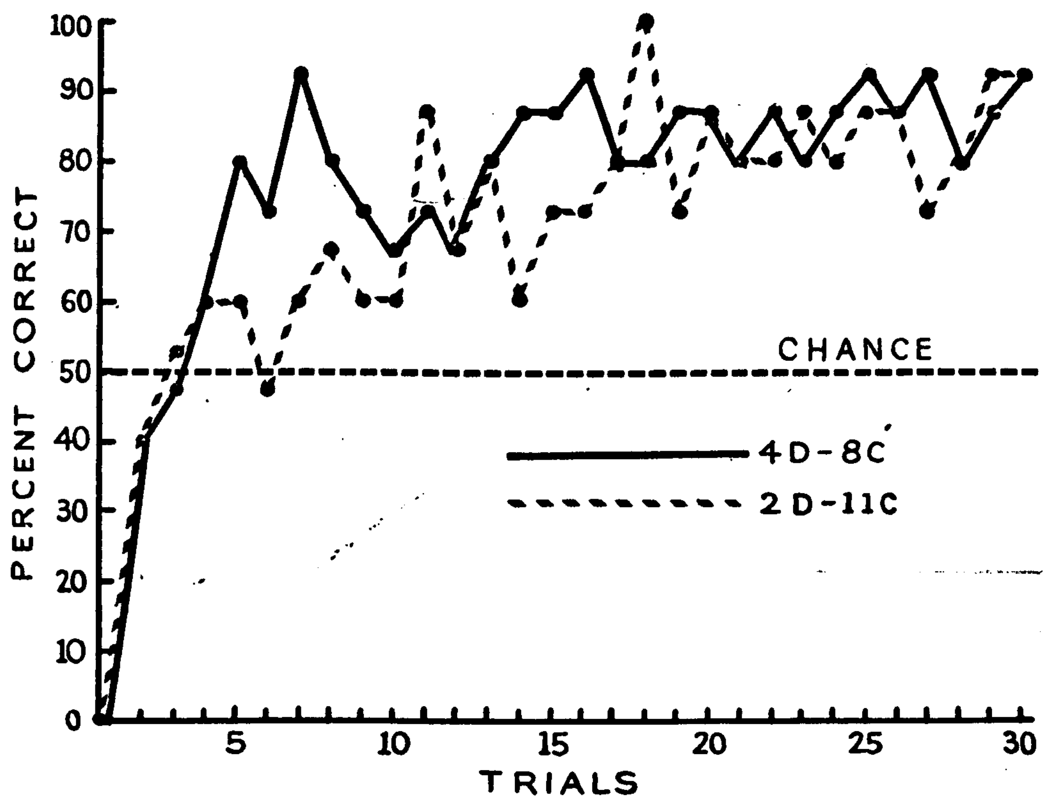
PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
2D-8C
2D-11C
4D-11C
4D-8C
CHANCE
1
2
3
4
5
6
7
8
9
10
11
12
13
TRIALS

Figure 11. Backward reversal curves of groups combined according to treatment variables. The right end of each curve is based on from 19 to 22 <u>S</u>s. The left end of each curve is based on from 11 to 14 <u>S</u>s, i.e., less than half of their respective samples.



PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
CHANCE
TWO-DIMENSION Ss
FOUR-DIMENSION Ss
8 TRIAL CRITERION Ss
11 TRIAL CRITERION Ss
10
5
0
TRIALS BEFORE CRITERION

Figure 12. Forward reversal curves of the two treatment groups predicted to show the greatest and least plateau effects. Increased dimensionality and an easy criterion are supposed to lead to the greater plateau.



100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
5
10
15
20
25
30
TRIALS
CHANCE
4D-8C
2D-11C

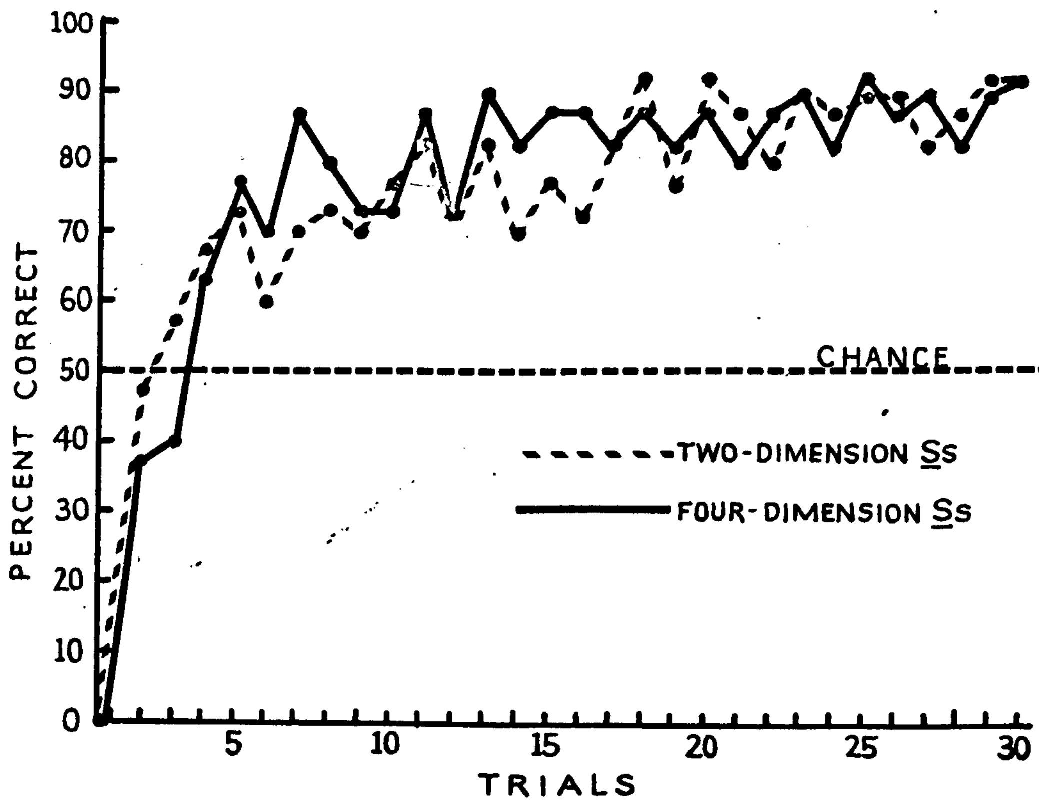
that the apparent plateau is due to anything more than random error in the data.

The forward reversal curves in Figures 13 and 14 are derived from groups combined according to treatment variables. Neither of these figures offers any substantial support for the plateau predictions.

In Figure 15 are both a backward and a forward curve. The forward curve is based on all 60 Ss combined. The backward curve is based upon those 41 Ss who needed more than one trial for reversal. These two curves provide a dramatic illustration of how misleading a forward curve may be. Instead of a gradual rise to criterion, the backward curve indicates only 54% correct just two trials before the criterial run.

