"ERRORLESS" DISCRIMINATION LEARNING AND TRANSFER IN THE MENTALLY RETARDED

THESIS FOR THE DEGREE OF M. A.

MICHIGAN STATE UNIVERSITY

LAWRENCE JOSEPH KRIPS

1971

ABSTRACT

"ERRORLESS" DISCRIMINATION LEARNING AND TRANSFER IN THE MENTALLY RETARDED

By

Lawrence Joseph Krips

There are many theories postulating the learning process involved in acquired a discrimination. Acquisition usually included S- responding, a method which does not allow that stimulus to retain its neutral value, a factor which may be detrimental to the total discrimination learning process.

Several studies have attempted not permitting this negative cue value to become assigned to S-. This type of learning is usually achieved by physically not allowing an S- response to occur; by not presenting S- during acquisition of S+ responding; or by using an early introduction of S- in the presence of S+ with the assistance of fading.

The present study is concerned with a modification of Terrace's errorless approach applied to a visual two discriminanda problem and transfer with twenty-nine moderately retarded children. The fading variable was a gradual increase or decrease along the brightness dimension. Data was collected on the number of S- responses emitted and the latency difference between S+ and S- responses.

The experiment had three sections. Original learning was composed of two groups learning a discrimination of a projected sketch of a man (S+) and a projected sketch of a

dog (S-). The control group learned with each stimulus at equal intensity for 100 trials. The experimental group faded in S- in four incremental phases during 100 trials. In transfer learning, a word replaced its respective sketch. For this task the experimental group evenly and randomly divided into those who would be presented a typical discrimination situation and those in which S would be decrementally faded out and then S' incrementally faded in. The posttest was simply a 25 trial (full illumination) presentation of OL material to the experimental group.

As expected, more Ss learned OL material by fading than by a typical presentation. The fading, however, showed evidence of producing more errors than the non-fading method. Despite this unexpected finding, transfer by fading also produced less failures than the typical task. The experiment was further confounded by transfer fading Ss acquiring the original discrimination faster than non-fading transfer Ss.

The posttest strongly indicated no response disruption of OL if the transfer material was acquired by a non-fading approach. No significant differences were found for either experimental group during this recall section.

The variance between this study's findings and Terrace's errorless learning approach were analytically discussed in terms of both other studies and of inherent factors of the present experiment. The relation between this investigation and programmed instruction material was also discussed.

"ERRORLESS" DISCRIMINATION LEARNING AND TRANSFER IN THE MENTALLY RETARDED

By

Lawrence Joseph Krips

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Psychology

1971

ACKNOWLEDGMENTS

The author would like to express his appreciation to Dr. M. Ray Denny and Dr. Rilling who served as thesis committee members. Special appreciation is extended to Dr. Lester Hyman, the chairman of this committee whose guidance and critique were invaluable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

"Errorless" Discrimination Learning and Transfer in the Mentally Retarded

Lawrence Joseph Krips

Michigan State University

In discrimination problems, S's response to the S- is often considered a necessary part of the task. Through nonreward, the response becomes (at least in part) extinguished. This process, however, also neccessitates the assigning of cue value(s) to the S- by S; S- is not neutral. It has been postulated (Holland, 1960; Hull, 1939; Hull, 1949; Terrace, 1963), however, that this value placed on the S- by S may be detrimental to overall discrimination learning. The effect of wrong responses has been postulated to be deleterious to subsequent behavior. In this investigation, an errorless (i.e., little or no S- responses) procedure was utilized with MR children for a discrimination task and a transfer task.

Discrimination learning is most generally defined as learning not to generalize. A response is made to one stimulus and not to another stimulus even though these two stimuli may have some properties in common. That is, S learns not to generalize the cues of one stimulus to that of another, and therefore, avoids indiscriminate (chance) responding to both stimuli. The response made to S+ is correlated with reinforcement, and the response to S- is correlated with nonreinforcement. S+ is then said to have acquired functional control over the response associated

with it.

The answer to the question of how the organism does, in fact, learn a simple discrimination problem has been attempted by several theories. Single unit S-R theories postulate a relationship between the visual stimulus and an overt response. Two-stage theories also assume that an approach response is made to one of the discriminative cues, and further assumes that S first attends to the relevant stimulus dimensions before the occurrence of instrumental learning. That is, an orienting response must be made before any learning can occur.

One theory that breaks down the process still further is elicitation theory (Denny and Adelman, 1955; Denny, 1966). According to this theory, an array elicits covert dimensional responses which produce feedback cues to which overt responses are conditioned. A consistent contiguous relationship of reward (or nonreward) with a stimulus elicits the response. Backchaining is said to occur from the goal cues (S+) back to cues in the starting region. The operant response is associated with kinesthetic cues which in turn act as an S+ for further responses in a chain.

A teaching device based on the above theory has been designed by Denny. The Multiple Differential Response and Feedback Apparatus (MUDRAFA) is basically an errorless procedure tool. Although an incorrect response may be attempted by S, the apparatus, by the use of barriers, does not allow completion of the response thereby eliminating

back-chaining of unreinforced stimuli and eventual learning of a response to an incorrect cue. Crutch cues and immediate knowledge of results are consistent with elicitation theory postulates as leading to rapid learning.

Another theory concerning acquisition of a simple discrimination problem was proposed by Hull (1939, 1943, 1950, 1952) and Spence (1936, 1937, 1937a). Their theories of discrimination learning are based on the following five postulates (Hall, 1966):

> 1. Every reinforced trial leads to an increment in excitatory strength for a given stimulus and its reinforced response.
> 2. Every nonreinforced trial results in an inhibitory increment to a given stimulus and its nonreinforced response.
> 3. Both excitatory and inhibitory tendencies generalize to stimuli along a stimulus continuum.
> 4. There is the algebraic summation of excitatory and inhibitory increments which result in ...
> 5. A discriminatory response based upon these algebraic summations.

According to these postulates then, if there are no generalized inhibitory tendencies, there should be better performance to S+ than if inhibition were present. In other words, conditioning not involving discrimination learning (i.e., differential conditioning) will result in superior response operation to the desired stimulus. Gynther (1957) gave support to this prediction by differentially training one group of Ss and non-differentially training another. The group to which the negative stimulus was never presented (non-differentially trained) during acquisition emitted more conditioned responses to S+ during test trials (both stimuli presented together) than the other group.

Another prediction that may be made from the Hull-Spence theory concerns the similarity of the form of the discriminanda used. The greater the dissimilarity of the stimuli, the more readily the discrimination should be learned. That is, the generalized inhibitory strength to S+ and the generalized excitatory strength to S- should become increasingly weaker as S+ and S- become more dissimilar. Hanson (1959), using four groups of pigeons each with different and increasingly dissimilar stimuli (light hues) found that the amount of training to reach the discrimination criterion varied directly with the quantitative difference between S+ and S-. That is, as the difference grew smaller, the required amount of training grew larger.

Terrace (1963, 1963a, 1966) seems to have utilized both the above predictions in his work on discrimination learning without errors. The Hull-Spence theory assumes that discrimination learning can take place only when a response is made to S- so that this response may be nonreinforced and, therefore, gain inhibitory strength. Terrace, however, has provided evidence that (at least with pigeons) no response has to be made to S- in order for a discrimination to be learned. This idea would be in agreement with Denny's elicitation theory in that prevention of incorrect responding also prevents the formation of an S-R complex. Terrace's concept is an extension of (but not the same as) Gynther's differentially-non-differentially trained Ss. In Gynther's investigation, one group was conditioned to respond to S+

during the absence of S-. When presented with both S+ and S-, they responded more to S+ than the group who had been given a simultaneous discrimination problem during acquisition. This result provides evidence for discrimination learning without errors (i.e., without responses to S-) since indeed making a response to S- was impossible during acquisition.

Since experimental evidence has shown that better responding occurs when acquisition is by an errorless procedure than by a typical simultaneous method, and since the Hull-Spence theory predicts that as the form difference between stimuli in a discrimination problem becomes smaller, the amount of training time grows larger, it follows that if one starts with greatly dissimilar stimuli and progressively decreases the difference(s), a discrimination task may be learned in this manner. This concept is similar to that of fading as used in programmed instruction (PI) (Holland, 1960; Skinner, 1958; Lumsdaine, 1969; Silberman, 1962). Terrace utilized both the errorless learning idea and the fading technique to teach pigeons a red-green discrimination in the presence of both S+ and S-. When Terrace speaks of errorless learning or learning without errors he is not necessarily referring to 100% correct responding. The terms refer to approaching perfect responding, that is, making fewer responses to S- using this technique than would be made in a simple discrimination problem. It is possible, however, that responding within Terrace's framework would allow more errors than an elicitation approach since the latter can provide effectively for no errors.

Terrace divided his Ss into four groups. In the constant group the brightness and duration of S+ and S- was the same both initially and at the end of training. The progressive group, however, was initially presented an S+ and S- of different brightness and duration and, in three incremental phases, was brought up to the brightness and duration of the stimuli given to the constant group. These two groups were also each divided into early and late groups to which S- was introduced either early, during the first conditioning session, or after a number of weeks of training had already occurred in the presence of S+. All four groups were presented the same S+ (red) and S- (green) of different wavelengths. The red-green discrimination was successfully acquired without the occurrence of any errors in twelve out of twelve cases in the early progressive group. Although the other groups did not fare as well, it was found that early presentation was better than late, and progressive better than constant. Terrace utilized a similar procedure for a transfer task. A vertical (S+) and horizontal (S-) line discrimination were superimposed on red and green backgrounds respectively. The red and green were then faded out. If fading was not used, or if an abrupt transfer was attempted, errors occurred. Even if the red-green discrimination were learned without errors, the performance was permanently impaired in terms of response latency if the

transfer was learned with errors.

An application of errorless discrimination learning with human Ss was made by Moore and Goldiamond (1964). The stimuli used were inverted isosceles skeleton triangles. The Ss were six male and female preschool children ranging in age from three to six years. The task was a delayed, three-choice matching to sample problem. The matches were rotations of the sample except for S+ which was the same as the sample. When the sample was withdrawn and only the S+ match illuminated, correct responding was easily learned. Fading-in the brightness of the negative stimuli was done in seventeen incremental steps. Varying the time of introduction of the fading technique provided evidence for Terrace's concept of errorless learning. Although the task was a difficult one for the preschoolers, the problem was learned with fading and best learned (at least from the aspect of economy of time) by early progressive fading. It is interesting to note that even if an S was originally presented the matches at full illumination (with its resultant chance responding), the introduction of fading immediately provided correct responding, an excellent example of stimulus control.

