

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

THE EFFECTS OF MULTIPLE-EXEMPLAR TRAINING AND STIMULUS
VARIABILITY ON GENERALIZATION TO LINE GRAPHS

by

SAL V. PLUCHINO

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

1998

UMI Number: 9830752

Copyright 1998 by
Pluchino, Sal Vincent

All rights reserved.

UMI Microform 9830752
Copyright 1998, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized
copying under Title 17, United States Code.

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

© 1998

SAL VINCENT PLUCHINO

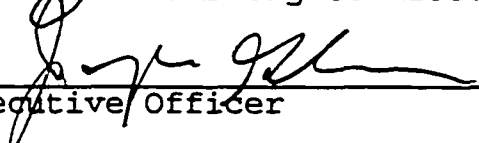
All Rights Reserved

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

3/27/98
Date

3/31/98
Date


Chair of Examining Committee


Executive Officer

Lanny Fields Ph.D.

Nancy S. Hemmes Ph.D.

Robert N. Lanson Ph.D.

Gene S. Fisch Ph.D.

Eric G. Heinemann Ph.D.

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

THE EFFECT OF MULTIPLE-EXEMPLAR TRAINING AND STIMULUS
VARIABILITY ON GENERALIZATION TO LINE GRAPHS

by

Sal V. Pluchino

Advisor: Professor Lanny Fields

Behavior analysts often assess the effectiveness of interventions by visual inspection of line graphs. The dissertation identified factors that influence accurate graph reading and generalization to novel graphs. Each stimulus was a line graph that had a left and a right panel called A and B, respectively. Each panel contained ten data points that were distributed around a regression line. The regression line in panel A always had a 0° slope. The regression line in panel B had slopes that varied from 0° to 18° in 2° increments. The dispersion of data points around the regression line in both panels was either 3 or 6 standard deviations (SD). The slope of the regression line in panel B defined each stimulus.

College students were trained to discriminate between 0° (S-) and 18° (S+) stimuli presented in 3 SD form. Single-exemplar training was conducted with one S+ and one S-. Multiple-exemplar training was conducted with forty-eight 18° S+'s and forty-eight 0° S-'s. The forty-eight S+'s differed by the location of data points, as did the forty-eight S-'s. Single-exemplar overtraining was conducted with one S+ and one S-, but the total number of trials matched that used in multiple-exemplar training.

Generalization tests were conducted using 0° through 18° stimuli in 3 SD and 6 SD form. Discriminative control

by the S+ and S- was disrupted after training or overtraining with single exemplars but was maintained after multiple-exemplar training. Disruption resulted from both decrements in S+ responding and increments in S- responding. More pronounced disruptions of stimulus control were seen when these stimuli were presented in 6 SD form than in 3 SD form.

A signal detection analysis of performance during 3 SD tests showed there was no bias regardless of prior training regimen. During 6 SD tests, there were substantial biases after single-exemplar and multiple-exemplars training but no bias after single-exemplar overtraining.

Slopes of generalization gradients were steeper for test stimuli presented in 3 SD form than 6 SD form. Slopes of generalization gradients were steepest after multiple-exemplar training.

Multiple-exemplar training was the only procedure that resulted in the maintenance of discriminative control and the induction of dimensional control during generalization tests.

Acknowledgment

This research was supported partially by ARI Contract MDA903-90-C-0132 and RACO Contract DASW01-96-X-0009.

Table of Contents

Title Page.....	i
Copyright Page.....	ii
Approval Page.....	iii
Abstract.....	iv
Acknowledgment.....	vi
Table of Contents.....	vii
List of Tables.....	viii
List of Figures.....	xi
Introduction.....	1
Experiment 1.....	19
Method.....	19
Results.....	35
Discussion.....	46
Experiment 2.....	48
Method.....	49
Results.....	52
Discussion.....	129
General Discussion.....	142
Appendix A.....	146
Appendix B.....	147
References.....	148

List of Tables

Table 1.	The Discrimination Index, the Percent of S+ Trials That Occasioned the DIFFERENT Response and the Percent of S- Trials that Occasioned the DIFFERENT Response to 0° and 18° Stimuli (3 <u>SD</u> and 6 <u>SD</u>) During Single-Exemplar Training and Generalization Testing for Experiment 1 (Averaged Data). Standard Error Is Shown in Parentheses.....	39
Table 2.	The Discrimination Index (DI), the Percent of S+ Trials that Occasioned the DIFFERENT Response and the Percent of S- Trials That Occasioned the DIFFERENT Response to 0° and 18° Stimuli (3 <u>SD</u> and 6 <u>SD</u>) During Single-Exemplar Training and Generalization Testing for Experiment 1 (Single-Participant [Part.] Data).....	40
Table 3.	Probability of the DIFFERENT Response to 3 <u>SD</u> Stimulus Slopes (Top) and 6 <u>SD</u> Stimulus Slopes (Bottom) for Experiment 1.....	45
Table 4.	The Average Discrimination Index, the Percent of S+ Trials that Occasioned the DIFFERENT Response and the Percent of S- Trials that Occasioned the DIFFERENT Response to 0° and 18° (3 <u>SD</u> and 6 <u>SD</u>) Stimuli During Discrimination Training and Generalization Testing for Experiment 2. Training Included Single-Exemplar Training (SET), Single-Exemplar Overtraining (SEOT), and Multiple-Exemplar Training (MET).....	56
Table 5.	Analysis of Variance (ANOVA) of the Discrimination Index for Training Group (T) and Variability of Test Stimuli (V) for Experiment 2.....	58
Table 6.	Post-Hoc, Pairwise-Comparisons of the Discrimination Index for the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) and Variability of Test Stimuli (3 <u>SD</u> and 6 <u>SD</u>) for Experiment 2.....	59
Table 7.	Chi-Square Analysis of the Discrimination Index During Experiment 2. The Number of Participants Who Achieved a $\geq 90\%$ Discrimination Index in the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) to 3 <u>SD</u> Stimuli (Top) and 6 <u>SD</u> Stimuli (Bottom) Is Shown.....	63

Table 8.	Analysis of Variance (ANOVA) of the Percent of S+ Trials that Occasioned the DIFFERENT Response for Training Group (T) and Variability of Test Stimuli (V) for Experiment 2.....	68
Table 9.	Post-Hoc, Pairwise-Comparisons of the Within-Group Percent of S+ Trials That Occasioned the DIFFERENT Response for Experiment 2. The Percent of S+ Trials in the 3 SD Form That Occasioned the DIFFERENT Response Is Compared to the Percent of S+ Trials in the 6 SD Form That Occasioned the DIFFERENT Response for the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET).....	69
Table 10.	Chi-Square Analysis of the Number of Participants Who Emitted the DIFFERENT Response for $\geq 90\%$ S+ Trials in Experiment 2. Data for the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar (MET) Training Group to 3 SD Stimuli (Top) and 6 SD Stimuli (Bottom) Are Shown.....	73
Table 11.	Analysis of Variance (ANOVA) of the Percent of S- Trials That Occasioned the DIFFERENT Response for Training Group (T) and Variability of Test Stimuli (V) for Experiment 2.....	78
Table 12.	Post-Hoc, Pairwise-Comparisons of the Percent of S- Trials That Occasioned the DIFFERENT Response for the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) and Variability of Test Stimuli (3 SD and 6 SD) for Experiment 2.....	79
Table 13.	Chi-Square Analysis of the Number of Participants Who Emitted the DIFFERENT Response for $\leq 10\%$ of S- Trials in Experiment 2. Data for the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) to 3 SD Stimuli (Top) and 6 SD Stimuli (Bottom) Are Shown.....	84

Table 14.	Mean d 's to 0° and 18° Stimuli (3 <u>SD</u> and 6 <u>SD</u>) for the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) During Generalization Testing for Experiment 2. The Standard Error Is Shown in Parentheses.....	86
Table 15.	Participants' d 's in Groups SET, SEOT and MET for Experiment 2.....	89
Table 16.	Median Response Bias (β) to 0° and 18° Stimuli (3 <u>SD</u> and 6 <u>SD</u>) For the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) During Discrimination Training and Generalization Testing for Experiment 2.....	92
Table 17.	Participants' Betas in SET, SEOT, and MET Groups for Experiment 2.....	95
Table 18.	Slopes of Generalization Gradients for Participants (Part.) with Generalization Slopes Values > 0 to 3 <u>SD</u> and 6 <u>SD</u> Stimuli for the Three Training Groups (SET, SEOT, MET) for Experiment 2.....	103
Table 19.	Proportion of Variance accounted for (r^2) of Slopes of Generalization Gradients for Participants (Part.) with Generalization Slope Values > 0 to 3 <u>SD</u> and 6 <u>SD</u> Stimuli for Three Training Groups (SET, SEOT, MET) for Experiment 2.....	123
Table 20.	Analysis of Variance (ANOVA) of Generalization Slopes for Participants with Slope Values > 0 to 3 <u>SD</u> and 6 <u>SD</u> Stimuli. The Effects of Training Group (T) and Variability of Test Stimuli (V) Are Shown for Experiment 2.....	124
Table 21.	Post-Hoc, Pairwise-Comparisons of Generalization Slopes for Participants with Slope Values > 0 in the Single-Exemplar Training Group (SET), the Single Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) and Variability of Test Stimuli (V) (3 <u>SD</u> and 6 <u>SD</u>) for Experiment 2.....	125

List of Figures

- Figure 1.** Six exemplars of graphs used as stimuli. The slope of the data path in the left and right panels of each graph is 0° . The three graphs to the left were constructed with 3 standard deviations (3 SD) while the three graphs to the right were constructed with 6 standard deviations (6 SD).....22
- Figure 2.** Six exemplars of graphs used as stimuli. In all graphs, the slope of the data path in the left panel is 0° while it is 10° in the right panel. The three graphs to the left were constructed with 3 standard deviations (3 SD) while the three graphs to the right were constructed with 6 standard deviations (6 SD).....24
- Figure 3.** Six exemplars of graphs used as stimuli. In all graphs the slope of the data path in the left panel is 0° while it is 18° in the right panel. The three graphs to the left were constructed with 3 standard deviations (3 SD) while the three graphs to the right were constructed with 6 standard deviations (6 SD).....26
- Figure 4.** Stages 1, 2, and 3 of preliminary training. In all graphs the slope of the data path in the left panel is 0° while it is 18° in the right panel. In Stage 1, no standard deviation is possible since no data points are shown. In Stage 2, 0 standard deviation is present as all data points fall on a straight line and the data points are shown. In Stage 3, 1 standard deviation is present and the data points are shown.....32
- Figure 5.** The probability of the DIFFERENT response to 3 SD and 6 SD stimuli during generalization tests as a function of stimulus slope during Experiment 1.....43
- Figure 6.** The mean discrimination index for the single-exemplar training group (SET) the single-exemplar overtraining group (SEOT) and the multiple-exemplar training group (MET) to 3 SD and 6 SD stimuli for Experiment 2. The vertical lines represent ± 1 SE.....54

- Figure 7. The discrimination index (DI) for individual participants after single-exemplar training (SET), single-exemplar overtraining (SEOT), and multiple-exemplar training (MET) in Experiment 2. The DIs to 3 SD stimuli are presented in the top panel while DIs to 6 SD stimuli are presented in the bottom panel. Filled symbols indicate $\geq 90\%$ criterion level.....61
- Figure 8. The percent of S+ trials that occasioned the DIFFERENT response for the single-exemplar training group (SET), the single-exemplar overtraining group (SEOT), and the multiple-exemplar training group (MET) to 0° and 18° stimuli, (3 SD and 6 SD), for Experiment 2. The vertical lines represent ± 1 SE.....65
- Figure 9. The percent of S+ trials that occasion the DIFFERENT response for individual participants after single-exemplar training (SET), single-exemplar overtraining (SEOT), and multiple-exemplar training (MET) for Experiment 2. Performance to 3 SD stimuli is presented in the top panel while performance to 6 SD stimuli is presented in the bottom panel. Filled symbols indicate $\geq 90\%$ criterion level.....71
- Figure 10. Percent of S- trials that occasioned the DIFFERENT response for the single-exemplar training group (SET), the single-exemplar overtraining group (SEOT), and the multiple-exemplar training group (MET) to 0° and 18° stimuli, (3 SD and 6 SD), for Experiment 2. The vertical lines represent ± 1 SE.....75
- Figure 11. Percent of S- trials that occasion the DIFFERENT response for individual participants after single-exemplar training (SET), single-exemplar overtraining (SEOT), and multiple-exemplar training (MET) for Experiment 2. Performance to 3 SD stimuli is presented in the top panel while performance to 6 SD stimuli is presented in the bottom panel. Filled symbols indicate $\leq 10\%$ criterion level.....81
- Figure 12. Mean d's after single-exemplar training (SET), single-exemplar overtraining (SEOT), and multiple-exemplar training (MET) to 3 SD and 6 SD stimuli for Experiment 2. The vertical lines represent ± 1 SE.....87

- Figure 13.** Median bias after single-exemplar training (SET), single-exemplar overtraining (SEOT) and multiple-exemplar training (MET) to 3 SD and 6 SD stimuli for Experiment 2.....93
- Figure 14.** Hypothetical generalization gradients that show identical performance (probability of the DIFFERENT response) to 0° and 18° stimulus slopes but differ in response to intermediate stimulus slopes (2,° . . . 16°). Generalization gradient slopes in graphs A to D are .332, .222, .132, and .071 respectively. R² values are .94, .99, .43, -.01 respectively.....100
- Figure 15.** Generalization gradients for individual participants in the single-exemplar training group for Experiment 2. The probability of the DIFFERENT response during generalization tests to 3 SD stimulus slopes is shown. The participant number and the generalization gradient slope are shown above each graph. Participants not in the subset (those who demonstrated generalization gradient slopes ≤ 0 to 3 SD or 6 SD stimulus slopes are indicated by an asterisk.....104
- Figure 16.** Generalization gradients for individual participants in the single-exemplar overtraining group for Experiment 2. The probability of a the DIFFERENT response during generalization tests to 3 SD stimulus slopes is shown. The participant number and the generalization gradient slope are shown above each graph. Participants not in the subset (those who demonstrated generalization gradient slopes < 0 to 3 SD or 6 SD stimulus slopes) are indicated by an asterisk.....107
- Figure 17.** Generalization gradients for individual participants in the multiple-exemplar training group for Experiment 2. The probability of the DIFFERENT response during generalization tests to 3 SD stimulus slopes is shown. The participant number and generalization gradient slope are shown above each graph.....110

- Figure 18.** Generalization gradients for individual participants in the single-exemplar training group for Experiment 2. The probability of the DIFFERENT response during generalization tests to 3 SD stimulus slopes (dots) and 6 SD stimulus slopes (circles) is shown. The participant number and generalization gradient slopes to 3 SD and 6 SD stimulus slopes are shown above each graph respectively. Participants not in the subset (those who demonstrated generalization gradient slopes ≤ 0 to 3 SD or 6 SD stimulus slopes) are indicated by an asterisk.....113
- Figure 19.** Generalization gradients for individual participants in the single-exemplar overtraining group for Experiment 2. The probability of the DIFFERENT response during generalization tests to 3 SD stimulus slopes (dots) and 6 SD stimulus slopes (circles) is shown. The participant number and generalization gradient slopes to 3 SD and 6 SD stimulus slopes are shown above each graph respectively. Participants not in the subset (those who demonstrated generalization gradient slopes ≤ 0 to 3 SD or 6 SD stimulus slopes) are indicated by an asterisk.....116
- Figure 20.** Generalization gradients for individual participants in the multiple-exemplar training group for Experiment 2. The probability of the DIFFERENT response during generalization tests to 3 SD stimulus slopes (dots) and 6 SD stimulus slopes (circles) is shown. The participant number and generalization gradient slopes to 3 SD and 6 SD stimulus slopes are shown above each graph respectively.....119
- Figure 21.** Mean generalization gradient slopes for the single-exemplar training group (SET), the single-exemplar overtraining group (SEOT), and the multiple-exemplar training group (MET) to 3 SD and 6 SD stimulus slopes for Experiment 2. Only participants in the subset are included (i.e., participants with generalization gradient slopes > 0 to 3 SD and 6 SD stimulus slopes). The vertical line represents ± 1 SE.....126

The Effects of Multiple-Exemplar Training and Stimulus Variability on Generalization to Line Graphs

General Introduction and Overview

Behavior analysts often assess the effectiveness of an intervention by visual inspection of data plotted in line graphs. Jones, Weinrott, and Vaught (1978) have questioned the accuracy of interpretation of graphic data by visual inspection. These researchers point out that different analysts often come to different conclusions after viewing the same graphs. This might be rectified by establishing a graph reading repertoire by intradimensional discrimination training with many S+ and S- exemplars.

Participants were trained to detect changes in one graphic characteristic--the slope of the data path. Some participants were trained with a limited stimulus set, as is typically used in intradimensional discrimination training. Other participants were trained with an extensive stimulus set, as is typical of concept formation studies.

The goals of the present research were to discover how discrimination training regimens influenced the accuracy of graph reading. Specifically three different methods were used to train college students to discriminate between graphs based on changes in slope from baseline to treatment phases. To put the current study in context, three pertinent areas will be reviewed: (a) interpretation of graphs by visual inspection, (b) discrimination training and stimulus generalization, and (c) concept formation.

Interpretation of Graphs by Visual Inspection

Visual inspection of line graphs has been the method most commonly used to evaluate data in the study of

behavior analysis (Kratochwill & Brody, 1978). In single-subject designs, functional control by a variable is assessed by examining changes in patterns of data (Kazdin, 1982). For example, visual inference is made from changes in data from baseline to treatment phases when such data are presented graphically (Barlow & Hersen, 1984). When a data set is complex, visual inspection can disclose its intricate structure (Cleveland, 1993).

Line graphs from a single-participant design study can be defined by three basic characteristics. First is "level", which is the value on the vertical (Y) axis around which a set of data points converges. In visual interpretation of graphs, level is measured by an absolute value (mean, median) on the Y-axis. Second is "slope" which is the overall direction of the data path (increasing, decreasing, or flat) also called "trend". Specifically, slope is the angle between the data path and the X-axis measured in degrees. Third is "variability" which refers to the extent that measures of behavior under the same conditions differ from one another. In graph reading, variability refers to the dispersion of scores from the regression line (DeProspero & Cohen, 1979; Cooper, Heron, & Heward, 1987). Other features that characterize graphs and can influence interpretation include number of data points in each phase, variability and overlap of baseline and treatment data, serial dependency of data points, and position of the function in the frame (Fisch & Schneider, 1993; Jones, Vaught, & Weinrott, 1977; Matyas & Greenwood, 1990; Sidman, 1960).

In single-subject design the experimental phases are

designated alphabetically such that the "A" represents the baseline phase and the "B", "C", "D", and "E" represent specific treatment phases. For example, A-B-A-C-A, represents baseline-treatment 1-baseline-treatment 2-baseline.

Jones et al. (1978) asked professors and graduate students to evaluate meaningful changes in adjacent phases of graphs taken from the Journal of Applied Behavior Analysis. The graphs had different configurations (A-B, A-B-A-B, A-B-C-B, A-B-A-C-A-D-E-A) and each test item was a pair of adjacent phases. The test items were presented, in random order, to participants who judged them for meaningful changes in level (e.g., yes, no, or unsure). The participants did not show a consensus as the intercorrelation of responses, between the judges, was low (median $r = .39$). In addition, agreement and disagreement between participants' visual inferences and time-series analysis were calculated. For example, agreement was scored if visual inference was scored as "yes" and time-series analysis was significant ($p < .05$). An agreement proportion (agreement proportion = agreements / (agreements and disagreements)) was obtained for each judge. The agreement proportion between visual inference and quantitative, time-series analysis was low (i.e., 0.60), only somewhat better than chance levels (0.50). The researchers found that high levels of serial dependency between adjacent data points were correlated with reduced agreement between reviewers.

In a related study, DeProspero and Cohen (1979) sought to identify the effects of certain graphic characteristics

on interpretation of data through visual inspection. A-B-A-B reversal graphs were constructed in which the pattern of level between phases (e.g., no level changes in A's, and level changes in B's), magnitude of level shifts, variability (within-phase dispersion of scores about the mean), and slope of data path across phases were manipulated. Reviewers of behavioral journals rated each graph on a 100 points scale, with respect to the degree of "experimental control" demonstrated. Interpretations of these graphs were most influenced by the pattern of level shifts. That is, high ratings were given for "ideal" graphs that showed ideal patterns, such as stable baselines (A's), with level increases in the treatment phases (B's). The magnitude of level shifts also affected interpreters' ratings of graphs--the larger the level shift the higher the rating of "experimental control." Even ideal graphs with large level shifts were rated lower when a data set showed (a) high variability within a phase, and (b) changes in slope. This study found low interrater agreement ($r = 0.61$), and replicated the findings reported by Jones et al. (1978). Both studies show that inferences drawn from visual inspection by a panel of reviewers were not particularly reliable. The authors speculated that each reviewer may have weighed the above-mentioned parameters of graphs differently from other reviewers when interpreting the same graphs.

Furlong and Wampold (1982) had editors of the Journal of Applied Behavior Analysis sort A-B panel graphs that contained changes in slope, level, or both slope or level. These graphs had no slope in the baseline phase (panel A)

and all changes occurred in the treatment phase (panel B). The experimenters manipulated the graphs in three ways. In the standard transformation the effect size (i.e., the values of the slope and level) and variability (normally selected deviates of each data pattern) were held constant. In the scaling transformation, the effect size and random variability were both multiplied by a constant, while the Y-axis remained unchanged. This produced a larger absolute effect size compared to the standard transformation; however, when compared to variability, the relative size of the intervention effect was the same. In the variation transformation the effect size was left unaltered but the random variability was multiplied by a constant.

Sets of these three type graphs were sent to journal reviewers who were asked to sort them into classes (no number given) that demonstrated "similar experimental effect." The majority of reviewers sorted the graphs in a manner that was correlated with changes in level, changes in slope, or both. The classifications of a minority of reviewers were correlated with variability of the data. These reviewers therefore attended more to the absolute size of the treatment effect and did not systematically compare the size of the treatment effect to the variability of the data. Therefore graphs that were mathematically equivalent (i.e., standard and scaling transformation graphs) were judged as different.

Matyas and Greenwood (1990) examined the effects of magnitude of level shifts, variability (random error added or subtracted to each data point) and serial dependency (autocorrelation among temporally adjacent data points) in

interpretation of A-B graphs. Graduate students usually evaluated the graphs with respect to the presence of change in level, slope, or changes in both level and slope. The results showed high false-alarm rates (i.e., reporting an effect when it was absent) were correlated with increases in variability and serial dependency. Miss rates (failing to report an effect when it is present) were relatively lower than false-alarm rates.

Fisch and Schneider (1993) performed a systematic replication of Matyas and Greenwood (1990) and also tested the effects of two additional parameters on graph reading: The labeling of the Y-axis as either "Proportion in %" or "Rate" (responses/minute) as opposed to "Score" used by Matyas and Greenwood (1990). The placement of the function near the top, middle, or bottom of the frame that contained the graph was also manipulated. Graduate students in Psychology were asked to indicate the presence of level changes, slope changes, or both, in A-B line graphs. Only 35% of the graphs with level changes and only 19% of graphs with slope changes were correctly identified. Only 10% of graphs that had both level and slope changes were correctly identified. In contrast, 62% of the graphs that had neither level or slope effects were correctly identified. Therefore both low hit rates and high false-alarms rates were found. Placement of the graph in the frame influenced accuracy of description, although the axes remained unchanged. Specifically, when changes in level only were present, accuracy (percent correctly identified) was highest when the function was located in the bottom of the frame (54%) and intermediate when it was located in the top of the

frame (42%). Accuracy was lowest when the function was located in the middle of the frame (8%). Finally, accuracy and graphic interpretation were not influenced by the Y-axis label.

The results of Fisch and Schneider (1993) confirmed the findings of Greenspan and Fisch (1992) and Matyas and Greenwood (1990). These studies showed that slope and, to a lesser extent, level changes were difficult for reviewers to detect. In addition, misses and false alarms increased when variability in the data was high and serial dependency was present.

Summary and interpretation. Line graphs are complex stimuli that vary on many parameters that include: number of data points, number of phases, magnitude of level shifts, magnitude of slope changes, position of the graph in the frame, level of variability, and serial dependency. Interpretation of graphs based on visual inspection may be ambiguous. The outcomes of many studies show low-interrater agreement and accuracy in the identification of changes in slope and level.

Given the low hit rates and high false-alarm rates found in their study, Matyas and Greenwood (1990) urged that graduate students receive training to improve their skills in the interpretation, by visual inspection, of data commonly presented in single-participant, time series designs. Furlong and Wampold (1981) also recommended such training to avoid misinterpretation of data assessed by visual inspection as well as the use of a "Visual Inference Checklist" to systematize the assessment process. Students may become more accurate in the interpretation of graphic

data if explicit training in visual inspection is given. Participants in these studies were not provided with any explicit discrimination training in graph reading prior to being asked to evaluate test graphs. The participants, as reviewers of behavioral journals and graduate students, undoubtedly had received training in single-participant design and graph reading prior to the experiment. The nature and extent of this training, however, were not known and had to vary across participants.

I surmise that the poor performance of interpreters of data based upon visual inspection can be attributed to inconsistent training with stimulus materials that do not foster the establishment of stimulus control by the appropriate parameters. In brief, studies on visual inspection of graphs include unspecified discrimination training in graphic interpretation, poor maintenance of discrimination skills (assuming prior training) and, by implication, poor generalization to the accurate interpretation of novel graphs. In the current study, participants were given explicit discrimination training in graph reading. Generalization testing with novel graphs followed training.

In training participants in graph reading, selecting an effective method is crucial. A brief review of discrimination training methods will identify some key training parameters and underscore their importance in the current research.

Discrimination Training and Stimulus Generalization

In a typical discrimination experiment, there are many presentations of two stimuli (S+ and S-). Responding in the

presence of S+ results in a different outcome than responding in the presence of S-. This differential reinforcement of S+ and S- responses results in a well-trained, basal discrimination. Next, the S+ and S- discrimination is maintained under reduced probabilities of reinforcement. Finally, the training stimuli and other stimuli from the same dimension (e.g., wavelength, frequency, angle) are presented in extinction conditions. Response rates to both trained and untrained stimuli are examined to determine if responding varies systematically as a function of the particular dimension.

In their landmark study, Guttman and Kalish (1956) provided aperiodic reinforcement for pecking a disk illuminated by a particular wavelength (S+) that alternated with a key and chamber blackout periods. During generalization tests, the S+ and adjacent wavelengths were presented serially and in random order under extinction conditions. Maximal responding was maintained by the S+ during tests that included many new stimuli. High response rates were demonstrated to a narrow range of stimuli adjacent to the S+ and rates declined progressively as wavelength differed increasingly from the S+. The steep slopes of the primary generalization gradients showed the establishment of dimensional control using a presence-only training procedure.

Honig, Boneau, Bernstein, and Pennypacker (1963) demonstrated dimensional control by line-tilt using a presence/absence procedure. Two groups of birds were trained to discriminate between the presence or absence of a vertical black line on a illuminated response key. For

one group (stimulus-positive) the presence of this line was correlated with reinforcement. For another group (stimulus-negative) the absence of the vertical line was correlated with reinforcement. Training for both groups alternated with blackout periods. Generalization tests were conducted by presenting the vertical line and adjacent line-tilts under extinction conditions. The results showed maximal responding to the vertical line for the stimulus-positive group that declined progressively as line-tilts differed from the vertical. For the stimulus-negative group, maximal responding was seen to line-tilts that differed most from the vertical. The minimal responding was seen to the vertical line for this group.

Jenkins and Harrison (1960) compared presence/absence training (S+/S-) to presence only training (S+) with the dimension of tonal frequency. This work was replicated by Newman and Baron (1965) with the dimension of line-tilt. In the Newman and Baron study, for two groups of pigeons, a vertical line on a green key occasioned reinforcement for pecking. For the presence/absence training group this stimulus was alternated with a green key (S-) that did not occasion reinforcement for pecking. For the presence only group no S- was presented. For both groups, no blackout periods were scheduled during training. During generalization tests with S+ and adjacent angles, the presence/absence group showed maximal responding to the (S+), and orderly generalization gradients. By contrast, the presence only group showed no differential responding by line-tilt. Similar results were reported by Jenkins and Harrison (1960) with the dimension of tonal frequency.

Hanson (1959) compared two procedures in training a wavelength discrimination with pigeons. In the presence/absence training, birds were reinforced for pecking a key illuminated with a particular wavelength (S+) alternated with key and chamber blackout periods. In the dimensional training procedure, birds were reinforced in the presence of one wavelength but not reinforced when a different wavelength (S-) illuminated the key. Maintenance of control by the S+ and dimensional control to adjacent wavelengths were demonstrated for both groups. The birds trained with the dimensional procedure, however, showed higher and steeper generalization gradients than birds trained with the presence/absence procedure.