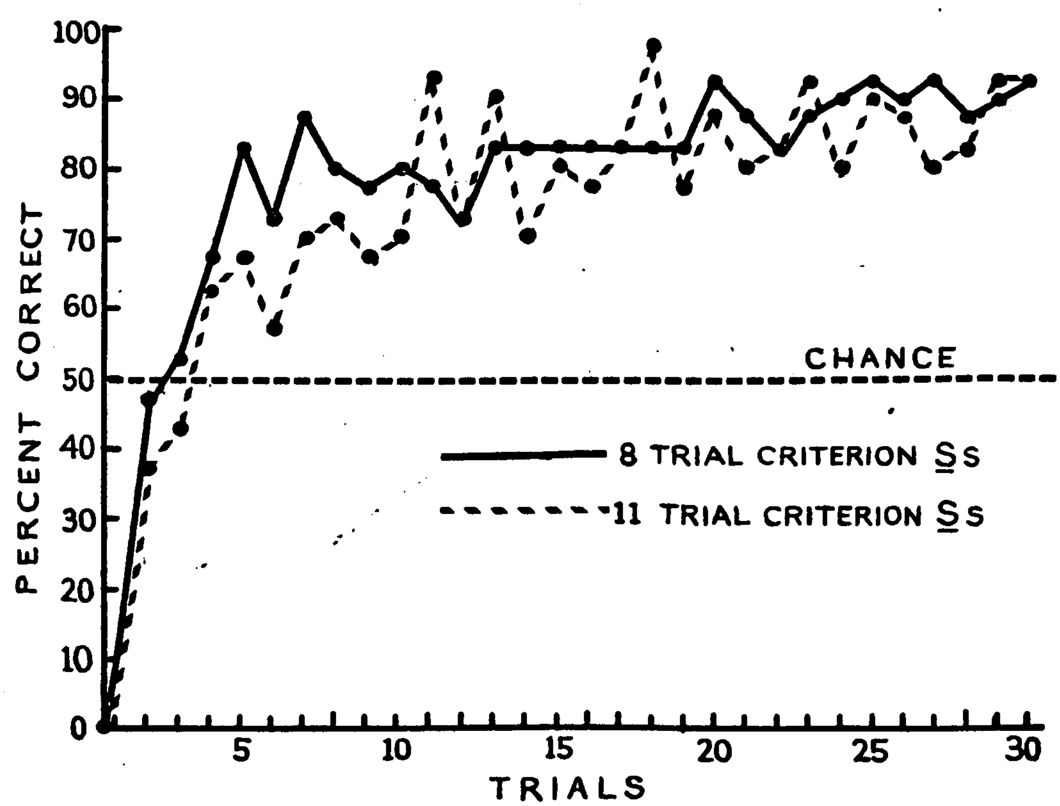
The reversal performances of the high and low MA groups are compared in Figures 16 and 17. The forward curves in Figure 16 would seem to indicate a difference in their rates of reversal; however, as this is a group curve it does not reveal the pattern of individual Ss. The backward curves in Figure 17 show that Ss from both groups performed near chance till the very end. What is more important for the purposes of this study is that there is no substantial indication that the curves in Figure 16 differ in their underlying functions; that is, both appear to be negatively accelerated throughout, with no inflection in either. Thus,

Figure 13. Forward reversal curves of groups combined according to stimulus-dimensionality. Each curve is based on the data of 30 Ss.



100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
5
10
15
20
25
30
TRIALS
CHANCE
TWO-DIMENSION Ss
FOUR-DIMENSION Ss

Figure 14. Forward reversal curves of groups combined according to their pre-reversal criteria. Each curve is based on the data of 30 Ss.



100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
CHANCE
8 TRIAL CRITERION Ss
11 TRIAL CRITERION Ss
5
10
15
20
25
30
TRIALS

Figure 15. Reversal curves based on the combined data of all four treatment groups. The forward curve is based on the data of 60 Ss. The backward curve is based on the data of 41 Ss. Placement and length of the backward curve were determined by the median number of trials-to-criterion in reversal.

