6-1973

# Rehearsal of the Incorrect Response in Frequency Discrimination

Russel L. Smith
*Kansas State College of Pittsburg*

2Y

# REHEARSAL OF THE INCORRECT RESPONSE IN FREQUENCY DISCRIMINATION

A Thesis Submitted to the Graduate Division in Partial Fulfillment of the Requirements for the Degree of Master of Science


By

Russel L. Smith





KANSAS STATE COLLEGE OF PITTSBURG

Pittsburg, Kansas

June, 1973

ACKNOWLEDGEMENTS

My sincere appreciation is extended to the members of my thesis committee, Mr. Gregory Gaustad, Dr. Ralph Wright, and Dr. Michael Heffernan for their guidance and helpful, constructive criticisms. My appreciation is also sent out to Miss Victoria Sanchez for her excellent typing skills, and for her untiring encouragement without which this project would have been a much more difficult task.

A note of special thanks goes to Mr. Gaustad for his stimulating conversation and understanding when moments of frustration slowed my progress.

TABLE OF CONTENTS

ABSTRACT

The frequency theory of V-D learning states that the difference in frequency of occurrence of the correct and incorrect items serves as a cue to Ss; the Ss make their choice based on the relative frequency of the items. The present study consisted of a transfer task with a 2X3 factorial design; on the first task, one half of the Ss learned a list in which the correct items were changed each trial (CC), while one half learned a standard V-D list. On the second task, one third of the CC group learned a list in which high frequency items were correct and low frequency items were incorrect (HL), one third of the CC group learned a list in which low frequency items were correct, and high frequency items were incorrect (LH), and one third of the CC group learned a St. V-D list. On the second task, one third of the St. V-D group learned an HL list, one third of the St. V-D group learned an LH list, and one third of the St. V-D group learned a St. V-D list. The study was a test of the hypothesis that Ss may be rehearsing the incorrect response in certain situations; if so some anomalous results in frequency theory could be explained. The results of the present study were inconclusive, and gave some evidence which supports the RIR mechanism as well as evidence not in support of the RIR mechanism.

LIST OF TABLES

LIST OF FIGURES

## REVIEW OF RESEARCH

The present study examined the frequency theory of verbal-discrimination (V-D) learning, which was formally proposed as the way learning takes place in V-D learning by Ekstrand, Wallace, and Underwood (1966). In V-D learning, pairs of items are presented to the subjects (*S*s), and the experimenter (*E*) arbitrarily designates one of the items of each pair as correct. The task of the *S* is a simple one; as each pair of items is presented, the *S* tells the *E* which item of the pair he chooses. Working with the V-D paradigm is a standard method of laboratory investigations of recognition memory; an example of a recognition memory task is the making of a choice on a multiple choice test. The *E* may inform the *S* as to whether he has chosen the correct item in one of two ways; in the anticipation method the *E* immediately informs the *S* after each choice or pair, while in the study-test method the *E* waits until the *S* has made his selection for all pairs before he reveals the correct item in each pair (Ekstrand et. al., 1966). V-D learning occurs when an increase in the number of correct responses over trials is evident, or when the subjects performance after an initial trial surpasses his level of performance that would be expected by chance.

According to frequency theory, as the S progresses through trials in the V-D experiment, he builds up frequency units for the items, and decides which item of a pair is correct based upon the number of frequency units accrued for the items of a pair. Ekstrand et al. proposed that within frequency theory there are two rules which the S may use; Rule I, which states that the S will choose the most frequent item of each pair as the correct item, and Rule II, which states that the S will choose the least frequent item of each pair as the correct item. The basic principle, the counting postulate, which was derived from considerations of the V-D task, is that the correct alternative in each pair acquires two frequency units on each exposure, whereas the incorrect alternative acquires only one frequency unit per exposure. Two critical aspects within the V-D task form the basis of the counting postulate, and there are a number of ways in which frequency units can be built up. First, V-D procedure involves the presentation of both items of the pair to the S; thus the S acquires one frequency unit for both items of the pair. Ekstrand et al. called this perception or recognition of an item a representational response (RR). Second, the correct item of each pair receives at least one more frequency unit since the S may be required to vocalize this item , and the E indicates which item of each pair is correct (e.g., the correct item may be presented alone, underlined, or designated by an asterisk). Ekstrand et al. called this pronunciation of the correct item the pronunciation response (PR). One additional frequency unit may be acquired when the correct alternative is shown to the S during feedback; this third mechanism Ekstrand et al. called the rehearsal of the correct response (RCR). One other possible mechanism whereby the S might acquire frequency units is an implicit response associated with one of the verbal units. An implicit response can occur when a word such as "dog" is presented in a list, and the S automatically, as a result of past experience with the language, associates the word "cat" with "dog"; if cat is presented somewhere else in the list, it would have already acquired one frequency unit as a result of the presentation of "dog" earlier. Ekstrand et al. called this fourth mechanism an implicit associative response (IAR). The IAR will facilitate learning if the associated word is correct, or it will hinder learning if the associated word is incorrect. One St. V-D task presents the two words "door" and "line" as a pair of items within a list, "door" being arbitrarily designated as correct by the E. When "door" and "line" are presented as a pair to the S, each acquires one frequency unit as a result of the RR mechanism. When the S vocalizes his choice (PR), the word he says will receive one more frequency unit; if the S chooses "door", the correct item now has a frequency ratio of 2:1 over the incorrect item. However, if the S chooses "line", the ratio is now 2:1 in favor of the incorrect item. When the S sees the correct item during feedback, the word "door" will accrue (because of the RCR) one more frequency unit; if the S has chosen "door", the ratio is now 3:1 in favor of door, but if he has chosen "line" the ratio is now 2:2. Over trials, the correct word "door" will accrue more frequency units than the incorrect word "line". In summary, frequency theory states that the S has the power to discriminate differences in frequency of verbal units; since the frequency values of correct alternatives generally exceed the frequency values of incorrect alternatives over trials, the S performs successfully a V-D task by selecting the item of a pair that has accrued the greatest number of frequency units.

Wallace (1972) presented an excellent summary and explanation of the results of many tests of the frequency theory which have been performed. In general, the results of these tests have supported frequency theory, but some of the results have pointed to difficulties with the theory. Many of the studies have involved a transfer task, where the S is first presented with one task (e.g., a V-D list, a free-recall list, etc.) followed by presentation of a second V-D task, performance on the second task being of primary interest. Positive transfer would be indicated by a high level of performance on the second list, while negative transfer would be indicated by a low level of performance, the performance being compared to performance on a standard V-D list.

The major data supporting the frequency theory consist of the differeential effects on performance resulting from manipulations of correct and incorrect item frequencies. Underwood, Jesse, and Ekstrand (1964) manipulated the frequencies in transfer tasks by pairing items in the first list with new correct or new incorrect items in the second list. The experiment consisted of three conditions. In one condition (R), the correct items in the second list were the same words which had been the correct items in the first list, and each correct item was paired with a new incorrect item which had not appeared in the first list. In a second condition (W), the incorrect items in the second list were the same as the incorrect items in the first list, each incorrect item being paired with a new correct item which had not appeared in the first list. In a third condition (C), the Ss learned a second list composed of new words for both correct and incorrect alternatives. The Ss were informed as to the nature of the second list construction in all conditions. Group R performed well initially and continued to perform well. Group W pwerformed well initially (relative to Group C), but showed very slow improvement; Group C reached criterion, in fact, sooner than the Ss in Group W. The argument, according to frequency theory, is that during early trials on the second list there is a substantial frequency difference, for Group W, in favor of the incorrect items. The Ss perform well in the early trials by discriminating differences in frequency and selecting the items which register relatively lower in frequency value (i.e. Rule II). As the correct items accrue frequency units faster across trials than do the incorrect items, the discrimination as to frequency becomes more difficult for Group W, which accounts for the better performance of Group C. Group W, however, according to frequency theory, should have returned to a chance level of performance, which did not occur; explanation of the failure to obtain these results will be discussed at a later point.

Kanak and Dean (1969) have reported results which support frequency theory; the study involved the investigation of mechanisms for V-D learning under various transfer conditions, as did the Underwood et al. (1964) study. The Kanak and Dean (1969) study involved two experiments; Experiment I tested extensions of Osgood-type empirical laws to V-D transfer, and Experiment II tested mechanisms in corresponding vs. noncorresponding (or re-paired) paradigms. In general, the two experiments supported the conceptualization of V-D transfer effects as the result of positive transfer mechanisms accruing from frequency-based rules and negative transfer mechanisms based on incidentally learned associations between wrong and right items.

Ekstrand et al. (1966) reported results which support frequency theory. They manipulated frequency within a single list in two ways: in one procedure, specific items were presented in two pairs, and in the other procedure a given word was presented in one pair and a strong associate of that word in another pair (refer to IAR). In the latter procedure, the rationale was that the presentation of a specific word might increase the frequency value of its associate through the implicit occurrance of associative responses. The repeated words (or associated words) were always correct items in their respective pairs in one set of conditions (R), while the repeated words (or associated words) were always incorrect in another set of conditions (W). According to frequency theory the R conditions should be superior in performance to the W conditions, and the results showed that the R conditions were superior to the W conditions, in accordance with that prediction.