The type of training prior to generalization testing affected the likelihood of establishing control by a particular dimension. Single-stimulus training (S+ presentations only with no blackout periods or intertrial intervals) was least likely to establish control by a given dimension (Jenkins and Harrison, 1960). The experimental stimulus and incidental stimuli (constant features of the experimental space) were all present when the response was reinforced. Although an experimental stimulus occasioned a response, it did not come to control it. A procedure more likely to induce stimulus control involved random presentation of a single stimulus (S+) with non-reinforced periods (e.g., blackout periods). Similarly, the presence of the S+ (vertical black line of a illuminated key) could be alternated with its absence (illuminated key only as S-) (Dinsmoor, 1995). Beyond this presence/absence training, dimensional training with two stimuli (S+ and S-) located

at different points on a dimension was most likely to establish control by that dimension (Jenkins and Harrison, 1960; Newman and Baron, 1965).

Amount of training, in addition to type of training, also can influence generalization to new stimuli. Using an operant procedure, Sherman and Pierrel (1961) trained groups of rats on a auditory discrimination using a presence/absence procedure. The presence of an auditory tone (S+) occasioned reinforcement for bar pressing while its absence (S-) signaled periods of non-reinforcement. Groups of rats were trained for 6, 13, or 27 days with this procedure. During generalization tests, the slopes of the generalization gradients became steeper after increased amounts of training.

Using a free operant procedure, Hearst and Koresko (1968) trained groups of pigeons on a line-tilt discrimination using a presence/absence procedure. The presence of a vertical line on the key signaled the availability of reinforcement for key pecking, while its absence signaled non-reinforcement. Groups of pigeons were trained for 2, 4, 7, or 14 days. During generalization tests to the vertical and adjacent stimuli, more training resulted in higher levels of stimulus control that were reflected by steeper slopes.

In these studies mentioned, when novel stimuli were introduced during generalization testing, the trained discriminations were not disrupted. Such outcomes, however, are not inevitable. Heinemann, Avin, Sullivan, and Chase (1969) trained pigeons to discriminate between two intensities of white noise in a 3-key, matching-to-sample

procedure. A trial began with the illumination of the center key and simultaneous presentation of one of two intensities of white noise. A peck on the center key caused it to go off and illuminated both side keys. A peck to the right key was reinforced in the presence of the lower sound intensity while the peck on the left key was reinforced in the presence of a higher sound intensity. During generalization tests with the training stimuli and adjacent sound intensities stimulus control by the training stimuli declined for most subjects.

Comparison of stimulus generalization and graph reading studies. Stimulus generalization studies include extensive training with single S+'s and S-'s. These stimuli are simple in the sense that they lie on different points on the same dimension and all other stimulus dimensions were held constant. Generalization tests included both training stimuli and adjacent stimuli on the same dimension. The outcomes usually showed maintenance of the baseline discrimination, systematic control to intermediate stimuli and replicability.

In contrast to stimulus generalization studies with infrahumans cited, graph reading studies with humans differ in the number of stimulus dimensions, training and outcome.

The test stimuli in graph reading studies are complex; they vary on many dimensions simultaneously such as slope, level, and variability. In addition, discrimination training prior to tests was not conducted, but at best it was assumed to have occurred. The outcomes included poor discrimination of components of graphs. Finally, low-interrater agreement was found. It should be noted,

however, that in the cited animal studies, variability between subjects during generalization testings was also found.

In the present study the procedures used in stimulus generalization studies were applied to those of graph reading. Reviewers were trained to discriminate between S+ and S- graphs that differ in value on a particular dimension--the slope of the data path.

Concept Formation

Stimulus generalization studies typically include many training trials with a small number of discriminanda and many test stimuli to assess generalization. During generalization tests, a narrow range of stimuli closest in dimensional value to the S+ evoke responses comparable to that produced by the S+ (Newman and Baron, 1965). In graph reading studies, however, generalization to a wide range of stimuli (e.g., novel graphs with similar characteristics) is desirable. One way of producing generalization to a wide range of new stimuli involves training with a broad range of discriminative stimuli, as is done in traditional studies in concept formation.

In a prototypical concept-formation experiment, Herrnstein (1979) presented pigeons with slides of trees and slides without trees. Forty of these slides were positive (S+) exemplars that included full or partially obstructed views of trees. Forty of these slides were negative exemplars (S-) that did not contain trees. Pecking a response key was reinforced when positive exemplars were present but was not reinforced in the presence of negative exemplars. Birds demonstrated a discrimination between S+

and S- exemplars by the fifth day of training.

This performance may have reflected the correlation of reinforcement with these particular exemplars used in training, control by classes of stimuli which shared S+ or S- properties, or both. To assess these possibilities, new exemplars of each potential S+ and S- category were presented. All birds responded to the new exemplars at levels comparable to those occasioned by the exemplars presented during training. Control by stimulus class or concept formation (Keller and Schoenfeld, 1950) was demonstrated by generalization to novel stimuli within the potential stimulus class of trees, and the discrimination between novel exemplars of trees and non-trees.

In studies in concept formation, the number of exemplars necessary for participants to form a stimulus class was explored. Herrnstein and de Villiers (1980) trained pigeons to discriminate slides of fish and non-fish with a set of 80 slides (40 of each type). During testing, the birds discriminated novel slides of fish and non-fish ($p = .74$) well above chance levels ($p = .50$) but below training levels ($p = .81$).

Malott and Siddall (1972) demonstrated that pigeons needed from 6-36 positive and negative exemplars, in total, to discriminate and generalize to photographs of people, faces, and human torsos.

Using a simultaneous matching-to-sample format for training and testing, Wright, Cook, Rivera, Sands, and Delius (1988) trained two groups of pigeons to discriminate between colored pictures of objects, presented in pairs, (e.g., airplane, ball, duck, hammer, etc.) into the

response classes of same/different. Two experimental conditions were explored. In the two-stimulus group, training was conducted with only 2 stimuli. In the trial-unique group, training was conducted with 152 stimuli. During generalization tests with 20 novel picture stimuli, the trial-unique group showed performance comparable to that seen to the training stimuli (i.e., \geq 80% correct to both training and testing stimuli). The two-stimulus group, by contrast, performed the matching-to-sample task with the novel picture stimuli at virtually chance levels (i.e., 51% correct). Repeated attempts to induce concept formation by additional training were also unsuccessful.

Previous experiments showed how the number of training exemplars influence categorization. In a series of studies, Bhatt, Wasserman, Reynolds, and Knauss (1988) examined the effect of repetition of exemplars in concept formation. Two groups of pigeons were trained on a four-category (i.e., cats, flowers, persons, man-made objects) classification. The repeat group was trained with a limited, repeated-stimulus set (10 exemplars per category). The nonrepeat group received training with a large, non-repeating stimulus set (500 exemplars per category). During testing, both groups classified novel exemplars well above chance (25%) levels, (e.g., repeat group 70%; non-repeat group 64%). This study demonstrated repetition of exemplars was not necessary to demonstrate concept formation, as concept formation was established with nonrepeating training sets.

Homa, Cross, Cornell, Goldman, and Schwartz (1973) showed that the number of exemplars presented during

training influenced classification of new exemplars. In their study, college students were trained on three complex dot patterns represented by 3, 6 or 9 exemplars of each. During generalization testing, transfer to novel exemplars (measured as percent correct) increased as a direct function of the number of exemplars presented during training. Like results were seen in other studies (Homa, 1978; Homa & Vosburgh, 1976).

In a similar study, Omohundro (1981) trained college students to classify geometric shapes represented by 4, 8, or 12 exemplars in each category. During generalization testing, classification of novel exemplars was enhanced by increasing the number of exemplars representing each category during training.

Another parameter that also affect concept formation is the variability of the test stimuli. Posner and Keele (1968) produced geometric letter and nonsense shapes by connecting 9 dots in a 30 x 30 matrix with straight lines. They then increased the variability of each form by displacing each dot from its original position by low, medium or high increments. Three groups of college students were trained with sets of 12 stimuli, 3 exemplars each of the 4 shapes. One group was trained with the low-variability set, the second with the medium-variability set and the third group received no training and thereby served as the control group. Next, all groups were tested with a set of novel, high-variability forms, representing each of the four shapes. The medium-variability group made the fewest errors. The control group made the most errors, while the error rate of the low-variability group was

intermediate to the other two groups. The results suggested that as the variability of the training stimuli increased so did the ability of participants to classify novel stimuli that were even more variable.

Homa, Sterling, and Trepel (1981) extended the findings of Homa et al. (1973). In their study, the training stimuli (shapes made from dot patterns) were made more variable by displacing each dot from its original position by one to five increments. During testing with novel exemplars, college students showed the best categorization of shapes whose variability was closest to that presented during training.

Summary and interpretation. The findings of studies in multiple-exemplar training and concept formation have implications for procedures intended to improve accuracy in the interpretation of novel graphs by visual inspection. Graphs are complex stimuli varying on many dimensions simultaneously. Graphs are therefore more representative of the type of stimuli presented in concept formation studies than the simpler stimuli presented in many stimulus generalization studies. Procedures used in concept formation include training with multiple exemplars of the S+ and S-. Outcomes include maintenance of the trained discriminations, discrimination between sets of new stimuli that are different, and generalization to new stimuli that are similar. These outcomes are highly desirable for those who interpret data by visual inspection.

The goal of the present research was to discover how different discrimination training regimens influenced participants' accuracy in graph reading. Students were

trained to discriminate between graphs that differed in value on one dimension--the slope of the data path. Maintenance of the discrimination and generalization to graphs with intermediate slopes were also explored.

Experiment 1

The first purpose of Experiment 1 was to assess if single-exemplar training would be sufficient to establish a discrimination between graphs that differ in the slope of their data paths. The second purpose was to determine if this discrimination could be maintained with test graphs of the same and increased variability. The third purpose was to determine if dimensional control by intermediate slopes could be established.

Method

Participants

Twelve undergraduate students, (7 women and 5 men, 18-25 years of age) were recruited from Introductory Psychology classes at Queens College/CUNY and served as participants. No participant had any familiarity with the research. Each participant received partial course credit for completing the experiment. Credit did not depend on performance. Completion of the experiment required one session of approximately 2 hours.

Apparatus

The experiment was controlled by IBM compatible, micro-computer. Responses were made by touching specific keys on the computer keyboard and were recorded by the computer. Participants were seated in a room at a table facing the screen of the microcomputer. Experimental stimuli were presented on a 13" monochrome VGA monitor.

Stimuli

Each stimulus was composed of a graph presented on the computer monitor. The X-axis was 8.75 inch (22.23 cm) and the Y-axis was 4.5 inch (11.43 cm). Each stimulus had a "A" and "B" panel. The X-axis was divided into two equal parts by a line perpendicular to the X-axis. The region to the left of the divider was the A panel while the region to the right was the B panel. Each panel contained ten data points. Adjacent data points in each panel were connected by straight lines, but a line did not connect data points across the A and B panels.

The stimuli were produced by a computer program that generated regression lines for the slopes in the A and B panels. A normal distribution was computed around each data point. The program then generated a unit normal distribution around each data point. It then randomly selected some value from the unit normal distribution which was to be used as the actual data point. If the standard deviation was within 2% of the specified standard, the specific set of data points were made into a graph and saved for future presentation in the experiment. If the data points differed by more than 2% of the specified standard deviation, a new set of data points was generated until the 2% criterion was met again. While a mathematical regression line was used to produce the data points, the regression line was never presented in a stimulus panel. The standard deviation (SD) of the data points was always the same in the A and B panel of a given stimulus. All stimuli were constructed with the same level in both the A and B panels as defined by Jones et al. (1977). These

researchers define change in level as discontinuity in the data path at the interruption point (i.e., the end of the data path in panel A and the beginning of the data path in panel B).

Procedure

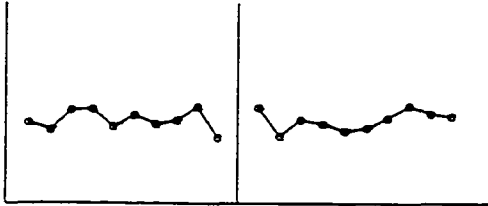
Stimulus manipulation. The stimuli were manipulated in two ways: (a) the slope of the data path, and (b) the variability of the data points around the regression line. The slope of the data path in the B panel was varied from 0° to 18° in increments of 2° , while the slope in the left panel was always 0° . Figure 1 presents six exemplars of the 0° stimulus, Figure 2 presents six exemplars of the 10° stimulus, while Figure 3 presents six exemplars of the 18° stimulus.

The variability of the data points around the regression line was manipulated. This is illustrated by comparing stimuli on the left of Figures 1, 2, and 3 with those on the right. In each figure the stimuli on the left were constructed with 3 standard deviations (3 SD) while those on the right were constructed with 6 standard deviations (6 SD).

For a given stimulus with the same slope and variability, comparable exemplars were produced by generating different sets of data points around the regression line. A comparable set of stimuli was the same with regard to slope and standard deviation but not identical across graphs because of the unique placement of data points. For example, the three stimuli on the left in Figure 1 are comparable exemplars. The three stimuli on the right in Figure 1 are also comparable exemplars.

Figure 1. Six exemplars of graphs used as stimuli. The slope of the data path in the left and right panels of each graph is 0° . The three graphs to the left were constructed with 3 standard deviations (3 SD) while the three graphs to the right were constructed with 6 standard deviations (6 SD).

0°, 0°,
3-Standard Deviations



0°, 0°,
6-Standard Deviations

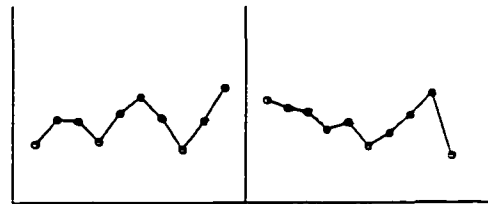
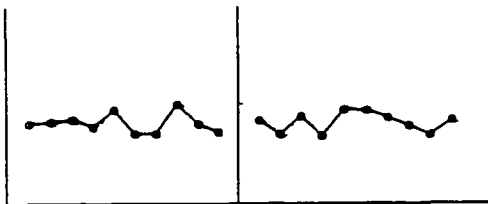
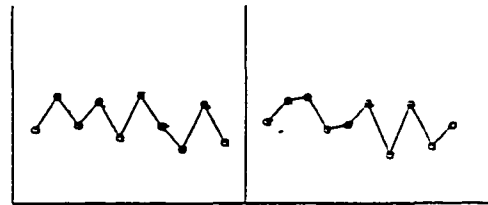
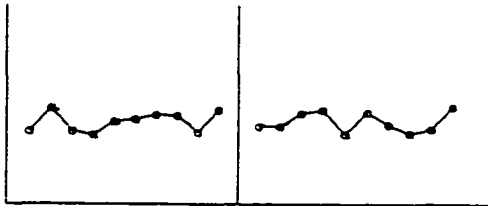
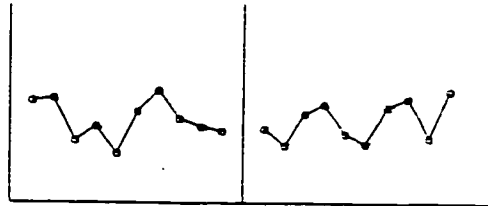
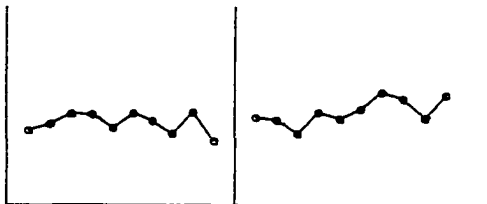


Figure 2. Six exemplars of graphs used as stimuli. In all graphs, the slope of the data path in the left panel is 0° while it is 10° in the right panel. The three graphs to the left were constructed with 3 standard deviations (3 SD) while the three graphs to the right were constructed with 6 standard deviations (6 SD).

0°, 10°,
3-Standard Deviations



0°, 10°,
6 Standard Deviations

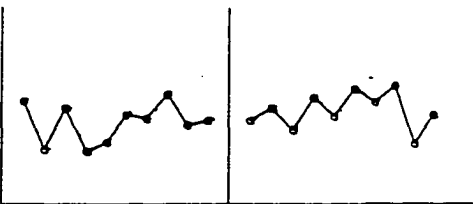
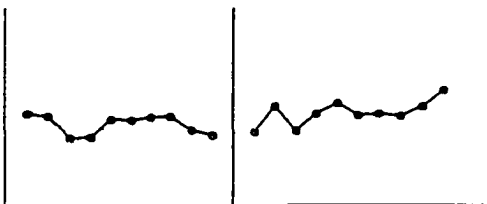
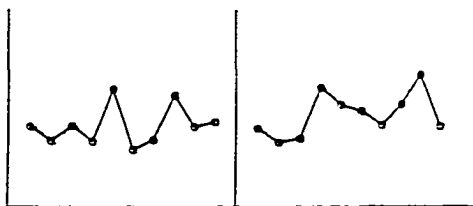
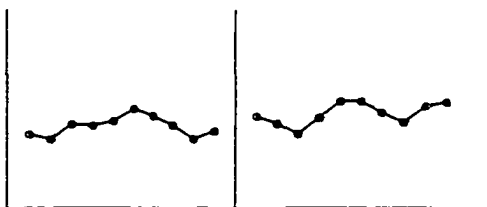
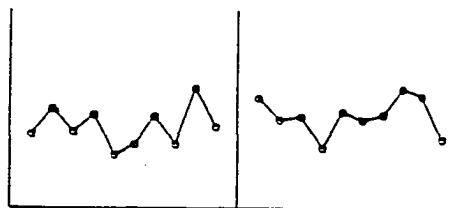
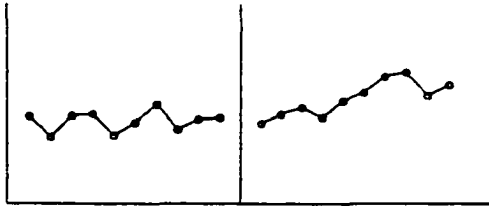


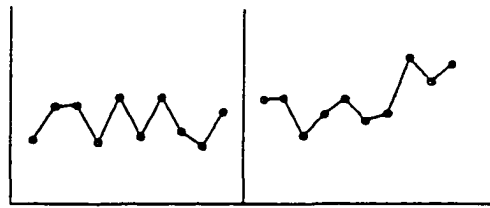
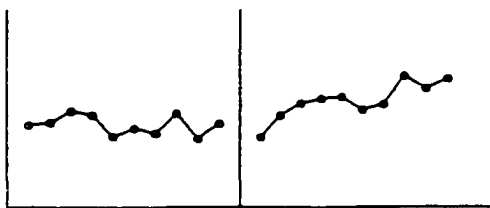
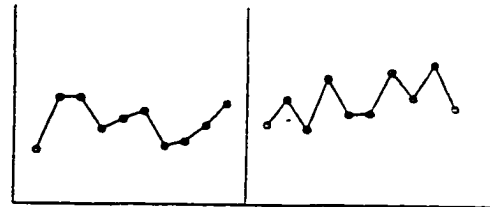
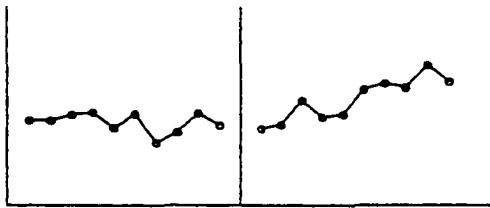
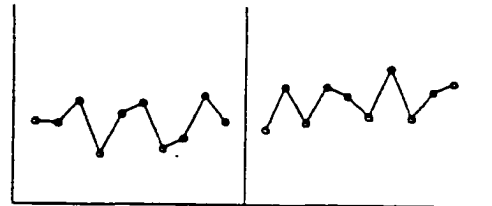
Figure 3. Six exemplars of graphs used as stimuli.

In all graphs the slope of the data path in the left panel is 0° while it is 18° in the right panel. The three graphs to the left were constructed with 3 standard deviations (3 SD) while the three graphs to the right were constructed with 6 standard deviations (6 SD).

0°, 18°, 3-Standard Deviations



0°, 18°, 6 Standard Deviations



Instructions. The participants were read the following instructions:

Soon on the screen you will see two forms, one on the right and one on the left. If you think the form on the right is not different from the form on the left press the number "1" key. If you think the form on the right is different from the one on the left press the number "2" key. After you press either the number "1" or number "2" key press the "ENTER" key. Any Questions?

The computer will tell you how you did. It may say "RIGHT" if you answered right. When that happens you will press the "R" key. It may say "WRONG" if you answered wrong. When that happens you will press the "W" key. Sometimes it won't tell you if you answered right or wrong it will just say "END". When that happens you will press the "E" key. Any questions?

Try to get as many "RIGHTS" as you can. Work at your own pace. Don't rush ahead and get a lot wrong but don't just sit here and do nothing either. Any questions?

I will now leave the room and let you work alone. If you have any questions, knock on the door and I will come in. Please begin.

Trial structure, contingencies and responses within a trial. Each trial began with a presentation of a graph on the screen. After 1 s, the number "1" adjacent to the word "NOT DIFFERENT" and the number "2" adjacent to the word "DIFFERENT" appeared below the graph. Participants pressed

the number 1 key to indicate that the B panel was not different from the A, while they pressed the number 2 key to indicate that they were different. If the slopes in the A and B panels were equal, pressing the number 1 key (not different) led to the appearance of the word "RIGHT" on the screen. If the slopes were not equal, pressing the number 2 key (different) also resulted in the appearance of the word "RIGHT" on the screen. If the slopes were equal and the participant responded on the number 2 key (different) or the slopes were not equal and the participant responded on the number 1 key (not different) the word "WRONG" appeared on the screen. The presentation of the words "RIGHT" and "WRONG" was defined as informative feedback since it reflected the accuracy of a participant's response. On some of the trials, however, responses did not produce informative feedback. At these times, the word "END" appeared after either a 1 or 2 response. This was defined as noninformative feedback since it did not reflect the accuracy of a participant's response, but merely signaled the end of a trial.

The stimuli remained on the screen while the 1 or 2 response was made and also while the feedback word (RIGHT, WRONG, END) was presented. The feedback message remained on the screen until the participant pressed the "R" key if the word was RIGHT was present, the "W" key if the word WRONG was present, or the "E" if the word END was present. After a response to the feedback message was made, the screen cleared and a new trial began. The intertrial interval was 4 s.

Trial block structure and feedback. In all stages of training and testing, trials were grouped into blocks. The number of trials per block is described in each training and testing condition below. Trials within a block were presented in random order, without replacement. Experimental conditions changed when a criterion of 100% was achieved. For example, during single-exemplar training, the criterion was set at 100% correct on a block for each reference stimulus (0°, 18°, 3 SD). Performance relative to this criterion determined a participant's progression through the levels of informative feedback. Achieving this criterion resulted in a reduction of informative feedback on the next block. Training always began with 100% of trials occasioning informative feedback. If the criterion was achieved, only 75% of trials occasioned informative feedback on the next block. If the criterion was not achieved, the block was repeated.

If a participant's current level of informative feedback was 75% and the participant met the criterion, informative feedback would be set at 50% on the next block. If sub-criterion performance occurred, a participant would repeat that block at the 75% informative feedback level. If the criterion was not achieved at that second attempt, the participant would "back up" to higher informative feedback levels (e.g., 75% to 100%) on the next block. This advancement and backing up through the feedback levels continued until the training criterion was achieved at the 0% informative feedback level.

Preliminary training. The purpose of preliminary training was to establish a discrimination based on slope, and to familiarize participants with the different types and levels of feedback for each response. Preliminary training was conducted in three stages. The complexity of the stimuli was systematically increased from stage to stage. In Stage 1 the dimension of slope was manipulated. Straight lines (0 SD) without data points were presented with either 0° or 18° slope in the B panel, while the slope in the A panel remained at 0°. In Stage 2, the presence of data points was manipulated. Data points were added to these straight lines (0 SD) with the slopes identical to those of Stage 1. In Stage 3, the location of data points was manipulated. The standard deviation was increased to 1 SD in both the A and B panels. The data points were then offset from the regression line. The regression line was not shown in the graph. The data points were connected by a series of straight lines. In all stages, the slope in the B panel was either 0° or 18° while it was always 0° in the A panel. An example of Stage 1, 2, and 3 stimuli is presented in Figure 4 from top to bottom respectively.

Participants began preliminary training at Stage 1 at the 100% informative feedback level. If the criterion (100% on a block of 8 trials) was achieved, only 50% of trials occasioned informative feedback on the next block. Sub-criterion performance at the 50% level of feedback resulted in an increase in informative feedback on the next block, (e.g., 100%). When the mastery criterion was achieved at the 50% level, informative feedback was eliminated on the next block. If the criterion was achieved at the 0%

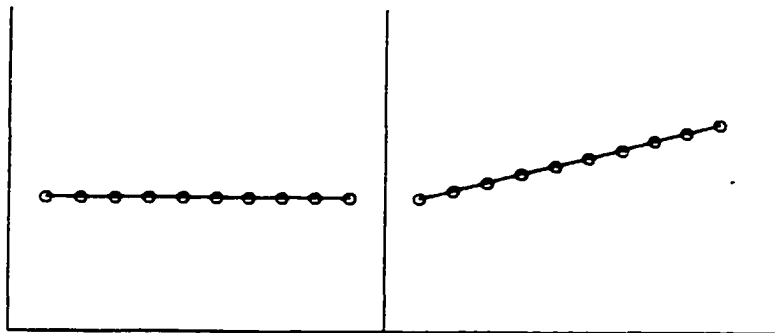
Figure 4. Stages 1, 2, and 3 of preliminary training.

In all graphs the slope of the data path in the left panel is 0° while it is 18° in the right panel. In Stage 1, no standard deviation is possible since no data points are shown. In Stage 2, 0 standard deviation is present as all data points fall on a straight line and the data points are shown. In Stage 3, 1 standard deviation is present and the data points are shown.

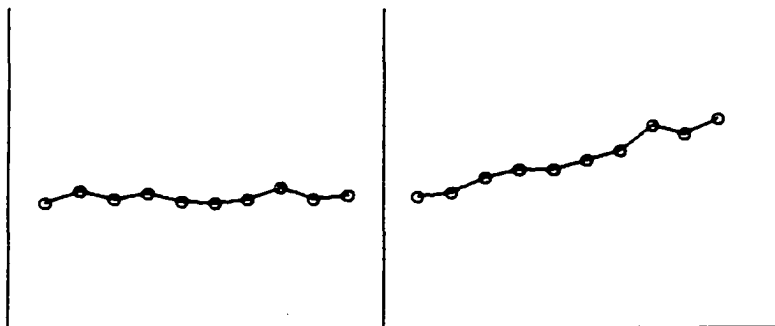
Stage 1
0°, 18°
0 Standard Deviations



Stage 2
0°, 18°
0 Standard Deviations



Stage 3
0°, 18°
1 Standard Deviation



informative feedback level for one block, training with Stage 2 stimuli began the next block at the 100% feedback level.

During Stages 2 and 3 informative feedback started at the 100% level and was reduced or increased as described in Stage 1. There were three levels of feedback in each stage (100%, 50%, and 0%). Criterion performance was 100% correct on all blocks. There were 8 trials in each block, 4 trials of the 0° stimulus and 4 trials of the 18° stimulus.

Single-exemplar training. Single-exemplar training immediately followed Stage 3 of preliminary training. During single-exemplar training a discrimination was established between one exemplar of 0° stimulus and one exemplar of the 18° degree stimulus. Both stimuli were presented with 3 SD variability. These particular exemplars were referred to as the "reference stimuli." There were 24 trials in a block, 12 trials of the 0° stimulus and 12 trials of the 18° stimulus.

During the first training block, participants received 100% informative feedback. If the criterion was achieved, informative feedback was reduced to 75% of trials on the next block. If the criterion was not achieved, participants repeated the block at the 100% informative feedback level. If the participant was at a reduced level of informative feedback (e.g., 75%) and did achieve the criterion, the amount of informative feedback was further reduced on the following block (e.g., 50%). If the criterion was not achieved at the reduced level (e.g., 75%), informative feedback was increased on the next block (e.g., 100%). After participants achieved the mastery criterion at the 0%

informative feedback level, they began the generalization test.

Generalization test. During the generalization test participants were presented a total of 120 trials, arranged in 6 blocks. Each block contained 20 trials. These 20 trials were composed of one exemplar of each of the ten slopes (0° , 2° , . . . 18°) in 3 SD form and one exemplar of these same ten slope stimuli in 6 SD form.

All stimuli were presented in random order in each block, without replacement. The particular exemplars presented during generalization testing were not presented previously in the study. Specifically, the reference stimuli (0° and 18° , 3 SD) were replaced by comparable exemplars (i.e., same slopes and standard deviation but a different configuration of data points). No informative feedback followed any trial.

Results

Dependent Variables

The effects of single-exemplar training were measured during the generalization tests. Specifically, the maintenance of the discrimination that had been established between the 0° and 18° (3 SD) stimuli during training was examined. A discrimination index (DI) was computed from the performance occasioned by the 0° (S-) and 18° (S+) stimuli during training and also during generalization tests. Each DI was determined by subtracting the percentage of S- trials that occasioned the DIFFERENT response from the percentage of S+ trials that occasioned the DIFFERENT response. This difference was called the discrimination index. The operation of subtraction served as a correction

for "guessing".