Terrace's work then, has shown that earlier introduction of S- in the presence of S+ with the added aid of a fading technique (the progressive group) provides greater stimulus control in a visual discrimination problem than: 1) a fading, late introduction of S-; 2) a constant, early introduction of S- (a typical discrimination problem); or

3) constant, late introduction of S- (nondifferentially, then differentially trained as in Gynther's experiment). The same results occurred in a transfer situation based on the original stimuli.

In the present investigation, the fading technique of errorless learning was applied to a visual discrimination problem using moderately retarded children. The use of an early progressive procedure might be one way of compensating for the MR's often hypothesized inhibition deficit and/or attentional difficulties in discrimination problems. This study is then, a modification of Terrace's work. For our purposes, fading is defined as a gradual increase or decrease along a given dimension of a stimulus. The independent variables are brightness and time of introduction of S-. The latter is not the same as the early and late introduction of S- as defined by Terrace. Terrace assumed that the presentation of S- in accompaniment with S+ was an effective stimulus even though its illumination was far lower than that of S+. We do not make this assumption since we do not in fact "know" when the S- is attended to by S. We do assume, however, that at some time during the brightness increment of S-, this stimulus does become an effective one. It is further assumed that the attentional effectiveness of S- is not the same for all Ss in the fading group. We are not attempting to determine the moment of such effectiveness. This moment is assumed to randomize out over all Ss. The differentiation between brightness and time of

introduction of S- is made not for the sake of measurement of these factors as independent variables but simply to point out that fading in this experiment actually has two components only one of which is quantifiable.

The dependent variables are the number of responses emitted to S-, and the latency difference of responses to S+ and S- both during and after the acquisition of the discrimination.

In light of the above discussion, the following hypotheses are set forth:

1. The use of a fading technique during the acquisition of a discrimination task will result in better learning of the discrimination than if this technique is not used. That is, a. The fading group will emit less responses to S- than a non-fading group; and b. the difference between S+ and S- latencies of response will be greater for the fading group than the non-fading group. Both a and b will occur during the acquisition and during testing of the discrimination.

2. Even though a visual discrimination problem is acquired with a fading technique, a transfer problem of the original task not based on fading transfer will result in poorer learning than if transfer is attempted with fading. That is: a. there will be a lesser number of responses to the new negative stimulus (S'-) by the transfer

fading group than by the abrupt transfer group; and b. the (S'+)-(S'-) latency difference will be less for the non-fading group both during acquisition and testing than those of the transfer fading group.

3. If a discrimination problem is learned with the aid of a fading technique, and if a transfer problem is also learned with fading, then on a review posttest of OL, not only will less responses be made to S-, but also the latency difference of (S+)-(S-) responses will be greater than if transfer were abrupt even though OL was based on fading.

Method

Subjects. The Ss were 31 moderately retarded children from the Hope School in Jackson, Michigan. The children lived at home and were bussed to and from school as are children from any normal school. The criterion for participation in this experiment was a child's inability to read even though he may have had some knowledge of letters. This prerequisite was measured by the teachers at the school. From this population, Ss were randomly assigned to the different groups as outlined in the experimental design below.

Only two children were discarded from the experiment. One of these was non-cooperative and the other became ill. Only 29 Ss were used, therefore, to test the hypotheses.

Apparatus. Please see the appendix for a complete description of the apparatus and the stimuli.

Experimental Design. There were three main parts to the experiment (see Table 1). The first two sections consisted of fading and non-fading groups for the given tasks of original learning (OL) and transfer learning (TL). Only the fading group (Fo) of OL were used for the transfer task. These Ss were randomly divided into non-fading (FoNFt) or fading (FoFt) transfer. These two groups were then given a posttest on OL. We see, therefore, that NFo was a control group utilized only for the original discrimination. The Fo Ss were the experimental groups for OL. Fo Ss were also the only ones used for the remainder of the experiment. During TL, FoNFt was contrasted with FoFt. Both were then

Table 1-Experimental Design

Original Learning

NFo (N=10)

$S- = S+$ brightness (trials 1-100)

(end group NFo)

Fo (N=19)

$S- \neq S+$ brightness (i.e., $S- < S+$)
phase 1: $S- = \frac{1}{4}S+$ (trials 1- 25)
phase 2: $S- = \frac{1}{2}S+$ (trials 26- 50)
phase 3: $S- = 3/4S+$ (trials 51- 75)
phase 4: $S- = S+$ (trials 76-100)

Transfer Learning

FoNFt (N=6)

$S' = S$ (trials 1- 50)
$S' = 1$ (trials 51-100)
$S = 0$

FoFt (N=6)

$S' = \frac{1}{4}S$ (trials 1- 11)
$S' = \frac{1}{2}S$ (trials 12- 22)
$S' = 3/4S$ (trials 23- 32)
$S' = S$ (trials 33- 43)
$3/4S' = S$ (trials 44- 53)
$\frac{1}{2}S' = S$ (trials 54- 64)
$\frac{1}{4}S' = S$ (trials 65- 75)
$S' = 1$ (trials 76-100)
$S = 0$

Posttest

Fo (N=12)

$S- = S+$ (25 trials)

compared on original learning disruption after the transfer task by means of a posttest.

During OL, the only difference between the two groups was that the S- (a picture of a dog) initially presented to Fo was not of the same intensity as the picture of the man (S+). Group Fo, however, was "brought up" to the luminosity of the other stimulus in three incremental steps.

The transfer task involved the transfer of the original discrimination to that of the words (S') the line drawings symbolized. FoNFt was simply presented the word MAN (S'+) and the word DOG (S'-) at maximum intensity, superimposed over the appropriate and equally intense pictures. The drawings were then abruptly removed leaving only the words. FoFt, on the other hand, transferred by fading-in the word MAN over the picture of the man, and the word DOG over the picture of the dog. The words were initially superimposed on the proper picture at maximum intensity (G13) and the word at partial (G3) intensity. The brightness level was increased in three phases (i.e., from G3 to G7 to G10 to G13) so that at the end of the incremental fading, both the words and pictures were at full illumination. The pictures were then faded out by using the same incremental steps as the fading-in, but reversed. At the end of this series, therefore, the words remained at full intensity, and the pictures did not appear. The task at this point was the discrimination between the words MAN and DOG.

It is important to note that in the above design, all

fading steps (both incremental and decremental) for all groups have the same psychophysical intensity differences. Also, all the groups within a given learning task are presented with the same stimuli by trial 76. NFo have the same stimuli initially as Fo have by trial 76 during OL. FoNFt have the same stimuli during the first 50 trials as does FoFt for trials 33-43. The last 50 trials for FoNFt consist of the same stimuli as FoFt for trials 76-100.

During the posttest, both groups (i.e., FoNFt and FoFt) were again presented with the original picture discrimination problem. Both S+ and S- were of equal intensity.

According to this design, it was expected that group Fo would learn better (in terms of the dependent variables) than NFo during OL. For the transfer task, FoNFt was expected not to perform as well as FoFt. It was further hypothesized that the transfer task would be more disruptive on a posttest of OL for FoNFt than for FoFt. That is, although all Ss in these two groups learned the original task "without errors," those that learned the transfer with errors (FoNFt) would not only show poorer performance on the transfer task than the other transfer group (FoFt), but would also on a review posttest show poorer discrimination and make more errors on the original stimuli than both themselves (i.e., FoNFt during OL) and FoFt during OL and during the posttest.

Procedure. After a listing of all possible Ss for the experiment was compiled, students were randomly assigned to

groups. Group Fo was to have approximately 2N and group NFo-1N, since the former were to be evenly and randomly divided for the transfer task.

As their turn came, each *S* was taken from the classroom by *E* and accompanied outside to a trailer in which the apparatus was housed. An attempt was made by *E* to be as friendly and unimposing a figure as possible. Conversation, a friendly smile, and a helpful hand were accorded to all *Ss*.

Pretraining. Once inside the trailer, *Ss* were shown the room in which they were to be seated. It was explained that they were going to be playing a candy game. The apparatus was already projecting the pretraining stimuli, and the *Ss* were seated in front of and facing the discriminanda board. The door was closed with *E* standing behind the seated *S*. The following instructions were given: "I will show you how to play the game now. In front of you are two pictures, and you have to guess which one is correct. You make a guess simply by pressing the window with the correct picture with your finger. Go ahead and make a guess." If *S* understood the instructions, he pressed one of the windows and was told that he was playing the game correctly and that he made either a correct or incorrect response. In this pretraining series only verbal reinforcement (e.g., good, that's correct, you are doing well, etc.) or non-reinforcement (e.g., no, that is wrong, etc.) was given. If *S* did not understand the instructions, seemed bewildered and did not make a response, the instructions were repeated. If a response was still not made, E took S's finger and pressed the appropriate window. Verbal reinforcement was given as if S had responded correctly by himself. This process continued until S did respond without prompting. All Ss went through the pretraining series until it was obvious to E that S understood the concept involved (five correct responses in succession).

Discrimination training. Upon completion of the pretraining procedure, the door was opened by E, and S was told, "Now that you know how to play the game, you can play by yourself. Every time you press the correct picture you will get an M&M in the tray in front of you (E pointed to the tray). Here is a bag for you (a small paper bag was opened and placed on the shelf below the reward tray) to put your candy in if you do not want to eat them now. We will begin playing in about one minute. I will tell you when to start, and you can play the game just like before." The door was then closed, and a new slide tray inserted into the projector with the appropriate original learning stimuli. E seated himself in front of the control panel, making sure that the timer was at zero. The proper response record sheet showing the position of the stimuli to be presented was placed in front of E to be marked appropriately as S responded. S was then told, "You can start playing now. Press a window." If at any time during the course of the experiment, an S did not make a response within ten seconds of the beginning of a trial, he was reminded to press a window. This command was repeated every ten seconds until a response was made.

The first few slides presented to all Ss was not part of the training sequence. They were simply the words MAN and DOG presented randomly on the screens[1] and responded to without reward, to determine the presence of stimulus preference. If such selectivity was shown (i.e., not chance responding), S was discarded from the experiment.