Underwood and Freund (1969) extended the experiment of Ekstrand et al. (1966). They presented specific items in 1, 2, 3, or 6 different pairs of a 12-pair list; with the repeated items always correct or always incorrect. V-D performance improved as the number of pairs with the same correct word increased; as the number of pairs with the same incorrect word increased, however, V-D difficulty first increased, then decreased. These results were generally in accord with frequency theory expectations although the specific point of decrease in difficulty was not in line with predictions.

The results of an experiment done by Yelen (1969) provided further support for frequency thoery; she introduced new incorrect alternatives on successive trials. In accordance with frequency theory expectations, the group with new incorrect alternatives performed better than a control group which learned a standard V-D task. Since the new incorrect alternatives had accrued no frequency units at the time of presentation to the Ss, and the incorrect alternatives in the standard V-D list, though not as numerous as the correct alternatives, had accrued some frequency units, these results would be expected.

Kausler and Farzanegan (1969) used transfer conditions to study whether a selection strategy learned in a first task transfered to a second task in V-D learning. The items in the two lists were not related between the lists. However, a strategy which the Ss learned during the first list (based on the pre-experimentally established word-frequency, according to Lorge and Thorndike 1944, attributes of wrong and right items in the list) was found to transfer to performance on the second list. For a transfer task, the Ss received either an HL (high frequency correct, low frequency incorrect), LH (low frequency correct, high frequency incorrect), or HL-LH (mixed) list as the second list. Kausler and Farzanegan reasoned that a generalized strategy to select H in an HL first list should yield highly efficient performance on a second HL list, even though the items were different in the lists (the relationship was A-B, C-D). Also, a generalized strategy to select L items in an LH first list should transfer to a second LH list. Groups which received either HL to LH or LH to HL conditions paralleled reversal-shift conditions, and if such shifts do occur, transfer under these conditions should have been very efficient. Transfer from or to a mixed list, however, should not be enhanced by a selection strategy, with inefficient performance on the second list, relative to an HL-LH to HL-LH condition, being the expected outcome.

In the transfer task, HL to HL, the Ss performed at a nearly perfect level; the LH to HL (reversal-shift condition) required a slightly greater number of trials than HL to HL, although the number was not significant. HL-LH to HL (not expected to benefit by transfer of a selection strategy) required a significanly greater number of trials than Group HL-HL, and more trials than Group LH to HL; the latter number was not significant. LH to LH failed to show a facilitative transfer effect. Both Group LH to LH and Group HL to LH failed to show a significant advantage over Group HL-LH to LH, which was probably due to the fact that Rule II was harder to learn.

Although the above evidence offers much support for frequency theory, some anomalous findings have been reported which offer evidence unfavorable to frequency theory (Underwood et. al., 1964; Raskin, Boice, Rubel, and Clark, 1968; Underwood and Freund, 1968; 1969). The difficulties reported in these experiments, all of which have conditions where frequencies of incorrect items were initially higher than correct items, develop from the counting postulate. That the correct item in each pair receives two frequency units per trial and the incorrect item in each pair one frequency unit is a basic assumption of frequency theory. This assumption has been useful in generating predictions that have received general confirmation; however, frequency theory has failed to account for certain results relative to the point of maximum difficulty in lists with wrongs initially higher in frequency than rights. Frequency theory predicts that at some point in the list, when wrong items are initially higher in frequency, the correct items must catch up in frequency units with the incorrect items. Since the S rehearses the correct alternative, at this point when frequency units for correct and incorrect items are equal, the frequency cue should be ineffective and performance should be at a chance level. This return-to-chance level of performance has not been found. In the Underwood et al. (1964) experiment, for example, Group W began second-list learning with an initial frequency advantage in favor of the wrong items. Group W was eventually overtaken by a control group; however, Group W did not show a return-to-chance level of performance.

In the Raskin et al. (1968) experiment, a transfer procedure was used where one group had their List-1 right items switched to List-2 wrong items and List-1 wrongs switched List-2 rights. The groups failed to show a return-to-chance level of performance.

Underwood and Freund (1969) varied the number of different words used as incorrect alternatives (2, 4, 6, or 12) within a 12-pair list. According to the counting postulate, the point of maximum difficulty should have occurred when six different words (each word presented in two pairs) were used as incorrect alternatives, since each correct word would have gained two frequency units per trial as well as each incorrect word. The results showed that the point of maximum difficulty occurred when four different words (each word presented in three pairs) were used as incorrect items, not in accord with expectations from frequency theory. Underwood et al. explained this failure of the theory by stating that possibly when a frequency differential is initially indiscriminate (as it might have been for both lists), the S turns to other characteristics or attributes of the words to establish the discrimination.

Underwood and Freund (1968) presented Ss with either 0, 2, or 5 free-recall learning trials (to build additional frequency units) before presenting a V-L task; then within the V-D list, the free-recall words appeared as either correct or incorrect. When the free-recall words appeared as the correct items within the V-D list, performance was essentially perfect on all trials. With two free-recall trials and the free-recall words appearing as incorrect within the V-D list, performance was similar to that of a control group. However, with five free-recall trials and the free-recall words appearing as incorrect within the V-D list, performance was initially very good but improvement across trials was slight. With performance initially very high with the latter group, one would assume that the free-recall trials transferred positively to the V-D list, making the incorrect words initially higher in frequency. According to the frequency theory, the frequency of the correct words should have caught up with the frequency of the incorrect words in the latter condition, and performance should have returned to a chance level.

Paul (1966) investigated verbal discrimination reversal (reversal meaning the right items on the first task become the wrong items on the second task or vice versa) as a function of overlearning, and the precentage of items reversed; he presented the Ss with either 0% (no trials after criterion was reached) or 50% (50% of the number of trials to criterion for the S) post criterion trials on the original 12 pairs of items. The Ss were then presented with a reversal condition consisting of the original 12 pairs of items with either 100%, 75%, 50%, or 25% of the original correct items designated as incorrect. The results showed that the mean number of errors to reversal criterion for the groups receiving 25% and 100% of the items reversed was very low, while the mean number of errors for the groups receiving 50% and 75% of the items reversed was considerably greater. According to frequency theory, it would be expected that the mean number of errors to criterion on the second (reversal) task would increase as a function of the number of items reversed. As a St. V-D task, the first list would have been learned through the application by the Ss of Rule I (choosing the most frequent item of the pair). When the items were reversed, however, Rule II (choosing the least frequent item of the pair) would have been more applicable. When only 25% of the items were reversed, the Ss performed well, indicating that an insufficient number of items had been reversed to cause confusion as to whether to apply Rule I or Rule II. When 50% of the items were reversed, however, it seems that confusion as to which rule to apply did exist. Rule I and Rule II would have been equally applicable when 50% of the items were reversed; when 75% of the items were reversed, however, it seems that confusion should not have existed. Rule II would have been more applicable, but confusion possibly existed because the Ss entered the situation with a preference for Rule I. This preference apparently interfered with the application of Rule II. When 100% of the items were reversed, performance was very good, indicating that the Ss were able to apply Rule II successfully.

Newman, Suggs, and Averitt (1972) studied the use of Rule I and Rule II in a single-list situation; further studying the effects of designating an item as "right" or "wrong" on study trials, and also of requiring the Ss to designate either the "right" or "wrong" items on test trials. In the study, two experiments were used. In the first experiment, the right item of each pair was repeated for half of the Ss, and the "wrong" item of each pair was repeated for the other half of the Ss. On test trials, half of the Ss in each of these two treatment groups were asked to indicate the "right" item of each pair; the other half of each treatment group was asked to indicate the wrong item of each pair. The four groups were designated as RR, RW, WR, and WW. The reasoning behind the experiment was that, for the Ss in Groups RR and WW, use of Rule I would be the more appropriate since the S was to respond on the test with the more frequent item of each pair, whereas for the Ss in Groups RW and WR use of Rule II would be the more appropriate, since the S was to respond on the test with the less frequent item of each pair. Results indicated that performance during V-D training was not affected by whether the "right" or "wrong" item of a pair was repeated during the study trials, nor was performance affected by whether the S had to designate the "right" or the "wrong" item during test trials. As long as a discriminable difference in frequency existed between items of a pair, the S responded correctly by selecting the more frequent item of the pair. The results appeared to suggest that performance may be better with Rule I than with Rule II. An alternative interpretation was proposed, which was based on the assumption that the Ss rehearse covertly the items they expect to have to designate on the test. Were this so, then it may have been that Rule I was the more appropriate for the Ss in all four groups since there would be more covert rehearsal for those items the Ss were told they would be tested for, and the better performance by the Ss in the RR and WW groups would then be due to a greater difference in frequency units between items of each pair than in the other two groups.