A DI of 100 indicates that the DIFFERENT response was emitted during all presentations of the 18° stimulus and never in the presence of the 0° stimulus. This value represents optimal discriminative performance.

A DI of 50 may result in different ways. For example, a participant could emit the DIFFERENT response to all presentations of the 18° stimulus but also emit this response to half the presentations of the 0° stimulus. Alternately, a DI of 50 could result if a participant emitted the DIFFERENT response during only half the 18° stimulus presentations and none of the 0° stimulus presentations.

A DI of 0 indicates that the DIFFERENT response was emitted for the same percentage of trials to the 18° stimulus as it was to the 0° stimulus during testing. Thus a DI of 0 represents non-differential responding to the stimuli arrayed along the dimension of slope. When the 0° and 18° stimuli occasioned the DIFFERENT response occasions the equally, responding can be high (e.g., 100% vs. 100%), intermediate (e.g., 50% vs. 50%) or low (e.g., 10% vs. 10%).

A DI of -50 indicates that the DIFFERENT response was emitted more often in the presence of the 0° stimulus than it was in the presence of the 18° stimulus during testing. This DI of -50 represents a loss of stimulus control by a partial reversal of responding to S+ and S-. This DI value of -50 could occur in different ways. First, the percentage of 18° stimulus presentations that occasion the DIFFERENT response could be intermediate while the percentage of 0°

stimulus presentations that occasion the DIFFERENT response could be high (e.g., 50% and 100% respectively). Second, the percentage of 18° stimulus presentations that occasion the DIFFERENT response could be low while the percentage of 0° stimulus presentations that occasion the DIFFERENT response could be intermediate (e.g., 0% and 50% respectively).

A DI of -100 indicates that the DIFFERENT response was emitted to all presentations of the 0° stimulus but never in the presence of the 18° stimulus. This represents a complete reversal from training levels--a very low percentage of the DIFFERENT response on S+ trials (i.e., 0%) and a very high percentage of the DIFFERENT response on S- trials (i.e., 100%).

In summary, the discrimination index summarizes each participant's performance by combining the percentage of S+ trials and the percentage of S- trials that occasion the DIFFERENT response in a systematic manner. The scale ranges from +100 to -100.

Discriminative control was also measured by examining the two components of the discrimination index directly, namely the percentage of S+ trials the occasioned the DIFFERENT response and the percentage of S- trials that occasioned the DIFFERENT response. Since the percentage of S+ and S- trials that occasion the DIFFERENT response could vary independently, examining these components could provide useful information that would supplement the discrimination index. Therefore, separate examinations of both group and individual results as measured by the discrimination index, the percentage of S+ trials that

occasion the DIFFERENT response and the percentage of S-trials that occasion the DIFFERENT response were made.

Discrimination Index

As seen in Table 1, when the 3 SD stimuli (0° and 18°) were presented during generalization tests, the DI dropped sharply from training levels. A Wilcoxon matched-pairs signed-ranked test showed the decline was significant ($T[11] = 0, p < .01$).

When the 6 SD stimuli (0° and 18°) were presented during generalization tests, the DI was at a chance level. The difference between the DIs occasioned by 3 SD and 6 SD stimuli was significant ($t[11] = 2.67, p < .05$) as found by a dependent-pairs, t -test. The analysis of the averaged discrimination index is representative of the individual participant data that are listed in Table 2.

Responses to S+

When the 18° stimuli (S+) were presented in 3 SD form, the percentage of trials that occasioned the DIFFERENT response dropped moderately from training to testing levels. A Wilcoxon matched-pairs signed-rank test showed this decline was significant ($T[8] = 0, p < .05$). When the 18° stimuli (S+) were presented in 6 SD form during generalization tests, the percentage of trials that occasioned the DIFFERENT response was moderate. A dependent-pairs, t -test showed that the difference in the percentage of S+ trials that occasioned the DIFFERENT response to 3 SD and 6 SD stimuli was not significant ($t[11] = 2.08, p > .05$).

Table 1

The Discrimination Index, the Percent of S+ Trials That Occasioned the DIFFERENT Response and the Percent of S- Trials That Occasioned the DIFFERENT Response to 0° and 18° Stimuli (3 SD and 6 SD) During Single-Exemplar Training and Generalization Testing for Experiment 1 (Averaged Data). Standard Error is Shown in Parenthesis.

Condition	SD	Discrim. Index	Percent S+ Trials With DIFFERENT Response	Percent S- Trials With DIFFERENT Response
Training	3	100 (0)	100 (0)	0 (0)
Testing	3	26 (11.9)	79 (6.2)	53 (10.2)
	6	-9 (13.7)	63 (7.8)	72 (9.7)

Table 2

The Discrimination Index (DI), the Percent of S+ Trials That Occasioned the DIFFERENT Response and the Percent of S- Trials That Occasioned the DIFFERENT Response to 0° and 18° Stimuli (3 SD and 6 SD) During Single-Exemplar Training and Generalization Testing During Experiment 1 (Single-Participant [Part.] Data).

Part.	Training		Generalization Test		
	SD	DI	DI	Percent S+ Trials With DIFFERENT Response	Percent S- Trials With DIFFERENT Response
18	3	100	100	100	0
14		100	83	100	17
19		100	67	100	33
12		100	50	83	33
2		100	33	67	33
17		100	33	50	17
16		100	0	100	100
3		100	0	83	83
4		100	0	83	83
15		100	-17	83	100
22		100	-17	67	83
20		100	-17	33	50
18	6		-17	50	67
14			100	100	0
19			-17	67	83
12			-33	67	100

Table 2--Continued

Part.	Training		Generalization Test		
	<u>SD</u>	DI	DI	Percent S+ Trials With DIFFERENT Response	Percent S- Trials With DIFFERENT Response
2	6		17	33	17
17			-17	67	83
16			0	100	100
3			-17	67	83
4			-83	17	100
15			-50	50	100
22			0	83	83
20			0	50	50

Responses to S-

When the 0° stimuli (S-) were presented in 3 SD form during generalization tests, the percentage of trials that occasioned the DIFFERENT response rose substantially from testing to training levels. A Wilcoxon matched-pairs signed-rank test showed that this increase was significant ($T[11] = 0, p < .01$).

When the 0° stimuli (S-) were presented in 6 SD form during generalization tests, the percentage of trials that occasioned by the DIFFERENT response was high. A dependent-pairs, t -test showed that the difference in the percentage of S- trials that occasioned the DIFFERENT response to 3 SD and 6 SD stimuli was not significant ($t[11] = 2.04, p > .05$).

Primary Generalization Gradients

Figure 5 illustrates the averaged generalization gradients to stimuli in the 3 SD and 6 SD condition. Compared to training levels, the probability of the DIFFERENT response to the 18° stimuli (S+) decreased while the probability of the DIFFERENT response to the 0° stimuli (S-) increased. In addition, there was no change of responding occasioned by the intermediate stimuli in both the 3 SD and 6 SD condition. The probabilities of the DIFFERENT response to intermediate stimuli, in the 3 SD and 6 SD condition, were in a narrow 30 point range, from 50 to 80. The averaged functions are representative of single participant data that are listed in Table 3.

Figure 5. The probability of the DIFFERENT response to 3 SD and 6 SD stimuli during generalization tests as a function of stimulus slope during Experiment 1.

Experiment 1

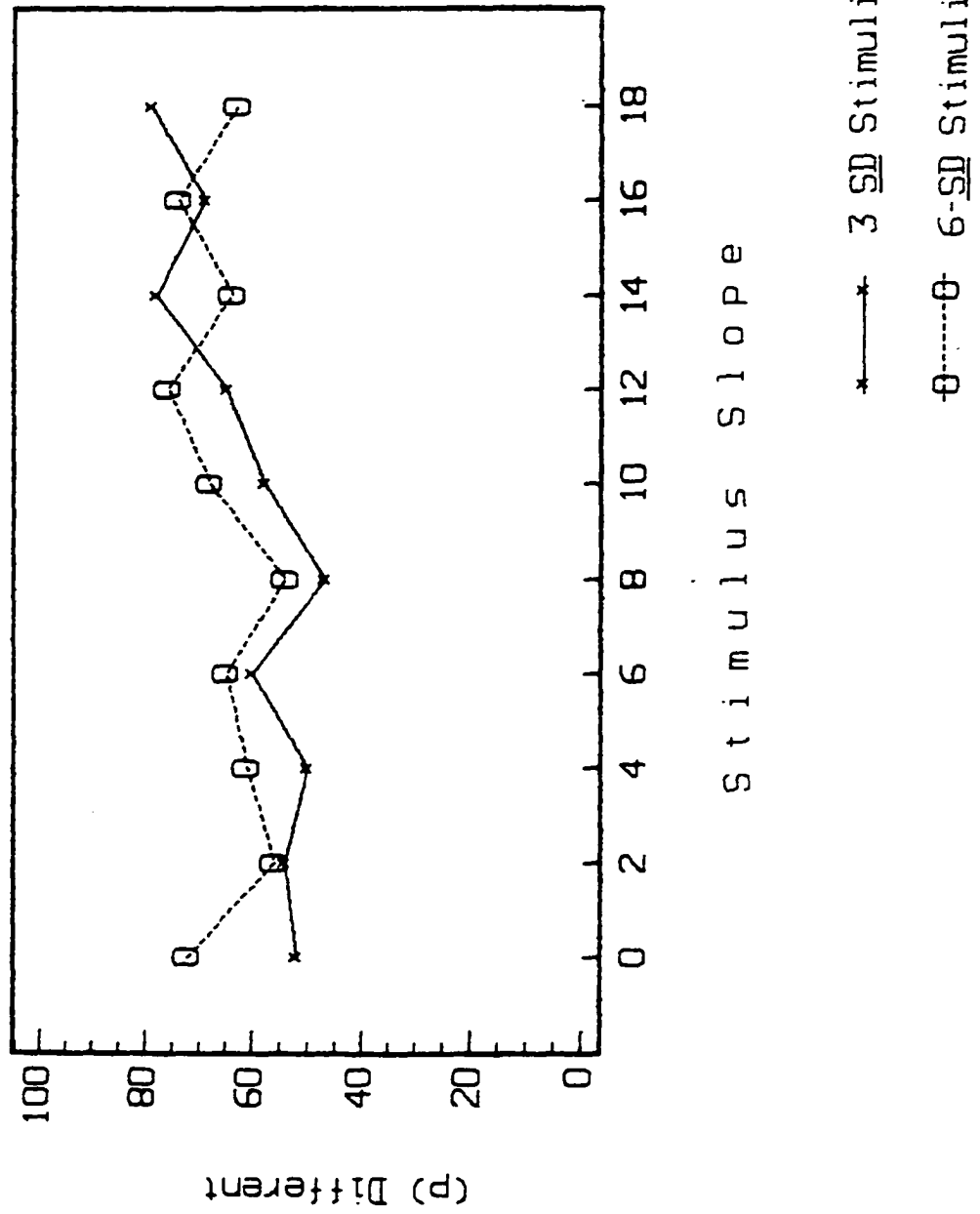


Table 3
Probability of the DIFFERENT Response to 3 SD Stimulus
Slopes (Top) and 6 SD Stimulus Slopes (Bottom) for
Experiment 1 (Single-Participant Data).

Stimulus Slopes	Participant											
	2	3	4	12	14	15	16	17	18	19	20	22
0	33	83	83	33	17	100	100	17	0	33	50	83
2	50	67	83	50	17	67	100	50	0	33	50	83
4	17	83	33	50	17	67	100	0	33	100	17	83
6	67	83	83	67	0	83	100	17	17	67	33	100
8	33	100	0	50	33	67	100	0	0	33	67	83
10	33	100	83	50	50	50	100	17	33	67	50	67
12	50	50	83	83	83	33	100	50	50	83	50	67
14	67	50	83	83	100	83	100	50	67	83	67	100
16	33	83	83	50	83	83	100	50	67	50	83	67
18	67	83	83	83	100	83	100	50	100	100	33	67
0	17	83	100	100	0	100	100	83	67	83	50	83
2	67	33	100	83	33	50	100	17	17	50	59	67
4	67	83	83	67	33	67	100	17	33	83	17	83
6	67	67	67	100	33	67	100	67	50	50	50	67
8	33	33	67	67	50	83	100	0	33	50	33	100
10	33	67	67	83	83	67	100	50	67	50	50	100
12	50	83	83	67	100	83	100	50	50	83	67	100
14	50	33	67	100	100	83	100	67	33	50	17	67
16	83	33	83	83	100	83	100	50	67	67	50	83
18	33	67	17	67	100	50	100	67	50	67	50	83

Discussion

The results showed that differential reinforcement was effective in establishing a discrimination between a single S+ and a single S-. After training, participants discriminated between graphs with 0° (S-) and 18° (S+) slopes, in the 3 SD condition, even in the absence of informative feedback. During generalization tests, there was a clear loss of discriminative control to the 0° and 18° stimuli.

The disruption of stimulus control seen in the current study was unexpected. These results raise several issues regarding sources of stimulus control established in the current study and suggest a number of interfering sources.

1. The nature of the stimuli presented in the cited stimulus generalization studies with animals and the stimuli in the present study differ. In those animal studies, the stimuli were simple as they differed on only one dimension (e.g., wavelength, intensity, line-tilt). In the present study, the graphic stimuli were complex differing in many dimensions simultaneously (e.g., slope, location of data points, variability). The more complex nature of the graphs as stimuli may have been responsible for the decrement in discriminative control during generalization tests.

2. During generalization tests, the reference stimuli (0° and 18°, 3 SD) were replaced with comparable exemplars--stimuli that had the same slope and standard deviation as the reference stimuli but differed in the location of the data points. These results suggest that participants were under stimulus control by the location of

data points rather than the slope of the data path. Perhaps disruption of stimulus control was due to this difference between the training and testing stimuli.

3. The inclusion of the intermediate stimuli (2° , 4° , . . . 16°) in 3 SD form may have been responsible for the disruption of stimulus control. Disruption of a trained discrimination by the introduction of test stimuli, while rare, was reported by Heinemann, Alvin, Sullivan, and Chase (1969). They trained pigeons to discriminate between two intensities of white noise. Generalization tests were conducted with the training stimuli and other, adjacent-intensity stimuli. The results showed the trained discrimination deteriorated substantially for most birds.

Fields et al. (1997) and Fields, Landon-Jimenez, Buffington, and Adams (1995) reported the disruption of conditional discriminations, in an equivalence experiment, when many new emergent-relations probes were introduced concurrently. Therefore, although stimulus control can be adequately demonstrated to a stimulus set during training, the introduction of novel stimuli during testing can disrupt the trained discrimination.

4. Inclusion of trials of higher variability stimuli (0° , 2° . . . 18° , 6 SD) may have been responsible for the disruption of stimulus control. During training, a discrimination between 0° and 18° stimuli, both in 3 SD form, was established. Since both the S+ and S- presented during training had the same variability (3 SD), responses may have come under partial control of this particular level of variability, in addition to the configuration of the data points and the slope of the data path. During

generalization tests, the variability of the stimuli was either 3 SD or 6 SD, which changed unsystematically from trial to trial. This change of variability may have been responsible for the loss of stimulus control to the training stimuli seen during generalization tests. A similar disruption of a trained discrimination was reported by Homa and Little (1995). In their study 10-year-old children were trained on three categories of forms, represented by 3, 6, or 9 exemplars of each. An immediate generalization test followed that included the training forms and new forms, all presented under extinction conditions. Correct categorization of the training forms declined during the test, particularly when the training categories were defined by only a few (i.e., 3 or 6) exemplars.

Experiment 2

In Experiment 1, single-exemplar training was not sufficient to maintain differential responding to 0° and 18° stimuli during generalization tests and induce systematic responding to intermediate stimuli. With single-exemplar training, the S+ and S- differed in both the stimulus slope and the positions of the data points. Therefore, responses may have been under partial or complete stimulus control of the position of data points with little regard for the stimulus slope differences.

In Experiment 2, the effects of three training regimens were explored. Single-exemplar training included training with a single S+ and a single S- (few stimuli, few trials), multiple-exemplar training included training with 48 S+'s and 48 S-'s (many stimuli, many trials).

Single-exemplar overtraining included repeated training with a single S+ and a single S- (few stimuli, many trials). The single-exemplar overtraining regimen served as a control for the multiple-exemplar training regimen by equating both groups for number of trials but not number of stimuli.

The focus of Experiment 2 was training with multiple exemplars (comparable stimuli with the same slope and standard deviation but composed of different data configurations). Other researchers (Homa et al., 1973; Posner and Keele, 1968) have demonstrated that the ability to correctly classify new information improves with the number of instances classified in the original learning task. It was plausible that multiple-exemplar training would produce the following desired effects:

(a) maintenance of the trained discrimination (0° and 18° , 3 SD) during generalization tests (b) establishment of dimensional control by intermediate slopes (2° . . . 16°) in the same variability (3 SD) as that used in the training stimuli, and (c) the maintenance of discriminative control by training slopes (0° and 18°) and dimensional control to intermediate slope (2° , . . . 16°), when these stimuli are presented in a higher variability form (6 SD) than that presented during training.

Method

Participants

Forty-two undergraduates students (22 women and 20 men, 18-45 years of age) served as participants with the same arrangements as described in Experiment 1.

Apparatus

The apparatus and stimuli were identical to those described in Experiment 1.

Procedure

Preliminary training and single-exemplar training.

Participants were randomly assigned to one of three groups designated by the method of training: single-exemplar training, multiple-exemplar training and single-exemplar overtraining. Participants in all groups received preliminary training and single-exemplar training as described in Experiment 1. Participants in the multiple-exemplar and single-exemplar overtraining groups received additional training as detailed below.

Multiple-exemplar training group. Following preliminary training and single-exemplar training, participants began multiple-exemplar training and were presented with a total of 96 unique exemplars. Forty-eight were exemplars of the 0° stimulus and 48 were exemplars of the 18° stimulus, all in 3 SD form. There were 24 exemplars in a block. Twelve were exemplars of 0° stimuli and 12 were exemplars of 18° stimuli. All trials in a block were presented in random order, without replacement. In total four blocks were presented in a predetermined order. There were five levels of informative feedback (100%, 75%, 50%, 25%, 0%) and participants advanced and backed up through these feedback levels as described in single-exemplar training during Experiment 1. The mastery criterion was 100% correct on each block.

Single-exemplar overtraining group. Participants in the multiple-exemplar group received training with

additional exemplars, which necessitated the presentation of additional trials. Any effect of multiple-exemplar training could then be attributed to the additional trials, the additional exemplars, or both. To control for the additional number of trials, a single-exemplar overtraining group was developed.

First, the number of trials a participant in the multiple-exemplar training group needed to meet the criterion with 96 exemplars was quantified. Next, a participant in the single-exemplar overtraining group was given this same number of trials with two exemplars as their counterpart in the multiple-exemplar training group. In short, participants in the multiple-exemplar training group and the single-exemplar overtraining group were matched by number of trials but differed by number of exemplars.

There were five levels of informative feedback (100%, 75%, 50%, 25%, 0%) and participants advanced and backed up through these levels as described during single-exemplar training in Experiment 1.

Generalization test. Following training, participants in all groups were given a generalization test. No informative feedback was presented on any trial. This test was composed of 200 trials, divided into 10 blocks. Each block of 20 trials included the following:

1. one each of the reference stimuli, 0° , 18° ,
3 standard deviations
2. one each of slopes 2° , 4° , . . . 16° , 3 standard
deviations

3. one each of slopes 0° , 2° , . . . 18° , 6 standard deviations

All stimuli were presented in a block, in random order, without replacement.

Results

Dependent Variables

The number of training trials in each group was compared. The effects of the three training regimens on the maintenance of the trained discrimination (0° and 18° , 3 SD stimuli) were measured. The following sections examine the experimental results expressed as the discrimination index, the percentage of S+ trials that occasioned the DIFFERENT response, and the percentage of S- trials that occasioned the DIFFERENT response. With each of these measures, the averaged performance of each training group, and also the individual participant data, are presented.

Two additional statistics d' , a measure of sensitivity, and β , a measure of response bias, were computed for each training group. Finally, the primary generalization gradients of individual participants are examined and compared graphically and quantitatively.

Number of Trials During Training

The number of trials needed to establish discriminative control in each training group was examined. The discrimination between the reference stimuli (1 S+ and 1 S-) was acquired rapidly for participants in all training groups. The mean number of trials needed for mastery was 44 for the single-exemplar training group, 45 for the single-exemplar overtraining group, and 51 for the multiple-exemplar training group.

Once the discrimination between the reference stimuli was learned, the participants in the single-exemplar overtraining group and multiple-exemplar training group received additional training. Participants in the multiple-exemplar training group required, on average, 518 additional trials of training to meet criterion with the 48 S+'s and 48 S-'s. Participants in the single-exemplar overtraining group were matched to those on the in the multiple-exemplar training group. Therefore participants in the single-exemplar training group also received 518 additional trials, on average, but with 1 S+ and 1 S-. Therefore, participants in the single-exemplar overtraining group received extensive overtraining.

Discrimination Index

Group data. Figure 6 illustrates the discrimination index (DI) occasioned by the 0° and 18° stimuli during the generalization tests. The vertical line in each bar illustrates plus or minus one standard error. At the end of discrimination training, the 0° and 18° stimuli in 3 SD form occasioned DIs of 100 for all three training regimens.

During generalization tests with 3 SD stimuli, participants showed moderate decrements in performance following single-exemplar training and single-exemplar overtraining but a minimal decrement following multiple-exemplar training (see Table 4).

For each of the training conditions, the decrements in discrimination performance was greater to the 6 SD stimuli than to the 3 SD stimuli. Compared to the 3 SD test levels, the average DIs to 6 SD stimuli was 74% lower after both single-exemplar training and single-exemplar overtraining.

Figure 6. The mean discrimination index for the single-exemplar training group (SET), the single-exemplar overtraining group (SEOT), and the multiple-exemplar training group (MET) to 3 SD and 6 SD stimuli for Experiment 2. The vertical lines represent ± 1 SE.

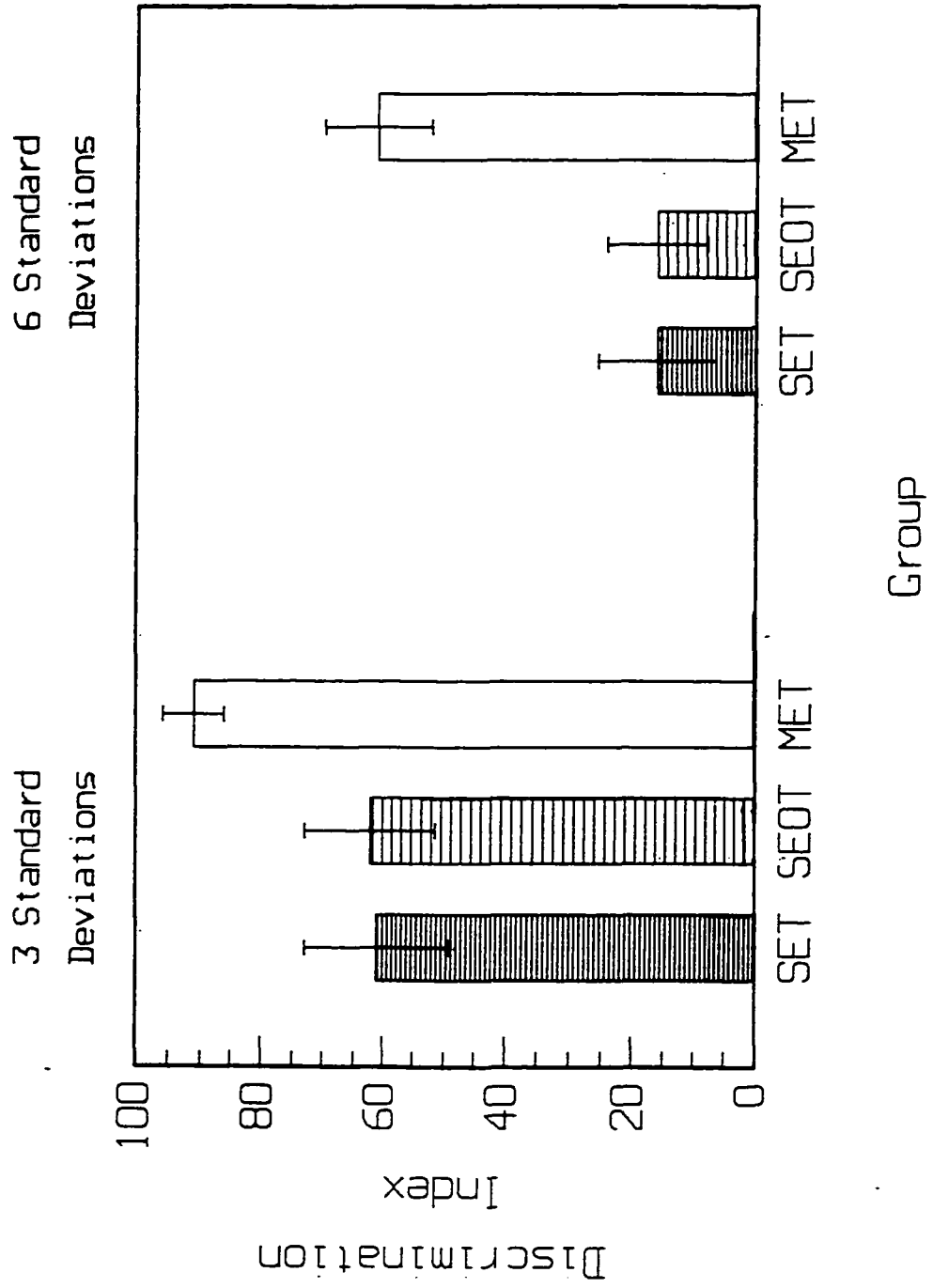


Table 4

The Average Discrimination Index, the Percent of S+ Trials That Occasioned the DIFFERENT Response, and the Percent of S- Trials that Occasioned the DIFFERENT Response to 0° and 18° Stimuli (3 SD and 6 SD) During Discrimination Training and Generalization Testing for Experiment 2. Training Included Single-Exemplar Training (SET), Single-Exemplar Overtraining (SEOT), and Multiple-Exemplar Training (MET).

Condition		Discrimination Index		Percent S+ Trials With DIFFERENT Response		Percent S- Trials With DIFFERENT Response	
		3 SD	6 SD	3 SD	6 SD	3 SD	6 SD
Training	All	100		100		0	
Testing	SET	61	16	85	61	24	46
	SEOT	62	16	77	66	15	50
	MET	91	61	94	79	3	18

By contrast, following multiple-exemplar training, the DI to 6 SD stimuli was only 33% lower.

A 2 x 3 ANOVA (Variability of Test Stimuli, [3 SD and 6 SD] x Training Group) with repeated measures was conducted (see Table 5). There were significant main effects for variability of test stimuli ($F[1, 39] = 76.01$, $p < .0001$), and for training group ($F[2, 39] = 7.53$, $p < .01$). The interaction between variability of test stimuli and training group was not significant ($F[2, 39] = 1.23$, $p > .1$).

Post-hoc, pairwise-comparison tests (Fisher's LSD) were conducted to identify the specific conditions that accounted for the main effects. The results of the pairwise-comparisons are listed in Table 6.

First, the effect of each training procedure on the performance occasioned by the 3 SD stimuli during generalization tests, was compared. Participants who received multiple-exemplar training showed significantly higher DIs than participants who received single-exemplar training ($t = 3.75$, $p < .01$) or single-exemplar overtraining ($t = 3.57$, $p < .01$). There were no significant differences between participants who received single-exemplar training and those who received single-exemplar overtraining.

Secondly, the performance of participants in each training group to the 6 SD stimuli, during the generalization tests, was compared. Participants who received multiple-exemplar training showed significantly higher DIs than participants who received single-exemplar training ($t = 5.62$, $p < .001$) and single-exemplar

Table 5
Analysis of Variance (ANOVA) of the Discrimination Index
for Training Group (T) and Variability of Test Stimuli (V)
for Experiment 2.

<u>Source</u>	<u>df</u>	<u>MS</u>	<u>F</u>	<u>p</u>
<u>Between</u>				
Training (T)	2	12760.72	7.53	0.0017
Sub. w. Grp.	39	1695.70		
<u>Within</u>				
Var. Stim. (V)	1	34001.18	76.01	< 0.0001
T x V	2	551.19	1.23	0.3028
V x Sub w. Grp.	39	447.34		

Table 6
Post-Hoc, Pairwise-Comparisons of the Discrimination Index for the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT) and the Multiple-Exemplar Training Group (MET) and Variability of Test Stimuli (3 SD and 6 SD) For Experiment 2.

		3 SD			6 SD		
		SET	SEOT	MET	SET	SEOT	MET
3 SD	SET						
	SEOT	0.18					
	MET	3.75**	3.57**				
6 SD	SET	5.62***					
	SEOT		5.74***		0.09		
	MET			3.75**	5.62***	5.53***	

* p < .05
 ** p < .01
 *** p < .001

overtraining ($t = 5.53, p < .001$). There were no significant differences in DI's between participants who received single-exemplar training and those who received single-exemplar overtraining.