A correct response was followed by delivery of an M&M and only intermittent verbal reinforcement (approximately 1/5 variable ratio schedule) for all Ss during the first 75 trials of any given session. For any trial in which an incorrect response occurred, only the presentation of the next slide was provided, no reward or verbalization being given. During criterion trials (76-100) of OL and TL and during the posttest (25 trials) only candy was given for correct responses.

The amount of time spent with each S in this experiment was relatively small. A set of one hundred trials was run in one session. Group NFo Ss, therefore, were started and finished in one day. Fo Ss, however, took at least three days to run. Criterion performance was 20 S+ responses out of the last 25 trials. If this level was not achieved by any given S during either OL or TL, then either one or two

1 All stimuli were randomly assigned to one of the two screens in all groups. The assignments were such, however, that the same stimulus never appeared in the same screen for more than three successive trials.

repetitions of one hundred trials were given to that S on the material failed. One set was presented on day one following the failure and the other set on day two after the failure, if criterion was not met during the first repetition. If by this procedure, an S could not meet criterion on the original discrimination, he was not allowed to continue and was deleted fron the experiment. If criterion by the above procedure was not met during transfer learning, however, S was allowed to go on to the posttest. Using this technique, the maximum time for each S in group Fo was seven days.

No more than one hundred trials were given on any one day except for the first day in which a pretraining series was also administered as described above. An attempt was made to allow no days to intervene between sessions. Weekends were the exception, of course, since there was no school on those days.

All NFo Ss were run before group Fo Ss. The latter weee put through the experiment as they were randomly picked. That is, some Ss were finished before others had begun, and different Ss could be on different parts of the experiment during any given day.

Results

Evidence in support of hypotheses one and two are provided by a comparison of the number of Ss learned versus Ss not learned between groups during OL and TL (see Table 2). On the original discrimination, 36.8% of the fading group Ss failed versus 50% of the non-fading Ss not learning the discrimination. During the transfer task, all Ss in the fading group (FoFt) learned the discrimination within the first one hundred trials whereas only four out of six FoNFt Ss learned within that period. On the second presentation of one hundred trials, only the remaining two FoNFt Ss were run, one of which learned, the other did not learn in three hundred trials. At best, then, the FoNFt group had 16.6% failures while FoFt had no failures. Considering only the first one hundred trials on TL increases the rate of failure for FoNFt to 33.3% while FoFt remains with no failures.

According to the above gross analysis, therefore, we have strong support for hypotheses one and two. A more detailed analysis, however, places certain qualifications upon this conclusion.

All groups were analyzed in terms of S+ and S- latencies. The mean latencies for block of five trials were determined. The latency difference between the two types of responses was then found (i.e., $\bar{X}(S+)-\bar{X}(S-)$). Thus one statistic represents two varying measures. When a discrimination has been acquired, a characteristic range of this statistic was found. For example, if no S- responses are

Table 2-Per Cent Failure

Original Learning

	NFo (6F, 4M)	Fo (10F, 9M)
Learned:	5 (3F, 2M)	12 (5F, 7M)
Not learned:	5 (3F, 2M)	7 (5F, 2M)
Failure:	50%	36.8%

Transfer Learning

	FoNFt (3F, 3M)	FoFt (2F, 4M)
Presentation I		
Learned:	4 (2F, 2M)	6 (2F, 4M)
Not learned:	2 (1F, 1M)	----------
Failure:	33.3%	0%
Presentation II		
Learned:	1 (1F)	
Not learned:	1 (1M)	
Failure:	50%	
Presentation III		
Learned:	-----	
Not learned:	1 (1M)	
Failure:	100%	
Total		
Learned:	5 (3F, 2M)	6 (2F, 4M)
Not learned:	1 (1M)	----------
Failure:	16.6%	0%

made $\bar{X}(S+)-\bar{X}(S-)$ is simply a measure of $\bar{X}(S+)$ response latencies. Since correct responding is our definition of a learned discrimination, we may assume that this level is definitive of $\bar{X}(S+)-\bar{X}(S-)$ learning. Even if S- responses have been made, the difference between the latencies of S+ and S- can still be in the direction and magnitude of just S+ latency measures. No matter what this level is, it is probable that no difference between the latencies (i.e., $\bar{X}(S+)-\bar{X}(S-)=0$) or a difference in the direction of longer S- latencies (i.e., $-x$ seconds) represents no learning or learning of the incorrect stimulus.

Allowing for an occassional error, an analysis of $\bar{X}(S+)-\bar{X}(S-)$ latencies associated with either no or one S- response per five trials was made. The mean latency difference was found to be +13.2 seconds with $\sigma=2.6$ seconds. Three standard deviations either side of this mean would include most trials associated with "errorless" responding. The range is, therefore, 5.4 seconds to 21.0 seconds. That is, responding consistently in this range is representative of learned responding. Corroborating evidence is found in the corresponding S- frequency distribution. We may say that for our population, $\bar{X}(S+)-\bar{X}(S-)$ latencies above the +5.4 second level is an operational definition of discrimination.

The first analysis we can make is within the control group. Five of the <u>S</u>s learned and five did not learn. An examination of the latency differences (see figure 1) shows

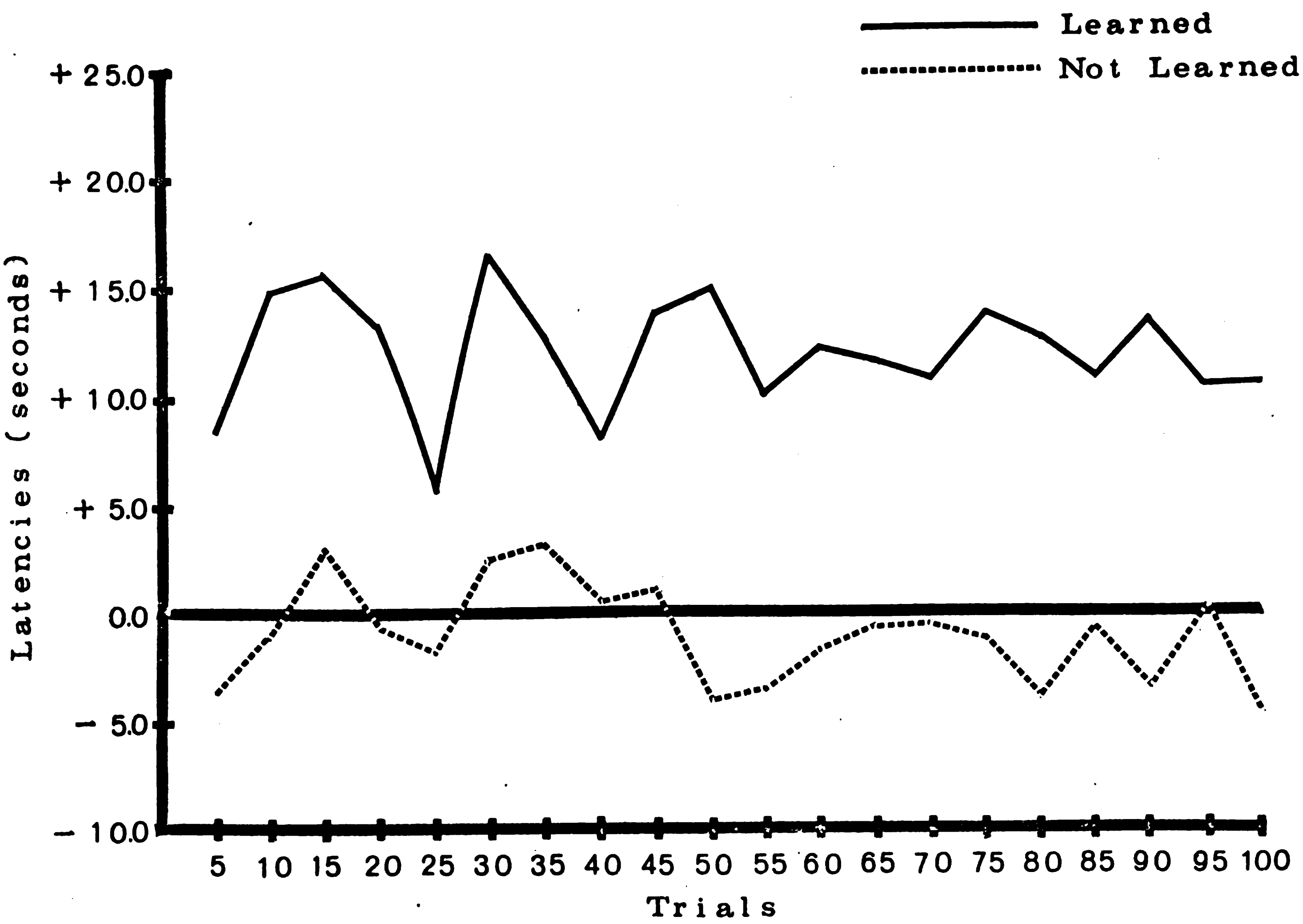

Figure 1—$\bar{X}$(S+)—$\bar{X}$(S−) Control Group Latencies during OL

the difference between the two groups. The learned group consistently had higher latency differences (above +5.4 seconds) than the not learned Ss (below 5.4 seconds). Failing Ss oscillate about the zero difference between S+ and S- latency. They did not discriminate the difference between the two stimuli. The S- frequency curves (see figure 2) corroborate this conclusion. It is interesting to note that, roughly speaking, the higher the latency curves the lower the S- frequency curves and vice versa. For example, during acquisition, the learned control group made no errors during trials 26-30 and 41-50. A glance at the latency chart would predict that such would be the case. The relationships within and between these two graphs may be used as a template for analysis of the rest of the groups since we have here an absolute visual difference between Ss that learn and Ss that do not learn a discrimination under typical discrimination circumstances.