Experiment II was a study to evaluate the assumption that the Ss rehearse covertly the items they expect to have to designate on the test. The Ss were presented each item of a pair only once on study trials, they were instructed as to which item of the pair ("right" or "wrong") they were to learn, and they were tested, half of the Ss for the same items, half for the other items. The rationale behind the experiment was that if frequency units accrue only as a result of overt occurrence of the items, then all groups will perform at chance level; if, however, the Ss rehearse covertly those items they expect to be tested for, more frequency units should accrue for those items and the Ss should perform at better-than-chance level. Also, if Rule II is as easy to use as Rule I, the performance of the four groups should be equivalent. The results showed that, first, instructing the S to learn the "wrong" items had the same effect as instructing him to learn the "right" items. Second, having the S designate the "wrong" item on a test had the same effect as having him designate the "right" items; these results replicate those of the first experiment. Third, Groups RR and WW performed at a better-than-chance level and these results are in accord with the assumption that Ss rehearse covertly the items they are instructed to learn. If Ss do rehearse covertly the items they are instructed to learn, then the better performance of these Ss than that of the other two groups (RW and WR) suggests that performance is better in situations in which Rule I is more appropriate than in situations in which Rule II is more appropriate.

The present experiment was a test of the possibility that a fifth mechanism, whereby the Ss rehearse the incorrect alternative (RIR) along with the correct alternative in certain situations, is responsible for the anomolous results. Specifically in V-D tasks having initially high-frequency items incorrect, Ss may use RIR as well as Rule II; RIR and Rule II are probably more difficult to use than Rule I because Rule I is more commonly used in every-day life. If Rule II is more difficult to use, Ss may not return to chance as happened in Underwood et al. (1964), Raskin et al. (1968), etc. In the field of cognitive psychology, there is some evidence to support the RIR in Rule II.

Freibergs and Tulving (1961) found that it was more difficult for Ss to infer the relevant attributes of a conceptual category when they were forced to proceed from negative instances, which means the Ss were forced to learn what the concept was not as opposed to what the concept was. However, the experiment showed that this difficulty can be largely overcome through extensive practice on many different problems where negative instances are used.

If the RIR operates, groups which have learned to use Rule II should perform well when the second list has low-frequency items correct, and should have trouble when the second list has high-frequency items correct. In the present study one group of Ss learned a changing correct (CC) list where new correct items were presented with each trial, the incorrect items remaining the same. This group should learn to apply Rule II, and select the least frequent item of the pair as the correct item, and also may learn to rehearse the incorrect alternative. A second group of Ss learned a standard V-D (St. V-D) list, in which both the correct and incorrect items remained the same in all trials. A transfer procedure was involved, where 1/3 of the Ss in each of the two groups above learned an HL (high frequency correct, low frequency incorrect) list, 1/3 an LH (low frequency correct, high frequency incorrect) list, and 1/3 a St. V-D (Standard V-D) list. The HL group should use Rule I and the RCR and do poorly for the CC-HL transfer task, and do well for the St. V-D-HL transfer task, while the LH group should use Rule II and the RIR and do well for the CC-LH transfer task, and do poorly for the St. V-D-LH transfer task. The St. V-D-St. V-D group served as a control group.

# METHOD

Subjects.

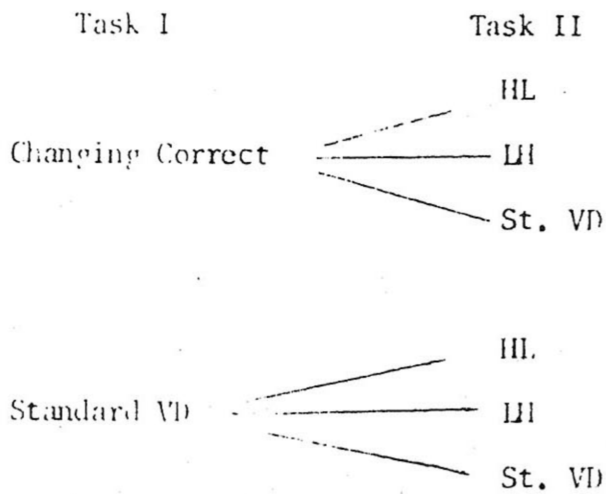
The Ss were 90 undergraduate students of Kansas State College of Pittsburg, enrolled in a general psychology course. The Ss were naive to V-D learning procedures. The Ss were volunteers who received extra course credit for participating in the experiment.

Design.

The design was a 2 X 3 factorial with the type of transfer task being one variable and level of frequency of correct and incorrect items being the other variable. Type of transfer involved either presenting the Ss with a CC list or St. V-D list for task 1. The level of frequency of correct and incorrect items in the second list was the task 2 variable. The second lists the Ss learned consisted of one list having the incorrect words as high frequency and the correct words as low frequency (LH), a list having the correct words as high frequency and the incorrect words as low frequency (HL), and a St. V-D list composed of medium frequency words. One-half of the Ss learned a CC list, one-half learned St. V-D list. Transfer tasks were CC-HL, CC-LH, CC-St. V-D, St. V-D-HL, St. V-D-LH, St. V-D-St. V-D. Figure I is a schematic representation of the design.

Figure I

Schematic Representation of Design



Lists.

Three pools of five-letter words were selected randomly, one pool being low frequency words, one pool being medium frequency words, and one pool being high frequency words. Low frequency words were selected randomly from words which occurred 20 - 30 times per million in the general count as given by Thorndike and Lorge (1944). Medium frequency words were selected randomly from words which occurred 30 - 100 times per million in the same general count by Thorndike and Lorge (1944). High frequency words were selected randomly from words which occurred 1000 or more times in either the Lorge Magazine Count or the Lorge-Thorndike Semantic Count (1944). All randomization was done with a table of random numbers which appears in Edwards (1953). After the pools of words were selected, the number of words required to form a list were randomly selected from the appropriate pool (i.e. the HL list was formed by selecting high frequency words from the high frequency list and low frequency words from the low frequency list). Whether a word was to be correct or incorrect was also determined randomly.

Procedure.

A 2:2 rate of presentation with the anticipation method was used. During the feedback interval, both items were seen by the Ss in all conditions, with the correct item being designated by an asterisk. The lists were recorded on videotape and then played back on a TV monitor. The lists were presented to the Ss in random order, and the top-bottom position of the items on the tape was randomly determined. The CC list differed form the other lists in that a new correct item was paired with the incorrect item each trial; the method of presentation was the same, however. For the other lists, the same pairs of correct and incorrect items were presented each trial, but the order of presentation was varied. All lists included 15 pairs of items.

The Ss were shown the lists until they reached a criterion of three perfect trials. The Ss were instructed to say aloud the alternative which they thought was correct, and they were instructed to respond to each pair on each trial.

The conditions (conditions meaning the groups, such as CC-HL, CC-LH, etc) were block randomized, which means that the Ss were assigned to conditions in random order, each condition being assigned one S before the next block of conditions was assigned a S. The Ss were assigned to a condition as they appeared at the testing site. Thus, the assignment of the Ss to conditions was random. There were 15 Ss in each of the six conditions.

# RESULTS

An analysis of variance (Winer, 1962) was conducted for each performance measure (trials to criterion and errors to criterion) for both task 1 and task 2. The analyses of variance on the first task were done as a function of the assigned second task; the Ss were assigned to the same group, for statistical analysis purposes, that they were in on the second task. Although the conditions that the Ss were in were not really separated until the second task, if an S was assigned to test in the HL group and if he learned the CC task first, he was considered in the CC-HL subgroup for the purpose of the statistical analysis of the first task performance. The analysis of variance on the first task was done to determine if further analyses could be done on task 2, since if there were differences among subgroups on task 1, differences in task 2 could be attributed not to task 2 performance but to task 1 performance. Table I is a summary of the analysis of variance for trials to criterion on the first task. The analysis of variance for trials to criterion on the first task showed a significant difference between the two categories of transfer (the two categories of transfer were the CC task and the St. V-D task), $F(1, 84) = 12.12$, $p < .01$. The analysis showed no significant difference among the types of second task,

TABLE I

Trials to Criterion

Summary of Analysis of Variance

First Task

| Source of Variation | SS | df | MS | F |
|---|---|---|---|---|
| A (category of transfer) | 80.27 | 1 | 80.27 | 12.12** |
| B (type of second list) | 20.82 | 2 | 10.41 | 1.57 |
| AB | 17.64 | 2 | 8.82 | 1.33 |
| Within Cell | 556.26 | 84 | 6.62 | |
| Total | 674.99 | 89 | | |

** p 01

$F(2, 84) = 1.57$, $p > .05$, and no significant interaction between category of transfer and type of second task, $F(2, 84) = 1.33$, $p > .05$. Table II is a summary of the analysis of variance for errors to criterion on the first task. The analysis of variance for errors to criterion on the first task showed no significant difference between the two categories of transfer, $F(1, 84) = .13$, $p > .05$, no significant difference among the types of second task, $F(2, 84) = 1.55$, $p > .05$, and no significant interaction between category of transfer and type of second task, $F(2, 84) = .58$, $p > .05$.