The DIs within each training group were compared for the 3 SD and 6 SD stimuli. Participants in all three groups showed better discriminative performance to the 3 SD test stimuli than to the 6 SD test stimuli. These differences in performance were significant for the single-exemplar training group ($t = 5.62, p < .001$), the single-exemplar overtraining group ($t = 5.74, p < .001$) and the multiple-exemplar training group ($t = 3.75, p < .01$).

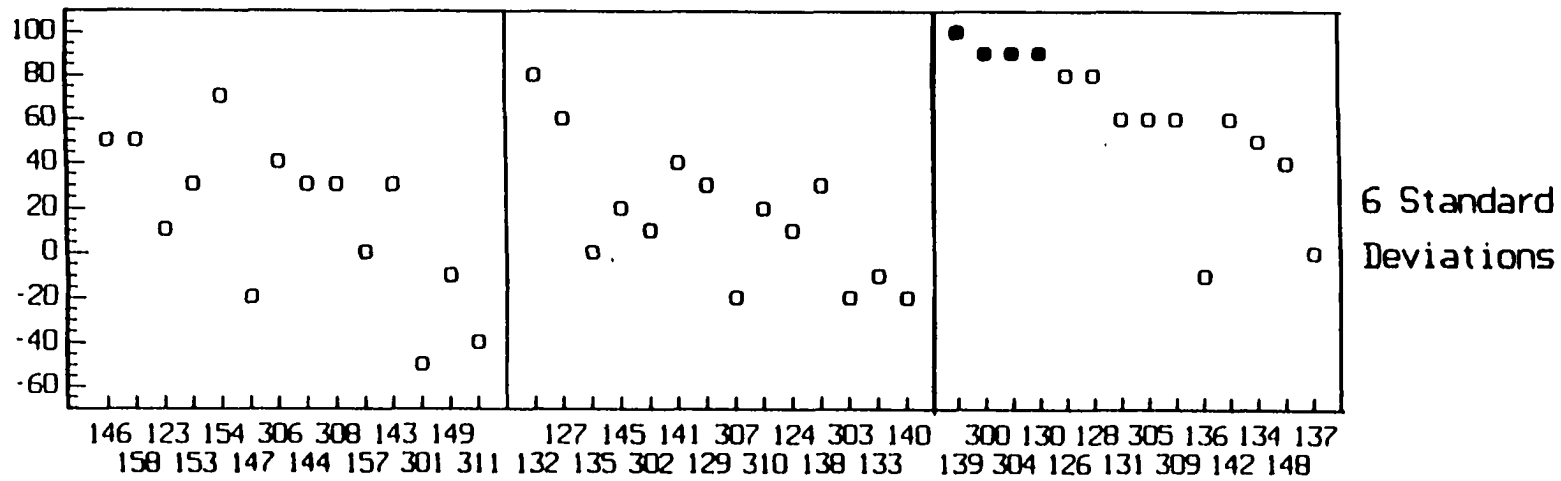
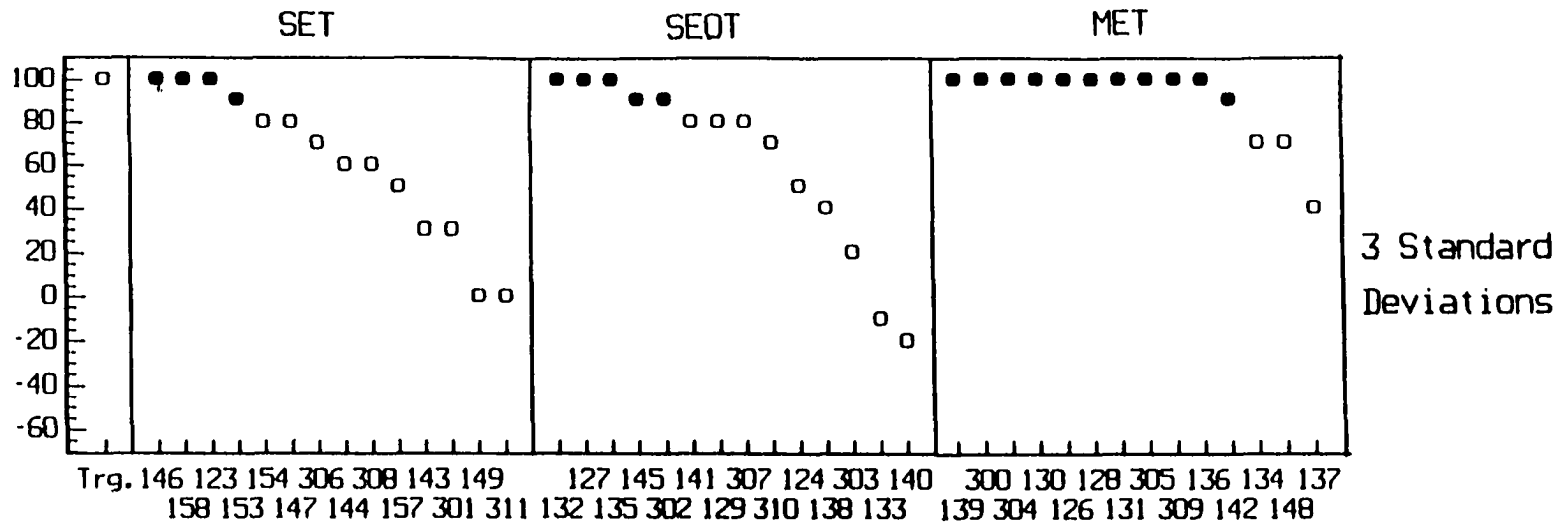
Single-participant data. The data in Figure 7 represent the DIs for the individual participants in each training group to the 0° and 18° stimuli during generalization testing. The filled symbols represent participants with DIs $\geq 90\%$. During discrimination training, all participants showed a DI of 100 to 0° and 18°, 3 SD stimuli.

During generalization testing with 3 SD stimuli, a DI of $\geq 90\%$ was obtained for 29% of participants in the single-exemplar training group, 36% of participants in the single-exemplar overtraining group and 79% of participants in the multiple-exemplar training group. A Chi-square analysis of the differences between the multiple-exemplar training group and the other training groups was significant ($X^2[2, N = 42] = 8.21, p < .05$), based on $\geq 90\%$ criterion, as shown in Table 7.

During generalization testing with 6 SD stimuli, 29% of participants in the multiple-exemplar training group

Figure 7. The discrimination index (DI) for individual participants after single-exemplar training (SET), single-exemplar overtraining (SEOT), and multiple-exemplar training (MET) in Experiment 2. The DIs to 3 SD stimuli are presented in the top panel while the DIs to 6 SD stimuli are presented in the bottom panel. Filled symbols indicate $\geq 90\%$ criterion level.

D i s c r i m i n a t i o n I n d e x



P a r t i c i p a n t s

Table 7

Chi-Square Analysis of the Discrimination Index for Experiment 2. The Number of Participants Who Achieved a \geq 90% Discrimination Index in the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) to 3 SD Stimuli (Top) and 6 SD Stimuli (Bottom) is Shown.

SD	Group			Chi-Square	p
	SET	SEOT	MET		
3	4	5	11	8.21	0.0165
6	0	0	4	8.84	0.0120

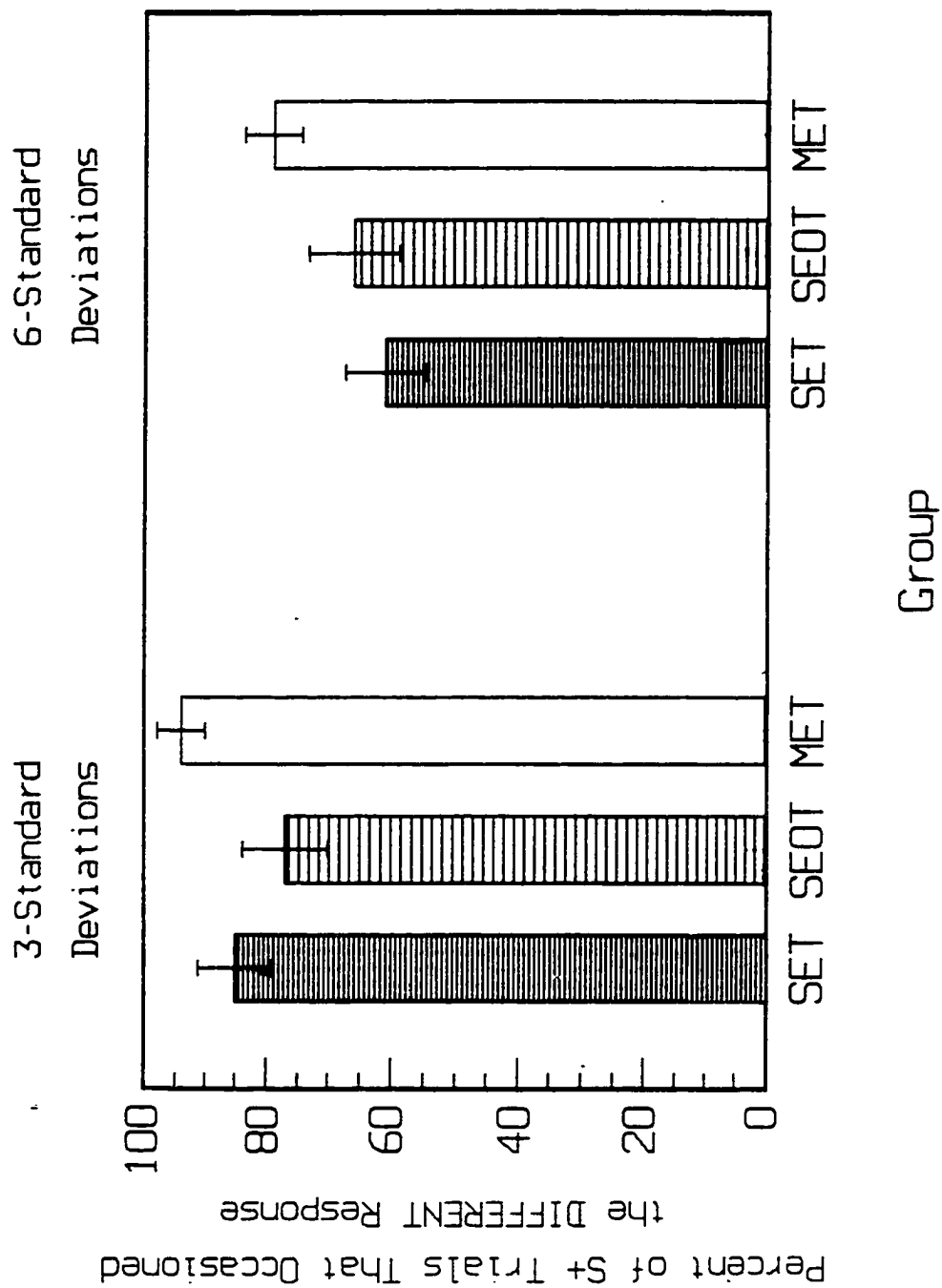
achieved a DI \geq 90% but no participants in either the single-exemplar training or the single-exemplar overtraining group met this criterion. A Chi-square analysis showed significant differences between training groups as to the number of participants who met this criterion ($X^2 [2, N = 42] = 8.84, p < .05$), as shown in Table 7. The performance of participants who achieved \geq 90% criterion to 3 SD and 6 SD stimuli was compared. A higher percentage of participants in each training regimen achieved the \geq 90% criterion to the 3 SD stimulus rather than to the 6 SD stimulus. A Chi-square analysis was not performed on this difference since it would violate the assumption of independence of observation.

Responses to S+

Group data. Figure 8 illustrates the percentage of trials that occasioned the DIFFERENT response to the 18° stimuli (S+) in 3 SD form during generalization tests. The vertical line illustrates plus or minus one standard error. At the end of discrimination training, the 18° stimuli in 3 SD form occasioned the DIFFERENT response for 100% of trials, for all three training groups.

When the 18° stimuli (S+) were presented in 3 SD form during generalization tests, the percentage of S+ trials that occasioned the DIFFERENT response declined in all groups from training levels. The average percentage of S+ trials that occasioned the DIFFERENT response declined 15% in the single-exemplar training group, 23% in the single-exemplar overtraining group, and 6% in the multiple-exemplar group, (see Table 4).

Figure 8. The percent of S+ trials that occasioned the DIFFERENT responses for the single-exemplar training group (SET), the single-exemplar overtraining group (SEOT), and the multiple-exemplar training group (MET) to 0° and 18° stimuli, (3 SD and 6 SD), for Experiment 2. The vertical lines represent ± 1 SE.



For a given training condition, the lower-variability (3 SD) stimuli presented during generalization tests occasioned less disruption of discriminative control than did the higher-variability (6 SD) stimuli. Compared to the 3 SD test levels, the average percentage of S+ trials that occasioned the DIFFERENT response to the 6 SD stimuli was 28% lower for the single-exemplar training group, 14% lower for the single-exemplar overtraining group and 16% lower for the multiple-exemplar training group.

A 2 x 3 ANOVA (Variability of Test Stimuli, [3 SD and 6 SD] x Training Group) with repeated measures was conducted with the percentage of S+ trials that occasioned the DIFFERENT response during the generalization tests. There was a significant main effect for variability of test stimuli ($F[1, 39] = 18.38, p < .001$), but no effect for training group, as shown in Table 8.

Post-hoc, pairwise-comparison tests (Fisher's LSD) were conducted to identify the specific conditions that accounted for the main effect. The results of the pairwise-comparisons are listed in Table 9. For participants in single-exemplar training group, a significantly higher percentage of S+ trials occasioned the DIFFERENT response to the 3 SD stimuli than to the 6 SD stimuli, ($t = 3.60, p < .01$). For participants in the multiple-exemplar training group, a significantly higher percentage of S+ trials occasioned the DIFFERENT response to the 3 SD stimuli than to the 6 SD stimuli, ($t = 2.18, p < .05$). Participants in the single-exemplar overtraining group did not differ significantly in their percentage of responses to 3 SD and 6 SD stimuli.

Table 8

Analysis of Variance (ANOVA) of the Percent of S+ Trials That Occasioned the DIFFERENT Response for Training Group (T) and Variability of Test Stimuli (V) for Experiment 2.

<u>Source</u>	<u>df</u>	<u>MS</u>	<u>F</u>	<u>p</u>
<u>Between</u>				
Training (T)	2	1825.00	2.61	0.0863
Sub. w. Grp.	39	699.08		
<u>Within</u>				
Var. Stim. (V)	1	5504.76	18.38	0.0001
T x V	2	308.34	1.03	0.3666
V x Sub w. Grp.	39	299.45		

Table 9

Post-Hoc, Pairwise-Comparisons of the Within-Group Percent of S+ Trials That Occasioned the DIFFERENT Response for Experiment 2. The Percent of S+ Trials in 3 SD Form That Occasioned the DIFFERENT Response is Compared to the Percent of S+ Trials in the 6 SD Form That Occasioned the DIFFERENT Response for the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET).

		3 SD		
		SET	SEOT	MET
6 SD	SET	3.60**		
	SEOT		1.64	
	MET			2.18*

* $p < .05$
 ** $p < .01$

Single-participant data. The data in Figure 9 represent the percentage of S+ trials (18° stimuli in 3 SD form) that occasioned the DIFFERENT response for individual participants in each training group. The filled symbols represent participants who made the DIFFERENT response for $\geq 90\%$ of S+ trials. During discrimination training, all participants emitted the DIFFERENT response on 100% of S+ trials.

During generalization tests to the 18° stimuli in 3 SD form, $\geq 90\%$ S+ trials occasioned the DIFFERENT response for 79% of participants in the single-exemplar training group, 50% of participants in the single-exemplar overtraining group and 86% of participants in the multiple-exemplar training group. A Chi-square analysis showed no significant differences between groups ($X^2[2, N = 42] = 4.90, p > .05$) as measured by the percentage of S+ trials that occasioned the DIFFERENT response (see Table 10).

During generalization tests to the 18° stimuli in 6 SD form, $\geq 90\%$ of S+ trials occasioned the DIFFERENT response for 7% of participants in the single-exemplar training group, 29% of the participants in the single-exemplar overtraining group and 43% of participants in the multiple-exemplar training group. A Chi-square analysis showed no significant differences between training groups as to the number of participants who met this criterion ($X^2[2, N = 42] = 4.68, p > .05$), as shown in Table 10.

The percentage of participants who responded as DIFFERENT for $\geq 90\%$ of S+ trials, to the 3 SD stimuli and the 6 SD stimuli, was compared. A higher percentage of participants in all training groups met the $\geq 90\%$ criterion

Figure 9. The percent of S+ trials that occasioned the DIFFERENT responses for individual participants after single-exemplar training (SET), single-exemplar overtraining (SEOT), and multiple-exemplar training (MET) for Experiment 2. Performance to 3 SD stimuli is presented in the top panel while performance to 6 SD stimuli is presented in the bottom panel. Filled symbols indicate $\geq 90\%$ criterion level.

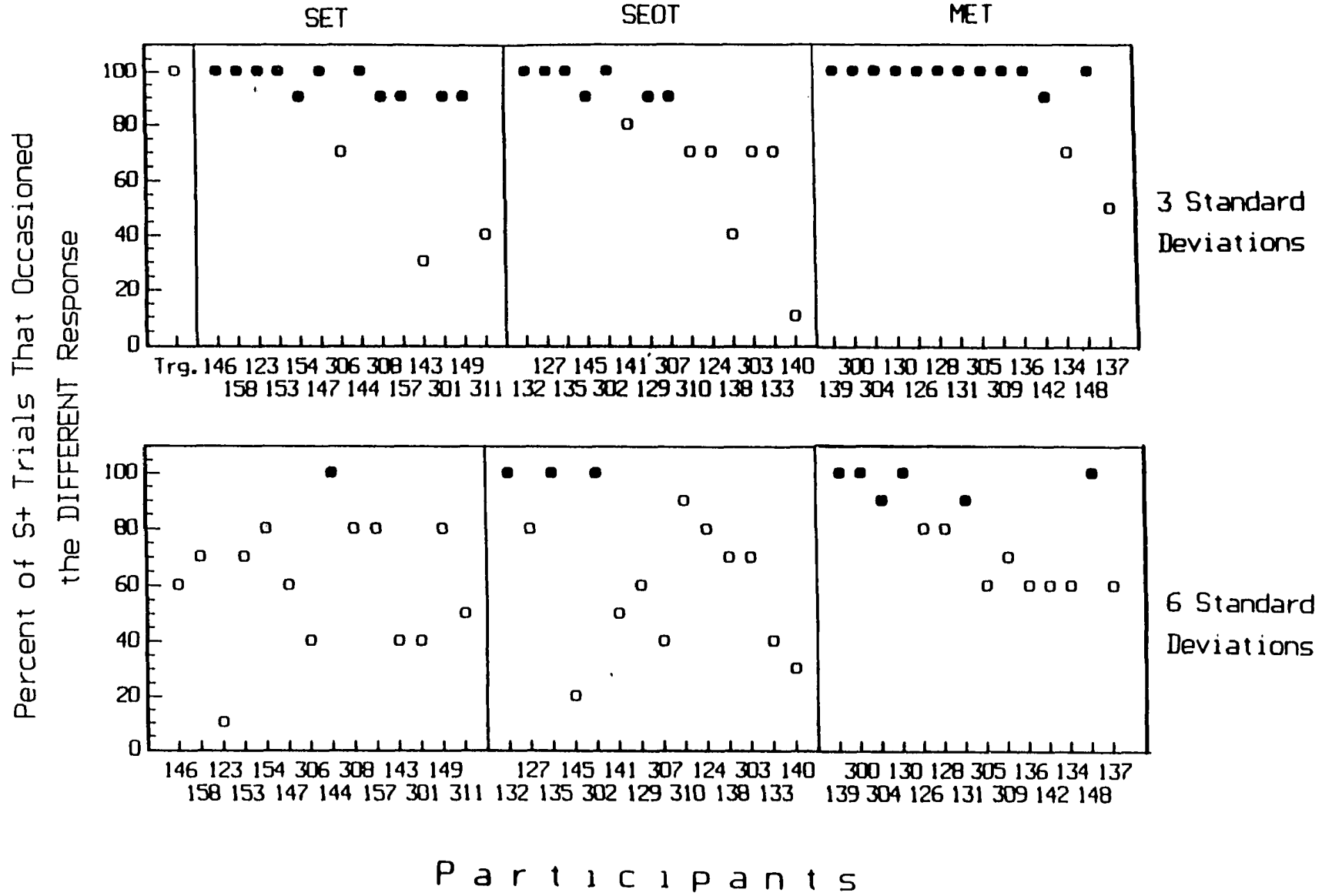


Table 10

Chi-Square Analysis of the Number of Participants Who Emitted the DIFFERENT Response for $\geq 90\%$ of S+ Trials in Experiment 2. Data for the the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) to 3 SD Stimuli (Top) and 6 SD Stimuli (Bottom) are Shown.

<u>SD</u>	<u>Group</u>			<u>Chi-Square</u>	<u>p</u>
	<u>SET</u>	<u>SEOT</u>	<u>MET</u>		
3	11	7	12	4.90	0.0863
6	1	4	6	4.68	0.0963

to the 3 SD stimuli than to the 6 SD stimuli (see Table 10).

Responses to S-

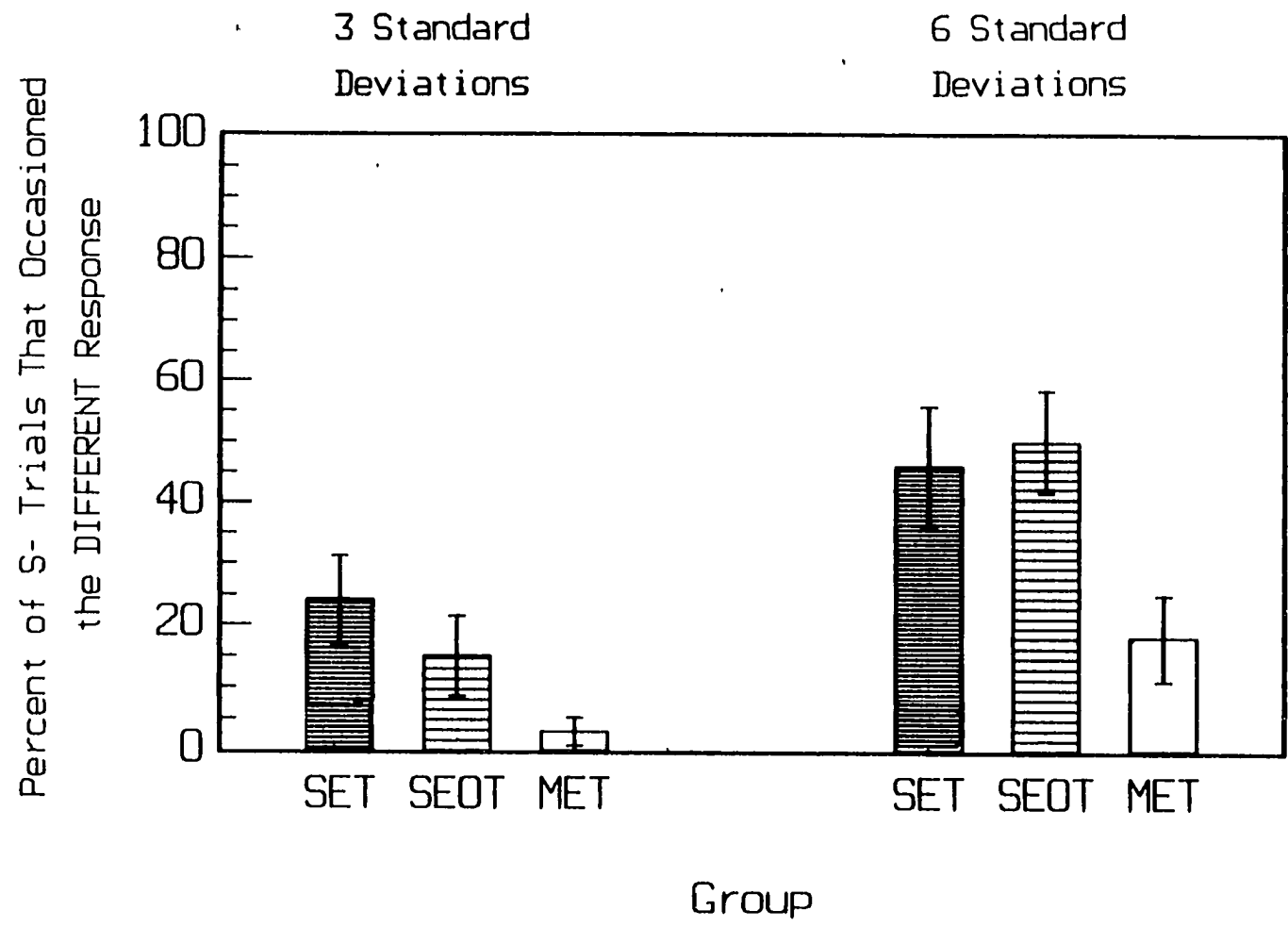
Group data. Figure 10 illustrates the percentage of trials that occasioned the DIFFERENT response to the 0° stimuli (S-), in 3 SD form, during generalization tests. The vertical line in each bar illustrates plus or minus one standard error. At the end of discrimination training the 0°, stimuli in 3 SD form occasioned the DIFFERENT response for 0% of trials, for all three training groups.

When the 0° stimuli (S-) were presented in 3 SD form during generalization tests, participants in all groups showed an increase in the percentage of S- trials that occasioned the DIFFERENT response. A large increase was seen for the single-exemplar training group, with an intermediate increase for the single-exemplar overtraining group, and a small increase for the multiple-exemplar training group (see Table 4).

When the 0° stimuli (S-) were presented in 6 SD form during generalization tests, a higher percentage of S- trials occasioned the DIFFERENT response for both the single-exemplar training group and the single-exemplar overtraining group than for the multiple-exemplar training group (see Figure 10). The percentage of S- trials that occasioned the DIFFERENT response to the 6 SD stimuli for both the single-exemplar group and the single-exemplar overtraining group, was more than 2.5 times that of the multiple-exemplar training group (see Table 4).

When the 0° stimuli (S-) were presented in lower-variability form, (3 SD), a lower percent of trials

Figure 10. The percent of S- trials that occasioned the DIFFERENT responses for the single-exemplar training group (SET), the single-exemplar overtraining group (SEOT), and the multiple-exemplar training group (MET) to 0° and 18° stimuli (3 SD and 6 SD) for Experiment 2. The vertical lines represent ± 1 SE.



occasioned the DIFFERENT response than when these stimuli were presented in higher-variability form (6 SD). This effect was seen during generalization tests for all training groups. A 2 x 3 ANOVA (Variability of Test Stimuli, [3 SD and 6 SD] x Training Group) with repeated measures was conducted with the percentage of S- trials that occasioned the DIFFERENT response, during generalization testing. There was a significant main effect of variability of test stimuli ($F[1, 39] = 36.34, p < .0001$) and also for training group ($F[2, 39] = 4.42, p < 0.05$). The interaction between variability of test stimuli and training group, as measured by the percentage of S- trials that occasioned the DIFFERENT response, was not significant ($F[2, 39] = 2.06, p > .1$) as shown in Table 11.

Post-hoc, pairwise-comparison tests (Fisher's LSD) were conducted to identify the specific conditions that accounted for the main effects. The results of these pairwise-comparisons are presented in Table 12.

First, the percentage of S- trials (18°) that occasioned the DIFFERENT response in the 3 SD condition for participants in each training group was compared, and is presented in Table 4. A significantly lower percentage of S- trials occasioned the DIFFERENT response for participants who received multiple-exemplar training than participants who received single-exemplar training ($t = 3.10, p = < .01$). Responses of participants in the multiple-exemplar training group did not differ significantly from those in the single-exemplar overtraining group, nor did the single-exemplar

Table 11

Analysis of Variance (ANOVA) of the Percent of S- Trials That Occasioned the DIFFERENT Response for Training Group (T) and Variability of Test Stimuli (V) for Experiment 2.

<u>Source</u>	<u>df</u>	<u>MS</u>	<u>F</u>	<u>p</u>
<u>Between</u>				
Training (T)	2	4996.43	4.42	0.0186
Sub. w. Grp.	39	1129.76		
<u>Within</u>				
Var. Stim. (V)	1	12144.04	36.34	< 0.0001
T x V	2	686.91	2.06	0.1417
V x Sub w. Grp.	39	334.16		

Table 12

Post-Hoc, Pairwise-Comparisons of the Percent of S- Trials That Occasioned the DIFFERENT Response for the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) and Variability of Test Stimuli (3 SD and 6 SD) for Experiment 2.

		3 SD			6 SD		
		SET	SEOT	MET	SET	SEOT	MET
	SET						
3 SD	SEOT	1.34					
	MET	3.10**	1.76				
	SET	3.10**					
6 SD	SEOT		5.07***		0.62		
	MET			2.27*	3.93***	4.55***	

* p < .05
 ** p < .01
 *** p < .001

overtraining group differ significantly from the single-exemplar training group.