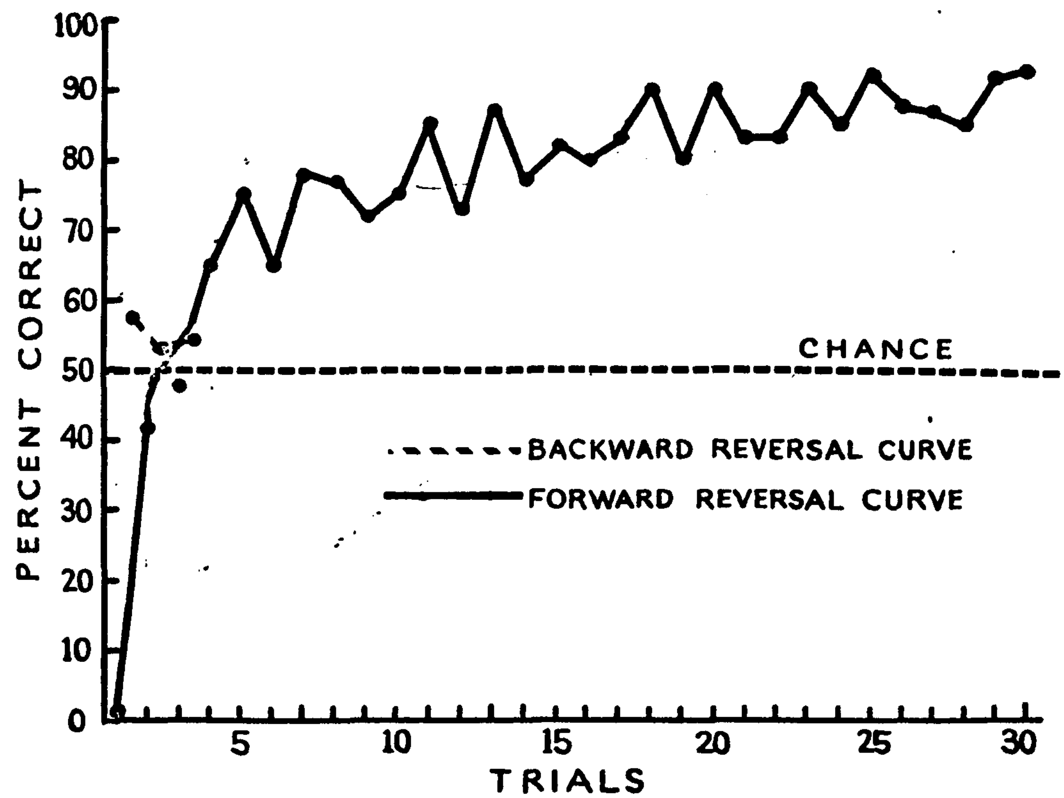
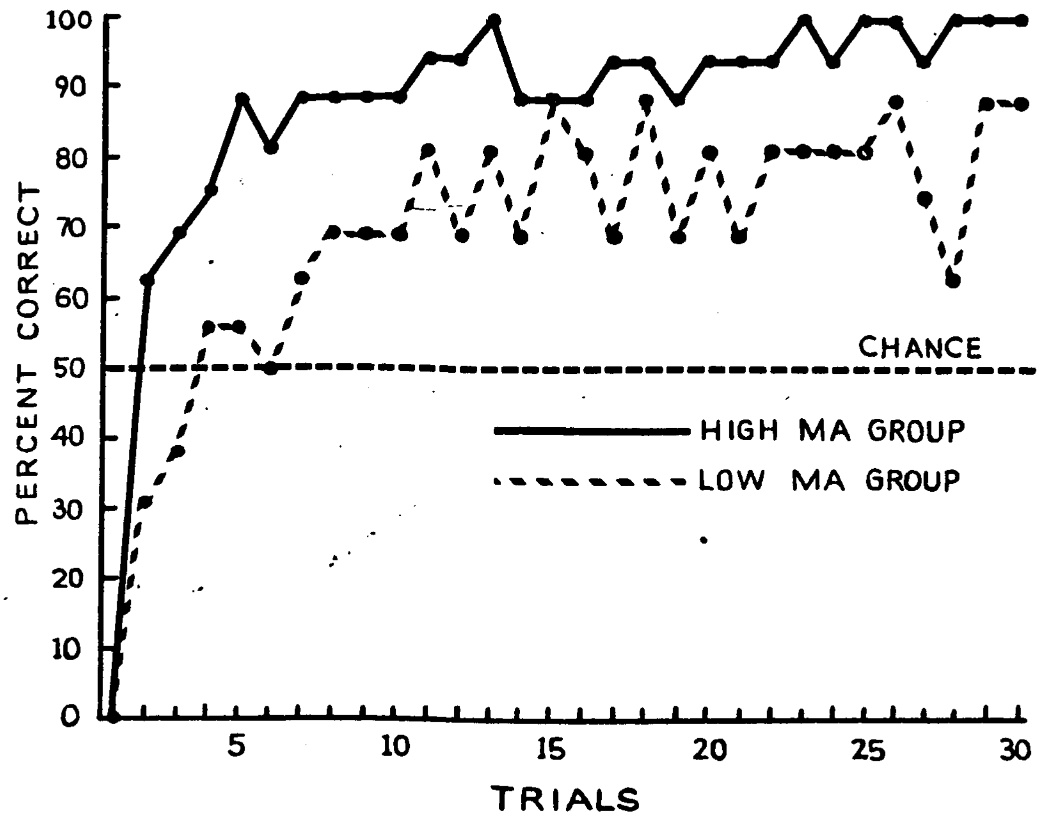
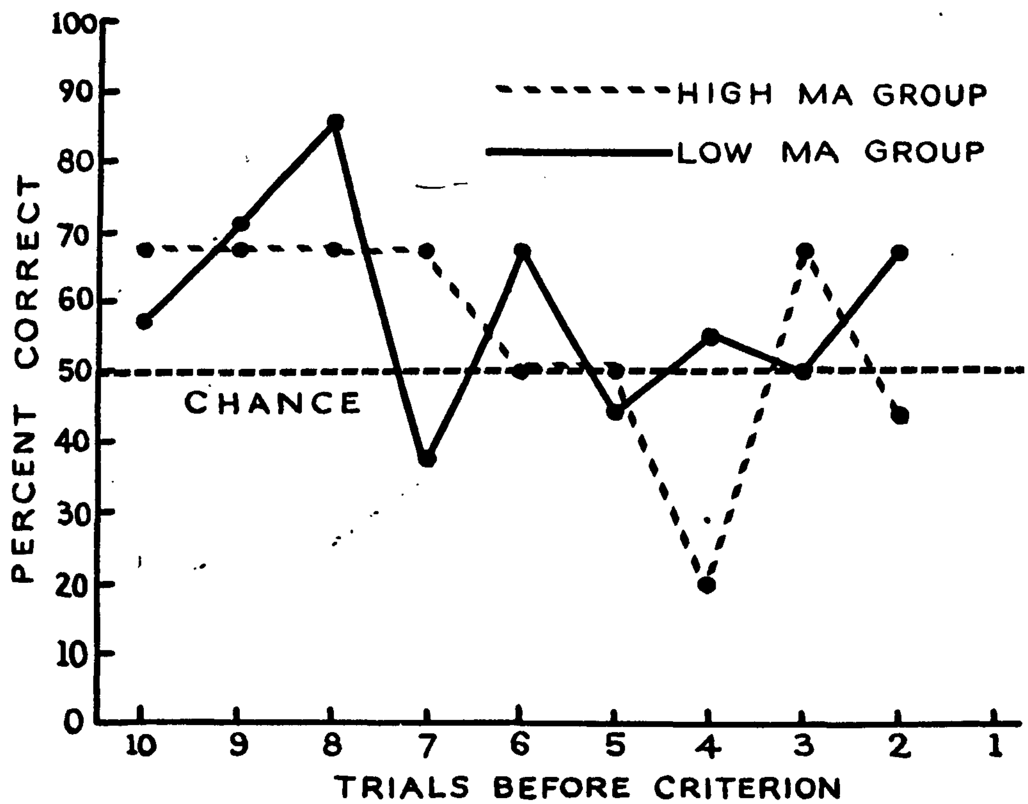

100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
CHANCE
BACKWARD REVERSAL CURVE
FORWARD REVERSAL CURVE
5
10
15
20
25
30
TRIALS

Figure 16. Forward reversal curves of high and low MA groups. Each curve is based on the data of 16 Ss.



100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
CHANCE
HIGH MA GROUP
LOW MA GROUP
5
10
15
20
25
30
TRIALS

Figure 17. Backward reversal curves of high and low MA groups. Each curve is based on the data of 16 Ss at its right end, and fewer Ss as it moves toward the left.



HIGH MA GROUP
LOW MA GROUP
PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
CHANCE
10
9
8
7
6
5
4
3
2
1
TRIALS BEFORE CRITERION

the fact that Ss in this research were of higher MA than the Zeaman and House Ss might not be relevant to the predicted mid-reversal plateau.

Mid-reversal plateaus. Two methods were devised to measure the size of the mid-reversal plateau, should it have manifested itself in the data. One procedure used individual data in an analysis of variance; the other compared group statistics.

Each individual S provides a sequence of right and wrong responses prior to meeting the criterion in reversal. This may be codified as a series of ones and zeroes. It is assumed that at the start of reversal training the probability of a correct response will be quite low. When the criterion is met, this probability should be high. If a plateau exists in the reversal curve, or merely if reversal learning is incremental, there should be a period when the probability of a correct response is intermediate. This being the case, fluctuations should occur between the zeroes and ones that represent the response sequence. The longer the plateau in the reversal curve, if there is one, the larger the number of fluctuations that would be expected.

The number of fluctuations may be computed in various ways. There are the total number of changes, or the number of changes from zero to one (wrong to right), or the number of runs. Happily, they all have additive relationships with each other and amount to the same thing. A "runs" test should indicate that, when the probabilities of two different events are equal, runs are more plentiful. The same is true for fluctuations.

Shepp and Turrisi (1969b) have suggested a different measure of plateau length. They count the number of trials between the first correct response and the last error in reversal. This is a grosser measure which logically should be subject to greater influence from chance factors. The use of runs or fluctuations is preferable.

Analysis of variance showed no significant differences between the four main treatment groups with regard to fluctuations or runs in the reversal data. This was also true when reciprocals were used. The analysis is summarized in Table 10. The means and medians are shown in Table 11.

When MA was added as a third factor, with four Ss in a group, a significant difference did occur, but only on the MA variable. This was the only instance of a significant difference on a plateau measure when dealing strictly with "learning" Ss. It was probably not unrelated to the fact that the 16 high MA Ss had a median of 2 trials-to-criterion in reversal, while the 16 low MA Ss had a median of 6.5 trials-to-criterion. Table 12 gives this analysis.