A Hartley test of the homogeneity of variance (Winer, 1962) was performed, for both trials to criterion and errors to criterion, on the second task. The trials to criterion test showed the variance not to be homogeneous among the different tasks, $F_{max\ .99}(6, 14) = 8.2$; the value obtained was $F = 10.14$. Although the variance was not homogeneous, this lack of homogeneity only existed between the two maximum variances, and all but two of the means were homogeneous; and even then the variance was almost homogeneous. Since the analysis of variance is a very robust - i.e. insensitive to violations of its assumptions- the lack of homogeneity between the two extreme means was not considered to be important. Also, there is some evidence that all of the tests of homogeneity are too sensitive.

TABLE II

Errors to Criterion

Summary of Analysis of Variance

First Task

| Source of Variation | SS | df | MS | F |
|---|---|---|---|---|
| A (category of transfer) | 36.11 | 1 | 36.11 | .13 |
| B (type of second list) | 836.35 | 2 | 418.17 | 1.55 |
| AB | 315.46 | 2 | 157.73 | .58 |
| Within Cell | 22,623.87 | 84 | 269.33 | |
| Total | 23,811.79 | 89 | | |

Table III is a summary of the analysis of variance for trials to criterion on the second task. The analysis of variance for trials to criterion on the second task showed a significant difference between the two categories of transfer, $F(1, 84) = 10.66$, $p < .01$, and a significant difference among the types of second task, $F(2, 84) = 3.90$, $p < .05$. The analysis showed no significant interaction between category of transfer and type of second task, $F(2, 84) = .48$, $p > .05$. Table IV is a summary of the analysis of variance for errors to criterion on the second task. The analysis of variance for errors to criterion on the second task showed a significant difference between the two categories of transfer, $F(1, 84) = 16.17$, $p < .01$, and a significant difference among the types of second task, $F(2, 84) = 8.42$, $p < .01$. The analysis showed no significant interaction between category of transfer and type of second task, $F(2, 84) = 2.59$, $p > .05$. The analysis for errors to criterion and trials to criterion showed the same effects, a difference between the two categories of transfer, a difference among the types of second task, and no interaction between category of transfer and type of second task.

Because the analysis of variance showed significant differences among the types of second task, a Tukey A test (Winer, 1962) was performed on the means to determine pre-

TABLE III

Trials to Criterion

Summary of Analysis of Variance

Second Task

| Source of Variation | SS | df | MS | F |
|---|---|---|---|---|
| A (category of transfer) | 134.44 | 1 | 134.44 | 10.66 ** |
| B (type of second task) | 98.42 | 2 | 49.21 | 3.90 * |
| AB | 12.30 | 2 | 6.15 | .48 |
| Within Cell | 1058.80 | 84 | 12.60 | |
| Total | 1303.96 | 89 | | |

** p .01

* p .05

TABLE IV

Errors to Criterion

Summary of Analysis of Variance

Second Task

| Source of Variation | SS | df | MS | F |
|---|---|---|---|---|
| A (category of transfer) | 3829 | 1 | 3829 | 16.17 ** |
| B (type of second task) | 3989 | 2 | 1994.5 | 8.42 * |
| AB | 1244 | 2 | 612 | 2.59 * |
| Within Cell | 19887 | 84 | 236.75 | |
| Total | 28949 | 89 | | |

** $p < .01$

* $p < .05$

cisely what means were different. For the trials to criterion analysis on the second task, the Tukey test showed the significant gap to be 2.12. Table V shows the mean trials to criterion for the second task. Among the groups which learned the CC first task, the St. V-D group did significantly poorer than the LH group (3.07), and significantly poorer than the HL group (2.20). The HL and LH groups were not significantly different from each other (.87). Among the groups which learned the St. V-D task first, the St. V-D group did significantly poorer than the HL group (2.14), but the St. V-D group did not do significantly poorer than the LH group (1.47). The HL and LH groups were not significantly different from each other (.67).

Table VI shows the mean errors to criterion for the second task. The Tukey A test on the mean errors to criterion showed the significant gap to be 9.27. Among the groups which learned the CC task first, the St. V-D group did significantly poorer than the LH group (23.67), and the St. V-D group did significantly poorer than the HL group (18.47). The HL and LH group performances were not significantly different from each other. Among the groups which learned the St. V-D task first, the St. V-D group performance was not significantly different from the LH group performance (5.47), nor from the HL group perfor-

TABLE V

Mean Trials to Criterion

Second Task

| | HL | LH | St. VD | |
|---|---|---|---|---|
| CC | 10.8 | 9.93 | 13 | 33.73 |
| St. VD | 7.86 | 8.53 | 10 | 26.39 |
| | 18.66 | 18.46 | 23 | 60.12 |

TABLE VI

Mean Errors to Criterion

Second Task

| | HL | LH | St. VD | |
|---|---|---|---|---|
| CC | 30.13 | 24.93 | 48.6 | 103.66 |
| St. VD | 17.47 | 20.8 | 26.27 | 64.54 |
| | 47.60 | 45.73 | 74.87 | 168.20 |

mance (8.80), although the St. V-D group almost did significantly poorer than the HL group. The HL and LH group performances were not significantly different from each other (3.33).

In the comparison of means between the two types of transfer, the trials to criterion analysis showed the CC-HL group performance to be significantly poorer than the St. V-D-HL group (2.94). The CC-LH group performance was not significantly different from the St. V-D-LH group performance (1.40). The CC-St. V-D group performance was significantly poorer than the St. V-D-St. V-D group performance (3.00). The errors to criterion analysis showed the CC-St. V-D group performance to be significantly poorer than the St. V-D-HL group performance (12.66). The CC-LH group performance was not significantly different from the St. V-D-LH group performance (4.13). The CC-St. V-D group performance was significantly poorer than the St. V-D-St. V-D group performance (22.33)

The average number of errors per trial per group was computed, and the results are shown graphically in Figure II. Figures III, IV, V, VI, and VII are taken from Figure II, and show various comparisons of the groups. Figure III shows the groups which learned the CC task first, while Figure IV shows the groups which learned the St. V-D task first. Figure V shows the CC-HL group compared to the St. V-D-HL group, Figure VI shows the CC-LH group compared to the St. V-D-LH group, and Figure VII shows the CC-St. V-D group compared to the St. V-D-St. V-D group. The tables of means for trials and for errors are shown in Figure VIII.

Task II

Average Error Per Trial Per Group

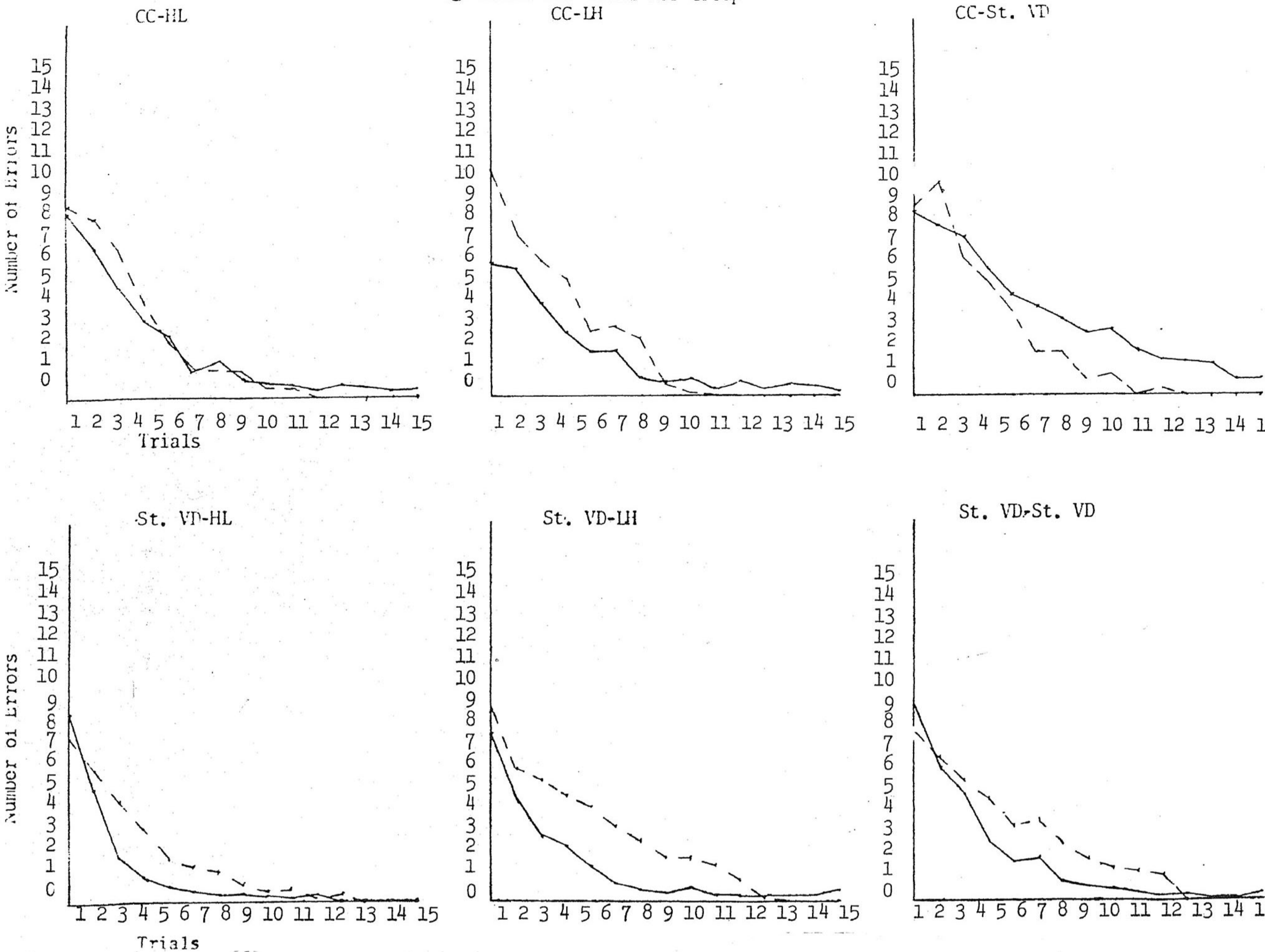



Figure II

Average Errors Per Trial Per Group

Figure III

Average Error Per Trial

(Groups Which Learned CC Task First)