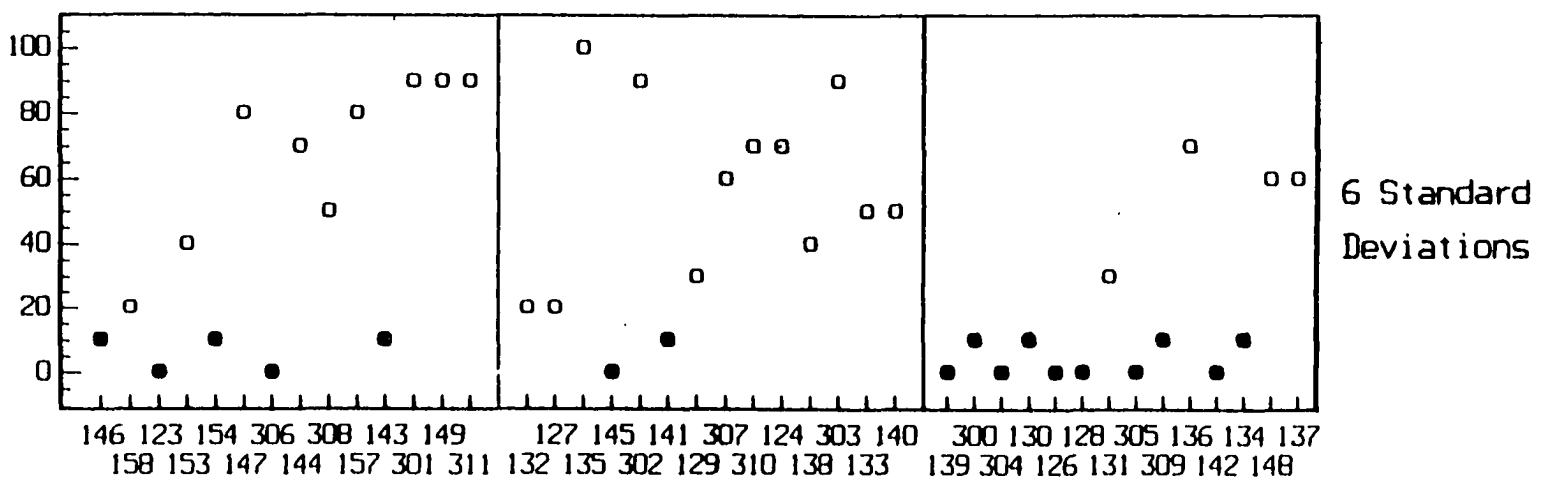
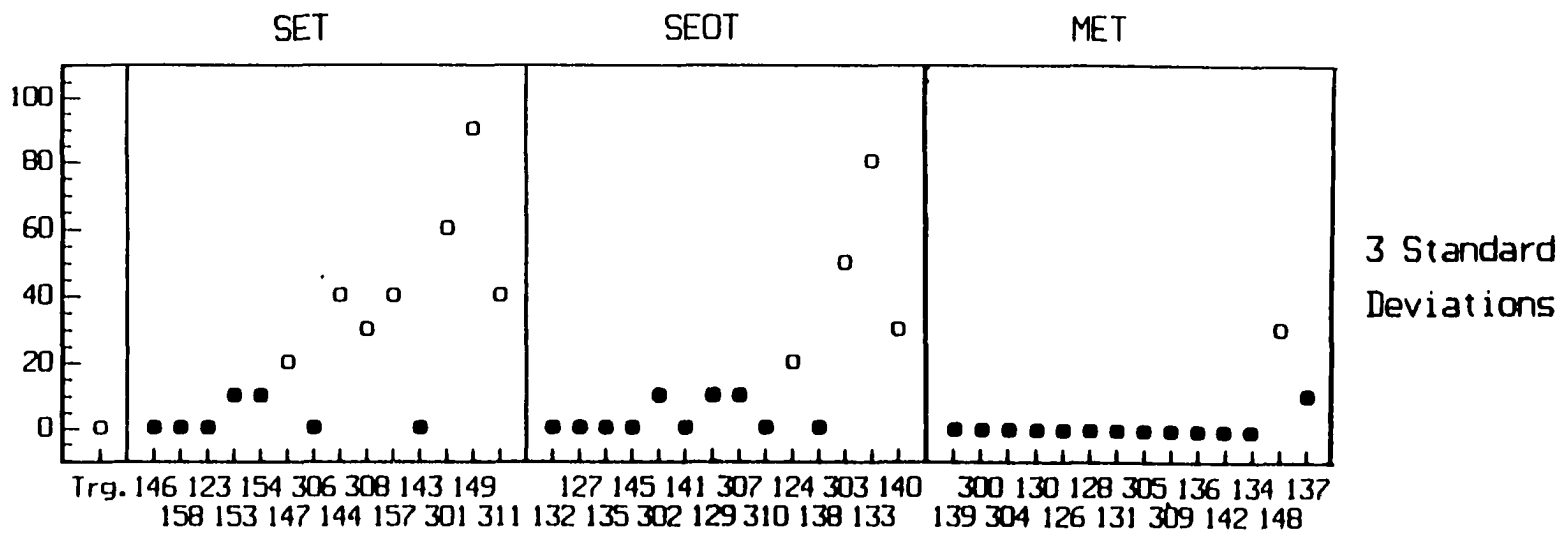
Second, the percentage of S- trials that occasioned the DIFFERENT response in the 6 SD condition for participants in each training group during the generalization test was compared and is presented in Table 4. For participants in the multiple-exemplar group, a significantly lower percentage of S- trials occasioned the DIFFERENT response than participants in the single-exemplar training group, ($t = 3.93$, $p < .001$) and the single-exemplar overtraining group, ($t = 4.55$, $p < .001$). There were no significant differences between participants who received single-exemplar training and those who received single-exemplar overtraining.

The previous two paragraphs compared the percentage of S- trials that occasioned the DIFFERENT response between groups. Here the percentage of S- trials that occasioned the DIFFERENT response of participants within each group, to the 0° stimuli (S-), in 3 SD and 6 SD form, was compared and are presented in Table 4. For participants in all three groups, a lower percentage of S- trials occasioned the DIFFERENT response to the 3 SD stimuli than to the 6 SD stimuli. These differences were significant for the single-exemplar training group ($t = 3.10$, $p < .01$), for the single-exemplar overtraining group ($t = 5.07$, $p < .001$) and for the multiple-exemplar training group ($t = 2.27$, $p < .05$).

Single-participant. The data in Figure 11 represent the percentage of S- trials (0° stimuli in 3 SD form) that occasioned the DIFFERENT response for individual

Figure 11. The percent of S- trials that occasioned the DIFFERENT responses for individual participants after single-exemplar training (SET), single-exemplar overtraining (SEOT), and multiple-exemplar training (MET) for Experiment 2. Performance to 3 SD stimuli is presented in the top panel while performance to 6 SD stimuli is presented in the bottom panel. Filled symbols indicate \leq 10% criterion level.

Percent of S- Trials That Occasioned
the DIFFERENT Response



P a r t i c i p a n t s

participants in each training group. The filled symbols represent participants who made the DIFFERENT response on $\leq 10\%$ of S- trials. During discrimination training, all participants emitted the DIFFERENT response on 0% of S- trials.

During generalization tests to the 0° stimuli (S-) in 3 SD form, $\leq 10\%$ of S- trials occasioned the DIFFERENT response for 50% of participants in the single-exemplar training group, 71% of participants in the single-exemplar overtraining group and 93% of participants in the multiple-exemplar training group. A Chi-square analysis showed these differences were significant ($X^2[2, N = 42] = 6.30, p < .05$), as shown in Table 13.

During generalization tests to the 0° stimuli (S-) in 6 SD form, $\leq 10\%$ of S- trials occasioned the DIFFERENT response for 36% of participants in the single-exemplar training group, 14% of participants in the single-exemplar overtraining group and 71% of participants in the multiple-exemplar training group. A Chi-square analysis showed that these differences were significant ($X^2[2, N = 42] = 9.65, p < .01$) as shown in Table 13.

The percentage of participants, who responded as DIFFERENT $\leq 10\%$ of S- trials to 3 SD stimuli and the 6 SD stimuli, was compared. A higher percentage of participants in all training groups met this criterion for the 3 SD stimuli than for the 6 SD stimuli.

Measure of Sensitivity (d')

An alternate measure of a participant's performance is d' . Derived from Signal Detection Theory, d' is a measure of a participant's sensitivity to the stimuli (Macmillan &

Table 13

Chi-Square of the Number of Participants Who Emitted the DIFFERENT Response for $\leq 10\%$ of S- Trials in Experiment 2. Data for the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) to 3 SD Stimuli (Top) and 6 SD Stimuli (Bottom) Are Shown.

SD	Group			Chi-Square	p
	SET	SEOT	MET		
3	7	10	13	6.30	0.0429
6	5	2	10	9.65	0.0079

Creelman, 1991). A d' score is defined in terms of z , the deviation of a specific score from the mean expressed in standard deviation units. Therefore, the mean of the transformed normal distribution is equal to 0 with a standard deviation of 1. Calculation of d' involves subtracting the z -score for the probability the DIFFERENT response to S+ from the z -score of the probability of the DIFFERENT response from S-. If the probability of the DIFFERENT response was equal in the presence of S+ and S- (i.e., perfect insensitivity) $d' = 0$.

Table 14 lists the mean d' s during for the single-exemplar training group, the single-exemplar overtraining group and the multiple-exemplar training group to the 0° and 18° stimuli in 3 SD and 6 SD form during generalization testing. These mean d' s are represented graphically in Figure 12 and are representative of the individual participant d' s listed Table 15.

During generalization testing to the 0° and 18°, 3 SD stimuli, the multiple-exemplar training group had the highest mean d' s. The single-exemplar training group and the single-exemplar overtraining group mean d' s were considerably lower than the d' for the multiple-exemplar training group.

During generalization tests to 0° and 18°, 6 SD stimuli, the multiple-exemplar training group had the highest mean d' . The mean d' s for the single-exemplar training group and single-exemplar overtraining group were substantially lower than the mean d' for the multiple-exemplar training group.

Table 14

Mean d's to 0° and 18° Stimuli (3 SD and 6 SD) for the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) During Generalization Testing for Experiment 2. Standard Error is Shown in Parentheses.

SD	Group		
	SET	SEOT	MET
3	2.49 (.42)	2.48 (.47)	4.09 (.28)
6	0.84 (.28)	0.66 (.28)	2.42 (.37)

Figure 12. Mean d 's after single-exemplar training (SET), single-exemplar overtraining (SEOT), and multiple-exemplar training (MET) to 3 SD and 6 SD stimuli for Experiment 2. The vertical lines represent ± 1 SE.

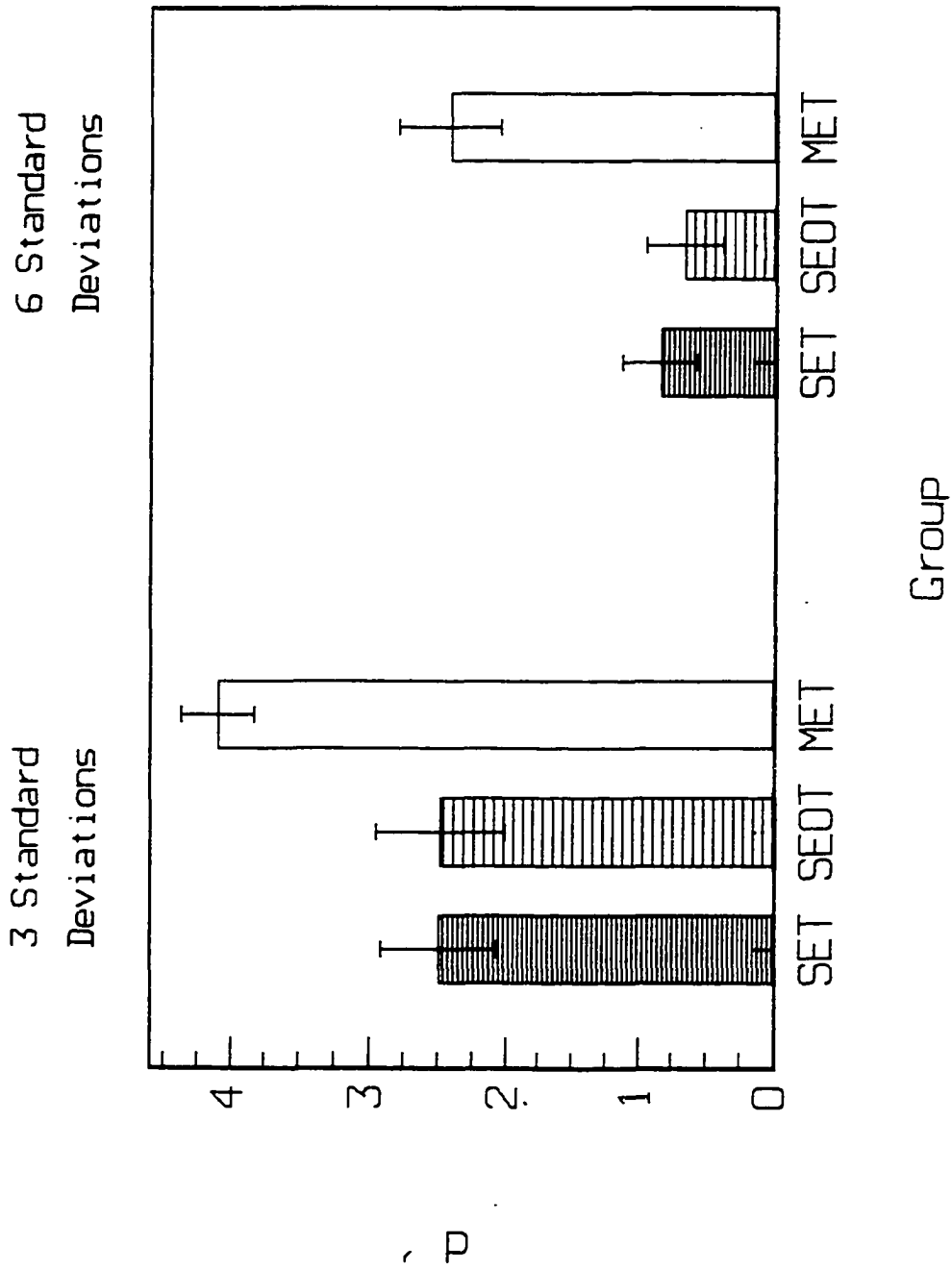


Table 15

Participants' d's in Groups SET, SEOT and MET for Experiment 2.

Group	Participant	3 SD	6 SD	6 SD ^a
SET	149	0.00	-0.44	-1.28
	311	0.00	-1.28	-0.59
	301	1.03	1.53	-0.44
	157	1.53	0.00	0.00
	308	1.87	0.84	0.77
	143	1.81	1.03	0.84
	154	2.56	2.12	1.03
	144	2.58	1.81	1.05
	306	2.85	2.08	1.36
	147	3.17	-0.59	1.53
	153	3.61	0.77	1.53
	123	4.66	1.05	1.81
	146	4.66	1.53	2.08
	158	4.66	1.36	2.12
SEOT	140	-0.76	-0.52	-0.76
	133	-0.32	-0.25	-0.52
	303	0.52	-0.76	-0.50
	124	1.36	0.32	-0.25
	138	2.08	0.77	0.00
	129	2.56	0.77	0.32
	307	2.56	-0.50	0.76
	310	2.85	0.76	0.77
	141	3.17	1.28	0.77
	145	3.61	1.49	1.05
	302	3.61	1.05	1.28

Table 15 -- Continued

Group	Participant	3 <u>SD</u>	6 <u>SD</u>	6 <u>SD</u> ^a
	127	4.66	1.68	1.49
	132	4.66	3.17	1.68
	135	4.66	0.00	3.17
MET	137	1.28	0.00	-0.27
	134	2.85	1.53	0.00
	148	2.85	2.08	1.53
	142	3.66	2.58	1.80
	126	2.66	3.17	1.80
	128	4.66	3.17	2.08
	130	4.66	3.61	2.58
	131	4.66	1.80	2.58
	136	4.66	-0.27	3.17
	139	4.66	4.66	3.17
	300	4.66	3.61	3.61
	304	4.66	3.61	3.61
	305	4.66	2.58	3.61
	309	4.66	1.80	4.66

^a This column rank orders the 6 SD d's for easier comparisons. This column does not correspond to the participant number on left.

In the previous two sections the d' s between training groups were compared. The d' s were also compared within each group. Relative to generalization test performance to 0° and 18° , 3 SD stimuli, performance to 6 SD stimuli was lowest for the single-exemplar overtraining group, intermediate for the single-exemplar training group and highest for the multiple-exemplar training group.

Response Bias (β)

Response bias (β) is a measure of the tendency of participants to favor one response over another, regardless of which stimulus is present. Unlike d' which measures a participant's sensitivity to the stimuli, the value of β reflects a participant's preference for one of two responses.

Calculation of bias involves dividing the normalized probability of the DIFFERENT response to S+ by the analogous measure of S-. In the current study, a $\beta = 1$ indicated that the NOT DIFFERENT response occurred with the same probability as the DIFFERENT response. Values of $\beta > 1$ indicated a bias of responding as NOT DIFFERENT. Values of $\beta < 1$ indicated a bias of responding of DIFFERENT.

Table 16 lists the median bias for the single-exemplar training group, single-exemplar overtraining group and multiple-exemplar training group to the 0° and 18° stimuli in 3 SD and 6 SD form. These data are presented graphically in Figure 13 and are representative of the individual participant data in Table 17. During training the median bias for all groups was equal to 1. Therefore participants did not favor one response over another.

Table 16

Median Response Bias (β) to 0° and 18° Stimuli (3 SD and 6 SD) for the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) During Discrimination Training and Generalization Testing for Experiment 2.

Condition	SD	Group		
		SET	SEOT	MET
Training	3	1.00	1.00	1.00
Testing	3	1.00	1.00	1.00
	6	1.59	0.98	1.54

Figure 13. Median bias after single-exemplar training (SET), single-exemplar overtraining (SEOT), and multiple-exemplar training (MET) to 3 SD and 6 SD stimuli for Experiment 2.

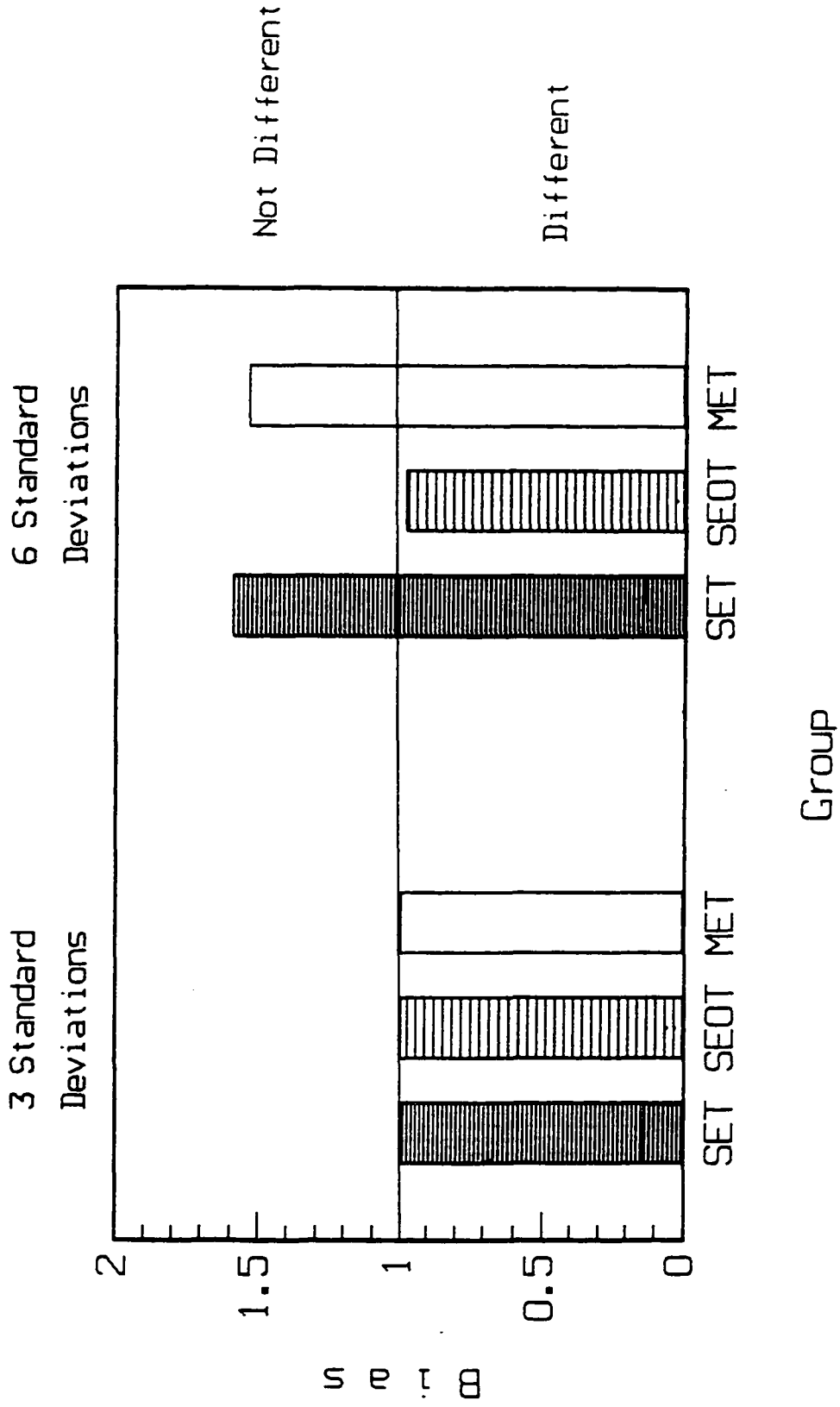


Table 17

Participants' Betas in Groups SET, SEOT and MET for Experiment 2.

Group	Participant	3 SD	6 SD	6 SD ^a
SET	144	0.07	0.08	0.08
	147	0.09	1.38	0.70
	153	0.15	0.90	0.90
	157	0.45	1.00	1.00
	301	0.45	2.20	1.24
	308	0.50	0.70	1.38
	123	1.00	6.67	1.59
	146	1.00	2.20	1.59
	149	1.00	1.59	2.20
	154	1.00	1.59	2.20
	158	1.00	1.24	2.20
	311	1.00	2.27	2.27
	143	13.20	2.20	6.66
	306	13.20	14.65	14.65
SEOT	302	0.15	0.15	0.09
	140	0.50	0.87	0.15
	303	0.87	1.98	0.50
	127	1.00	1.00	0.80
	129	1.00	1.11	0.87
	132	1.00	0.09	0.90
	135	1.00	1.00	0.97
	307	1.00	1.00	1.00
	124	1.24	0.80	1.00
	133	1.24	0.97	1.00
	145	6.66	10.62	1.11

Table 17--Continued

Group	Participant	3 <u>SD</u>	6 <u>SD</u>	6 <u>SD</u> ^a
	141	10.62	2.27	1.98
	310	13.20	0.50	2.27
	138	14.65	0.90	10.62
MET	148	0.08	0.07	0.07
	126	1.00	10.62	0.15
	128	1.00	10.62	0.15
	130	1.00	0.15	0.50
	131	1.00	0.50	1.00
	136	1.00	1.11	1.00
	139	1.00	1.00	1.11
	300	1.00	0.15	1.98
	304	1.00	6.66	2.20
	305	1.00	14.65	6.66
	309	1.00	1.98	10.62
	137	2.27	1.00	10.62
	142	6.66	14.65	14.65
	134	13.20	2.20	14.65

^a This column rank orders the 6 SD d's for easier comparisons. This column does not correspond to the participant number to the left.

During generalization tests to the 0° and 18°, 3 SD stimuli, the median bias remained unchanged from training levels for all groups.

During generalization tests to the 0° and 18°, 6 SD stimuli, the median response bias increased modestly for the single-exemplar training group, and the multiple-exemplar training group. Therefore, when presented with more-variable stimuli (6 SD rather than 3 SD), participants in these two groups responded more often as NOT DIFFERENT. Participants in the single-exemplar overtraining group, however, showed bias levels virtually unchanged from 3 SD to 6 SD stimuli.

Summary. The performance of participants to the 0° and 18° stimuli, in both 3 SD and 6 SD form was measured during generalization testing following single-exemplar training, single-exemplar overtraining and multiple-exemplar training. Performance measures included the discrimination index, the percentage of S+ trials that occasioned the DIFFERENT response, the percentage of S- trials that occasioned the DIFFERENT response, d' and β . From these results some general conclusions can be drawn.

In a given training group, participants demonstrated better control by the slope of the data path slope to 3 SD than to 6 SD stimuli. As measured by the DIs, participants in the multiple-exemplar training group showed significantly better maintenance of stimulus control to both 3 SD stimuli and 6 SD stimuli than those in the single-exemplar or single-exemplar overtraining groups. These differences in DIs were mostly the result of a much lower percentage of S- trials that occasioned the DIFFERENT

response for the multiple-exemplar training group. The differences in the percentage of S+ trials that occasioned the DIFFERENT response between groups, to 3 SD and 6 SD stimuli, were not significant. Performance measured by the discrimination index was in concordance with performance measured by d' . Single-exemplar training and the multiple-exemplar training resulted in a moderate bias to respond as NOT DIFFERENT for stimuli presented in more-variable (6 SD) form.

Training Effects on Generalization Gradient Slopes

Dependent measures. In the previous section, it was shown how the 0° and 18°, 3 SD stimulus slopes presented during training maintained discriminative control during generalization tests conducted after three different training regimens. During these generalization tests, participants were also presented graphs containing a wide range of intermediate stimulus slopes (2°, 4°, . . . 16°) in both 3 SD and 6 SD form. The relative effects of single-exemplar training, single-exemplar overtraining and multiple-exemplar training on the performance occasioned by these intermediate stimulus slopes were also examined.

Generalization gradient slopes. The generalization gradient slopes illustrate the rate of change of responding as a function of the stimulus slopes presented during the generalization tests. Generalization gradient slopes are therefore another dependent measure that could be influenced by the three training regimens.

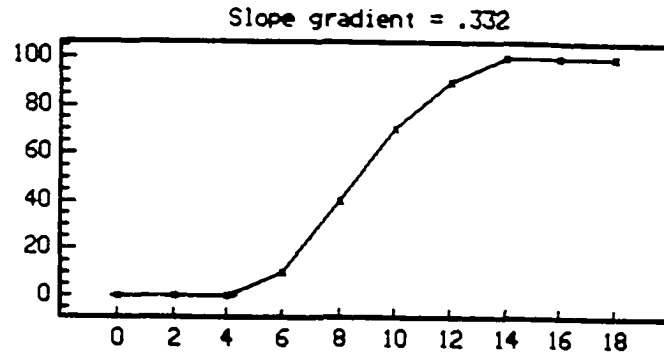
The generalization gradient slopes were quantified by the following procedure. First the probability of a "DIFFERENT" response to each of the 0°, 2° . . . 18°, 3 SD

and 6 SD stimulus slopes, presented in the generalization tests, was converted to a z-score. Next, the z-scores were graphed. Finally two regression lines were fitted to the normalized probability values occasioned by the stimulus slopes presented in the generalization tests. One regression line was for the 3 SD probability and the other was for the 6 SD probability. The equation of the regression line was ($y = mx + b$), in which the slope was specified by "m". In addition, r^2 , the percentage of variability accounted for by the regression line, was also calculated for each generalization gradient.

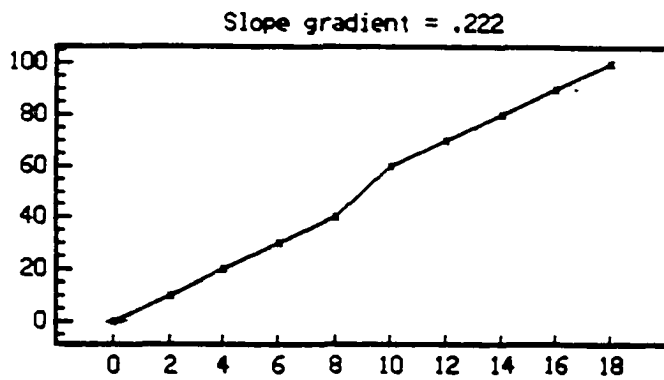
Figure 14 illustrates four hypothetical generalization gradients all constructed with DIs of 100 (i.e., asymptotic, discriminative performance to the 0° and 18°, 3 SD training stimulus slopes) but differing in response to intermediate stimulus slopes. Plot A shows a steep generalization gradient slope due to both good anchoring of response to extreme stimulus slopes and systematic responding to intermediate stimulus slopes. Plot B shows a moderate generalization gradient slope due to systematic responding to all intermediate stimulus slopes. Plot C shows a shallow generalization gradient slope due to non-differential responding to intermediate stimulus slopes. Plot D shows a very shallow generalization slope due to a reversal of responding to intermediate stimulus slopes. The generalization gradient slopes may vary widely even when the discrimination indices are identical. High DIs occasioned by the 0° and 18° training stimulus slopes do not necessitate steep generalization gradient slopes nor do they dictate the shape of the generalization gradients.