In comparing the control group and the experimental group on OL we find that of those Ss in both groups that learned the discrimination, acquisition lasted only until about trial 30 (for the control group) and trial 35 (for the experimental group). Contrary to expectation, however, fading acquisition was not "errorless" and, in fact, more errors were produced by the fading group throughout the 100 trials than by the control group. The upward acquisition slope of the latency graph and downward slope of the S- frequency chart are more characteristic of acquisition during

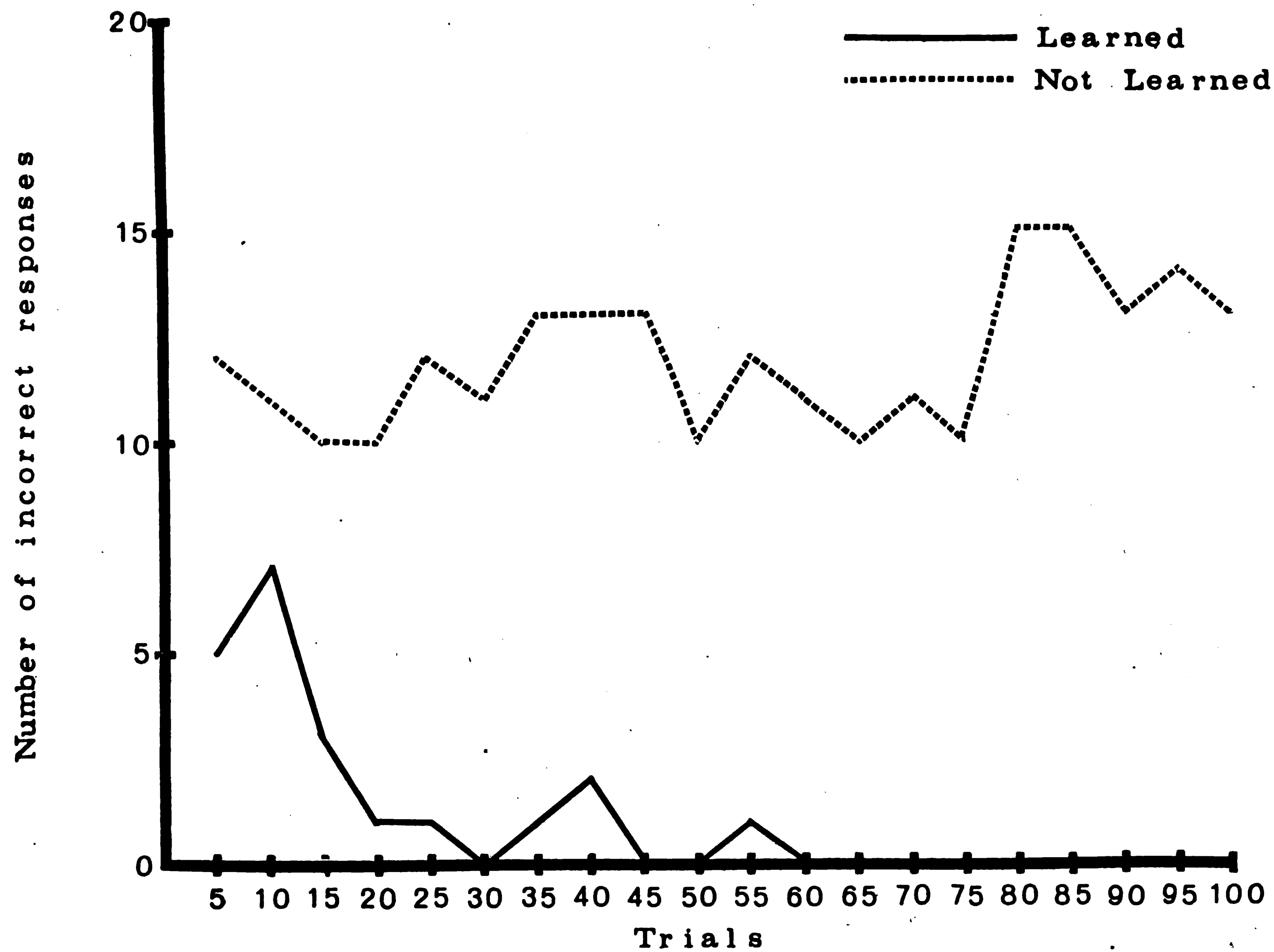

Figure 2—Mean number of incorrect responses for the Control Group during OL

a typical discrimination problem than an errorless approach. On the other hand, the control group shows a flatter graph of latency that would normally be more closely associated with "errorless" learning.

An examination of the original learning data for FoFt (see figures 3 and 4) further complicates the findings because of a marked difference between the two groups on the same task. Both groups show the upward acquisition slope in the latency curves and the downward slope in the incorrect frequency curves. The difference between the two groups is in the rate of this acquisition. FoFt learned the original discrimination faster than FoNFt. We can see that FoFt had definitely learned the discrimination by trials 21-25. FoNFt did not learn until trials 66-70. The two graphs in this case serve to corroborate each other and the above results could be easily determined by an examination of either set of data.

In comparing the data of the control group and the experimental groups during OL, we can see that the first hypothesis is not as strongly supported in these terms as was originally determined (see Table 3). That is, although fading in the S- may allow more Ss to learn a discrimination, such an approach does not necessarily provide better discrimination performance than a situation in which S+=S- throughout acquisition. The experimental groups had a total mean error of 16.7 as compared to the learned control group with 4.0 average errors. During acquisition this

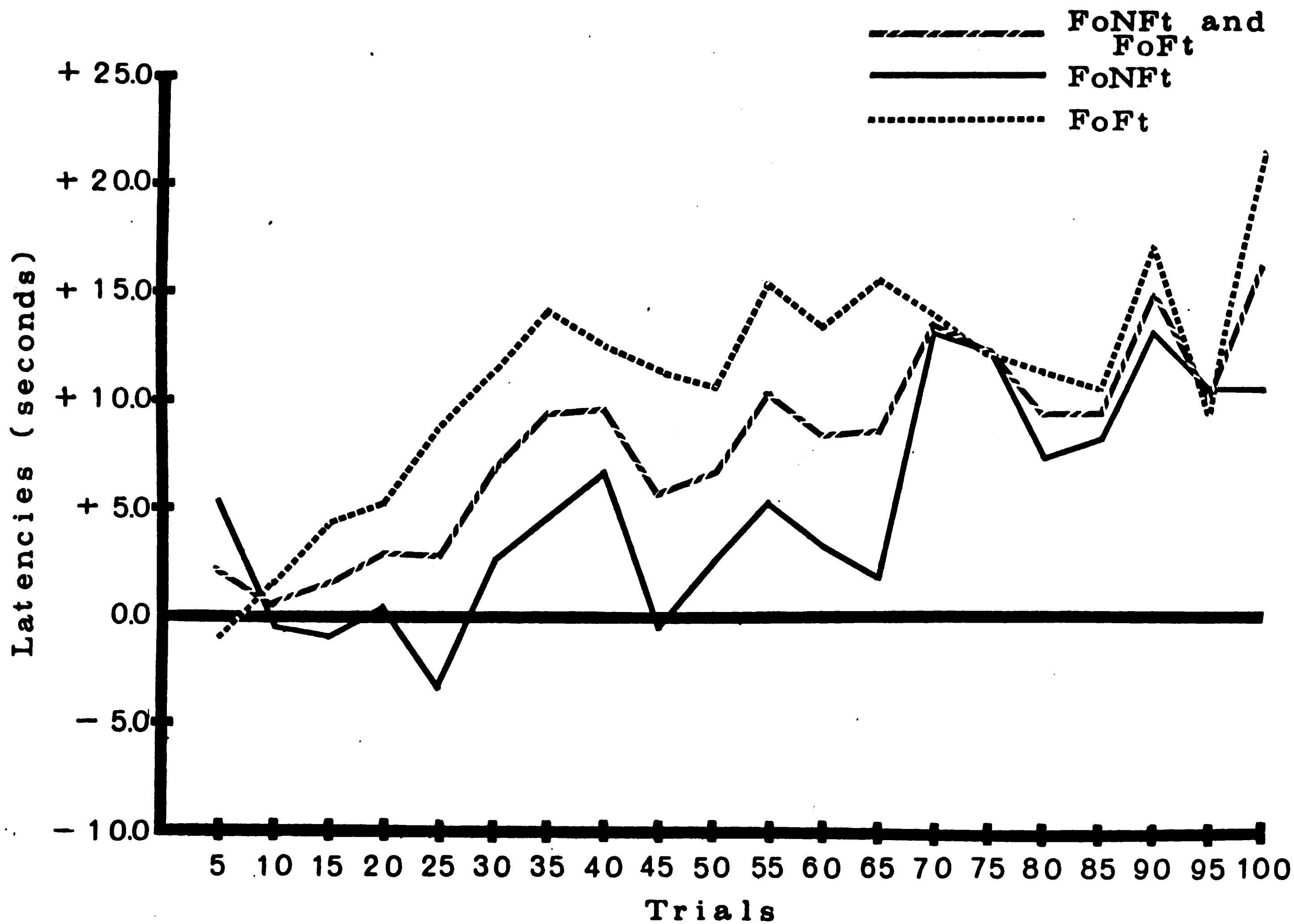

Figure 3–$\bar{X}$(S+)–$\bar{X}$(S–) Experimental Groups Latencies during OL.