CC-HL - - - -

CC-LH ——— - ——— -

CC-St. V-D ———

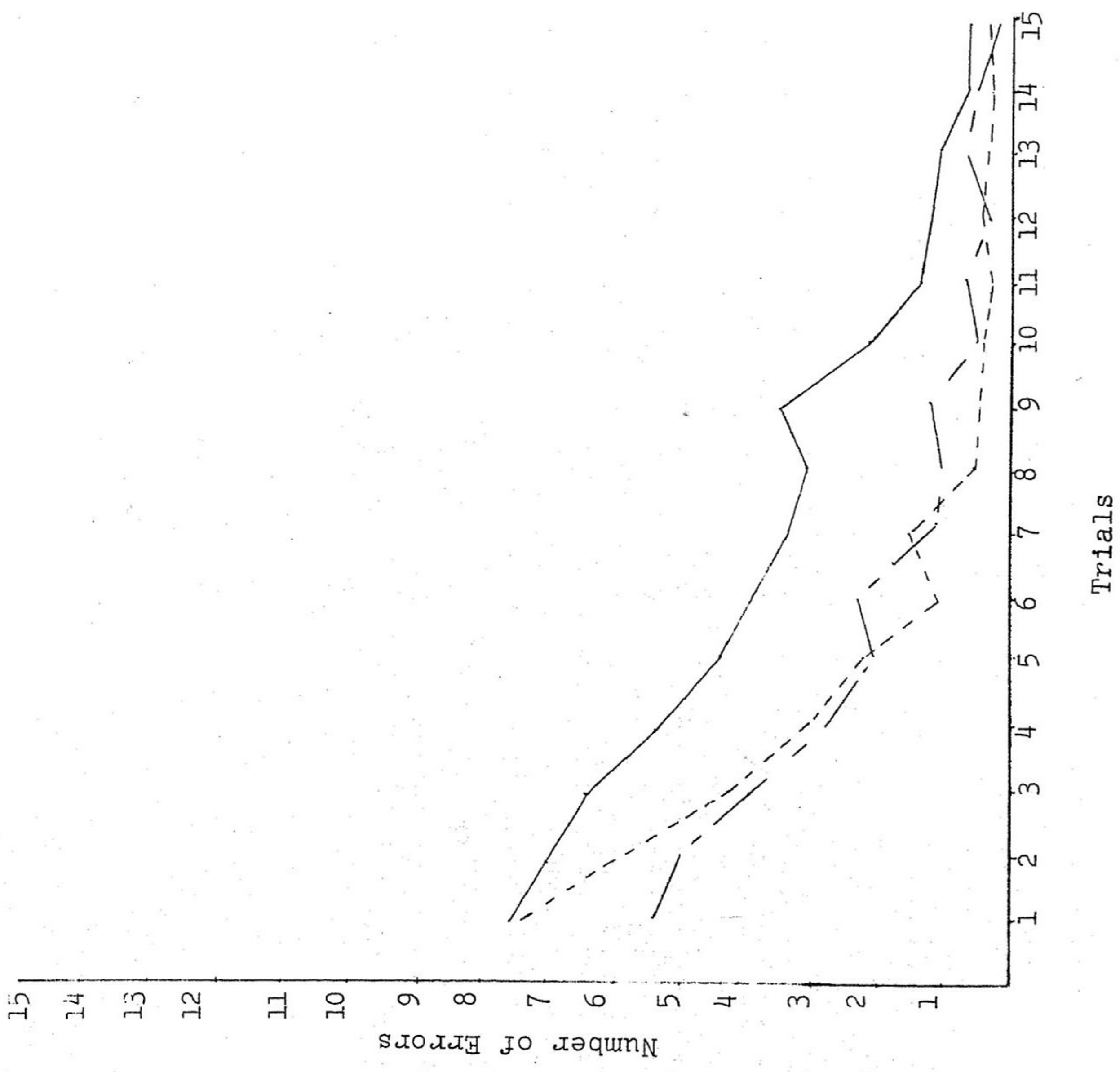


Figure IV

Average Error Per Trial

(Groups Which Learned St. V-D Task First)

St. V-D-HL -------

St. V-D-LH ________ - ________ -

St. V-D-St. V-D ______________

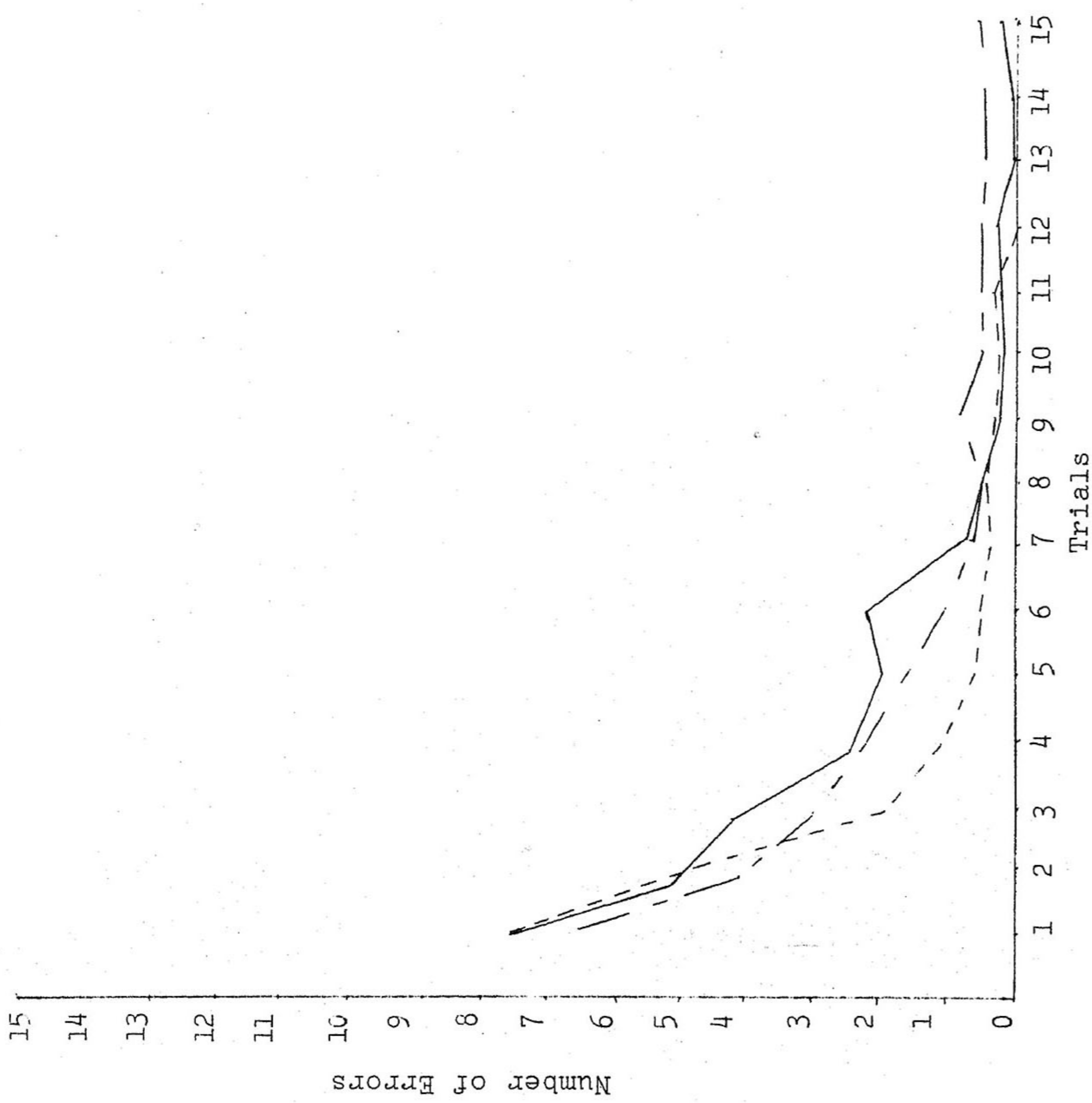


Figure V

Average Error Per Trial

(CC-HL Compared to St. V-D-HL)