Figure 14. Hypothetical generalization gradients that show identical performance (probability of the DIFFERENT response) to 0° and 18° stimulus slopes but differ in response to intermediate stimulus slopes (2° . . . 16°). Generalization gradient slopes in graphs A to D are .332, .222, .132, and .071 respectively. R^2 values are .94, .99, .43, -.01 respectively.

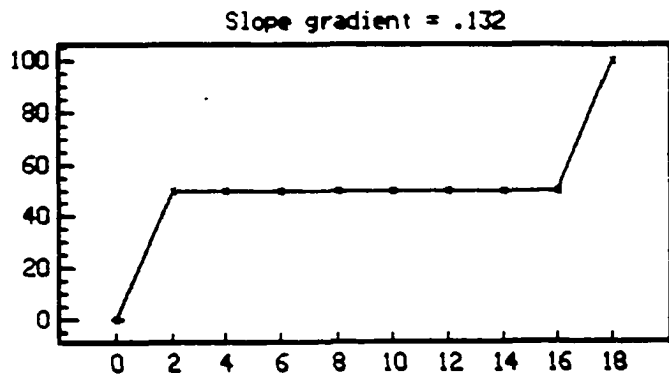
(P) D i f f e r e n t



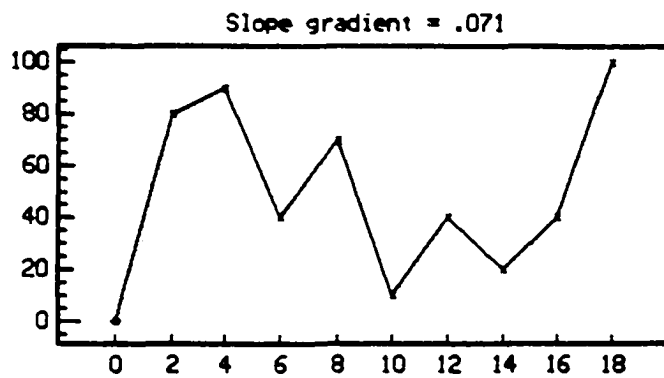
A



B



C



D

Stimulus Slope

Therefore, the effect of training regime on the generalization gradient slopes was examined.

Participant assignment. Of the 42 participants, 33 were selected (called the "subset") for an analysis of the generalization gradient slopes. The prerequisite for inclusion in the subset, slopes of the generalization gradient had to be > 0 in both 3 SD and 6 SD conditions. Subset participants are listed in Table 18. Participants not in the subset are listed in Appendices A and B.

Figures 15, 16, and 17 illustrates the generalization gradients of all participants to 3 SD stimulus slopes in the single-exemplar training group, single-exemplar overtraining group and the multiple-exemplar training group respectively. The generalization gradient slope values to 3 SD stimulus slopes are presented above each graph. The stimulus slopes, angles measured in degrees, are presented along the X-axis. An asterisk marks participants not included in the subset in the single-exemplar training group and the single-exemplar overtraining groups.

Within-Participant Generalization Gradient Slope Analysis

Comparison of generalization gradient slopes to 3 SD vs. 6 SD stimulus slopes. Participants' performances to 3 SD and 6 SD stimulus slopes were evaluated by examining their generalization gradients. Individual participant's gradients to both 3 SD and 6 SD stimuli were plotted on the same axes in the single-exemplar training group (Figure 18), the single-exemplar overtraining group (Figure 19), and the multiple-exemplar training group (Figure 20). The generalization slope values to 3 SD and 6 SD stimuli are presented above each graph respectively. Visual inspection

Table 18

Slopes of Generalization Gradients For Participants (Part.) With Generalization Slope Values > 0 to 3 SD and 6 SD Stimuli for Three Training Groups (SET, SEOT, MET Groups) for Experiment 2.

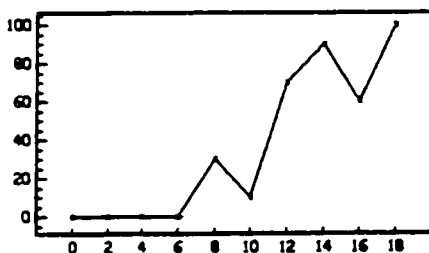
SET			SEOT			MET		
Part.	3 SD	6 SD	Part.	3 SD	6 SD	Part.	3 SD	6 SD
147	.119	.037	138	.045	.025	134	.116	.112
308	.125	.039	124	.053	.026	137	.121	.017
143	.139	.079	129	.134	.074	148	.206	.115
306	.144	.137	141	.182	.102	142	.218	.153
144	.149	.123	302	.195	.064	136	.224	.038
154	.161	.093	145	.219	.099	305	.254	.151
153	.177	.069	310	.240	.028	309	.259	.177
158	.229	.108	132	.293	.146	130	.268	.210
123	.248	.073	127	.306	.104	300	.295	.211
146	.271	.105				131	.299	.087
						126	.317	.186
						128	.330	.184
						304	.334	.214
						139	.336	.277
M	.176	.086		.185	.074		.256	.152
SE	.017	.011		.031	.014		.019	.019

Figure 15. Generalization gradients for individual participants in the single-exemplar training group for Experiment 2. The probability of the DIFFERENT response during generalization tests to 3 SD stimulus slopes is shown. The participant number and the generalization gradient slope are shown above each graph. Participants not in the subset (those who demonstrated generalization gradient slopes ≤ 0 to 3 SD or 6 SD stimulus slopes) are indicated by an asterisk.

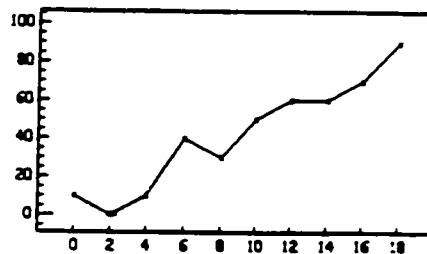
Single-Exemplar Training

(p) D i f f e r e n t

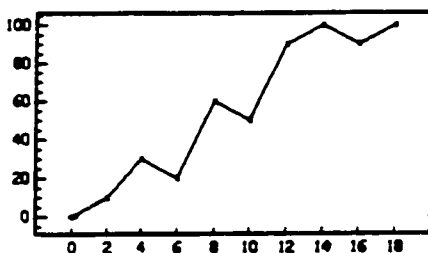
Participant 146
Slope .271



Participant 154
Slope .161



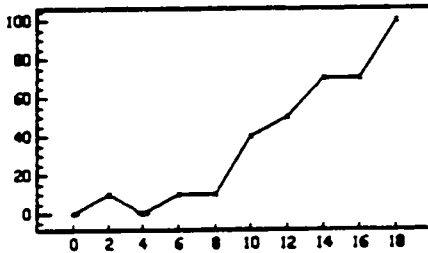
Participant 123
Slope .248



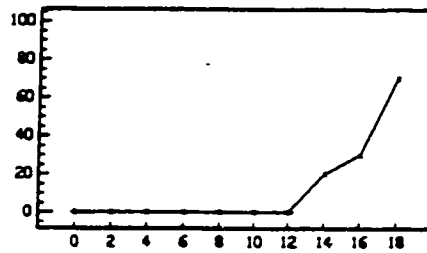
Participant 144
Slope .149



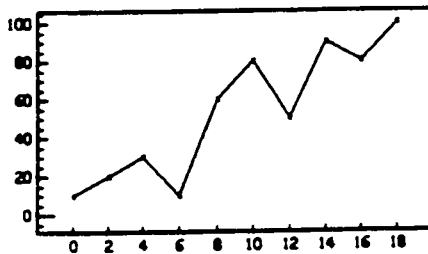
Participant 158
Slope .229



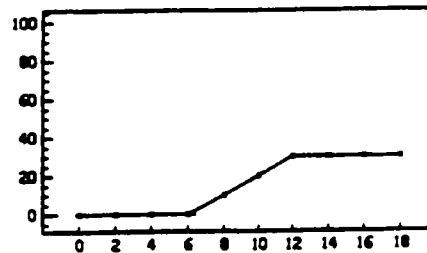
Participant 306
Slope .144



Participant 153
Slope .177



Participant 143
Slope .139



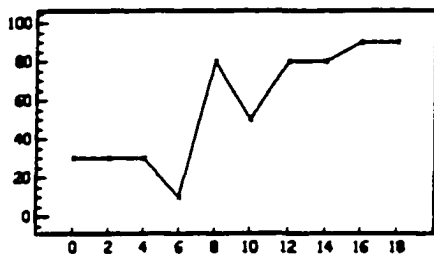
Stimulus Slope

→ 3 SD Stimuli

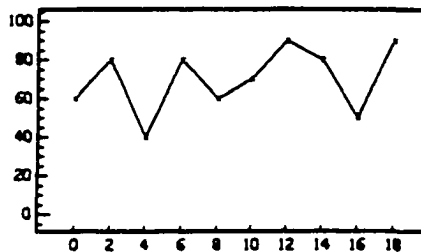
Single-Exemplar Training

D i f f e r e n t
 (p)

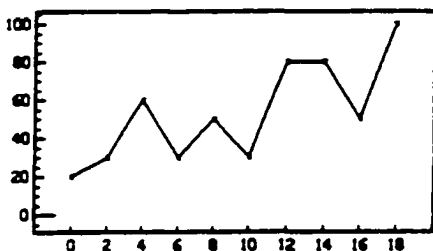
Participant 308
Slope .125



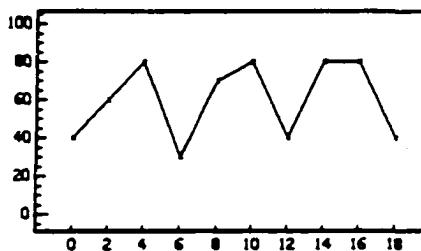
Participant 301 *
Slope .032



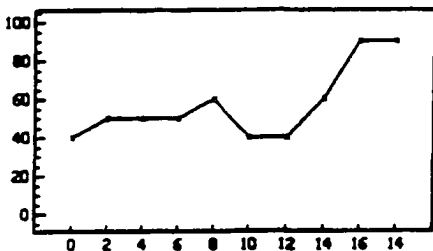
Participant 147
Slope .119



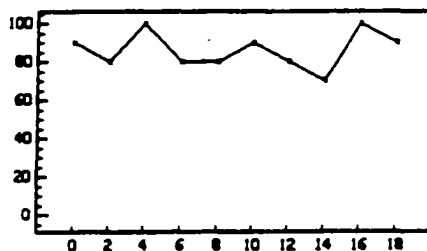
Participant 311 *
Slope .016



Participant 157 *
Slope .069



Participant 149 *
Slope -.021

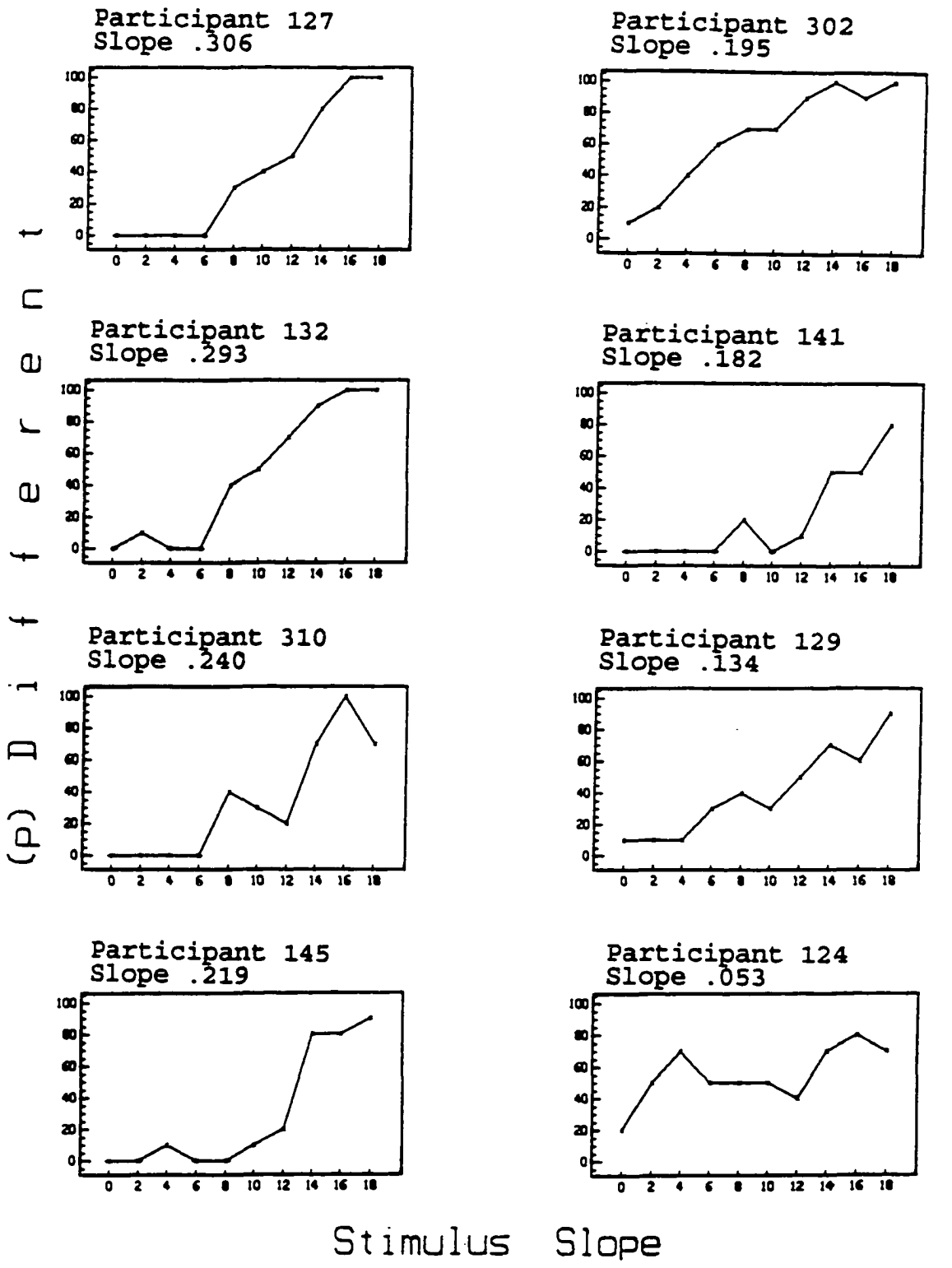


Stimulus Slope

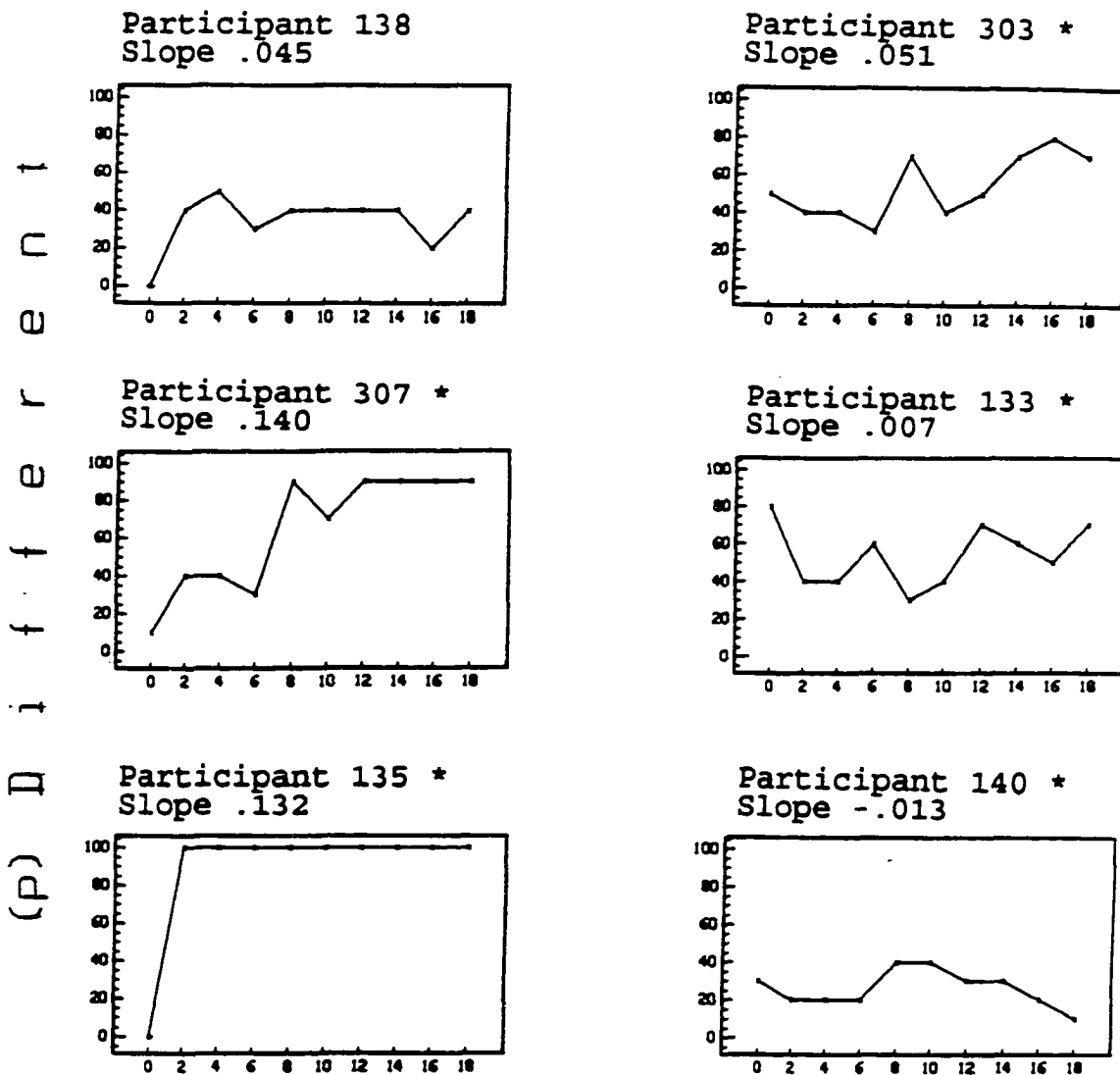
→ 3 SD Stimuli

Figure 16. Generalization gradients for individual participants in the single-exemplar overtraining group for Experiment 2. The probability of the DIFFERENT response during generalization tests to 3 SD stimulus slopes is shown. The participant number and the generalization gradient slope are shown above each graph. Participants not in the subset (those that demonstrated generalization gradient slopes ≤ 0 to 3 SD or 6 SD stimulus slopes) are indicated by an asterisk.

Single-Exemplar Overtraining



Single-Exemplar Overtraining

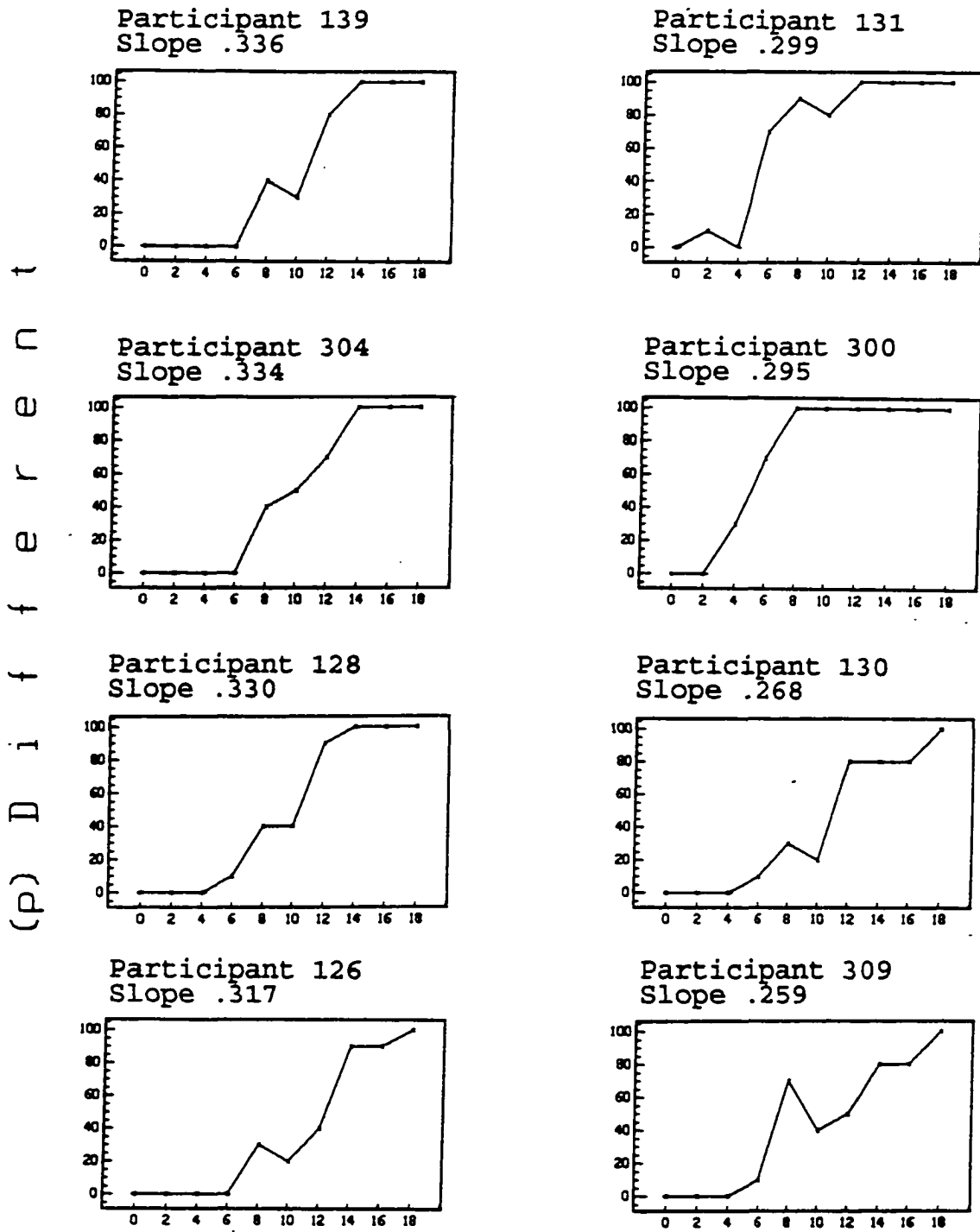


Stimulus Slope

← 3 SD Stimuli

Figure 17. Generalization gradients for individual participants in the multiple-exemplar training group for Experiment 2. The probability of the DIFFERENT response during generalization tests to 3 SD stimulus slopes is shown. The participant number and generalization gradient slope are shown above each graph.

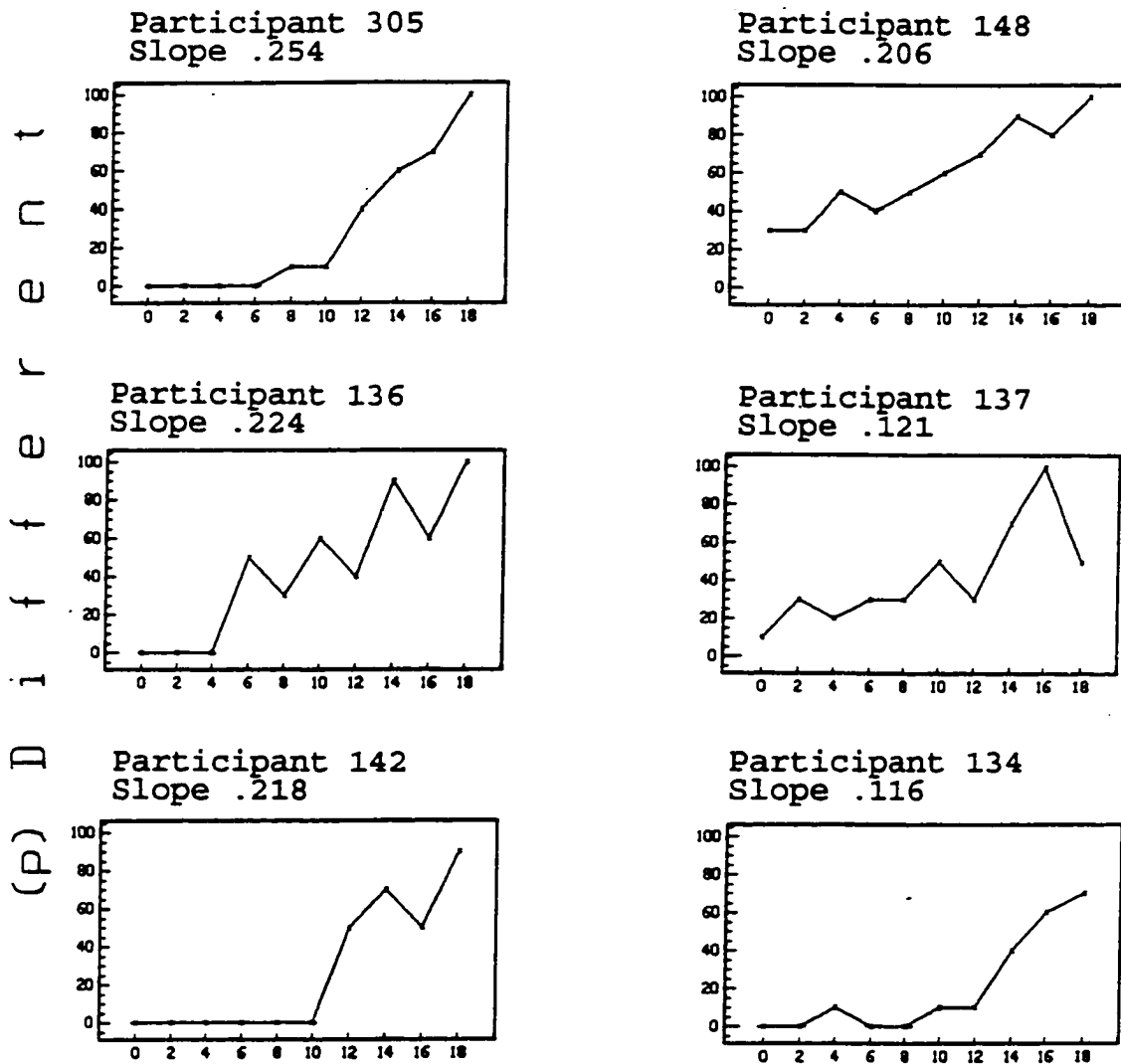
Multple-Exemplar Training



Stimulus Slope

← 3 SD Stimuli

Multiple-Exemplar Training

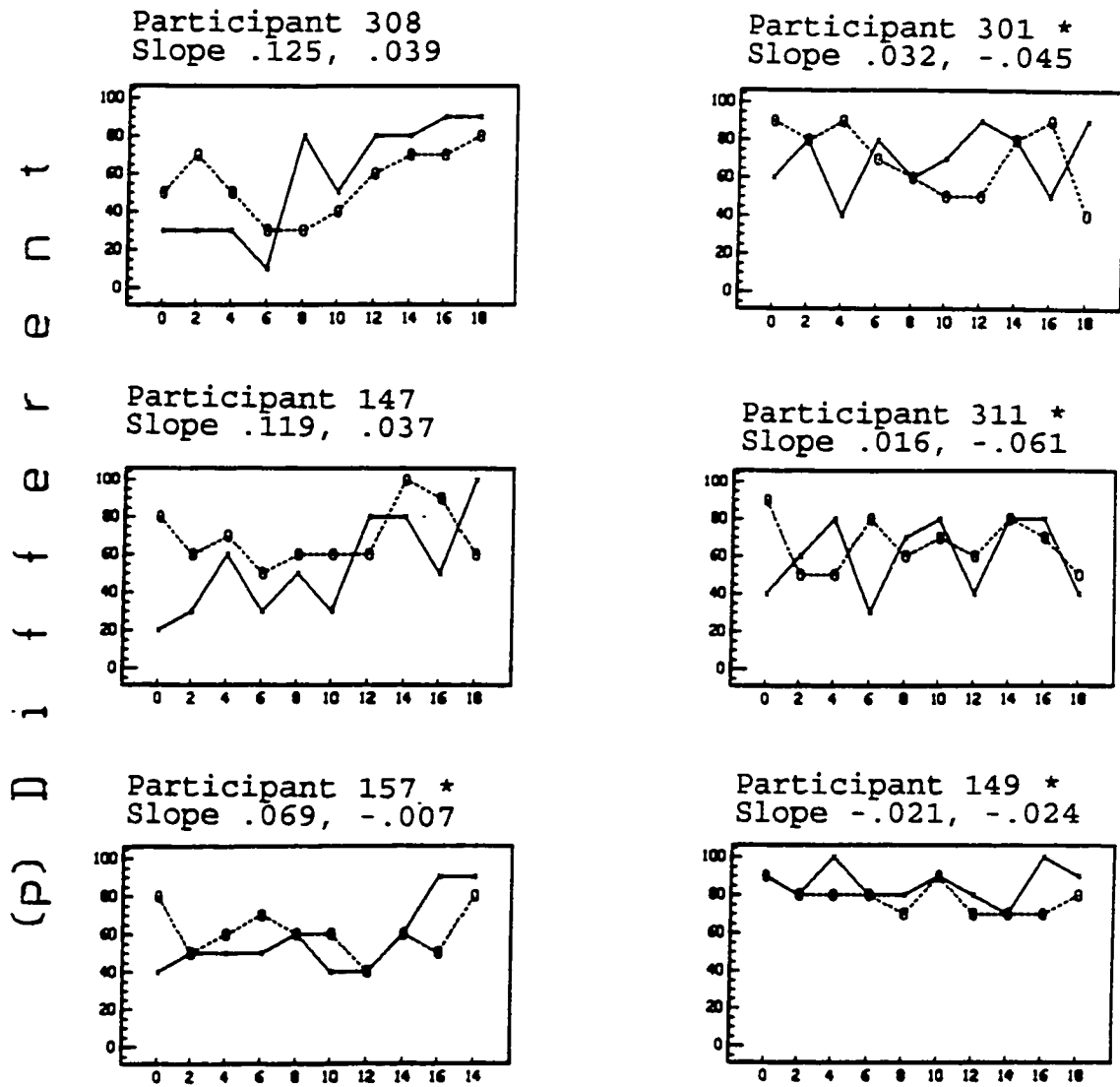


Stimulus Slope

← 3 SD Stimuli

Figure 18. Generalization gradients for individual participants in the single-exemplar training group for Experiment 2. The probability of the DIFFERENT response during generalization tests to 3 SD stimulus slopes (dots) and 6 SD stimulus slopes (circles) is shown. The participant number and generalization gradient slopes 3 SD and 6 SD stimulus slopes are shown above each graph respectively. Participants not in the subset (those who demonstrated generalization gradient slopes ≤ 0 to 3 SD or 6 SD stimulus slopes) are indicated by an asterisk.