Table 10

Analysis of Variance for Wrong-to-Right Fluctuations in Reversal Data

Source	SS	df	MS	F
Criteria (C)	81.66666	1	81.66666	2.42
Dimensions (D)	17.06667	1	17.06667	.51
C x D	.26667	1	.26667	.01
Within Groups	1891.73208	56	33.78093	
Total	1990.73208	59		

Table 11

Means and Medians of Runs in Reversal Data

Group	Means	Medians
Two Dimensions, Easy Criterion (2D-8C)	8.0	4
Two Dimensions, Hard Criterion (2D-11C)	12.4	6
Four Dimensions, Easy Criterion (4D-8C)	5.6	2
Four Dimensions, Hard Criterion (4D-11C)	10.5	4

Table 12

Analysis of Variance for Reciprocal of Wrong-to-Right Fluctuations in Reversal with MA as a Third Factor

Source	SS	df	MS	F
Mental Age (M)	.58401	1	.58401	4.28*
Criteria (C)	.29968	1	.29968	2.19
Dimensions (D)	.03025	1	.03025	.22
M x C	.02792	1	.02792	.20
M x D	.00687	1	.00687	.05
C x D	.02197	1	.02197	.16
M x C x D	.21886	1	.21886	1.60
Within Groups	3.27744	24	.13656	
Total	4.46700	31		

* $p < .05$

Trials-to-criterion scores and individual plateau measures are closely correlated. According to Zeaman and House (1963) it is the plateau effect that is responsible for some groups requiring more trials. In order to separate the effects of these two variables, analysis of covariance was used. It is recognized that this was not, strictly speaking, a legitimate use of the covariance technique, as both of the variables involved were dependent variables. Nevertheless, each of the two variables was used, in turn, as the dependent variable and as the covariable whose effect was to be partialed out of the variances. Since the groups did have correlated means on these two variables, the procedure adjusted both the among groups and the within groups variance estimates, and therefore did not necessarily increase the likelihood of significance. In fact, there were no significant differences for any main effect or interaction, regardless of which was the dependent variable.

The method for measuring the mid-reversal plateau in group data involves taking the heights of the points on the reversal curve and treating them as a set of scores. If there is a plateau in the reversal curve, the scores should cluster about their mean. If there is no such plateau, then the scores should be more widely dispersed. Thus, the variance of these scores should be a measure that is sensitive to the size of any plateau that might exist. Such variances may be compared by means of the F ratio. Inasmuch as the measures from a reversal curve are not independent, this must be regarded as a very approximate test.

The reversal functions which Zeaman and House used to illustrate the mid-reversal plateau were in the form of forward learning curves. Except for one computer-generated example where the parameters were chosen to maximize the plateau, the period of positive acceleration was generally within the first 10 or 20 trials of the forward reversal curve. Inspection of the present data also indicated that, if there was a plateau at all, it would be found within the early trials.

Variances were computed using Trials 2 through 11, and again using Trials 2 through 21. The first trial was not used because in all cases it had a value of zero and would have attenuated any differences. The variances of the two curves in each of Figures 12, 13, 14, and 16 were compared. No significant differences were found between the variances of any of the four pairs of curves. In the case of the dimensionality variable, the direction of the differences was opposite to that predicted by Zeaman and House. In the case of the MA variable, the direction of the differences was that the low MA group had the higher variance, indicating less of a plateau.

Having established that the four groups in this study did not differ significantly in their reversal performance, there still remains the question of whether any plateau was apparent in the data. Referring to Figure 15 which shows the forward reversal curve of all 60 Ss combined, if one were to limit one's view to just the first 10 trials, it might indeed apear that there was a plateau of sorts. However, upon broadening the perspective to include 20 or more trials, what appeared as a plateau could just as easily be the result of random error. The overall function would appear to be negatively accelerated throughout.

Stationarity in reversal. Tests for stationarity were performed on the backward data of variously constituted groups of Ss and on the data of individual Ss.

Tests on group data used the last 10 trials before the last error in reversal. The groups studied included each of the treatment groups, all four groups combined, the two-dimension Ss, the four-dimension Ss, the hard-criterion Ss, the easy-criterion Ss, the high MA Ss, and the low MA Ss. None of the resulting chi-square values were significant.

Only 23 Ss provided enough data in reversal for individual stationarity tests. Only 2 of these produced significant chi-squares. Although they exceed the number expected through alpha error, they could hardly be said to set a trend.

Pseudo-Learners

Is it possible that the 11 Ss who were disqualified as pseudo-learners were in fact slow learners? It could be argued that not all of these Ss deserved to be classified as pseudo-learners. Among this group may have been some who had achieved an intermediate stage of learning prior to reversal. If this were the case, it could account for at least some of the differences between the present findings and those of Zeaman and his associates. However, there is evidence to the contrary.

If the rate parameters are indeed similar for most S groups, as Zeaman and House assert, and if the so-called pseudo-learners had actually begun their final steep climb to mastery when they met their respective criteria, then it would be expected that they would show dramatic improvement in the post-criterial test session. But when the frequencies of correct responses in the last 30 trials before criterion were compared with those of the test session, using chi-squares, there were no significant differences, whether testing individual Ss or all 11 together. The mean proportion correct in the pre-criterial block of trials was .56, while in the test session it was .62.

A one sample t-test was used to determine if the pseudo-learners exceeded chance performance in the test session. With 10 degrees of freedom, the value of t was 3.38 and was significant at the 5% level. Also, 3 of the 11 pseudo-learners were found to have performed significantly above chance in the test session, based on individual chi-square tests.

Taking the 11 pseudo-learners as a group, stationarity tests were done using the last 30 trials of acquisition, and also using the 30 post-criterial trials that constituted the test session. Neither chi-square was significant, indicating no significant improvement within either block of trials.

It is recognized that the lack of significant differences does not establish the null hypothesis. While the improvement between the pre-criterial and post-criterial performances was slight, it could nevertheless represent a real gain. One more bit of information is relevant to the question.

Every one of the 60 Ss judged to be "true learners" began reversal with an error. On the second and third trials of reversal, still, less than half gave correct responses. On the other hand, 6 of the 11 pseudo-learners gave correct answers on the first reversal trial. Furthermore, considering only the 4 pseudo-learners who did best in the test session, in each of the first two trials of reversal there were exactly 50% correct responses.

The weight of the evidence indicates that most of the pseudo-learners were indeed correctly classified. There would seem to be no compelling reason to assume that any substantial learning had occurred prior to reversal.

As might be expected, nine of these Ss came from groups that were given the easier criterion. Eight came from groups exposed to four-dimensional stimuli. Most of them came from the lower end of the MA spectrum, with their median MA being less than five years. Therefore, it would seem reasonable to expect that studies which used easier criteria and lower MAs than did this study would produce a still greater proportion of such pseudo-learners.

It is relevant to ask what would have happened if these 11 Ss had not been discarded, but were grouped with the others. In trying to answer this question, it would not do merely to add them to the other groups. Instead, they must be used to replace Ss from the other groups. Only then would they have their appropriate weight in the revised data.