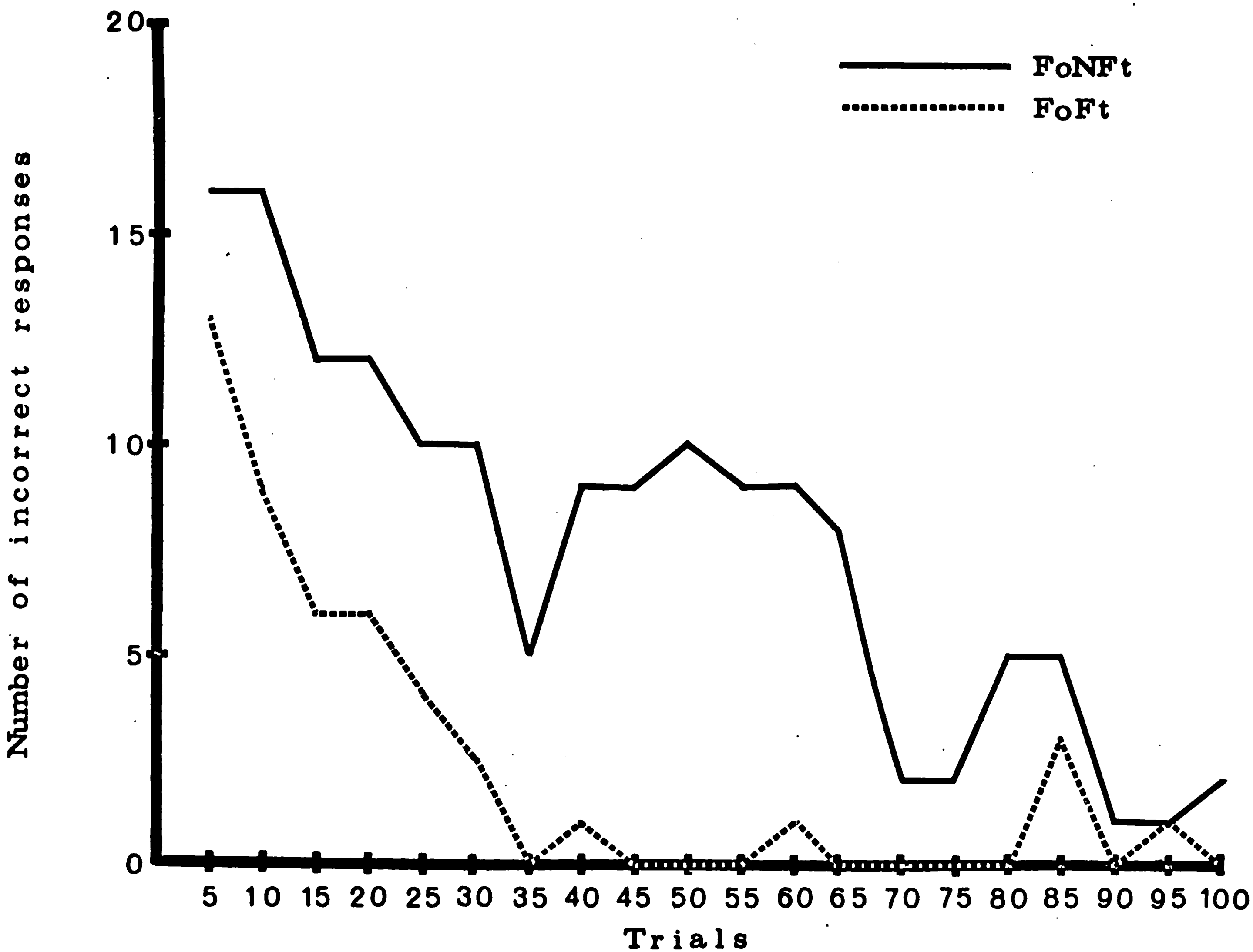

Figure 4—Mean number of incorrect responses for the Experimental Groups during OL

Table 3-Mean Number Incorrect Responses

Original Learning

trials	Control: Learned	Control: Not learned	Control: Total	Experimental: FoNFt	Experimental: FoFt	Experimental: Total
1- 25:	3.2	11.2	7.2	11.0	6.3	8.7
26- 50:	0.6	12.0	6.3	7.2	0.7	3.7
51- 75:	0.2	12.8	6.5	5.0	0.2	2.6
76-100:	0.0	14.0	7.0	2.3	0.7	1.5
1-100:	4.0	50.0	27.0	25.5	7.8	16.7

Transfer Learning

	FoNFt	FoFt
1- 25:	1.2	0.3
26- 50:	1.2	0.8
51- 75:	3.3	0.5
76-100:	3.5	0.8
1-100:	9.1	2.5

Posttest

FoNFt	FoFt
0.2	0.3

relationship held true throughout.

The contamination of results due to the differences between the two groups on the same task during OL necessitated an analysis of covariance of $\overline{X}(S+)-\overline{X}(S-)$ latencies between OL and TL for each group (see Table 4). The analysis was performed on the four blocks of twenty-five trials and on the whole one hundred trials. This analysis tells us the mean magnitude of $\overline{X}(S+)-\overline{X}(S-)$ latency deviation from the average between OL and TL and the direction of this deviation between these two tasks. For the first twenty-five trials, then, we may say that both groups varied with comparable magnitude in the same direction for the two tasks. Although both varied in similar ways during trials 26-50, the magnitude of this direction was greater for the FoFt group. Trials 51-75 provide a complete reversal of this situation. During criterion trials, the results again reverse with FoFt showing the greater magnitude of deviation. For the entire series of trials, however, we can see that FoFt varied negatively while FoNFt was positive. Since

Table 4-Analysis of Covariance of $\overline{X}(S+)-\overline{X}(S-)$ Latencies

trials	FoNFt	FoFt
1- 25:	+0.2	+0.4
26- 50:	+0.5	+3.3
51- 75:	+8.2	+0.6
76-100:	+0.7	+7.3
1-100:	+2.2	-1.5

both groups covaried in the same direction with comparable magnitude during the first twenty-five trials we may remove the effect of overlearning from OL to TL as a cause of TL learning differences.

As predicted by hypothesis 2, not only did the transfer fading group have less S- responses both initially, during acquisition, and test trials than FoNFt, but also the $\overline{X}$(S+)-$\overline{X}$(S-) latency difference was less for the non-fading group throughout the 100 trials (see figures 5 and 6). The only two marked exceptions from this prediction occurred during trials 41-45 and 76-80 in which FoFt did not make a discrimination between the two stimuli. Recovery was immediate, however, in both instances.

Group FoFt provided no slope of acquisitoon for either dependent variable whereas FoNFt did provide more of a slope, although not as steep as would be expected of a typical transfer situation. We may attribute this latter phenomenon to the OL based on fading.

The posttest, administered one day after completion of the transfer task and consisting of twenty-five trials of full presentation of the original problem, provided strong evidence for the null hypothesis. Both groups achieved a high level of proficiency in discriminating the stimuli (see figure 7). The actual number of incorrect responses for both groups was so low as to approach perfect performance. A total of three errors were made, one <u>S</u> of FoFt made two S- responses during trials 1-5, and one <u>S</u> of FoNFt made one