Single-Exemplar Training



Stimulus Slope

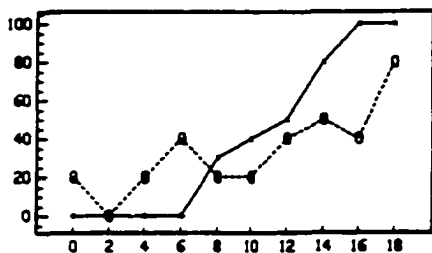
—●— 3 SD Stimuli
 ○---○ 6 SD Stimuli

Figure 19. Generalization gradients for individual participants in the single-exemplar overtraining group for Experiment 2. The probability of the DIFFERENT response during generalization tests to 3 SD stimulus slopes (dots) and 6 SD stimulus slopes (circles) is shown. The participant number and generalization gradient slopes to 3 SD and 6 SD stimulus slopes are shown above each graph respectively. Participants not in the subset (those who demonstrated generalization gradient slopes ≤ 0 to 3 SD or 6 SD stimulus slopes) are indicated by an asterisk.

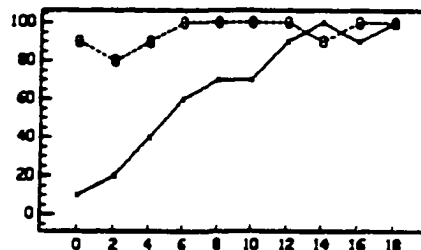
Single-Exemplar Overtraining

(P) D i f f e r e n t

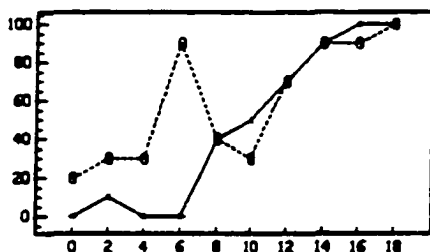
Participant 127
Slope .306, .104



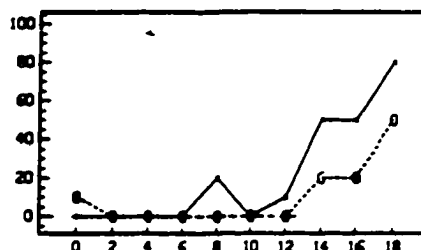
Participant 302
Slope .195, .064



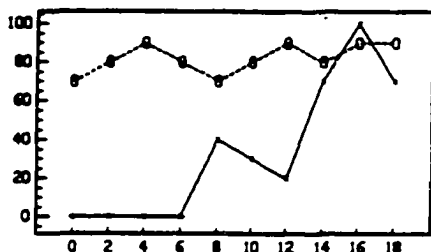
Participant 132
Slope .293, .146



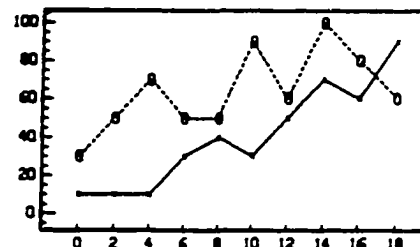
Participant 141
Slope .182, .102



Participant 310
Slope .240, .028



Participant 129
Slope .134, .074



Participant 145
Slope .219, .099



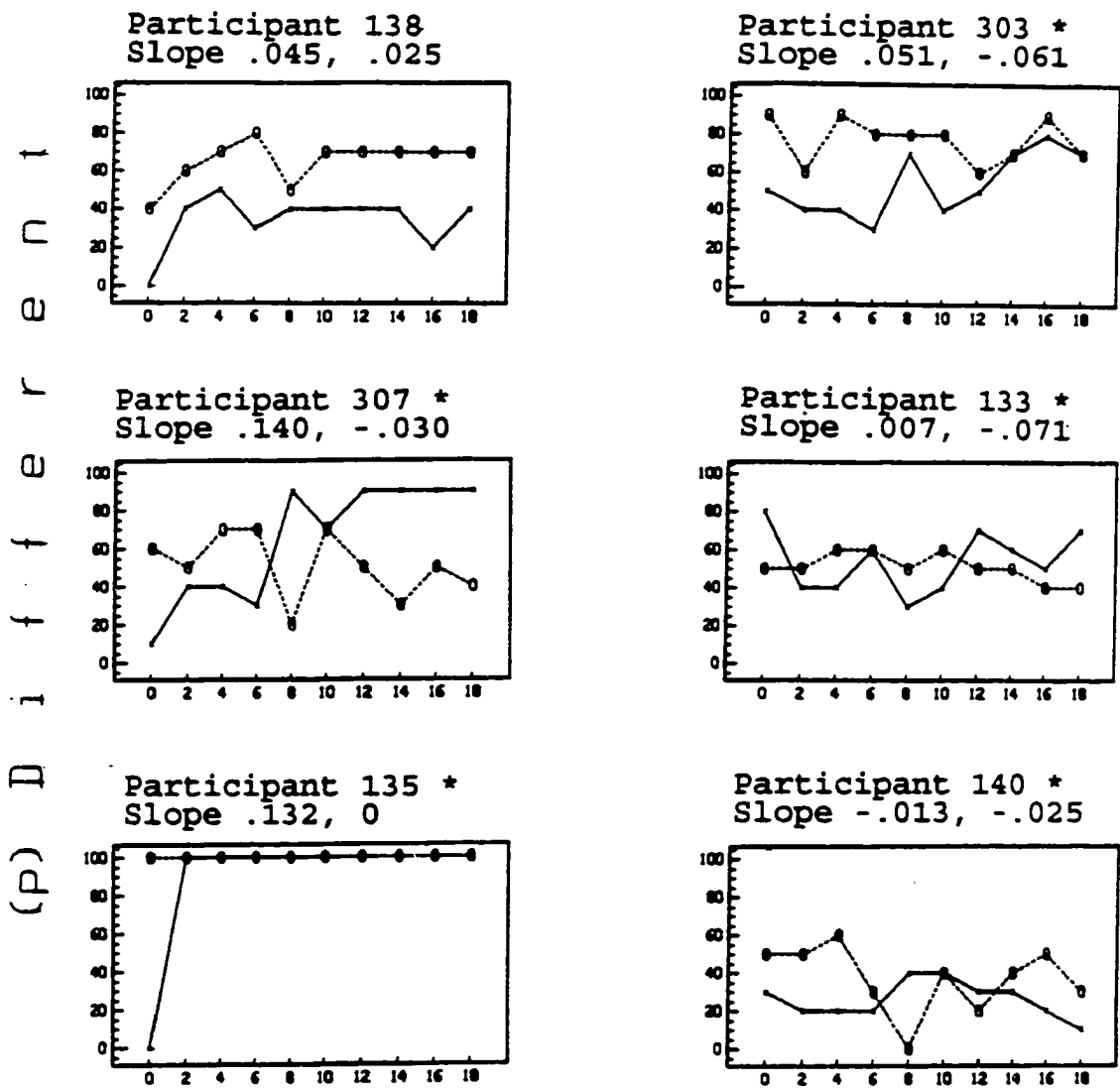
Participant 124
Slope .053, .026



Stimulus Slope

—●— 3 SD Stimuli
 -○- - - 6 SD Stimuli

Single-Exemplar Overtraining

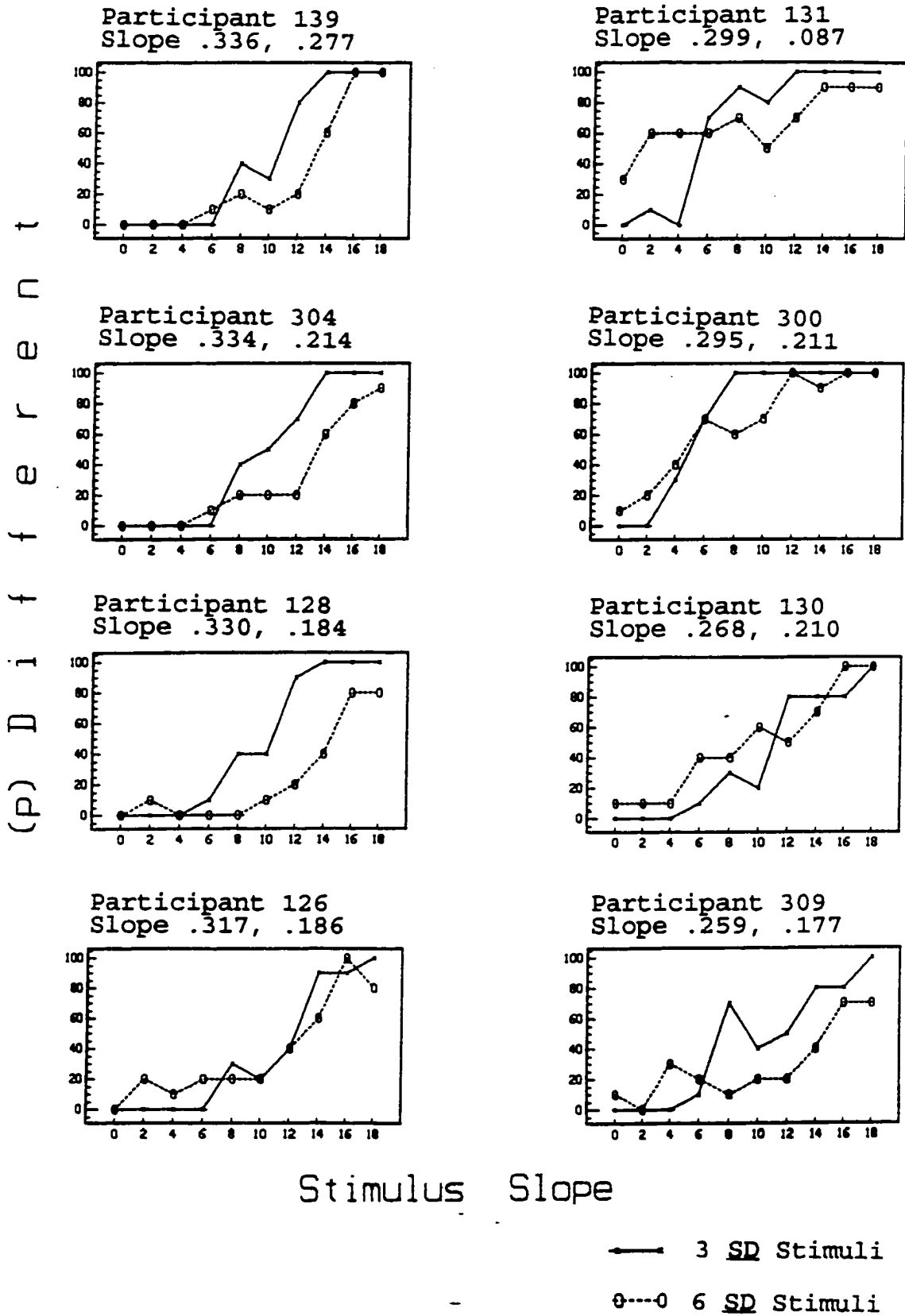


Stimulus Slope

- 3 SD Stimuli
- - 6 SD Stimuli

Figure 20. Generalization gradients for individual participants in the multiple-exemplar training group for Experiment 2. The probability of the DIFFERENT response during generalization tests to 3 SD stimulus slopes (dots) and 6 SD stimulus slopes (circles) is shown. The participant number and generalization gradient slopes to 3 SD and 6 SD stimulus slopes are shown above each graph respectively.

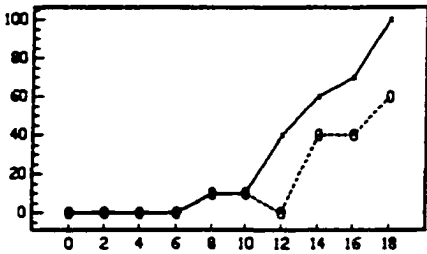
Multiple-Exemplar Training



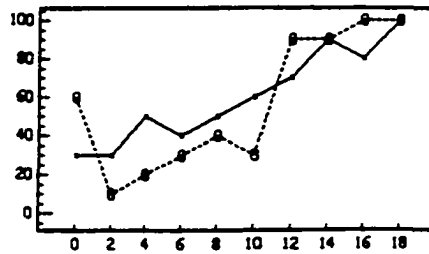
Multiple-Exemplar Training

(p) D i f f e r e n t

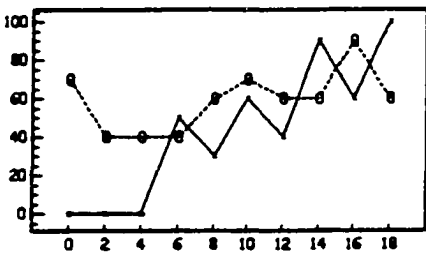
Participant 305
Slope .254, .151



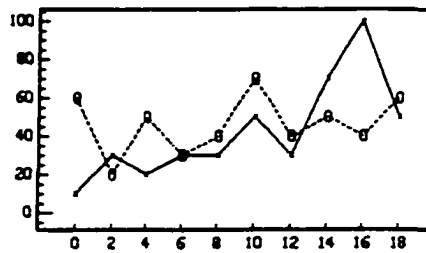
Participant 148
Slope .206, .115



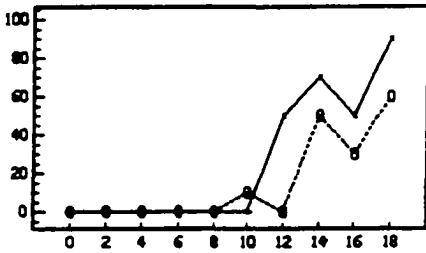
Participant 136
Slope .224, .038



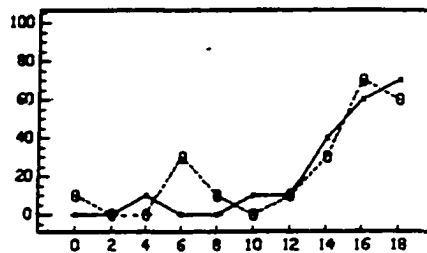
Participant 137
Slope .121, .017



Participant 142
Slope .218, .153



Participant 134
Slope .116, .112



Stimulus Slope

—●— 3 SD Stimuli
 - - - ○ - - - 6 SD Stimuli

of the data reveal generalization gradient slope differences under two levels of stimulus variability. For each subset participant, the slope of the generalization gradient in the 3 SD condition was steeper than the slope in the 6 SD condition.

These differences, however, can be seen more clearly by comparing the quantified, generalization gradient slope values for all subset participants. These generalization gradient slope values are listed in Table 18 and the corresponding r^2 values are listed in Table 19. For each subset participant, the value of the generalization gradient slope for 3 SD stimulus slopes was steeper than that attained for 6 SD stimulus slopes. An 2 x 3 ANOVA (Variability of Test Stimuli, [3 SD and 6 SD] x Training Group) with repeated measures confirmed that these generalization gradient slope differences were significant ($F[1, 39] = 162.26, p < .0001$), as shown in Tables 20. Post-hoc, pairwise-comparisons (Fisher's LSD) showed that the differences in the generalization gradient slope to 3 SD vs. 6 SD stimulus slopes were significant for the single-exemplar training group ($t = 5.51, p < .01$), the single-exemplar overtraining group ($t = 6.46, p < .01$) and the multiple-exemplar training group ($t = 7.49, p < .01$) as shown in Table 21.

Effects of Training Regimen on Generalization Gradient Slopes

Differential effects of training regimen on the generalization gradient slopes obtained with 3 SD and 6 SD stimulus slopes are illustrated in Figure 21 for subset participants. Each vertical line illustrates plus or minus

Table 19

Proportion of Variance Accounted for (r^2) of Slopes of Generalization Gradients For Participants (Part.) with Generalization Slopes Values > 0 to 3 SD and 6 SD Stimuli for Three Training Groups (SET, SEOT, MET) for Experiment 2.

SET			SEOT			MET		
Part.	3 SD	6 SD	Part.	3 SD	6 SD	Part.	3 SD	6 SD
147	.50	.02	138	.05	.15	134	.74	.41
308	.66	.15	124	.36	.04	137	.43	.05
143	.83	.50	129	.89	.20	148	.81	.62
306	.60	.67	141	.73	.42	142	.73	.65
144	.81	.69	302	.89	.27	136	.80	.15
154	.82	.78	145	.76	.56	305	.86	.69
153	.78	.50	310	.75	.23	309	.88	.61
158	.85	.71	132	.87	.61	130	.92	.85
123	.89	.25	127	.91	.52	300	.73	.87
146	.86	.24				131	.81	.73
						126	.92	.71
						128	.93	.70
						304	.90	.93
						139	.89	.83

Table 20

Analysis of Variance (ANOVA) of Generalization Slopes for Participants with Generalization Slope Values > 0 to 3 SD and 6 SD Stimuli. The Effects of Training Group (T) and Variability of Test Stimuli (V) are Shown for Experiment 2.

<u>Source</u>	<u>df</u>	<u>MS</u>	<u>F</u>	<u>p</u>
<u>Between</u>				
Training (T)	2	0.0503	9.59	0.0004
Sub. w. Grp.	39	0.0052		
<u>Within</u>				
Var. Stim. (V)	1	0.2158	162.26	< 0.0001
T x V	2	0.0008	0.60	0.5542
V x Sub. w. Grp.	39	0.0013		

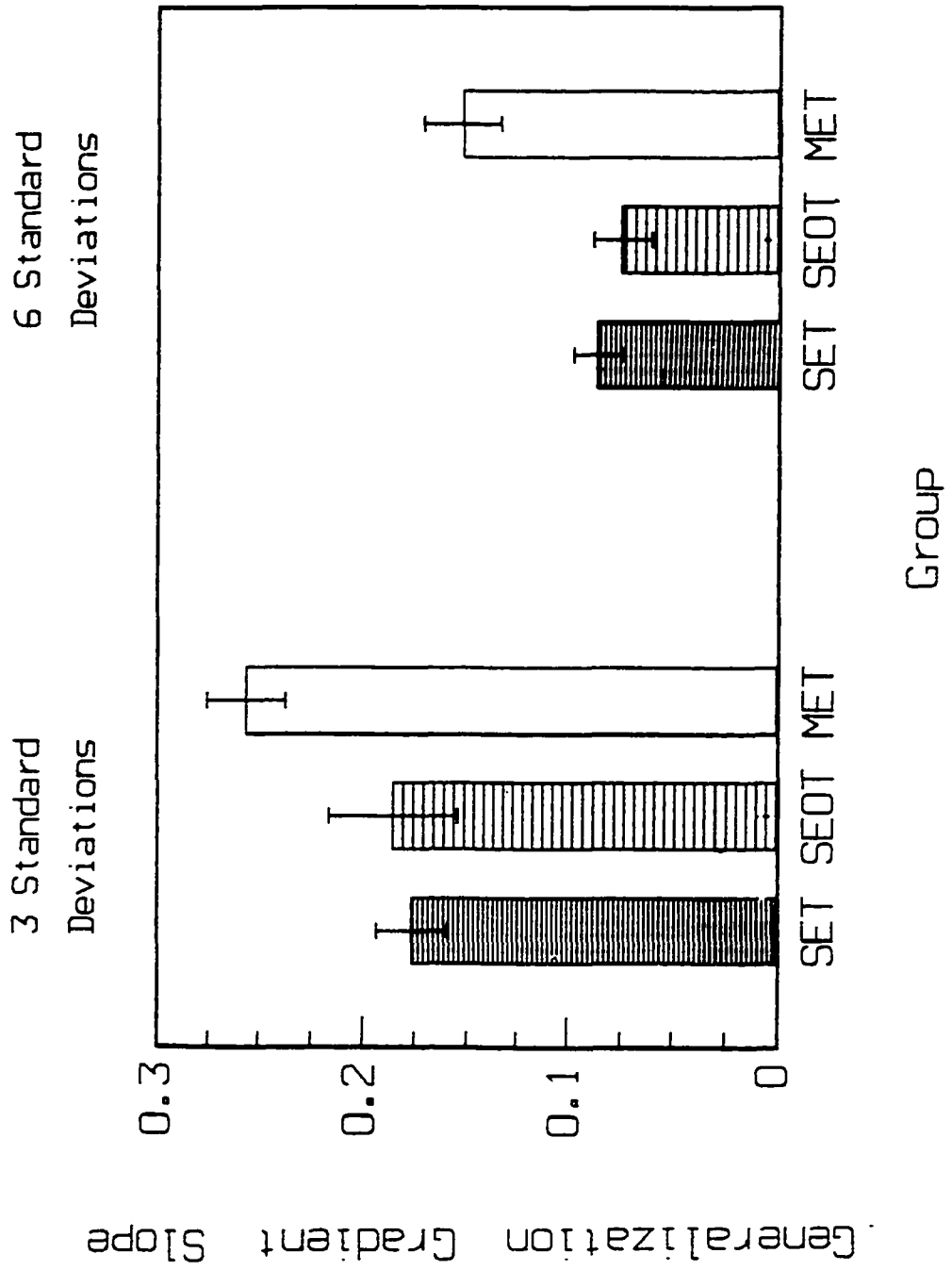
Table 21

Post-Hoc, Pairwise-Comparisons of Generalization Slopes for Participants with Slopes Values > 0 in the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET) and Variability of Test Stimuli (V) (3 SD and 6 SD) for Experiment 2.

		3 SD			6 SD		
		SET	SEOT	MET	SET	SEOT	MET
3 SD	SET						
	SEOT	0.54					
	MET	5.25**	4.51**				
6 SD	SET	5.51**					
	SEOT		6.46**		0.72		
	MET			7.49**	4.37**	5.01**	

* p < .05
** p < .01

Figure 21. Mean generalization gradients slopes for the single-exemplar training group (SET), the single-exemplar overtraining group (SEOT), and the multiple-exemplar training group (MET) to 3 SD and 6 SD stimulus slopes for Experiment 2. Only participants in the subset are included (i.e., participants with generalization gradient slopes > 0 to 3 SD and 6 SD slope stimuli). The vertical line represents ± 1 SE.



one standard error.

For 3 SD stimuli, generalization gradient slopes were steepest after multiple-exemplar training. The generalization gradient slopes after single-exemplar training and single-exemplar overtraining were similar, and were both shallower than those after multiple-exemplar training. An 2 x 3 ANOVA (Variability of Test Stimuli, [3 SD and 6 SD] x Training Group) with repeated measures of generalization gradient slope showed a significant effects of training group, $F(2, 39) = 9.59$, $p < .001$ (see Table 19). Post-hoc, pairwise-comparisons showed significant differences between the generalization gradient slopes obtained after multiple-exemplar training and single-exemplar training ($t = 5.25$, $p < .01$), or multiple-exemplar training and single-exemplar overtraining ($t = 4.51$, $p < .01$) (see Table 21). No significant differences were found between the generalization gradient slopes after single-exemplar training and the single-exemplar overtraining.

For 6 SD stimuli, the generalization gradient slopes were relatively steeper after multiple-exemplar training than after single-exemplar training and the single-exemplar overtraining. An 2 x 3 ANOVA (Variability of Test Stimuli, [3 SD and 6 SD] x Training Group) with repeated measures showed a significant effect of training group ($F[2, 39] = 9.59$, $p < .001$) (see Table 20). The post-hoc, pairwise-comparisons (Fisher's LSD) showed significant differences between the generalization gradient slopes after multiple-exemplar training and single-exemplar training ($t = 4.37$, $p < .01$), and multiple-exemplar training and single-exemplar overtraining ($t = 5.01$, $p < .01$), as shown in Table 21.

No significant differences were found between the generalization gradient slopes after single-exemplar training and the single-exemplar overtraining.

Summary. Regardless of training regimen, subset participants showed significantly steeper generalization slopes obtained in the 3 SD condition than those obtained in the 6 SD condition. Subset participants who received multiple-exemplar training showed significantly steeper generalization gradient slopes than those who received single-exemplar training or single-exemplar overtraining.

Discussion

Purposes of Experiment 2

The purposes of Experiment 2 were to determine how three training regimens influenced:

1. Acquisition of a baseline discrimination between graphs, called the reference stimuli (0° , 18°) in low variability form (3 SD).
2. Maintenance of the baseline discrimination between the reference stimuli (0° , 18°) in low variability form (3 SD).
3. Generalization to graphs identical in slope to the reference stimuli (0° , 18°) but in a higher variability form (6 SD).
4. Establishment of dimensional control by stimuli intermediate (2° , 4° , . . . 16°) in slope to the reference stimuli in low variability form (3 SD).
5. Establishment of dimensional control to stimuli intermediate (2° , 4° , . . . 16°) in slope to the reference stimuli, in higher variability form (6 SD).

Dependent measures. The performance of participants in Experiment 2 was chiefly measured in two ways: (a) performance to the reference stimuli (e.g., DI, d' , β), and (b) the generalization gradient slopes. The discrimination indices measured performance to the extreme stimulus slopes while the generalization gradient slopes chiefly measured performance to the intermediate stimulus slopes. Outcomes 1, 2, 3 were assessed by the discrimination index, d' , and β , while outcomes 4 and 5 were assessed by the generalization gradient slopes.

Summary of outcome. Training was effective for participants in all groups to establish a discrimination between the reference stimuli. Performance during generalization testing differed, however.

During generalization tests, maintenance of baseline discrimination of the reference stimuli in low-variability form, was better after multiple-exemplar training than after single-exemplar training or single-exemplar overtraining. Little difference was seen in the DI or d' between single-exemplar training, and single-exemplar overtraining.

Maintenance of discriminative control by graphs, identical to the reference stimuli but in the higher-variability form, was better after multiple-exemplar training than after single-exemplar training or single-exemplar overtraining, as measured by the DI and d' .

Establishment of dimensional control by intermediate stimulus slopes, in both low-variability form and high-variability form, was better after multiple-exemplar training than after single-exemplar training or single-

exemplar overtraining. There was a small increase in response bias (β) to more-variable stimuli for the single-exemplar training group and the multiple-exemplar training group. There was little difference in the generalization gradient slopes after single-exemplar training or single-exemplar overtraining.

The results of Experiment 2, represented by the discrimination indices, d' 's and generalization gradient slopes provide convergent evidence of enhanced stimulus control following training with multiple exemplars.

Separating the effects of variety of exemplars and number of training trials. Although all participants acquired the reference discrimination during training, the differential performance during generalization tests suggested that different characteristics of stimuli gained control of responding under each training regimen.

Multiple-exemplar training resulted in better maintenance of stimulus control to reference stimuli and better dimensional control to intermediate stimuli than single-exemplar training. This effect could be attributed to two factors that differed between single-exemplar training and multiple-exemplar training (a) variety of exemplars, and (b) numbers of trials.

Participants in the single-exemplar training group received training with only one S+ and one S-, while those in the multiple-exemplar training group received training with 48 S+'s and 48 S-'s. To reach the same mastery criterion in both groups, participants in the multiple-exemplar training group received, on average, 11.7 times the number of trials as those in the single-exemplar

training group. Therefore, the superior performance of the multiple-exemplar training group could be attributed to the increase in the variety of exemplars, the increase in the number of training trials, or both.

To separate the possible affects of these two factors single-exemplar overtraining was conducted as a control procedure. Participants in the single-exemplar overtraining group received as many trials with single-exemplars as those in the multiple-exemplar training group. Equating the number of trials used for training isolated the effects of variety of training exemplars on performance during generalization tests. Since maintenance of stimulus control demonstrated after multiple-exemplar training was superior to that observed after single-exemplar overtraining, this difference was attributable to training with the variety of exemplars and not to the large number of training trials.