CC-HL -------------

St. V-D-HL ————————

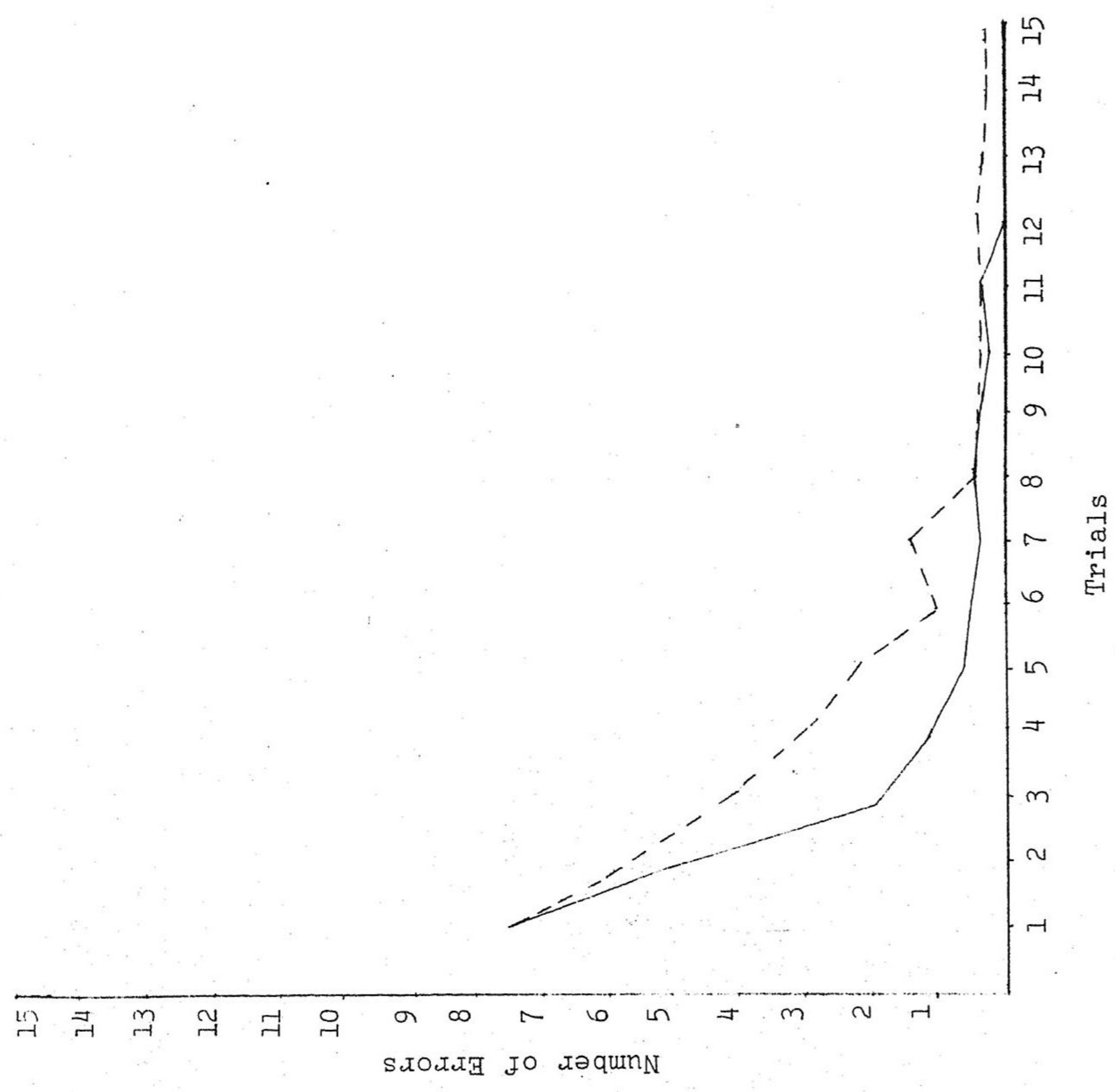


Figure VI

Average Error Per Trial

(CC-LH Compared to St. V-D-LH)

CC-LH ---------------

St. V-D-LH ___________

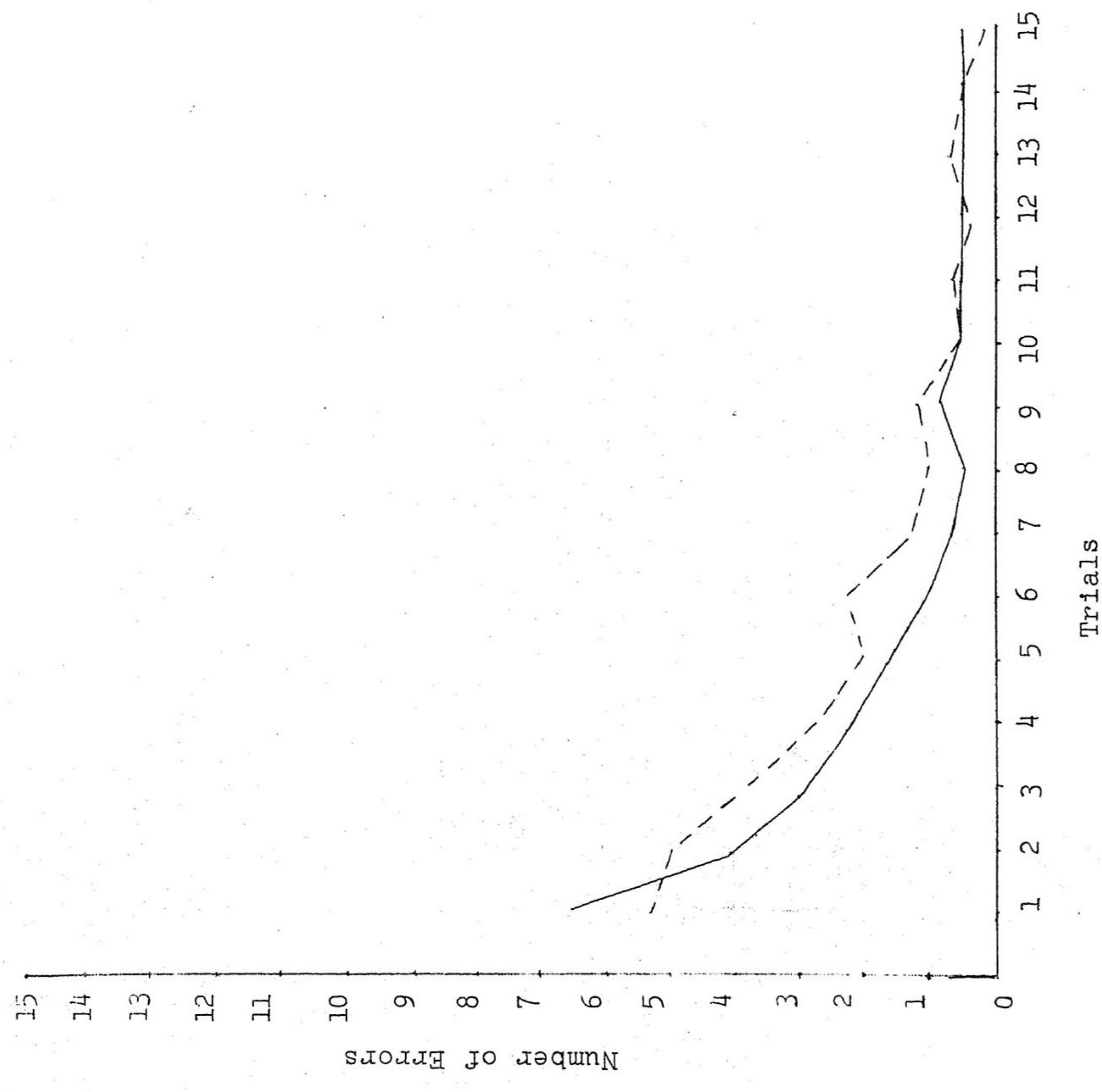


Figure VII

Average Error Per Trial

(CC-St. V-D Compared to St. V-D-St. V-D)