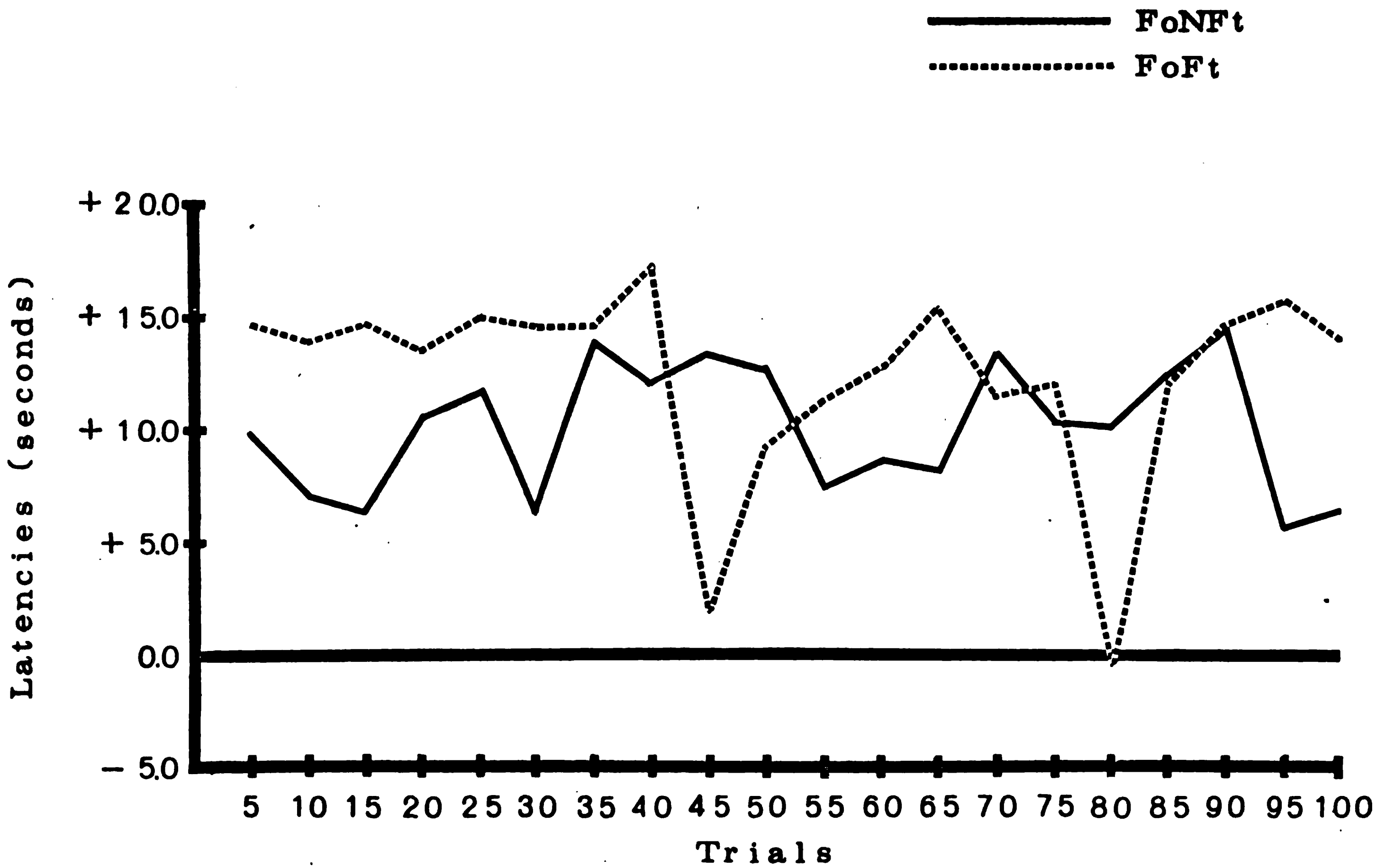

Figure 5—$\bar{X}$(S+)—$\bar{X}$(S−) Experimental Groups Latencies during TL

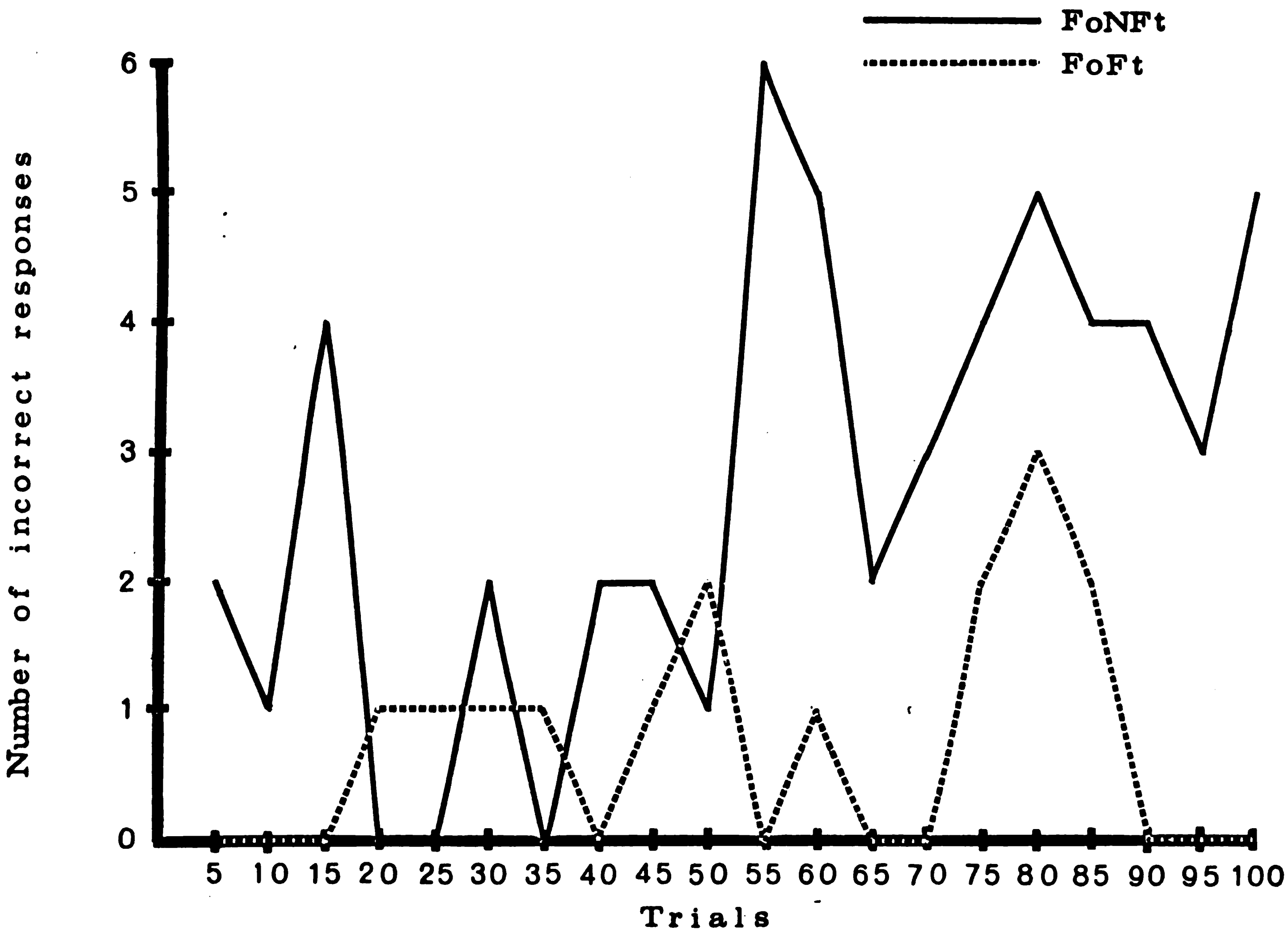

Figure 6—Mean number of incorrect responses for the Experimental Groups during TL

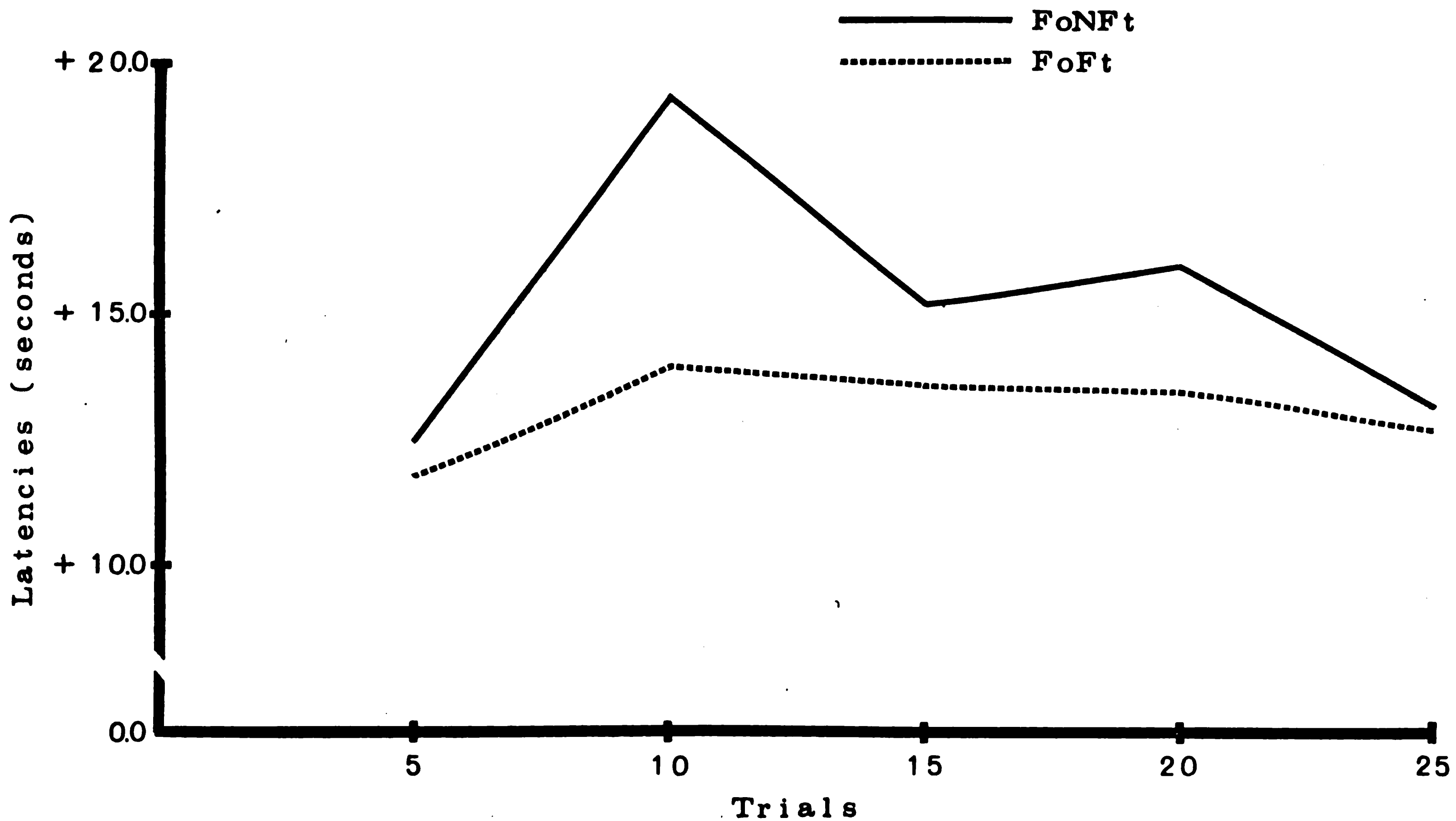

Figure 7–$\bar{X}$(S+)–$\bar{X}$(S–) Experimental Groups Latencies during the Posttest

error on trials 21-25.

Discussion and Conclusion

The results demonstrate that for a visual discrimination problem a fading-in of the brightness cue for a given stimulus allows more moderately retarded Ss to acquire this discrimination than if a simple simultaneous method is employed. The utilization of a fading-in technique, however, does not assure that less S- responses will be emitted than a non-fading approach. The $\bar{X}(S+)-\bar{X}(S-)$ latencies accordingly show acquisition for the fading group and no such slope for the non-fading group. That is, the typical discrimination paradigm may be a closer approximation to errorless learning than fading for a moderately retarded population during a non-transfer visual discrimination task.

During a transfer task, a fading transfer is also beneficial as to the number of people who can accomplish it. Using a fading technique, however, is not reflected as being significantly better with respect to either dependent variable.

The results further provide evidence of no response disruption of original learning if the intervening transfer task is accomplished with a non-fading approach. That is, although original learning is acquired by fading, an abrupt transfer task provides no significant differences from a fading transfer on recall of OL material as measured by the dependent variables.

As may be seen, the results vary significantly from what may be logically extrapolated from Terrace's studies.

The disagreement between theory and empirical testing occur in two main areas. First, fading the brightness cue is not the only or is not a prerequisite to errorless learning. Secondly, despite the fact that fading original learning is not errorless more Ss learn a transfer through fading than by abrupt transfer.

The failure of fading to provide errorless learning during OL may be due to one difference between Terrace's experiments and the present one. That is, the latter did not concern itself with the duration of the presentation of stimuli. There could be some importance attached to the length of viewing of stimuli. It may be that if the duration of the cues during presentation are longer, more attention (for example) would be paid to them. The assumption involved in this hypothesis is one that may not necessarily be borne out. Simply because the stimuli are presented for a longer period of time does not mean that an S will attend to them that much more. Evidence for this may be found in the experiment just presented. In it, the moment of effectiveness of S- as a stimulus was assumed to randomize out over all Ss. It could be that we have here a skewed sample which did not randomize between groups FoNFt and FoFt where, during OL, there was such a large difference between these groups on the same task. Essentially the same problem is involved in both concepts-how can an S be "forced" to attend to a stimulus. A longer period of presentation would not seem to solve the problem.

One way to assure the effectiveness of the stimuli is to have self-determining number of trials to criterion. That is, whether the stimulus is effective during any given trial is not important. What is important is that effectiveness does result after a number of trials. Allowing all Ss as many trials as needed to reach a given criterion (that is, to learn a discrimination) and then fading-in the next brightness level would (at least theoretically) be an answer. Adopting this technique to the present experiment would be a relatively easy matter. The same general design could be used, but instituting a criterion of five (5) consecutive correct responses before the next fading increment (or decrement). Utilization of this method would also solve the problem of variable overlearning trials that was inherent in the present experiment.

Another explanation of the discrepancy between the results of Terrace's and the present experiment may be due to factors inherent in the apparatus of this experimenter's investigation. Every trial may have been unwittingly reinforcing due to a satisfaction of S's manipulatory drive. Each S had control over the stimuli presented to him. He could shut off the projector and turn it on again simply by pressing a panel (either panel). The room was dark in the interim and the possibility of new stimuli were imminent. The sound of the projector was also under their control. All this could have served as a conditioned reinforcement. Another confounding problem evolves from this theory in that

S is asked to make not only a visual discrimination, but also a reinforcement differentiation. The greater reward occurred with an S+ (or S'+) response because this was accompanied by a primary physical (M&M) and secondary auditory (click of the M&M dispenser and occassional verbal reinforcement) reward. An objective determination had to be made as to which reward was greater. A subjective determination also was forced upon S in that he must determine if one set of reinforcement was better, worse, or made no difference.