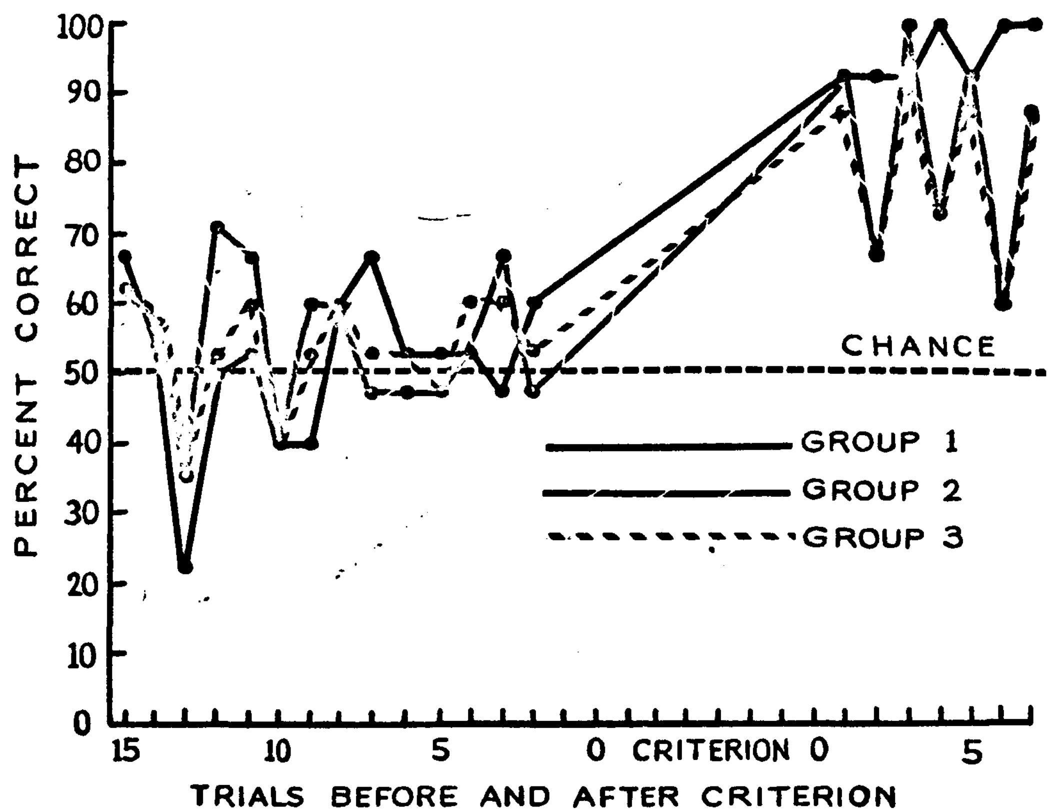
Acquisition. With regard to acquisition data, if backward curves were constructed in the manner suggested by Hayes

(1953) and possibly adopted by Zeaman and House (1963), it seems clear that the inclusion of the pseudo-learners would lower the success rate in post-criterial trials, and thereby make the slope less steep.

To illustrate this effect upon the backward acquisition curve, the data of just one group will be used. There were seven pseudo-learners, all with MAs less than six years, in the group with an eight-trial criterion and four stimulus-dimensions. In selecting which of the true learners to replace with pseudo-learners, two strategies present themselves. On the one hand, since this study used stratified sampling according to MA, the pseudo-learners should replace Ss of similar MA. They would be the lowest seven in the group. On the other hand, replacing the upper seven would yield a more homogeneous group, similar in MA to those studied by Zeaman and House. Both of these approaches were adopted. Figure 18 illustrates the effect of the pseudo-learners upon the backward acquisition curve of this group.

Reversal. Three of the pseudo-learners did not achieve criterion during reversal training. Therefore, it is impossible to say what their trials-to-criterion scores or plateau measures might ultimately have been. Fortunately, all three were from the same treatment group, which did

Figure 18. Backward acquisition curves using post-criterial trials. The three differently constituted groups were exposed to the same experimental treatment. Group 1 consists of the original 15 Ss judged to be true learners. Group 2 has the bottom 7 Ss in MA replaced by pseudo-learners. Group 3 has the top 7 Ss replaced.



100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
CHANCE
GROUP 1
GROUP 2
GROUP 3
15
10
5
0 CRITERION 0
5
TRIALS BEFORE AND AFTER CRITERION

have higher means than the other groups, so that it may safely be said that the differences between group means should have been larger than they were. For the following analyses, the pseudo-learners were substituted for the true learners who were closest to them in MA.

Analyses of variance were done using trials-to-criterion scores, plateau measures, and the reciprocal of each. There were no significant differences among the groups using reciprocals; however, using either the trials-to-criterion scores or the plateau measures, significant differences were found between the two-dimension and the four-dimension groups, with the four-dimension groups needing more trials or having longer plateaus. Table 13 presents the results of the analyses of variance for both trials-to-criterion and plateau measures.

Once again, analysis of covariance was used, albeit improperly, to separate the variables of trials and plateaus. Each of the variables took its turn as the dependent variable and as the co-variable. Under this procedure, there were no significant differences at all. The loss of significance between the regular analysis of variance and the analysis of covariance would reflect the fact that the groups had correlated means on these two variables.

Group measures of plateaus, as before, were obtained by computing the variances of the data points of the forward

Table 13

Analyses of Variance for Reversal Data

with Pseudo-Learners Substituted for True Learners

Source	SS	df	MS	F
Trials-to-Criterion				
Criteria (C)	17819.3	1	17819.3	2.54
Dimensions (D)	33512.1	1	33512.1	4.78*
C x D	19512.1	1	19512.1	2.78
Within Groups	392420.2	56	7007.5	
Total	463263.7	59		
Plateau Measures[1]				
Criteria (C)	1195.4	1	1195.4	2.76
Dimensions (D)	1773.3	1	1773.3	4.10*
C x D	1216.7	1	1216.7	2.81
Within Groups	24248.1	56	433.0	
Total	28433.5	59		

[1] The specific plateau measure that was used in this analysis was the number of changes from wrong to right. The same results would have occured with any of the plateau measures discussed.

* $p < .05$

reversal curves. Figures 19, 20, and 21 show the forward reversal curves which resulted when pseudo-learners were substituted for true learners. For each pair of curves, variances were computed using Trials 2 through 11 and again using Trials 2 through 21. Of the six pairs of variances, one pair differed significantly. Based on 20 trials, Ss who had the easier criterion in acquisition had a significantly lower variance in reversal than did Ss who had the more stringent criterion. In other words, the data points of the Ss with the easy criterion clustered together as would be expected if there were a plateau. The F ratio was 2.24, with 19 degrees of freedom in the numerator and in the denominator, which is significant at the 5% level.

Once again it is worthwhile to look at the results of the one treatment group that generated the greatest number of pseudo-learners. This was the group with the eight-trial criterion and four stimulus-dimensions. Figure 22 shows the forward reversal curves of this group, with and without the substitution of the pseudo-learners. Using just the first 10 trials, the variance of the original group of true learners is significantly larger than that of the group when pseudo-learners are substituted. This was the case whether the pseudo-learners replaced the top seven or the bottom seven of the original group. Table 14 shows the F values and variances.

Figure 19. Forward reversal curves of the two treatment groups predicted to show the greatest and the least plateau effects. In these curves pseudo-learners were substituted for true learners of similar MA. Each curve is based on the data of 15 Ss.