CC-St. V-D --------------

St. V-D-St. V-D ________

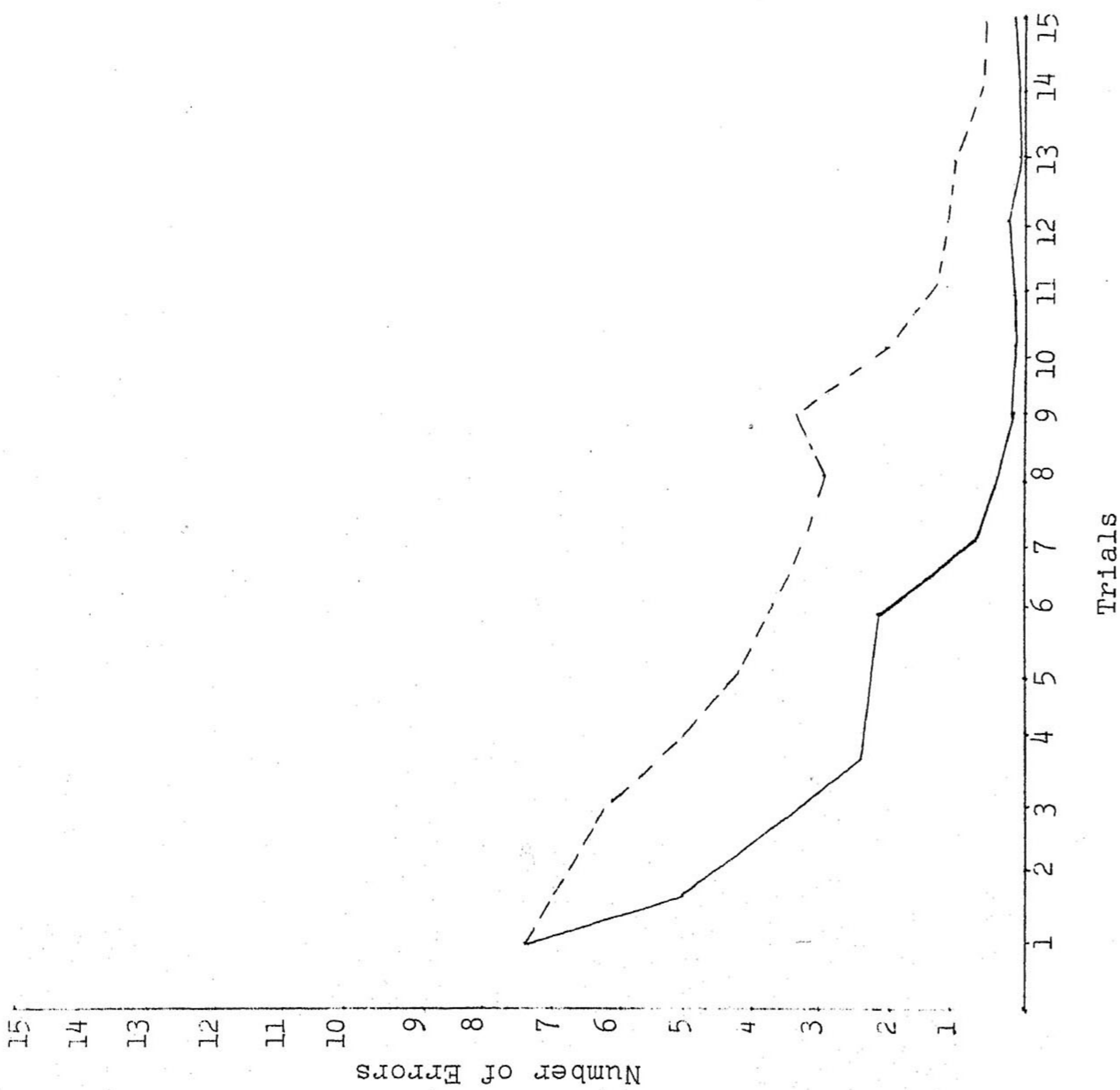
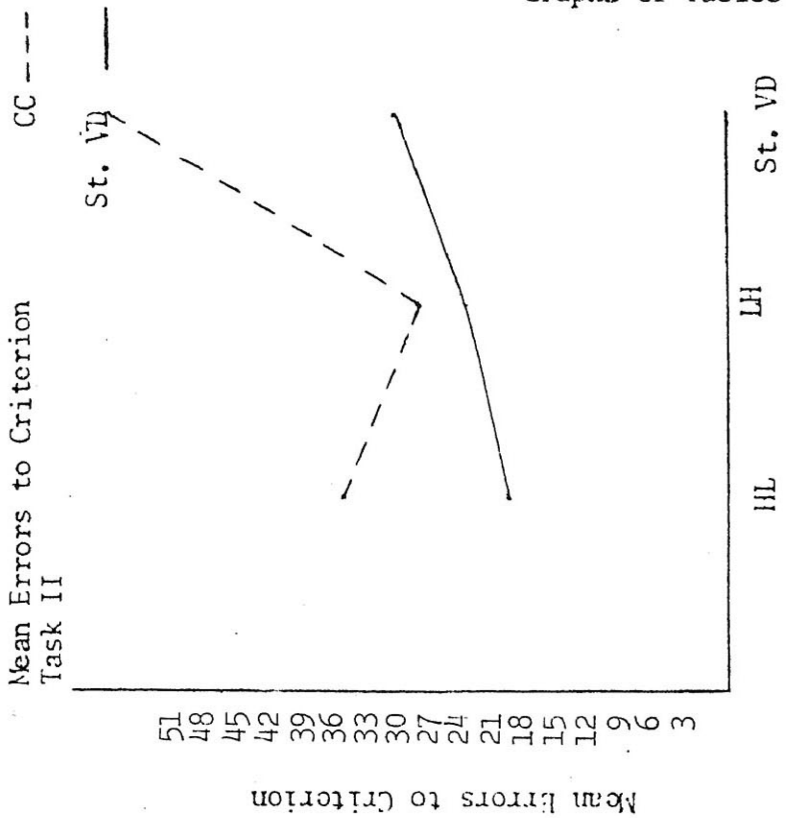
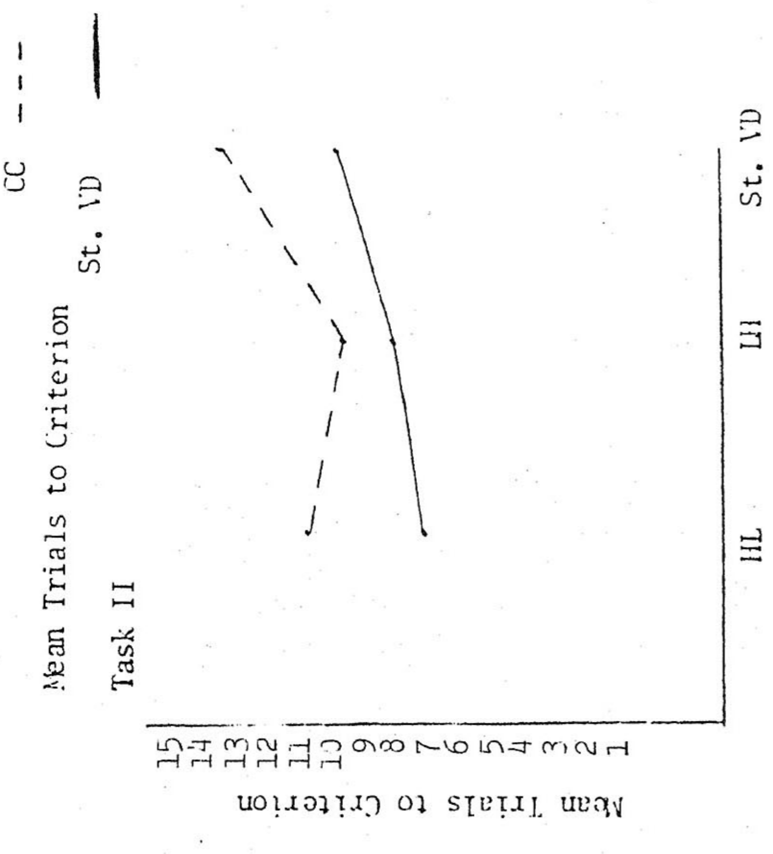


FIGURE VIII

Graphs of Tables of Means



# DISCUSSION

While many of the experiments done in V-D learning have supported frequency theory, there have been some anomalous results. A discussion of the supporting evidence and of the non-supporting evidence was presented above. Some examples of the anomalous results (Underwood et al., 1964; Raskin, Boice, Rubel, and Clark, 1968; Underwood and Freund, 1968; 1969) were the failure to return to a chance level of performance in some experiments and a point of maximum difficulty, when the number of different words used as incorrect alternatives in a list were varied, which was not in accord with expectations from frequency theory. If the Ss were using the RIR, these anomalous results could be explained, since the incorrect items would accrue frequencies at a faster rate than would be expected from traditional frequency theory.

The analysis of the results showed that some of the results supported the hypothesis of the RIR mechanism, while some of the results did not support the hypothesis of the RIR mechanism. The conclusions drawn as to whether the results supported the RIR hypothesis were based upon expectations as to positive and negative transfer mentioned earlier.

Since the analysis of variance on the first tasks showed no significant differences in the performance of the groups within the CC or St. V-D tasks, any differences which showed up among the second task groups with a similar first task was attributed to a differential positive or negative transfer effect. Although the analysis of variance showed a significant difference between the CC and St. V-D groups on the first task, this difference was expected; the comparison of second task means was a comparison of the differential effects of the transfer of learning which took place on the first task.