The problem raised here may be partially solved by the use of a non-automated WGTA. A simple sliding drawer with food wells covered by physical stimuli would somewhat lessen the secondary reinforcement inherent in the automated apparatus. This device would not, of course, provide the high degree of time measurement accuracy as the automated apparatus, but this fact would perhaps be compensated by a lesser degree of confounding.

A second factor, intrinsic to the present design that may have accounted for the results, was the ITI. Between trials a period of 6.8 seconds elapsed. This time length may have been too long for the population involved. A shorter ITI may produce less of a strain on their short term memory (STM). We had assumed that the carry over from trial to trial would be great enough for a discrimination to result. The results may be indicative of the invalidity of this assumption in that a long STM functioned to counteract the fading crutch. Any further experimentation in this area should, therefore, be preceded by a determination of the most beneficial time between trials for the given population. Not only may STM be strained, but too short an ITI could confuse the subject.

STM may also be a factor involved in the discrepancy of the experimenter's results and those of Moore and Goldiamond (discussed previously). In the present investigation STM was utilized when S's eyes traveled between the two windows and the ITI. Moore and Goldiamond also had STM between trials. But within the trial the time involved for STM was longer and forced. That is, an S had to attend to the sample until it was removed and then to the matches which were then presented. Since it was impossible to refer back to the sample, the S was forced to rely on his memory of the cues in the sample, That is, there was no reference back beyond the initial surveying of the sample stimulus. The differences between the STM in the two experiments may have been compensated by the differences in the I.Q.'s of the populations. We should remember, however, that they utilized younger Ss with higher I.Q.'s than the present study with older children but lower I.Q.'s. The mental ages, therefore, might be equatable. Since data on intelligence was not provided, no I.Q. comparisons between the two experiments may be made.

An analysis of intelligence data from the present experiment provide some insight into its problematic results (see Table 5). I.Q., M.A., and C.A. distributed in a highly equatable fashion between the control and experimental groups. Within the control group, the non-learners were those with a lower C.A.. In fact, the low C.A. measures of the non-learners in the control group and the deleted members of the experimental group were equatable as were the learners in the respective groups. That is, we may postulate a C.A. deficit as a contributing factor in not learning independent of method of acquisition. More specifically, those S̲s with a mean C.A. of 7.0 or less did not learn to respond to S+ to criterion no matter what visual crutches were utilized.

We can now understand why a greater percentage of the fading S̲s learned during OL. The S̲s were so distributed as to have a lower percentage of low C.A. S̲s in the fading group than in the non-fading group. The fading, therefore, had no effect.

During transfer, however, fading did provide less errors than no fading. This effect occured despite of an M.A.-C.A. deficit for the fading transfer group. We may conclude, therefore, that fading during a transfer situation does indeed assist in learning even though such learning is not based on errorless OL. This latter conclusion is our only remaining contradiction to Terrace's theory.

We may find a clue to this predicament by returning to an analytic comparison between the present study and that of Moore and Goldiamond. These investigators utilized

Table 5-Distribution of Measures of Intelligence Among Groups

Control Group

NFo (N=10)

$\overline{\text{I.Q.}}$=48
$\overline{\text{C.A.}}$= 7.4
$\overline{\text{M.A.}}$= 3.5

Not Learned (N=5)

$\overline{\text{I.Q.}}$=51
$\overline{\text{C.A.}}$= 7.0
$\overline{\text{M.A.}}$= 3.6

Learned (N=5)

$\overline{\text{I.Q.}}$=43
$\overline{\text{C.A.}}$= 8.2
$\overline{\text{M.A.}}$= 3.5

Experimental Group

Fo (N=19)

$\overline{\text{I.Q.}}$=47
$\overline{\text{C.A.}}$= 7.8
$\overline{\text{M.A.}}$= 3.7

Deleted (N=7)

$\overline{\text{I.Q.}}$=48
$\overline{\text{C.A.}}$= 6.7
$\overline{\text{M.A.}}$= 3.2

Learned (N=12)

$\overline{\text{I.Q.}}$=47
$\overline{\text{C.A.}}$= 8.4
$\overline{\text{M.A.}}$= 3.9

FoNFt (N=6)

$\overline{\text{I.Q.}}$=48
$\overline{\text{C.A.}}$= 9.5
$\overline{\text{M.A.}}$= 4.6

FoFt (N=6)

$\overline{\text{I.Q.}}$=46
$\overline{\text{C.A.}}$= 7.4
$\overline{\text{M.A.}}$= 3.4

seventeen (17) fading steps while the present study (as in Terrace's experiments) had three (3). The former study also accomplished good stimulus control. There was little disturbance in the ongoing learning process with seventeen fading steps. The present study did, however, find disturbance with the 3 "jumps."

This problem points out an important question involved in any study concerned with a fading variable. That is, after a fading-in step, how much of a new task is being presented to S? In the present study, the non-fading transfer group (FoNFt) was given an abrupt transfer (or very large size-of-step fade-in) after trial fifty. Their latencies were consequently disturbed and the number of incorrect responses immediately increased. It would seem that there is a negative correlation between the amount of disturbance, in so far as the number of incorrect responses are concerned, and the number of fading steps (in this case 1 vs. 7). For group FoFt only a few of the fading-in steps caused great latency disruption. The last fade-in (trial seventy-six) was the most notable exception. Obviously, the change in stimuli was too great. The fading-in caused a 7.5 second average latency. This figure is far different from any other latencies for either group during a fading-in step. The last fading-in for group FoFt can now be interpreted as much more of a presentation of a new task than previous "jumps." This process was similar to the one abrupt transfer step (trial fifty-one) of group FoNFt in which this latency equaled that of their first trial (a new task). Although disturbance occurred for FoFt during TL, there was less such disturbance. Therefore, fading during transfer assists the moderately retarded S to learn a visual discrimination.

A continuum of fading steps to find the maximum number of "jumps" needed for the least latency disturbance and number of S- responses would be one extrapolation of this study. It is expected that the greater the number of steps, the less the disturbance until an asymptote is reached. Beyond this point the profitability of further fade-in's would decline. This approach would be similar to that of Sidman and Stoddard (1966) in which a program is changed by the experimenters as it is evaluated by the Ss. The present experiment suggests that not only the number of incorrect responses be used in evaluation (to determine criterions before fading-in steps), but also the latencies before and after the "jump." Too low a latency difference below a predetermined standard would indicate a mistake in the program, and a need for less of a change. This approach could mean, therefore, that not all steps should have equal psychophysical and/or quantitative changes.

The relation between this study and programmed instruction has already been alluded to. The results of this experiment point the way toward a subjectively determined (by population and individual) variability of a general program for any given subject matter or theoretical study.

The unique capability of simultaneously teaching and evaluating the behavior of the mentally deficient is inherent in this approach. Terminal behavior and eventual potential may thusly be reached. The malleability of a given program, in addition, is such that many levels of retardation may be dealt with, without a satiation effect. The most succinct method may be used, provided no informational value is lost, by both profoundly and moderately retarded. The author, of course, realizes the limitations of this mode of learning presentation. The potentialities, however, are not known.

LIST OF REFERENCES

LIST OF REFERENCES

Denny, M. R. A theoretical analysis and its application to training the mentally retarded. Chapter in N. R. Ellis (Ed.), International Review of Research in Mental Retardation, Vol. 2, 1966.

Denny, M. R. and Adelman, H. M. Elicitation theory: I. An analysis of two typical learning situations. Psychol. Rev., 1955, 62, 290-296.

Gynther, M. D. Differential eyelid conditioning as a function of stimulus similarity and strength of response to the CS. J. exp. Psychol., 1957, 53, 408-416.

Hall, J. F. The Psychology of Learning. New York: Lippincott, 1966.

Hanson, H. M. Effects of discrimination training on stimulus generalization. J. exp. Psychol., 1959, 58, 321-334.

Holland, J. G. Teaching machines: An application of principles from the laboratory. J. exp. Anal. Behav., 1960, 3, 275-287.

Hull, C. L. The problem of stimulus equivalence in behavior theory. Psychol. Rev., 1939, 46, 9-30.

Hull, C. L. Principles of Behavior. New York: Appleton-Century-Crofts, 1943.

Hull, C. L. Stimulus intensity dynamism (V) and stimulus generalization. Psychol. Rev. 1949, 56, 67-76.

Hull, C. L. Behavior postulates and corollaries-1949. Psychol. Rev., 1950, 57, 173-180.

Hull, C. L. A Behavior system. New Haven: Yale, 1952.

Lumsdaine, A. A. Some issues concerning devices and programs for automated learning. In A. A. Lumsdaine and R. Glaser (Eds.) Teaching Machines and Programmed Learning. Washington: National Education Association, 1960.

Moore, R. and Goldiamond, I. Errorless establishment of visual discrimination using fading procedures. J. exp. Anal. Behav., 1964, 7, 269-272.

Sidman, M. and Stoddard, L. T. Programming perception and learning for retarded children. In Norman R. Ellis (Ed.) International Review of Research in Mental Retardation, Vol. 2, 1966.

Silberman, H. F. Self teaching devices and programmed materials. Rev. Educ. Res., 1962, 32, 179-193.

Skinner, B. F. Teaching machines. Science, 1958, 128, 969-977.

Spence, K. W. The nature of discrimination learning in animals. Psychol. Rev., 1936, 43, 427-449.

Spence, K. W. The differential response in animals to stimuli varying within a single dimension. Psychol. Rev., 1937, 44, 430-444.

Spence, K. W. Analysis of the formation of visual discrimination habits in chimpanzee. J. Comp. Psychol., 1937a, 23, 77-100.

Terrace, H. S. Discrimination learning with and without errors. J. exp. Anal. Behav., 1963, 6, 1-27.

Terrace, H. S. Errorless transfer of a discrimination across two continua. J. exp. Anal. Behav., 1963a, 6, 223-232.

Terrace, H. S. Stimulus control. In W. K. Honig (Ed.) Operant Behavior: Areas of Research and Application. New York: Appleton-Century-Crofts, 1966.