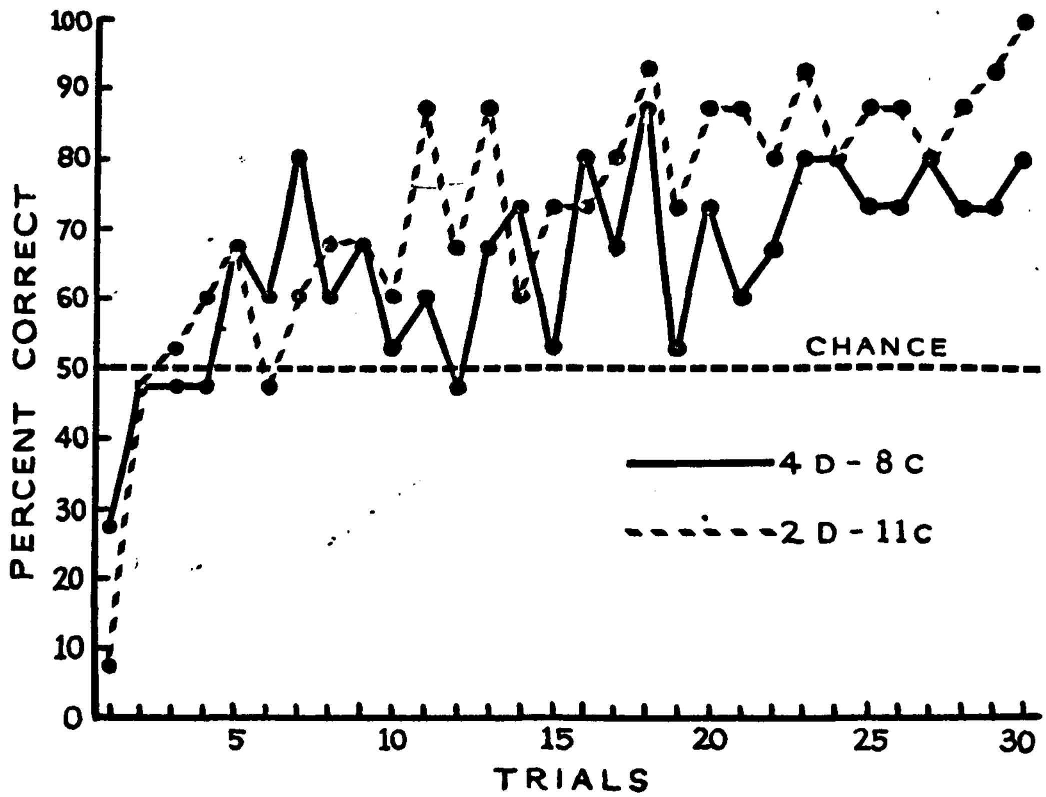

100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
CHANCE
4 D - 8 C
2 D - 11C
5
10
15
20
25
30
TRIALS

Figure 20. Forward reversal curves of groups combined according to stimulus-dimensionality, with pseudo-learners substituted for true learners of similar MA. Each curve is based on the data of 30 Ss.

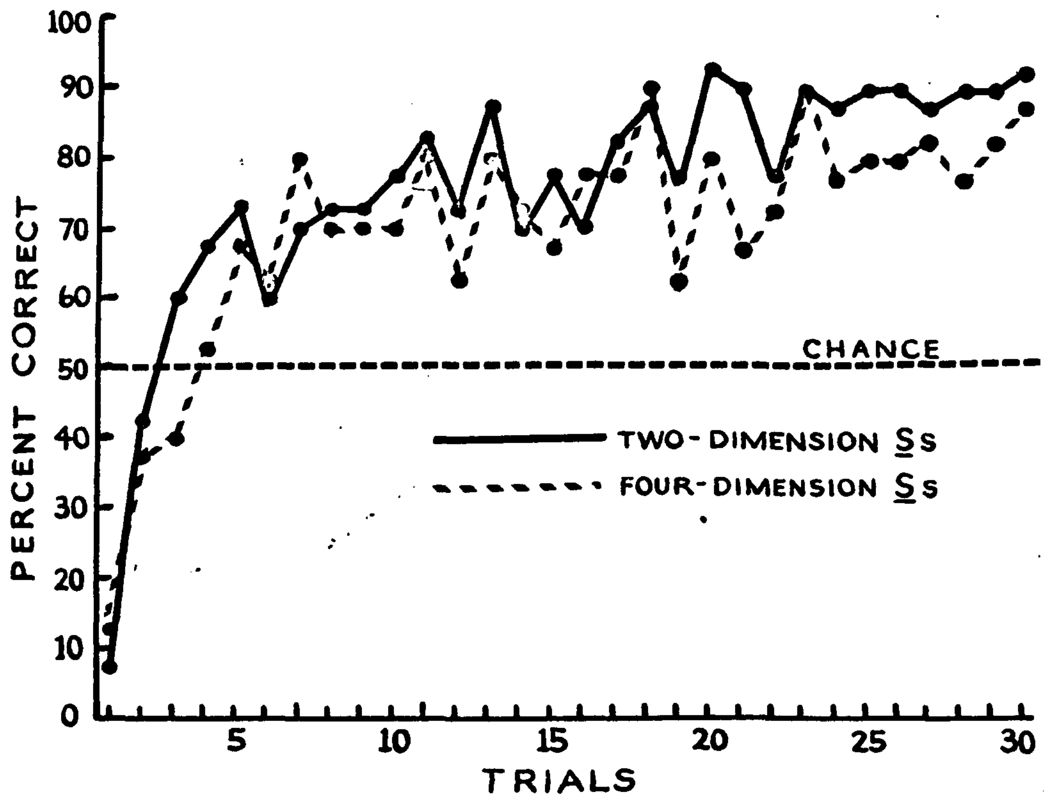
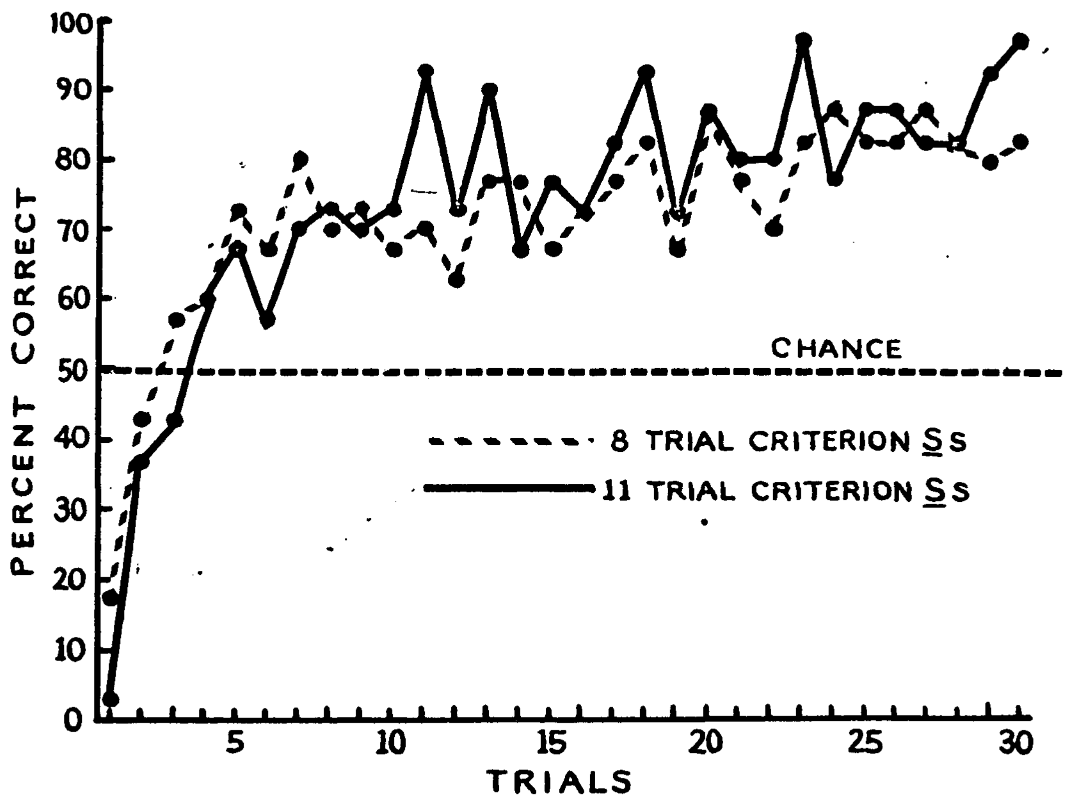

100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
5
10
15
20
25
30
TRIALS
CHANCE
TWO-DIMENSION Ss
FOUR-DIMENSION Ss

Figure 21. Forward reversal curves of groups combined according to their pre-reversal criteria, with pseudo-learners substituted for true learners of similar MA. Each curve is based on the data of 30 Ss.