Regarding Figures II through VII, on task 1 the CC-LH group appears to have had worse than chance performance, while all the other groups had performance close to chance on trial 1. Also, the CC-LH group on task 2 appears to be the only group which performed better than chance on trial 1, but had remarkably slow improvement on the remainder of the tasks. It appears that one of the major differences between the CC groups and the St. V-D groups is the shape (slope) of the curves, the slope being generally less for the CC groups than the St. V-D groups.

Since the analyses of variance on the second task showed that, for both errors to criterion and trials to criterion, significant differences existed among the types of second task, the various means were compared and some support for the hypothesis of the RIR mechanism was found. Although no significant difference was found between the CC-HL and CC-LH groups, both of these groups' performance was found to be significantly better than the CC-St. V-D group. Since the CC-St. V-D group right and wrong items were all of medium frequency, the Ss had no frequency differential to base their judgments upon at the outset. Although the CC-St. V-D group was not expected by the E to do so poorly, the fact that the performance was so poor can be explained in terms of the RIR mechanism. If the Ss were rehearsing the incorrect response, the frequency differences would not have accrued as they would have if the Ss were rehearsing the correct response. For example, suppose that two St. V-D words in the CC-St. V-D second task had been "throw" and "catch" and throw had been designated the correct item. If S chose "throw" and saw that it was correct during feedback, he may have rehearsed the incorrect response "catch". He would have rehearsed "catch" assuming that a new correct word would appear with the incorrect word "catch" on the next trial as had been the case with the first task the S learned, the CC list. Since the frequency differentials would not have accrued as they would have in a non-transfer St. V-D task, the performance of the Ss would have been poor. The CC-HL and CC-LH groups, of course, were composed of words of different frequency units (Lorge and Thorndike, 1944) at the outset on the second task, which apparently differentially effected their performance as compared to the CC-St. V-D group. The fact that the CC-St. V-D group did poorer than the control group St. V-D-St. V-D lends support to the RIR mechanism, since the St. V-D-St. V-D group was not experiencing the same negative transfer at the CC-St. V-D group.

Further support for the RIR mechanism came from the fact that the St. V-D-HL group did significantly better than the CC-HL group. Since the CC-HL group learned to choose the least frequent item (Rule II) and rehearse the more frequent (incorrect) response, they may have experienced negative transfer when they tried to learn a list where the correct items were of high frequency and the incorrect items were of low frequency (HL). If the Ss on the HL task tried to choose the low frequency item as correct and rehearse the high frequency item, they would experience difficulty. The St. V-D-HL group, on the other hand, may have experienced some positive transfer. The use of Rule I, choosing the most frequent item, is appropriate to a St. V-D task. The Ss, by rehearsing the correct alternative, build up frequency units for the correct alternative, and subsequently choose the most frequent item. When the St. V-D group was given a second task where the high frequency words were correct (HL), no difficulty arose; thus, the performance of the St. V-D-HL group was better than that of the CC-HL group.

Not all of the evidence supported the RIR mechanism. According to the RIR hypothesis, the CC-LH group would be expected to do better than the CC-HL group; this better performance of the CC-LH group did not occur. The CC-LH group was expected to perform well because of the expected positive transfer from the CC task to the LH task. The Ss learned to choose the least frequent item on task 1 (the CC task); since the LH task was composed of low frequency correct and high frequency incorrect items, the Ss were expected to have no difficulty in rehearsing the high frequency (incorrect) response and applying Rule II in choosing the least frequent (correct) response. One possible explanation for the failure of the CC-LH group to do better than the CC-HL group is that the CC-LH group may have been choosing the incorrect item because of some association which may have been operating with the right and wrong items. Even though the low frequency items were correct, and the CC-LH group had learned to choose the least frequent item, the Ss may have associated more value with the correct items than with the incorrect items. Thus, the correct items would have come to "carry more weight" than the incorrect items. Because the Ss had been taught to choose the least frequent item by using Rule II and rehearsing the incorrect response, the association process may have made the use of Rule II and RIR difficult by giving more weight to the low frequency item, in effect partially balancing out the frequency differential because of association. Another possible explanation for the failure of the CC-LH group to do better than the CC-HL group is that Rule II, less commonly used by the Ss, is simply harder to learn to use in an experimental situation than is Rule I. Since the CC-HL task required the Ss to use Rule I and the CC-LH task required the Ss to use Rule II, and since the Ss learned to use Rule II in the CC task, the pre-experimental bias toward the use of Rule I may have balanced out the effects of the first task (CC) training on the CC-LH group.

Further evidence which did not support the RIR hypothesis was the failure of the CC-LH group to do better than the St. V-D-LH group. The failure of the CC-LH group to do as well as expected was discussed above. The failure of the St. V-D-LH group to do as poorly as expected might be explained by association. Although there was a frequency differential at the outset of the second task (low frequency correct and high frequency incorrect), the correct items acquired more weight through an association mechanism. Since the Ss had learned to use Rule I in the first task (St. V-D), they would, after the first trial on the second task, begin to use Rule I and do well. Thus, the St. V-D-LH group would not do as poorly as expected.

Support was given to the operation of Rule I and the corresponding use of the RCR mechanism by the better performance of the St. V-D-HL group as compared to the St. V-D-St. V-D group. Although the Tukey A test showed the St. V-D-HL group to be not quite significantly better in performance than the St. V-D-St. V-D group in total errors, the St. V-D-HL group did do significantly better in trials to criterion. The E had expected some positive transfer in the St. V-D-HL group because Rule I was learned and the second task (HL) should have been easy for the Ss to use Rule I with for reasons discussed above. The St. V-D-St. V-D task had no frequency differential between right and wrong items at the outset, so little positive transfer was expected.

Some support for the association mechanism discussed above was given by the responses of the Ss to the question of how they learned the lists after they had completed the experiment. Most of the Ss said they had used association. Some support for an association mechanism has been presented in historical discussion of V-D learning, Kanak and Dean (1969) in the discussion of the results of an experiment done with verbal learning discussed an association mechanism. Wallace (1972) discussed multiple components analysis as an additional mechanism of verbal learning; the multiple components analysis states that frequency theory is incomplete and must add additional components in order to explain learning in the verbal-discrimination paradigm. Kanak and Dean (1969) proposed association between right and wrong items as an additional component.

From the observation of the results it appears that both the RIR and some association mechanism were operating. The RIR was not disproved as a possible explanation of anomalous results found in V-D experimentation, however, the evidence from the present study must remain inconclusive. More experimentation is recommended, with attempts being made to isolate both the RIR and association mechanism if they are in fact operating.

References

Edwards, Allen L. Experimental Design in Psychological Research. New York: Rinehart and Company, Inc., 1953.

Ekstrand, B. R., Wallace, W. P., and Underwood, B. J. A frequency theory of verbal-discrimination learning. Psychological Review, 1966, 73, 566-578.

Freibergs, V., and Tulvings, E. The effect of practice on utilization of information from positive and negative instances in concept identification. Canadian Journal of Psychology, 1961, 15, 101-106.

Kanak, N. J., and Dean, M. F. Transfer mechanisms in verbal discrimination. Journal of Experimental Psychology, 1969, 79, 300-307.

Kausler, D. H., and Farzanegan, F. Word frequency and selection strategies in verbal-discrimination learning. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 196-201.

Newman, Slater E., Suggs, Ralph E., and Averitt, Carol H. Use of Rule 1 and Rule 2 in verbal-discrimination learning. (Presented at a meeting of Psychonomic Society, St. Louis, November 1972) North Carolina State University, unpublished.

Paul, C. Verbal discrimination reversal as a function of overlearning and percentage of items reversed. Journal of Experimental Psychology, 1966, 72, 271-275.

Raskin, D. C., Boice, C., Rubel, E. W., and Clark, D. Transfer tests of the frequency theory of verbal discrimination learning. Journal of Experimental Psychology, 1968, 76, 521-529.

Thorndike, E., L., and Lorge, I. The Teachers Word Book of 30,000 Words. New York: Columbia University Press, 1944.

Underwood, B. J., and Jesse, F., and Ekstrand, B. R. Knowledge of rights and wrongs in verbal-discrimination learning. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 183-186.

Underwood, B. J., and Freund, J. S. Two tests of a theory of verbal-discrimination learning. Canadian Journal of Psychology, 1968, 22, 96-104.

Underwood, B. J., and Freund, J. S. Verbal-discrimination learning with varying numbers of right and wrong terms. American Journal of Psychology, 1969, 82, 198-202.

Wallace, W. P. 1972. Verbal discrimination. In Duncan, C. P., Sechrest, L., and Melton, A. W. (Eds.), Human Memory: Festschrift in Honor of Benton J. Underwood. New York: Appleton-Century-Crofts, 1972, pp. 283-315.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company, 1962.

Yelen, D. Verbal discrimination as a function of similarity and frequency of incorrect items. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 552-553.

# Appendix

## Examples of Words Used

## EXAMPLES OF WORDS USED

Medium Frequency

Anticipation Interval

SCOUT

ASIDE

Feedback Interval

SCOUT

ASIDE *

High and Low Frequency-High Frequency Correct (HL)

Anticipation Interval

BLACK

PERCH

Feedback Interval

BLACK *

PERCH