APPENDIX

APPENDIX

Apparatus. A semi-automated version of the two discriminanda Wisconsin General Testing Apparatus (WGTA) was used. The S sat in a semi-soundproof metallic gray cubicle facing a similarly colored wooden board. On the board were centered two transluscent screens which could be pushed by S in order to make a response. Centered beneath the screens was a reward tray under which was located a small shelf. These objects were placed so as to be within easy reach of S (see figure 8).

Behind the discrimination board were the power supplies and electronic equipment (not seen by S during the actual running of the experiment). A projector was aimed at the transluscent screens. An electric timer, accurate to 1/100 seconds, was connected to both the projector and to the screens. The presentation of the stimuli on the windows (trial initiation) was accompanied by the start of the electric timer. A trial was terminated by pressing one of the screens, thereby stopping the clock and projection of stimuli for 6.8 seconds (the interstimulus interval). This cutting-off procedure was accomplished by two microswitches, one placed behind each of the windows, which relayed to the clock, the projector, and the experimenter's panel lights. In this manner, the latency of responses was accurately measured. E was provided with a panel which indicated (by lights) which screen, left (L) or right (R),

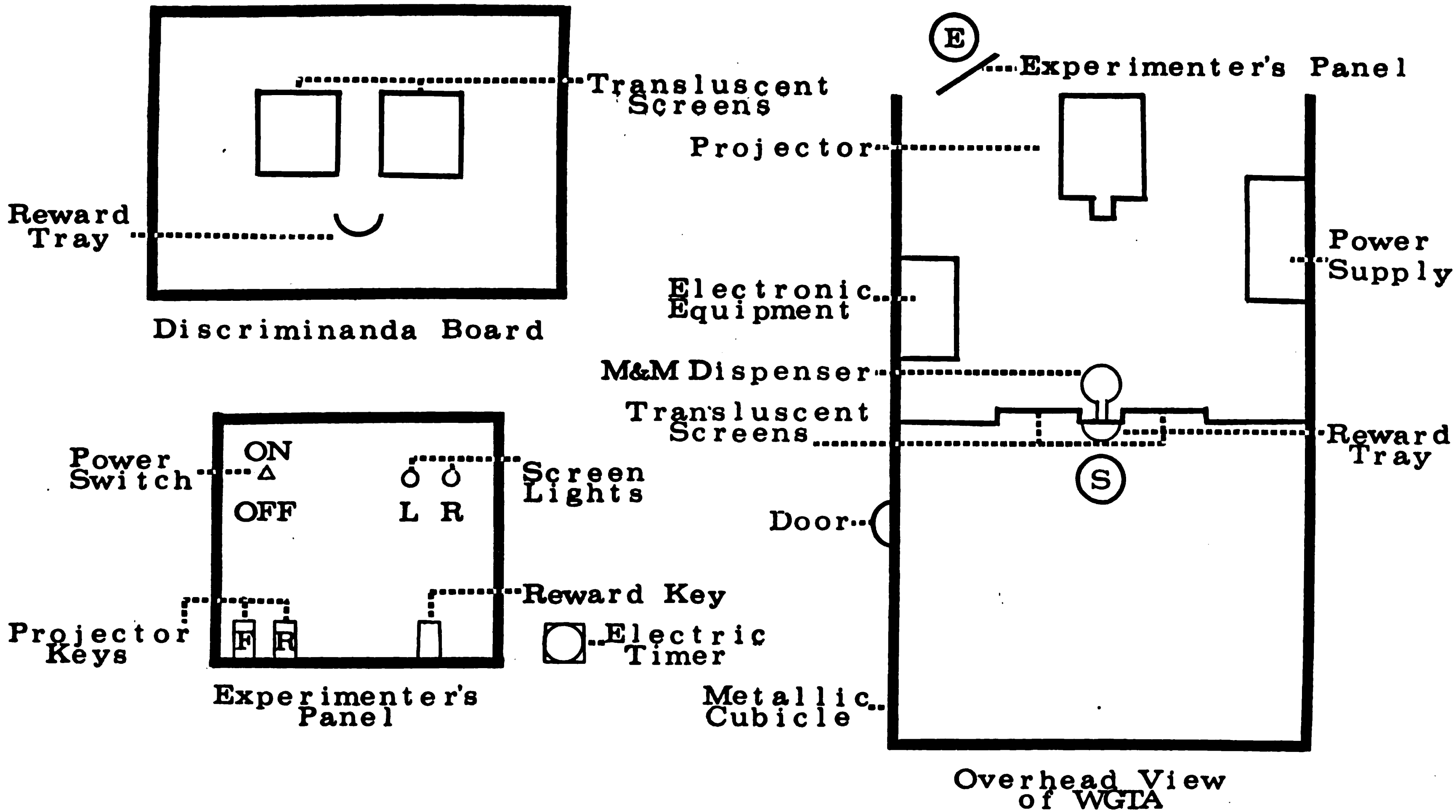

Figure 8—Wisconsin General Testing Apparatus

had been pressed by S. A panel key also allowed E to provide a reward from the M&M dispenser which was located in back of and above the screens. Two keys for manual operation of the projector, forward (F) and reverse (R) completed the apparatus.

Stimuli. The stimuli were projections of line drawings of a man striding, a dog standing on all fours, the word MAN in capital letters, and the word DOG also in capital letters (see figure 9). These stimuli were presented by means of 425 black and white slides. These slides were actually hand-mounted negatives of gray tempera color drawings on clear acetate. The darker the gray color against an almost white background, the more light was allowed through that area on a slide when projected. The independent variable of brightness was thereby controlled.

The grays were mixed from black and white tempera colors and matched by eye to a psychophysically determined gray series.[1] This continuum consists of 16 equidistant grays ranging from one unit above pure white (G1) to one unit below pure black (G16). G1 was used as the background upon which the forms were photographed. The drawings themselves utilized G4, G7, G10, and G13 of the series, each representing a higher degree of light intensity as a projected negative.

Four etchings were made of every object, one in each

1 This gray series was developed by Dr. William T. Stellwagen of Michigan State University.

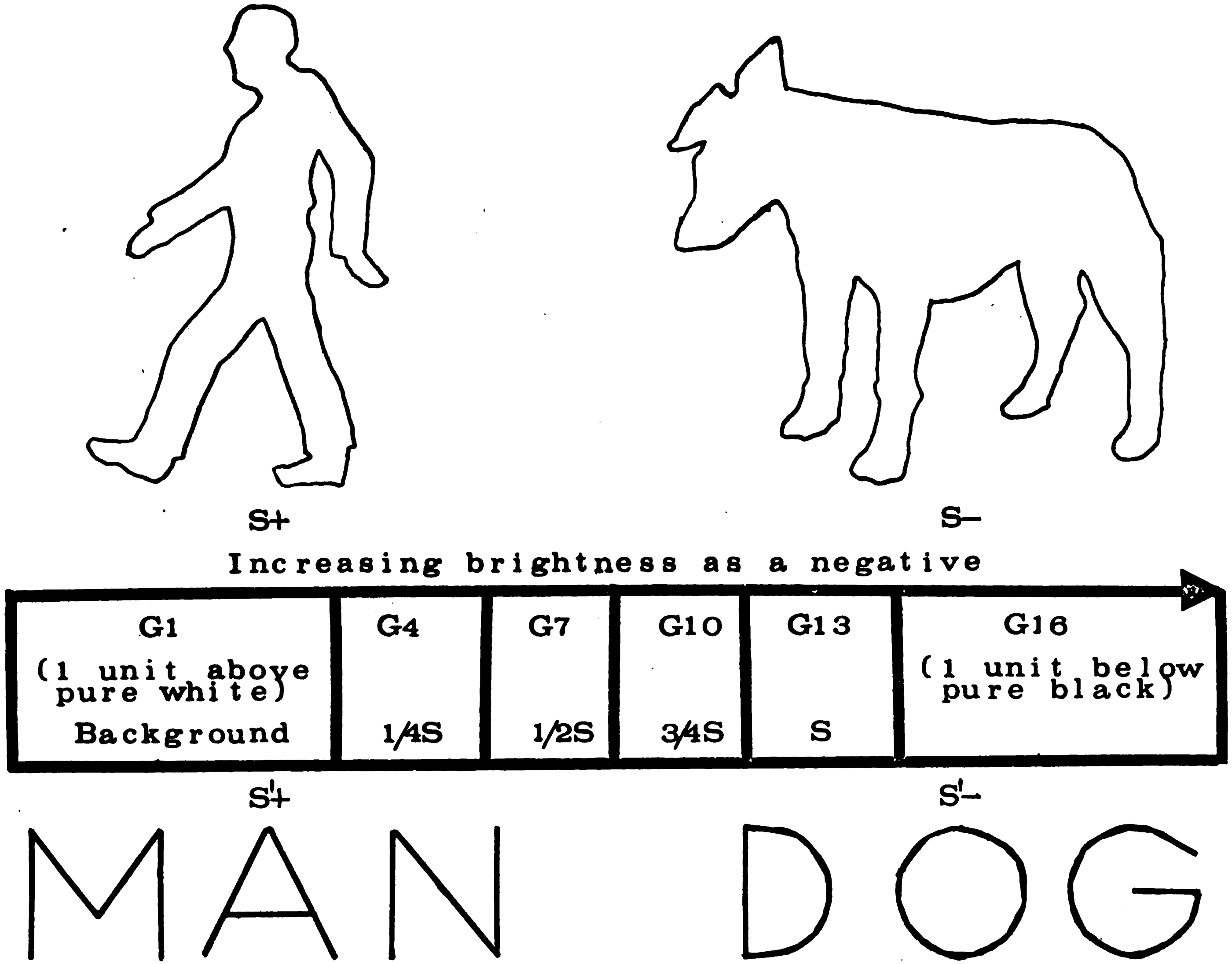

Figure 9–The Stimuli

group painted with a different color gray according to the above prescriptions. When photographed, one figure was simply overlayed (or not) onto its matching word according to the conditions of the experiment.

The photographs were taken with a copy set up and a Pentax camera with a 55mm. lens, loaded with Pan-X film ASA 32. Two 3,200° floodlights in 10" reflectors maintained an intensity of 375 foot candles at the site of the windows. The film was exposed for 1/15 second at a 5.6 stoppage. The height from the lens to the surface was 40", the drawings being separated by 16" from their centers. This same technique was used for the pretraining stimuli which consisted of a picture of jello and a picture of an airplane both taken from a magazine. In this case, however, indoor color film (ASA 25) was used.