100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
CHANCE
8 TRIAL CRITERION Ss
11 TRIAL CRITERION Ss
5
10
15
20
25
30
TRIALS

Figure 22. Forward reversal curves of differently constituted groups exposed to the same experimental treatment. Group 1 consists of the original 15 Ss judged to be true learners. Group 2 has the bottom 7 Ss in MA replaced by pseudo-learners. Group 3 has the top 7 Ss replaced.

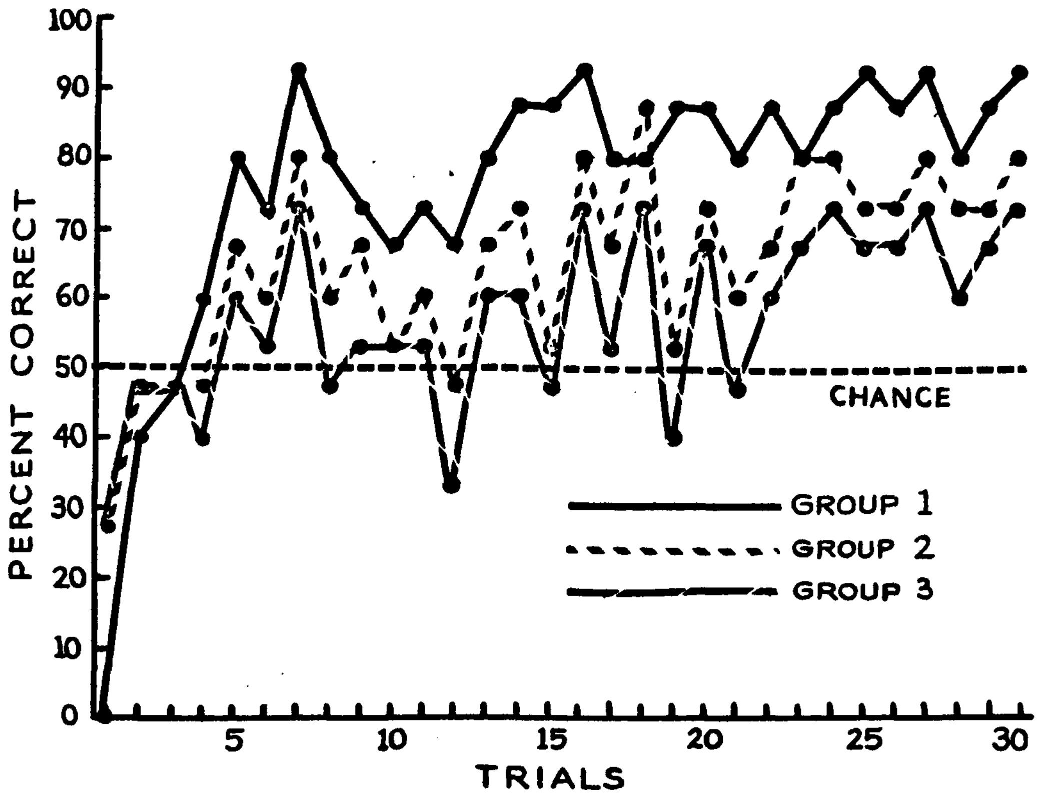

100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
5
10
15
20
25
30
TRIALS
CHANCE
GROUP 1
GROUP 2
GROUP 3

Table 14

Comparison of Variances from Figure 22

Using the First 10 Trials

Curve	Variance	df	F
Group 1: True Learners	4.022	9	(numerator)
Group 2: Mixed Ss	2.214	9	3.30*
Group 3: Mixed Ss	1.841	9	4.77*

* $p < .05$

Discussion

The major findings of the present research to be discussed were the apparent stationarity in the original acquisition data prior to solution, the absence of evidence of a mid-reversal plateau, the failure of the reversal data to be related to the experimental variables, and the role of pseudo-learners in producing mid-reversal plateaus.

Possibly the most important finding of this research pertained to stationarity. In both acquisition and reversal, as Ss neared their criterial runs the probability of a correct response did not increase. By and large, it stayed near the chance level until the last error was made. This would seem to contradict the notion that discrimination learning proceeds in a gradual or incremental fashion, and suggests instead an all-or-none learning process possibly involving hypothesis testing. There are currently several theoretical models of the all-or-none type from which to choose; see Bower (1967), Estes (1970), Levine (1966), Restle (1962), Shepp, Kember, and Anderson (1972), Trabasso and Bower (1968), and Theios (1963).

Although the results presented by Zeaman and House (1963) and Fisher and Zeaman (1973) would seem to support the incremental viewpoint, those results could well have been artifactual. First, it has already been noted that the Hayes (1953) method of constructing backward learning curves, if followed to the letter, necessitates that the resulting curves be continuous due to the interpolation between pre- and post-criterial trials. Second and more likely is the fact that Zeaman and his associates most frequently used a criterion of 20 out of 25 trials in a daily run. If the S solved the problem near the end of a daily run, the last trials of the day would have a high rate of success, but the S would not begin his criterial block of trials until the next day. It has been shown here that a similar effect may be obtained when the criterion is a run of correct responses.

The Zeaman and House theory predicts the occurence of a mid-reversal plateau if appropriate parametric conditions exist. In the present research, in which precautions were taken to "screen out" pseudo-learners, the plateau in the forward reversal curve failed to appear. Furthermore, there was no relation between trials-to-criterion in reversal and the experimental variables. One may be tempted to attribute this to the existence of improper parametric conditions. However, when the pseudo-learners were substituted for true learners, plateau measures indicated significant differences between the treatment groups in the predicted directions. The implication is that in those studies where a mid-reversal plateau was apparent the inclusion of pseudo-learners in the sample may have been responsible for it.

Fisher and Zeaman (1973) cite as empirical evidence of the plateau effect the work of Zeaman and House (1963) and especially that of Shepp and Turrisi (1969b). Both bodies of evidence suffer from the same problem, namely, the use of very weak pre-reversal criteria in order to produce the plateau. As noted before, Zeaman and House used a criterion of six successive correct responses in 25 trials, which has a chance probability of more than .15 in one session, and quickly rises to 1.00 with successive days of training. The fact that some Ss are non-learners in spite of this can be explained by the fact that few Ss respond in a truly random fashion even prior to solution.

Shepp and Turrisi (1969b) used a criterion of 9 correct in any 10 successive trials, given 40 trials a day for 10 days. The more stringent criterion of 9 consecutive correct in 40 trials has a daily probability greater than .03 (Grant, 1947). In 10 days the probability of 9 consecutive correct happening at least once is more than .27. The criterion used by Shepp and Turrisi would have a still greater probability.

Although Shepp and Turrisi (1969b) do not say how many of their Ss failed to maintain criterion during overtraining, in a different paper (1969a) they criticize another researcher who eliminated Ss that failed to maintain criterion. It was pointed out that this procedure could introduce a selectivity bias, since one of the groups was not overtrained at all. Still, it would be of interest to know what proportion of the Shepp and Turrisi (1969b) Ss failed to maintain a criterial level of performance. Of interest is their report that 6 of their 38 Ss failed to reverse, using the same easy criterion. In the absence of more information, one can only guess at how many Ss met the criterion only by chance.

It could be argued that in the present research every S received a modest amount of overtraining, consisting of 30 trials with 15 reinforced and 15 not. Although Ss with less than 26 correct in this test session were disqualified as pseudo-learners, the means of the treatment groups for MA and IQ did not indicate a selectivity bias to be operating.

From the perspective of the Zeaman and House theory, the overtraining of the test session may have brought the P_o values of the "true learners" too close to one for any plateaus to materialize in the reversal data, regardless of the criterial condition or stimulus-dimensionality. Thus, among the disqualified pseudo-learners would be precisely those Ss whose P_o values had not yet reached one, and who were at an intermediate stage of learning when reversal commenced. In that case, it would come as no surprise that the substitution of the pseudo-learners for true learners increased the magnitude of the plateau effect.

Specifically, 3 of the 11 pseudo-learners responded at a level significantly higher than chance during the test session, while the group as a whole gave 62% correct responses. Thus, it could be argued that not all of these Ss are properly "pseudo-learners." This argument loses some weight when one examines their data more closely. Stationarity was found for the 30 trials preceding the criterial run and again for the 30 trials following it. For none of the 11 Ss was there a significant gain in probability across the two sessions. On the first trial of reversal, as a group they were already at the chance level. While it is clear that the pseudo-learners did respond at a higher than chance level in the test session, it does not appear to be possible to find parameter values within the Zeaman and House model which would account for the behavior found. A more likely explanation is that the Ss were responding above chance through the use of "immediate memory." This explanation is similar to that of paired-associates learning with forgetting; however, since forgetting seems to be complete over longer time spans, a short-term memory mechanism seems to be more appropriate. All in all, then, there is no compelling argument for any change in the P_0 values of the pseudo-learners prior to reversal.

However, the argument that the true learners were subject to overlearning has yet to be answered. It is indeed a possibility that the 30 post-criterial trials, only 15 of which were reinforced, could have washed out any likely effects of the criterial variable. It is also a real possibility that the P_o values of the true learners were simply too high for the mid-reversal plateau to materialize. Nevertheless, it did materialize when the pseudo-learners were included in the analysis. At the least the present research casts suspicion upon the findings of research not employing safeguards against the inclusion of pseudo-learners. While the present controls were not ideal, the burden of replicating earlier findings with controls for pseudo-learners lies with the supporters of the Zeaman and House model.

As could be deduced from the stationarity data, reversal plateaus did indeed occur in the backward curves. Shepp and Turrisi (1969b) and Fisher and Zeaman (1973) cite this type of plateau as further evidence in favor of the Zeaman and House theory. Actually, a number of theories could predict backward curve plateaus.

Zeaman and House (1963) stated that the occurence of a plateau in the forward reversal curve would have probative value for their theory. It was this prediction that was relatively unique and unexpected. However, the occurence of a plateau in the backward data has no bearing on this point because forward and backward curves aren't comparable. In the forward curve, each point is equally influenced by each S, but in the backward curve the slow learners have a greater influence as the data of the fast learners quickly run out. What the backward plateau does indicate may be merely that some Ss require numerous trials between the time that they cease responding according to their original solution and the time that they discover the new solution. The status of P_o is not revealed. No stronger statement would seem justifiable.

In conclusion, it is suggested that both the gradual learning functions and the mid-reversal plateaus that are evidenced in the data of Zeaman and House (1963), Shepp and Turrisi (1969b), and Fisher and Zeaman (1973) are methodological artifacts. By and large, the present data support an all-or-none conception of discrimination learning, and, as such, argue against the gradual learning assumptions of the Zeaman and House model.

Summary

Acquisition and reversal in a discrimination task were studied under varied conditions of stimulus-dimensionality and pre-reversal criteria using retardates. Pre-solution performance in both acquisition and reversal was stationary and indicated that learning occured in an all-or-none fashion. It was suggested that data supporting the Zeaman and House model for discrimination learning were due to the inclusion of "pseudo-learners."

References

Atkinson, R. C., & Shiffrin, R. M. Human memory, a proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), The psychology of learning and motivation: Advances in research and theory. Vol. 2. New York: Academic Press, 1969, Pp. 89-195.

Bourne, L. E., Jr., & Restle, F. Mathematical theory of concept identification. Psychological Review, 1959, 66, 278-296.

Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), The psychol-of learning and motivation: Advances in research and theory. Vol. 1. New York: Academic Press, 1967, Pp. 299-325.

Estes, W. K. Learning theory and the new "mental chemistry." Psychological Review, 1960, 67, 207-223.

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, 1973, Pp. 169-256.

Grant, D. A. Additional tables on the probability of "runs" of correct responses in learning and problem solving. Psychological Bulletin, 1947, 44, 276-279.

Hayes, K. J. The backward curve: A method for the study of learning. Psychological Review, 1953, 60, 269-275.

Levine, M. Hypothesis behavior by humans during discrimination learning. Journal of Experimental Psychology, 1966, 71, 331-338.

Lovejoy, E. Analysis of the overlearning reversal effect. Psychological Review, 1966, 73, 87-103.

Restle, F. The selection of strategies in cue learning. Psychological Review, 1962, 69, 329-343.

Shepp, B. E., Kember, D. G., & Anderson, D. R. Selective attention and breadth of learning: An extention of the one-look model. Psychological Review, 1972, 79, 317-328.

Shepp, B. E., & Turrisi, F. D. Effects of overtraining on the acquisition of intradimensional and extradimensional shifts. Journal of Experimental Psychology, 1969, 82, 46-51. (a)

Shepp, B. E., & Turrisi, F. D. Effects of training procedures on transfer of observing and instrumental responses in discriminative reversals. Journal of Comparative and Physiological Psychology, 1969, 69, 522-527. (b)

Sidman, M. A note on functional relations obtained from group data. Psychological Bulletin, 1952, 49, 263-269.

Siegel, S. Nonparametric statistics. New York: McGraw-Hill, 1956.

Theios, J. Simple conditioning as two-stage all-or-none learning. Psychological Review, 1963, 70, 403-417.

Trabasso, T. Stimulus emphasis and all-or-none learning of concept identification. Journal of Experimental Psychology, 1963, 65, 398-406.

Trabasso, T., & Bower, G. H. Attention in learning: Theory and research. New York: Wiley, 1968.

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.

Autobiographical Statement

The candidate, Don Frederick Gonella, was born in Santa Monica, California on May 10, 1940. He attended Occidental College in Los Angeles and received a Bachelor of Arts degree in 1962. He then attended the California State University at Los Angeles and was awarded the degree of Master of Arts in 1965. While continuing his graduate education at the City University of New York, he has held teaching positions at Queens College, Upsala College, and Green Mountain College.