The superior dimensional stimulus control induced by multiple-exemplar training compared to single-exemplar overtraining was also seen by examining the generalization gradient slopes. After multiple-exemplar training, participants showed steeper generalization gradient slopes to the intermediate stimuli than after single-exemplar overtraining. Therefore the establishment of dimensional control to intermediate stimuli resulted from prior training with a variety of exemplars rather than the large number of training trials.

Effect of overtraining by single exemplars. The effect of overtraining with single exemplars can be assessed by comparing the single-exemplar training group with the single-exemplar overtraining group. After training with

these two regimens, the discrimination index and d' to 3 SD stimuli and also to 6 SD stimuli during generalization testing were comparable. In addition, the slopes of the generalization gradients, were similar after single-exemplar training and single-exemplar overtraining. Thus, there was no apparent benefit of overtraining with single-exemplars as measured by the discrimination index and d' .

Single-exemplar overtraining did, however, have an effect on bias. On tests with the more-variable 0° and 18° stimuli presented in 6 SD form, bias increased for both the single-exemplar training group and the multiple-exemplar training group, but not for the single-exemplar overtraining group. The single-exemplar overtraining group differed from the other two groups by one factor--the number of trials per stimulus during training. Participants in the single-exemplar overtraining group received, on average, 11.7 times the number of trials per stimulus as participants in the other two groups. This repeated training with a limited stimulus set (1 S+ and 1 S-) was responsible for keeping bias at low levels for the single-exemplar training group to more variable stimuli (0° and 18° , 6 SD). This analysis suggests that overtraining with multiple exemplars would result in reduced bias to more variable stimuli.

Theoretical analysis. In these experiments, the graphs were complex stimuli, varying in two dimensions simultaneously (a) the location of the data points, and (b) the stimulus slope. During training, both of these components were predictive of reinforcement. Training with different regimes may have resulted in changes in the

discriminative control acquired by each of these components of the discriminanda (Dinsmoor, 1995; Schwartz, 1989).

Variation in the degree of control acquired by the components of compound stimuli was first reported by Pavlov (1927). After a discrimination between two compound stimuli was formed, he reported that one component of the compound stimuli failed to elicit a conditioned response. Pavlov said that the stronger component "overshadowed" the weaker component.

Using a Pavlovian paradigm, Kamin (1968) trained groups of rats with a compound stimulus of two components, light and noise. Subsequent testing showed some control by each component. In contrast, pretraining with one component, before training with both components in compound, resulted in control by the pretrained component but "blocking" of control of the added element of the compound (Vom Saal and Jenkins, 1970). Kamin's study showed how pretraining with one component can prevent control by a second component, when both are subsequently trained in compound.

Reynolds (1961) trained pigeons with a compound-cue discrimination between a S+ (a white triangle with a red background) and a S- (a white circle with a green background). During testing in extinction with each of these four components, one pigeon responded almost entirely to the red test stimulus while another responded almost entirely to the white triangle. Therefore, participants can come under control of different components of a compound stimulus even after identical training.

Similar results were reported by Newman and Baron (1965) who trained pigeons to peck a key with a white vertical line on a green background. Subsequent tests by line-tilt showed flat generalization gradients. They concluded the birds were under stimulus control of the green background and failed to attend to the slope of the line presented during training.

Other researchers found stimulus control by both components of compound stimuli (Butter, 1963; Chase, 1968). Chase and Heinemann (1972) summarized previous findings and concluded that two factors determined the relative amount of stimulus control acquired by each component of a compound stimuli: (a) prior training with a component, and (b) relative discriminabilities (or intensities) of pairs of values of each of the two stimulus components. An additional factor is "predictiveness of reinforcement" from trial to trial by a stimulus component (Schwartz, 1989). Components of compound stimuli that best predict reinforcement come to control responding (Newman and Baron, 1965). If both components are equally predictive of reinforcement, the one that comes to control responding during testing is said to be more salient. The more salient component is said to overshadow the less salient component.

In the present study, the degree of control acquired by the stimulus slope varied with training regime (Fields, 1978).

During multiple-exemplar training, the 0° and 18° stimuli were presented with many different configurations of data points defining these two data paths. With 96 such comparable exemplars, the location of data points was quite

variable, and therefore may have become less predictive of reinforcement. In contrast, the stimulus slope remained constant across stimulus presentations and therefore highly predictive of reinforcement. Essentially the stimulus slope was a better predictor of reinforcement relative to the configuration of data points for those trained with multiple-exemplars.

Conversely, during single-exemplar overtraining, the 0° and 18° stimuli were each represented by only one configuration of the data points. Therefore the location of one or one data points was highly predictive of reinforcement as was the stimulus slope. The location of one or more data points may have been easier to detect, and therefore more salient than the stimulus slope.

A alternate interpretation suggests that participants may have come under the absolute property of the compound stimuli (i.e., the location of data points) while others came under the relative property of the compound stimulus (i.e., the stimulus slope). Wright, Cook, Rivera, Sands, and Delius (1988) posited that the type of control (absolute or relational) depends upon if the test stimuli were repeated or new. In a series of experiments, Wright, Cook, and Kendrick (1989) found that monkeys came under stimulus control of individual probe items when the same stimuli were presented repeatedly. These same participants demonstrated relational control, however, when transferred to sets of new stimuli. The authors concluded when the task does not require participants to deal with new stimuli, an absolute strategy may be best, since it requires less cognitive effort (for humans). When the task does require

that participants (nonhuman and human) deal with new stimuli frequently, a relational strategy is best suited. In summary, the data would suggest that after single-exemplar training, the location of the data points was more salient than the stimulus slope and thus was a better predictor of reinforcement. After multiple-exemplar training, the stimulus slope was more salient and thus best predicted reinforcement (Lovejoy and Russell, 1967; Schwartz, 1989).

Impact of theory on discrimination acquisition.

Participants mastered the discrimination between the reference stimuli in all three training regimens. Since these were compound stimuli the relative influence of each component, (the location of data points and the stimulus slope) was unknown in the current study. The assumption that different parameters of the stimuli gained control is supported by outcomes of the maintenance of the trained discrimination, generalization to comparable exemplars, and the demonstration of dimensional control. Therefore the two main measures during generalization tests (a) performance to 0° and 18° stimulus slopes in 3 SD and 6 SD form, and (b) generalization to intermediate stimulus slopes 2° , 4° , 8° , 16° in 3 SD and 6 SD form can be examined to reveal different sources of stimulus control.

Performance to 0° and 18° stimuli, 3 SD during generalization tests. Maintenance of control by the reference stimuli (0° and 18° , 3 SD) was good after multiple-exemplar training. A considerable decrement in performance was seen during testing that followed single-exemplar training or single-exemplar overtraining.

This disruption in performance to the reference stimuli during generalization testing is readily explained by the differential acquisition of stimulus control by specific dot pattern and slope during each of the training regimes. After single-exemplar training and single-exemplar behavior overtraining, the position of one or more data points acquired more control than stimulus slope. Conversely, after multiple-exemplar training, more stimulus control was acquired by the stimulus slope than the position of one or more data points, as the stimulus slope best predicted reinforcement (McIlvane, Serna, Dube, & Stromer, in press).

This is not a complete explanation, however. In all training regimes, the reference stimuli were presented during both training and testing phases. If responding was controlled by the location of data points, as posited for the single-exemplar training and single-exemplar overtraining group, why should the reference stimuli exert such poor maintenance of control during generalization tests?

During generalization tests, participants were presented with 180 configurations of data points all of which differed from the 0° and 18° reference stimuli. For the participants discriminating on the basis of location of data points, these novel configurations of data points may have been distractors which interfered with the discrimination of the reference stimuli.

Disruption of a trained discrimination, by the introduction of test stimuli, has been reported previously. Heinemann, et al. (1969) trained pigeons in a noise-

intensity discrimination and found substantial deterioration of stimulus control to the training stimuli during generalization testing. Fields et al. (1995, 1997) reported disruption of a conditional discrimination when many new emergent-relations probes were introduced concurrently. A similar disruption may have occurred for participants in the current experiment who were trained or overtrained with single exemplars.

During training, participants presented with multiple-exemplars, however, may have come to disregard the location of one or more data points in favor of the stimulus slopes. Therefore during testing, variation in location of data points would not be distracting for those trained with multiple-exemplars.

Performance to 0° and 18° stimuli, 6 SD. Better generalization to the 0° and 18° stimuli in the more variable 6 SD form was seen after multiple-exemplar training than after single-exemplar training and single-exemplar overtraining. There was no systematic relationship between the location of one or more data points from the stimuli presented in training to those presented during testing. Participants under stimulus control by the location of specific data points would be expected to show poor generalization to stimuli in which the location of data points varied even more than those during training. Such poor generalization was seen following single-exemplar training and also single-exemplar overtraining. Participants under stimulus control by stimulus slope would show good generalization to these more variable stimuli, since these were the same stimulus slopes of the data paths

(0° and 18°) presented during training. Good generalization to these more-variable, intermediate stimuli was found after multiple-exemplar training.

Performance to 2°, 4° 16° stimuli, 3 SD. The generalization tests presented to all participants were intradimensional since they included stimuli intermediate (i.e., 2°, 4°, 16°) to those presented during training (Guttman and Kalish, 1956; Honig, Boneau, Burstein, and Pennypacker, 1963). These stimuli were presented in low variability (3 SD) form. All stimuli in the intradimensional generalization consisted of stimulus slopes and data point configurations never before presented to participants. Essentially it tested for systematic responding as a function of the slope of the data path. Such a test was designed to reveal sources of stimulus control after the three training regimens. Participants under stimulus control by the particular location of one or more data points would do poorly in this test, since there is no systematic relationship between data point location in the training and testing stimuli. Conversely, participants under control by stimulus slope would do well in this test since they could generalize from the discrimination acquired during training--by stimulus slope--to the new stimulus slopes during generalization testing. Participants in the single-exemplar training group and the single-exemplar overtraining group showed poorer dimensional control to these intermediate stimulus slopes than those in the multiple-exemplar training group. This supports the view that participants in the single-exemplar training group and the single-exemplar overtraining group

were under greater stimulus control by the position of individual data points while participants in the multiple-exemplar training group were under greater control by the stimulus slopes.

Performance to 2°, 4° . . . 16°, 6 SD. During the intradimensional generalization test, participants under stimulus control by the location of data points would be expected to show poor generalization to the more varied (6 SD) intermediate stimuli. First, there was no systematic relationship between the position of individual data points during training and those presented during testing. Second, the data points in these intermediate stimuli were dispersed even more than those presented during training, making generalization by the position of individual data points from training to testing stimuli, that much more difficult.

Participants under control by stimulus slope during training, however, would be expected to show good generalization of control--by stimulus slope--to these new, intermediate test slopes. Some decrement in performance by to these "stimulus slope discriminators" would also be expected due to the increased difficulty in discriminating the slope of the data path from a more "noisy" data set.

In summary, the results of Experiment 2 favor control by data configuration following single-exemplar training and single-exemplar overtraining, but control by the slope of the data path after multiple-exemplar training.

General Discussion

Test Differences Between Experiment 1 and 2 for Participants Who Received Single-Exemplar Training.

In Experiment 1 participants were trained with the reference stimuli (0° , 18° , 3 SD). During generalization tests comparable exemplars (i.e., exemplars with the same slopes and variability but different configuration of data points) were presented. Stimulus control to these comparable exemplars was poor.

In Experiment 2, the reference stimuli were presented during single-exemplar training and again during generalization tests. Fair maintenance of stimulus control was seen to the reference stimuli during generalization tests.

Taken together, these results suggest that in both experiments, after single exemplar training, participants were under greater stimulus control by the location of data points than by slope. In Experiment 1, there was no systematic relationship between the location of data points in the training and testing stimuli, consequently generalization to the test stimuli was poor. Conversely, in Experiment 2, the same location of data points was used for training and testing, consequently there was fair maintenance of control seen to the reference stimuli during testing.

Alternate Sources of Stimulus Control

The previous section discussed sources of stimulus control by examining decrements in performance from training to testing. Different sources of stimulus control could also be present when performance is maintained from

training to testing.

The assumption of this study and the data show that performance is a function of the slope of the data path in panel B. However, control by the location of data points must also be considered. For adjacent points along the X-axis, there is the same variability of data points on the Y-axis for different stimulus slopes. It is possible that the absolute position of data points may control responding. Touchette (1969) trained boys, with mental retardation, in a line-tilt discrimination. Testing with probes (different segments of the training lines) revealed different sources of stimulus control across subjects. In short, although all participants were under control of the training stimulus, different aspects of the tilted lines may have gained control of responding.

Given the complexity and variability of the stimuli in the present study, it is more likely that the slope of the data path in the B panel or some segment of this data path controlled responding. Additional research would be needed to isolate the levels of stimulus control exerted by each of these stimulus parameters.

Responses to S+ and S-. During Experiments 1 and 2 one measure of a participant's performance was the discrimination index. This measure represented the percentage of S+ trials that occasioned the DIFFERENT response (correctly responding as DIFFERENT to the 18° stimuli) minus the percentage of S- trials that occasioned the DIFFERENT response (incorrectly responding as DIFFERENT to the 0° stimuli). High DIs resulted from a high percentage of DIFFERENT responses on S+ trials and a low

percentage of DIFFERENT responses on S- trials. Lower DIs were generally the result of a moderately lowered percentage of S+ trials occasioned the DIFFERENT response and a substantially elevated percentage of S- trials occasioned the DIFFERENT response. Participants, therefore, tended to respond to the slopes of data paths as different when, in fact, they were the same. This increase in the percentage of S- trials that occasion the DIFFERENT response (also call false alarms) was most pronounced when the variability of the test stimuli was high. Others researchers have also reported high false-alarm rates in the interpretation of data by visual inspection (Matyas & Greenwood, 1990).

A high percentage of false alarms have serious implications for making clinical decisions based solely on the interpretation of data presented graphically (Matyas & Greenwood, 1990). With a high percentage of false alarms, practitioners may incorrectly judge an intervention as effective when in fact it was not (a Type I error). This lack of conservatism in the interpretation of data presented graphically certainly bears continuing investigation for possible methods of remediation.

Prior reviews of the interpretation of graphs by visual analysis showed a high percentage of false alarms, a low percentage hits, low interobserver agreement, and difficulty in detecting changes in stimulus slope and level. High variability and serial dependency in a data set adversely affect interpretation of data by visual inspection.

Given these difficulties, explicit training in

interpretation of graphs by visual inspection could aid behavior analysis in detection the presence or absence of intervention effects. Such training, would include presentation of multiple exemplars of the particular dimension of interest. For example, first multiple-exemplar training with graphs to detect changes in slope could be implemented. Second, multiple exemplar training with graphs to detect changes in level could be trained. Third, training with both slope and level, in compound, could be trained using multiple exemplars. Finally, generalization testing with both dimensions would provide a basis to evaluate the training of each dimension. Initially, all this training could be performed with low-variability stimuli but the variability could be increased as the behavior analysts become more proficient in detecting these changes with low-variability stimuli.

The outcomes of the current study are in concordance with previous studies in the interpretation of graphic data by visual inspection. Differences in slopes between adjacent phases in graphs are difficult for observers to accurately identify. These difficulties increase as variability of the data increases. Similar results have been reported by other investigators (Greenspan & Fisch, 1992; Fisch & Schneider, 1993; Matyas & Greenwood, 1990).

In the present study multiple-exemplar training was shown to improve participants' discriminations of changes of slopes in line graphs when these stimulus slopes were presented in both low and high variability. Systematic responding to intermediate stimulus slopes was demonstrated best after multiple-exemplar training.

Appendix A

Slopes of Generalization Gradients For Experiments 2.

Participants with Slopes Values ≤ 0 to 3 SD or 6 SD Stimuli in the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET).

SET			SEOT			MET		
Part.	3 SD	6 SD	Part.	3 SD	6 SD	Part.	3 SD	6 SD
149	-.021	-.024	140	-.013	-.025			
311	.016	-.016	133	.007	-.017			
301	.032	-.045	303	.051	-.016		none	
157	.069	-.007	135	.132	.000			
			307	.140	-.030			

Appendix B

R² of Slopes of Generalization Gradients For Participants with Generalization Slopes Values \leq 0 to 3 SD or 6 SD Stimuli in the Single-Exemplar Training Group (SET), the Single-Exemplar Overtraining Group (SEOT), and the Multiple-Exemplar Training Group (MET).

SET			SEOT			MET		
Part.	3 SD	6 SD	Part.	3 SD	6 SD	Part.	3 SD	6 SD
149	.12	.15	140	-.06	.12			
311	-.09	-.07	133	-.11	.24			
301	.03	.12	303	.40	-.06	none		
157	.45	.11	135	.18	n/a			
			307	.74	.05			

References

- Barlow, D. H., & Hersen, M. (1984). Single case experimental design: Strategies for studying behavior change. Oxford, England: Pergamon Press.
- Bhatt, R. S., Wasserman, E. A., Reynolds, W. F. Jr., & Knauss, K. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel exemplars from four classes of natural and artificial stimuli. Journal of Experimental Psychology: Animal Behavior Processes, 14, 219-234.
- Butter, C. M. (1963). Stimulus generalization along one and two dimensions in pigeons. Journal of Experimental Psychology, 65, 339-346.
- Chase, S. (1968). Selectivity in multidimensional stimulus control. Journal of Comparative and Physiological Psychology, 66, 787-792.
- Chase S., & Heinemann E. G. (1972). Choice based on redundant information: An analysis of two-dimensional stimulus control. Journal of Experimental Psychology, 92, 161-175.
- Cleveland, W. S. (1993). Visualizing data. Summit, New Jersey: Hobart Press.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). Applied Behavior Analysis. Columbus Ohio: Merrill.
- DeProspero, A., & Cohen, S. (1979). Inconsistent visual analysis of intrasubject data. Journal of Applied Behavior Analysis, 12, 573-579.
- Dinsmoor, J. A. (1995). Tutorial, Stimulus control: Part II. The Behavior Analyst, 18, 253-269.
- Fields, L. (1978). Fading and errorless transfer in successive discriminations. Journal of the Experimental Analysis of Behavior, 30, 123-128.
- Fields, L., Landon-Jimenez, D. V., Buffington, D. M., & Adams, B. J. (1995). Maintained nodal-distance effects in equivalence classes. The Journal of the Experimental Analysis of Behavior, 64, 129-145.
- Fields, L., Reeve, K. F., Rosen, D., Varelas, A., Adams, B. J., Belanich, J., & Hobbie, S. A. (1997). Using the simultaneous protocol to study equivalent class formation: The facilitating effects of nodal number and size of previously established equivalence classes. Journal of the Experimental Analysis of Behavior, 67, 367-389.
- Fisch, G. S., & Schneider, R. (1993). Visual inspection of data: Placement of point-to-point graphs affects discrimination of trends and treatment effects. Proceedings of the American Statistical Association, proceedings on Statistical Graphics (in press).

Furlong, M. J., & Wampold, B. E. (1981). Visual analysis of single-subject studies in school psychology. Psychology in the School, 18, 80-86.

Furlong, M. J., & Wampold, B. E. (1982). Intervention effects and relative variations as dimensions in experts' use of visual inference. Journal of Applied Behavior Analysis, 15, 415-421.

Greenspan, P. & Fisch, G. S. (1992). Visual inspection of data: A statistical analysis of behavior. American Statistical Association 1992, proceedings Statistical Graph Section 79-82.

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. Journal of Experimental Psychology, 51, 79-88.

Hanson, H. (1959). Effects of discrimination training on stimulus generalization. Journal of Experimental Psychology, 58, 321-334.

Hearst, E., & Koresko, M. B. (1968). Stimulus generalization and amount of prior training on variable-interval reinforcement. Journal of Comparative and Physiological Psychology, 66, 133-138.

Heinemann, E. G., Avin, E., Sullivan, M. A. & Chase, S. (1969). Analysis of stimulus generalization with a psychophysical method. Journal of Experimental Psychology, 80, 215-224.

Herrnstein, R. J. (1979). Acquisition, generalization and discrimination reversal of a natural concept. Journal of Experimental Psychology: Animal Behavior Processes, 5, 116-129.

Herrnstein, R. J., & de Villiers, P. A. (1980). Fish as a natural category for people and pigeons. In G. H. Bower (Ed.) The psychology of learning and motivation, Vol. 14, pp. 59-95). New York: Academic Press.

Homa, D. (1978). Abstraction of ill-defined form. Journal of Experimental Psychology: Animal Learning and Memory, 4, 407-416.

Homa, D., Cross, J., Cornell, D., Goldman, D. & Schwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. Journal of Experimental Psychology, 101, 116-122.

Homa, D., & Little, J. (1995). The abstraction and long-term retention of ill-defined categories by children. Bulletin of the Psychonomic Society, 23, 325-328.

Homa, D., Sterling, S. & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. Journal of Experimental Psychology, 7, 418-439.

Homa, D. & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. Journal of Experimental Psychology: Human Learning and Memory, 2, 322-330.

Honig, W. K., Boneau, C.A., Burnstein, K.R., & Pennypacker, H.S. (1963). Positive and negative generalization gradients obtained after equivalent training conditions. Journal of Comparative and Physiological Psychology, 56, 111-116.

Jenkins, H. M., & Harrison, R. H. (1960). Effects of discrimination training on auditory generalization. Journal of Experimental Psychology, 59, 246-253.

Jones, R. R., Vaught, R. S., & Weinrott, M. (1977). Time-series analysis in operant research. Journal of Applied Behavior Analysis, 10, 151-166.

Jones, R. R., Weinrott M. R., & Vaught, R. S. (1978). Effects of serial dependency on the agreement between visual and statistical inference. Journal of Applied Behavior Analysis, 11, 277-283.

Kamin, L, J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.) Miami symposium on the prediction of behavior: Aversive stimulation. Miami, Florida: University of Miami.

Kazdin, A. E. (1982). Single-case research design: Methods for clinical and applied settings. New York: Oxford University Press.

Keller, F. S., & Schoenfeld, W. N. (1950). Principles of psychology. New York: Appleton-Century-Crofts.

Kratochwill, T., & Brody, G. (1978). Single subject designs: A perspective on the controversy over employing statistical inference and implications for research and training in behavior modification. Behavior Modification, 2, 291-307.

Lovejoy, E. & Russell, G. (1967). Suppression of learning about a hard cue by the presence of an easy cue. Psychonomic Science, 8, 365-366.

McIlvane, W. J., Serna, R. W., & Stromer, R. (in press). Stimulus control topography coherence and stimulus equivalence: Reconciling test outcomes with theory.

Macmillan, N. A. & Creelman, C. D. (1991). Detection theory: A users guide. Cambridge: Cambridge University Press.

Malott, R. W. & Siddall J. W. (1972). Acquisition of the people concept in pigeons. Psychological Reports, 31, 3-13.

Matyas, T. A., & Greenwood, K. M. (1990). Visual analysis of single-case time series: Effects of variability, serial dependency, and magnitude of intervention effects. Journal of Applied Behavior Analysis, 23, 341-351.

Newman, F. L., & Baron, M. R. (1965). Stimulus generalization along the dimension of angularity: A comparison of training procedures. Journal of Comparative and Physiological Psychology, 60, 59-63.

Omohundro, J. (1981). Recognition vs. classification of ill-defined category exemplars. Memory and Cognition, 9, 324-331.

Pavlov, I. P. (1927). Conditioned reflexes. London: Oxford University Press.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. Journal of Experimental Psychology, 77, 353-363.

Reynolds, G. S. (1961). Attention in the pigeon. Journal of the Experimental Analysis of Behavior, 4, 203-208.

Schwartz, B. (1989). Psychology of learning and behavior. New York: Norton.

Sherman, J. G., & Pierrel, R. (1961). Generalization of auditory intensities as a function of amount of discrimination training. Journal of the Experimental Analysis of Behavior, 4, 237-241.

Sidman, M. (1960). Tactics of scientific research: Evaluating experimental data in psychology. New York: Basic Books.

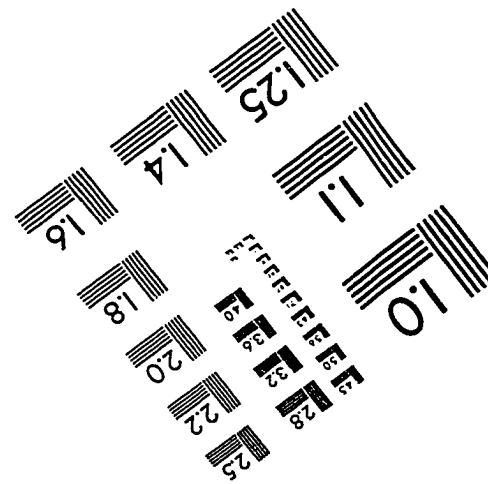
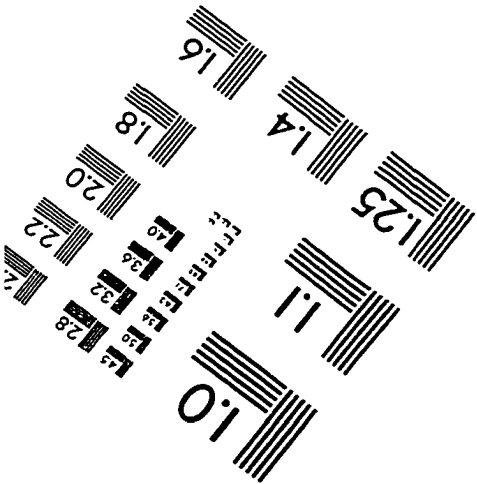
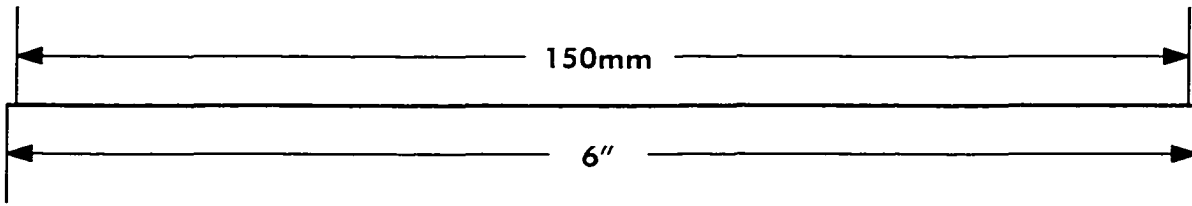
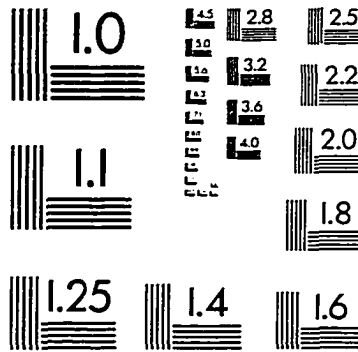
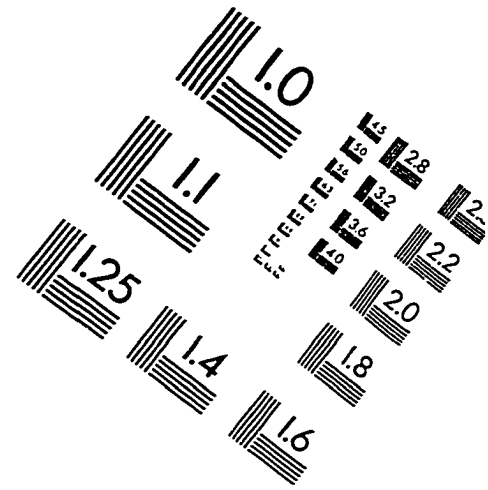
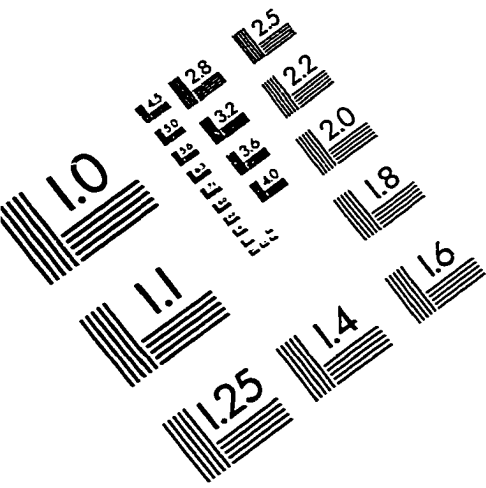
Touchette, P. E. (1969). Tilted lines as complex stimuli. Journal of the Experimental Analysis of Behavior, 12, 211-214.

Vom Saal, W. & Jenkins, H. M. (1970). Blocking and the development of stimulus control. Learning and Motivation, 1, 52-64.

Wright, A. A., Cook, R. G., & Kendrick, D. F. (1989). Relational and absolute stimulus learning by monkeys in a memory task. Journal of the Experimental Analysis of Behavior, 52, 237-248.

Wright, A. A., Cook, R. G., Rivera, J. J., Sands, S. F., & Delius, J. D. (1988). Concept learning in pigeons: Matching-to-sample with trial-unique video picture stimuli. Animal Learning and Behavior, 16, 436-444.

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE . Inc
 1653 East Main Street
 Rochester, NY 14609 USA
 Phone: 716/482-0300
 Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved