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

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

EFFECT OF NUMBER OF FORCED-CHOICE PRIMARY GENERALIZATION TEST TRIALS ON THE ESTABLISHMENT OF DIMENSIONALLY-BASED PERCEPTUAL CLASSES

by

KENNETH F. REEVE

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

1998

UMI Number: 9830758

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

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.

4/6/98

Date

Professor Lanny Fields
Chair of Examining Committee

4/8/98

Date

Professor Joseph Glick
Executive Officer

Professor Bruce L. Brown

Professor Claire L. Poulson

Professor Peter Balsam

Professor Thom Verhave

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

EFFECT OF NUMBER OF FORCED-CHOICE PRIMARY GENERALIZATION TEST TRIALS ON THE ESTABLISHMENT OF DIMENSIONALLY-BASED PERCEPTUAL CLASSES

by

Kenneth F. Reeve

Adviser: Professor Lanny Fields

In Experiment 1, 20 college students learned two identity conditional discriminations with two squares that were filled with pixels comprising either 23 or 77% of the square's interior. This was followed by 0, 152, 456, or 760 two-choice generalization test trials. The samples were 19 squares with fill percentages that ranged from 23-77%. Fill23 and Fill77 were comparisons. Next, a generalization test with a third comparison, Neither, was presented. During three-choice generalization testing, both the area under the generalization gradients, and the range of test fill percentages that occasioned the exclusive selection of Fill23 or Fill77, were direct functions of the number of prior two-choice generalization trials. Next, a discriminability test revealed that Fill23 and Fill77 were discriminable from the test fill values. Thus, a range of filled squares functioned as members of dimensionally defined perceptual classes of low- or high-fill stimuli. In Experiment 2, five new students received 760 two-choice generalization trials for fill percentage, followed by three-choice generalization testing. Students were then trained to select a glyph in the presence of Fill23 and another glyph in the presence of Fill77. Next, a three-choice generalization test with the 19 fill stimuli as samples and the glyphs as comparisons was presented. These gradients overlapped with three-choice generalization gradients previously obtained with fill patterns as comparisons. Finally, a discriminability measure showed that adjacent stimuli along the dimension of fill were discriminable from each other. Thus, each perceptual class functioned as a transfer network for a new response.

Acknowledgments

I am indebted to Dr. Lanny Fields for his active support and guidance throughout the course of this research. Gratitude is also extended to Drs. Claire L. Poulson and Bruce L. Brown for their valuable guidance, and to Drs. Peter Balsam and Thom Verhave for their insightful comments and observations. Finally, I wish to express my gratitude to my family and friends for their constant support and encouragement throughout this endeavor.

Financial support for this research was provided by U.S. Army Research Institute Contracts MDA903-90-C-0132 and DASW01-96-K-0009 to Dr. Lanny Fields.

Table of Contents

Approval Page ii

Abstract iii

Acknowledgments iv

Table of Contents v

List of Tables vi

List of Figures vii

Introduction 1

Method (Exp. 1) 19

Results (Exp. 1) 27

Discussion (Exp. 1) 36

Method (Exp. 2) 45

Results (Exp. 2) 48

Discussion (Exp. 2) 53

General Discussion 55

Tables 58

Figures 67

Bibliography 79

List of Tables

Table 1. Stimuli and Trial Configurations (Phase 1) 58

Table 2. Stimuli and Trial Configurations (Phase 4) 59

Table 3. Design of Experiment 1 60

Table 4. Individual Two-Choice Generalization Test Performances (Exp. 1) 61

Table 5. Quantification of Area Differences 62

Table 6. Quantification of Range Differences 63

Table 7. Individual Two-Choice Generalization Test Performances (Exp. 2) 64

Table 8. Individual Fill-Fill and Fill-Glyph Generalization Test Performances 65

List of Figures

Figure 1. Stimuli Used In Experiment 1 67

Figure 2. Instructional Prompts During Phase 1 Keyboard Familiarization 68

Figure 3. Two-Choice Primary Generalization Group Test Performances 69

Figure 4. Individual Student Performances for Selection of Fill77 During Three-Choice Generalization Testing 70

Figure 5. Individual Student Performances for Selection of Fill23 During Three-Choice Generalization Testing 71

Figure 6. Mean Area Under Generalization Gradients During Three-Choice Generalization Testing 72

Figure 7. Mean Range of Fill Stimuli That Occasioned Near Exclusive Selection of Fill23 or Fill77 73

Figure 8. Mean Discriminability Functions for Selection of Fill23 and Fill77 74

Figure 9. Mean Performances During Two-Choice Generalization Testing 75

Figure 10. Fill-Fill and Fill-Glyph Generalization Test Performances 76

Figure 11. Individual Range Data Obtained During Fill-Fill and Fill-Glyph Three-Choice Generalization Testing 77

Figure 12. Mean Performances for Selection of Each Endpoint Value During the Five Discriminability Tests 78

EFFECT OF NUMBER OF FORCED-CHOICE PRIMARY GENERALIZATION TEST TRIALS ON THE ESTABLISHMENT OF DIMENSIONALLY-BASED PERCEPTUAL CLASSES

by

KENNETH F. REEVE

When an organism learns to emit some behavior in the presence of a narrow range of stimuli, the behavior may also occur in the presence of a much larger range of novel stimuli, which are physically similar yet discriminable, without benefit of direct training (Keller & Schoenfeld, 1950). When this occurs, the entire set of stimuli is functioning as a perceptual class (Bourne, Dominowski, & Loftus, 1979; Cook, Wright, & Kendrick, 1990; Domjam, 1998; Fields, Adams, Buffington, Yang, & Verhave, 1996; Fields & Reeve, in press; Fields, Reeve, Adams, Brown, & Verhave, 1997; Lea & Ryan, 1984; Wasserman, Kiedinger, & Bhatt, 1988). Perceptual classes have also been referred to as similarity-based categories (Wasserman & DeVolder, 1993), open-ended categories (Herrnstein, 1990), generalization classes (Stemmer, 1980) and feature classes (McIlvane, Dube, Green, & Serna, 1993; Stromer & Mackay, 1997).

Stimulus domains of perceptual classes. Regardless of name, perceptual classes can be differentiated by the stimulus domain, or set of limiting properties, that characterizes the stimuli in the classes. One type of perceptual class, called a dimensional class (Fields & Reeve, in press; Fields, Reeve, Adams, Brown, & Verhave, 1997; Zentall, Jackson-Smith, & Jagielo, 1990), is defined by a range of discriminable values along some stimulus dimension (Herrnstein, 1984). Within a specific range along the dimension, each stimulus occasions the same response. Other stimuli outside of this range along the dimension do not occasion that response. Thus, ranges of stimuli that comprise dimensional classes differ quantitatively from each other along some dimension.

The dimension along which such a class is defined can be a simple physical one, such as length (Fields, Reeve, Adams, Brown, & Verhave, 1997), auditory click frequency (Migler & Millenson, 1969), visible light wavelength (Zentall et al., 1990), sound frequency (Njegovan, Ito, Mewhort, & Weisman, 1995; Risley, 1964), or object rotation (Wasserman, Gagliardi, Cook, Kirkpatrick-Steger, Astley, & Biederman, 1996). The dimension may also be specified in terms of a complex mathematical formula, such as a shape's "compactness" (Hrycenko & Harwood, 1980).

In another type of perceptual class, called a fuzzy class, the stimuli differ qualitatively from one another and the class-defining properties of the stimulus members are more difficult to identify and measure than they are in dimensional classes (Adams, Fields, & Verhave, 1993; Bhatt, Wasserman, Reynolds, & Knauss, 1988; Blough, 1990; Fields et al., 1996; Herrnstein, 1990; Lea, 1984; Medin & Smith, 1984; Rosch & Mervis, 1975; Wittgenstein, 1968). Specifically, the stimuli in fuzzy classes contain different combinations of discrete physical characteristics, or features, that occur across the class members (Cerella, 1979; Cook et al., 1990; Jitsumori, 1993, 1996; Herrnstein, 1984). No single feature, however, is necessary or sufficient to define membership in the class. Thus, stimuli in fuzzy classes differ from one another across a multiplicity of features and/or along many different dimensions.

Fuzzy categories have been referred to by a variety of names including basic level (Rosch, 1978; Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Bream, 1976), natural (Herrnstein, 1990; Lea & Ryan, 1984; Wasserman & DeVolder, 1993; Wittgenstein, 1968) ill-defined (Bourne et al., 1979; Homa & Little, 1985; Omohundro, 1981; Neisser, 1967) and probabilistic categories (Medin & Smith, 1984). Researchers have investigated fuzzy categories comprised of leaves (Cerella, 1979), fish (Herrnstein & de Villiers, 1980), people (Malott & Siddall, 1972), cars, cats, and flowers (Wasserman et al., 1988), locations on a college campus (Honig & Stewart, 1988), trees and water (Herrnstein, Loveland, & Cable, 1976), different musical pieces (Porter & Neuringer, 1985), and human-manufactured objects (Lubow, 1974), among others.

Researchers have proposed that responding specific to the members of a fuzzy category comes under the control of some unspecified combination of many of the potential features found across the class members (Herrnstein, 1990; Jitsumori, 1993, 1996; Lea & Ryan, 1984; Medin & Smith, 1984; Rosch & Mervis, 1975; Wasserman et al., 1988). That is, of the "many" possible stimulus features that occur across class members within a fuzzy class, a class exemplar must contain "some" undefined number of these features to evoke the same responding as other class members. Thus, the phrase "some of many" characterizes the stimulus domains of fuzzy classes because the specific features are not easily identified or quantified.

Because stimuli within fuzzy classes are characterized by the presence of some of many different features that occur across the class members, the class members cannot be arrayed along a simple dimension. Nevertheless, class members can be arrayed along psychometrically defined dimensions. One example that illustrates the creation of such a dimension involves the sorting of stimuli in a potential class according to their "typicality" (Bourne et al., 1979; Cook et al., 1990; Lea & Ryan, 1990; Rosch & Mervis, 1975; Smith, 1989). For example, when presented with different instances of trees, participants may be instructed to indicate how typical of "tree" each instance is.

Establishment and Assessment of Perceptual Classes

To infer that stimuli are functioning as members of a perceptual class, specific behavioral properties must be demonstrated. Each requirement is necessary but not sufficient by itself to infer that stimuli are functioning as members of a perceptual class. One requirement is that each stimulus in the set occasions a particular response, whereas stimuli in other sets do not occasion that response. In addition, the response specific to each stimulus in the set must occur without direct training for at least some of the stimuli in the set (Adams et al., 1993; Dinsmoor, 1995b; Fields et al., 1996; Fields & Reeve, in press; Keller & Schoenfeld, 1950; Lea, 1984; Herrnstein, 1990; Wasserman & DeVolder, 1993). Either simple or conditional discrimination training procedures have typically been used to train the discrimination between sets of stimuli and to test for transfer of responding to novel stimuli that were not used in training.

Simple discrimination training. During simple operant discrimination training, participants are exposed to at least two different stimuli, each of which is a member of a potential perceptual class. Each stimulus in a set serves as a discriminative stimulus (S+) for a particular response. The response common to class members is topographically based (Balsam, 1988; Guttman & Kalish, 1956; Jenkins & Harrison, 1960; Honig & Urcuioli, 1981). This may entail physically contacting the stimulus or emitting some other response such as a lever press, key peck, or verbal response. Regardless of its topography, during training the response specific to each S+ is reinforced only when it occurs following the presentation of that particular stimulus. Emitting the response following the presentation of S- stimuli results in no programmed reinforcement.

Typically, perceptual classes established with simple discrimination training procedures use many stimuli drawn from two or more potential classes. Studies have shown that an increase in the number of exemplars used during discrimination training results in a greater likelihood of the establishment of classes and increases the extent of the classes (Becker, 1971; Bhatt & Wright 1992; Cook et al., 1990; Engelmann, &

Carnine, 1982; Homa & Chambliss, 1975; Homa, Cross, Cornell, Goldman, & Swartz, 1973; Homa & Little, 1985; Homa, Sterling, & Treple, 1981; Omohundro, 1981; Mallott & Siddall, 1972; Stokes & Baer, 1977; Wright, Cook, Rivera, Sands, & Delius, 1988). This procedure is referred to as multiple exemplar training (Becker, 1971; Cook et al., 1990; Fields & Reeve, in press; Haring, Breen, & Laitenen, 1989; Homa & Little, 1985) or multiple stimulus discrimination training (Lea & Ryan, 1990).

Simple discrimination transfer testing. Once a discrimination has been learned between different sets of stimuli following simple discrimination training, it cannot yet be inferred that perceptual classes have been established. Novel physically similar stimuli must be presented under extinction conditions to determine the extent to which control of responding transfers to the novel stimuli (Cook et al., 1990; Zentall, Hogan, & Edwards, 1984). When the new stimuli occasion the same response as those stimuli used in training, and other stimuli do not occasion that response, such responding is considered to be class-consistent. This fulfills one of the requirement to infer that perceptual classes have been established (Adams et al., 1993; Bhatt et al., 1988; Cerella, 1979; Cook et al., 1990; Fields & Reeve, in press; Fields, Reeve, Adams, Brown, & Verhave, 1997; Herrnstein, 1990; Herrnstein & de Villiers, 1980; Herrnstein & Loveland, 1964; Herrnstein, Loveland, & Cable, 1976; Hrycenko & Harwood, 1980; Lea, 1984; Lubow, 1974; Porter & Neuringer, 1985; Wasserman et al., 1988; Zentall et al., 1984).

A study that used a multiple exemplar simple discrimination procedure to establish perceptual classes was conducted by Bhatt et al. (1988). First, pigeons were randomly exposed to 10 exemplars of four distinct potential fuzzy classes consisting of cats, flowers, people, and cars. For a particular class, each exemplar served as an S+ for pecks to a key located in one corner of a rectangular array. Correct pecks were reinforced while pecks to incorrect locations resulted in no reinforcement. This discrimination training continued until the pigeons reached an asymptote of responding correctly on about 75% of trials. Because this was significantly higher than the percentage of correct responding predicted by chance (25%), Bhatt et al. concluded that the pigeons had learned the discrimination between four potential classes.

Based on the performances during discrimination training alone, it cannot be inferred that perceptual classes had been established. The pigeons in the Bhatt et al. (1988) study may have simply learned many unrelated discriminations in which particular slides were correlated with reinforcement for pecking at a particular location (Herrnstein, 1990). Alternatively, pecking may have come under the control of the members of different perceptual classes. Transfer tests were used to determine which of these explanations accounted for the pigeons' performances.

After discrimination training was completed, 10 new slides depicting novel exemplars were presented for each of the four stimulus sets. If learning many unrelated discriminations had occurred, correct responding to the novel exemplars would not be expected to exceed chance responding of 25% correct. The results of the transfer test, however, showed that the pigeons responded correctly to the novel stimuli at levels significantly greater than chance, thus demonstrating that four distinct potential fuzzy classes had been established.

Conditional discrimination training. In addition to simple discrimination procedures, conditional discrimination procedures, such as matching to sample (Cumming & Berryman, 1965) have been used by a small number of researchers to investigate perceptual class formation (Fields, Adams, Brown, & Verhave, 1993; Fields et al., 1996; Fields, Reeve, Adams, Brown, & Verhave, 1997; Fields, Reeve, Adams, & Verhave, 1991). In this procedure, the response common to class members is selection based. To establish two potential perceptual classes, a participant is first exposed to a sample drawn from a potential perceptual class. This is followed by the presentation of two or more comparison stimuli. Selection of the positive comparison, which is drawn from the same potential class as the sample, is reinforced. Selection of other comparisons drawn from a potential class that is different from the sample results in no reinforcement. During subsequent trials, a stimulus drawn from a different potential perceptual class serves as the sample.

In contrast with simple discrimination training which typically uses multiple exemplar training, perceptual classes established with conditional discrimination training procedures have usually involved single exemplar training. One stimulus drawn from each potential class serves as both a sample and a comparison during conditional discrimination trials (Fields et al., 1993; Fields et al., 1996; Fields, Reeve, Adams, Brown, & Verhave, 1997; Fields et al., 1991). When each stimulus is presented as the sample, the selection of the comparison that is identical to the sample is reinforced. This procedure is referred to as identity matching to sample (Cumming & Berryman, 1965). No reinforcement is provided for selecting a comparison that is drawn from a different potential class than the sample. Once a discrimination has been learned between different sets of stimuli following conditional discrimination training, it cannot be inferred that perceptual classes have been established. As with simple discrimination procedures, a transfer test must also be conducted under the conditional discrimination format.

A conditional discrimination training procedure was used by Fields, Reeve, Adams, Brown, and Verhave (1997), to establish two dimensional classes with college students. Training was conducted with single exemplars drawn from each of two potential dimensional classes. The stimuli consisted of lines comprised of contiguous strings of ASCII character 176 (░) that varied in length from 1 to 25 units. The sample stimulus presented was either the 1- or 25-unit line. The two comparison stimuli were the same 1- and 25-unit lines. When students selected the comparison line that was identical to the sample, they were provided with informative feedback. As long as responding remained 100% correct during this conditional discrimination procedure, the level of feedback following a response was systematically reduced to 0% of all trials.

Conditional discrimination transfer testing. Once this discrimination was learned, students were presented with intermediate length sample lines that varied from 2 to 24 units. The comparisons were the 1- and 25-unit lines used in training. This forced-choice transfer test determined the extent to which selection of the 1- and 25-unit lines was occasioned by the test stimuli. The resulting generalization gradients showed that a range of the shorter test lines always occasioned the selection of the 1-unit line used in training. Likewise, a range of the longer test lines always occasioned the selection of the 25-unit line used in training. As the sample test lines became more intermediate in length, the choice of a given comparison decreased systematically.

As noted by Fields, Reeve, Adams, Brown, and Verhave (1997), one shortcoming with the use of forced-choice conditional discrimination trials is that they do not allow for independence of the classes; selection of one comparison is always the complement of the other. Because the test involved a forced choice in the Fields, Reeve, Adams, Brown, and Verhave (1997) study, if an intermediate length line did not appear to be perceptually similar to either comparison, a student could not assign that line to neither of the available comparisons.

While some lines may have been functioning as members of a perceptual class during the two-choice generalization test, it is possible that selection of the 1- and 25-unit lines could have been forced by the available response options. Students were required to "assign" each sample line to one of the two comparisons even when the sample did not appear to be perceptually similar to either comparison. (Fields et al., 1991; Fields et al., 1993). Thus, the range of lines that occasioned selection of the 1- and 25-unit lines may have overestimated the breadth of the perceptual classes.

To ameliorate this problem, a third comparison option, "Neither," was introduced on all test trials. As a result, both the range of shorter test lines that always occasioned the selection of the 1-unit line, and the range of longer test lines always occasioned the selection of the 25-unit line, were reduced in size. The inclusion of the Neither option allowed for the assessment of two dimensional classes that were functionally independent of one another, albeit narrower in scope than those potential classes observed during forced-choice testing,. Under the three-choice testing, some of the intermediate length line samples that previously occasioned selection of the 1- and 25-unit lines no longer did so. Instead, these intermediate lines now occasioned selection of the Neither comparison. Thus, the comparison selections occasioned by the ranges of lines under the forced-choice testing may have reflected the forced assignment rather than membership in a perceptual class.

Measures of discriminability among perceptual class members. Although class-consistent responding during the transfer test is necessary to infer the establishment of perceptual classes, it is not sufficient to do so. Such performances can be explained in two ways. One explanation is that the stimuli in such sets, while different physically, may not be discriminable from each other. That is, under a different context, it may be impossible to teach a participant to respond differentially among the different potential class members. Such a failure to discriminate may reflect the psychophysical limits of the sensory system of the organism (Lashley & Wade, 1946). Thus, the different stimuli in the set may occasion the same response simply because an organism is unable to discriminate among the stimuli. In such a case, referring to the stimuli used during discrimination training and the transfer test as members of a "class" is trivial because it is not indicative of the control of responding by stimuli comprising a class (Cook et al., 1990; Fields, Reeve, Adams, Brown, & Verhave, 1997; Lea, 1984; Wasserman et al., 1988).

The other account for class-consistent responding is that the stimuli within the set, while physically similar, are discriminable from each other and still occasion a common response in a given context. Demonstrating this within-set discriminability would fulfill the necessary and sufficient requirements that perceptual classes had been established because control of responding had transferred to discriminable, novel stimuli (Cook et al., 1990; Fields & Reeve, in press; Fields, Reeve, Adams, Brown, & Verhave, 1997; Lea, 1984; Wasserman et al., 1988).

Selection of the appropriate account requires an independent measure of within-set discriminability apart from the demonstration of class-consistent performances during the transfer test. Although only a small number of researchers have provided such measures of discriminability among stimuli in a perceptual class, it is important to note that it cannot be inferred definitively that perceptual classes have been established without such measures (Cook et al., 1990; Fields & Reeve, in press; Fields, Reeve, Adams, Brown, & Verhave, 1997; Lea, 1984; Wasserman et al., 1988). Three general procedures have been used to provide such discriminability measures. One involves the use of a conditional discrimination procedure (Fields, Reeve, Adams, Brown, & Verhave, 1997). The other two use simple discrimination procedures to establishment subcategories within perceptual classes (Wasserman et al., 1988), or to establish "quasi-" or "pseudo-categories" (Cook et al., 1990; Edwards & Honig, 1987; Herrnstein & de Villiers, 1980; Honig & Stewart, 1988; Wasserman et al., 1988).

Fields, Reeve, Adams, Brown, and Verhave, (1997) used a conditional discrimination training procedure to provide a measure of within-class discriminability following the establishment of potential dimensional classes consisting of short and long lines. To accomplish this, the selection of the Neither option was reinforced in the presence of the 2-24 unit sample lines previously presented during generalization testing. Reinforcement for selecting the 1-unit line used in training was provided only when the identical training line appeared as the sample. Likewise, selection of the 25-unit line was reinforced only if the identical line was presented as the sample. These contingencies maximized the discrimination between the intermediate length lines and the 1-unit and 25-unit lines (Blough & Blough, 1977; Hamilton & Coleman, 1933; Wright, 1972; Wright & Cumming, 1971).

Once such selections were occurring reliably, both training and all testing lines were presented under extinction conditions. When the 1-unit line used in training was presented as a sample, the students almost exclusively selected the 1-unit line comparison. When the 2- and 3-unit lines were presented as samples, students sometimes selected the 1-unit comparison and sometimes selected the Neither option. When longer intermediate lines were presented as samples, students almost exclusively selected Neither. This demonstrated that students could discriminate (a) the 1-unit line used in training from all the test lines, and (b) the 2- and 3-unit test lines from all other test lines.

When the 25-unit line was presented as a sample, students almost exclusively selected the 25-unit line. When the 24- and 23-unit lines were samples, students sometimes selected the 25-unit line and sometimes selected Neither. With decreases in the length of the sample, students almost exclusively selected Neither. Thus, students could discriminate (a) the 25-unit line from all the test lines, and (b) the 24- and 23-unit test lines from all other test lines.

Based on these discriminability measures, some of the stimuli that comprised the shorter dimensional class were discriminable from each other, and some of the stimuli that comprised the longer dimensional class were discriminable. Taken with the transfer test performances, the discriminability measures showed that selection-based responses transferred to stimuli that were novel and discriminable from those used during training. This provided an unequivocal demonstration that dimensional classes had been established in the study by Fields, Reeve, Adams, Brown, and Verhave (1997).

The second measure of discriminability involves the establishment of subcategories within perceptual classes (Wasserman et al., 1988). In such a procedure, stimuli from two potential perceptual classes are used to train a discrimination between the two potential classes, and to train a discrimination within each potential class. This procedure was used in a study by Wasserman et al., (1988, Exp. 1) to demonstrate whether pigeons could discriminate among the exemplars in fuzzy classes. Using a successive discrimination procedure with multiple exemplars, pigeons were trained to emit pecks to any of four different locations on a square array depending on the stimulus presented. On each training day, pigeons were presented with 40 slides that were the same as those used in the previously described Bhatt et al. (1988) study. Twenty of the slides were drawn from one of the four experimenter-defined sets (either cats, cars, flowers, or chairs). The other 20 slides were drawn from another of those sets. The two sets used, and the location of the correct pecking keys, were counterbalanced across birds. For each of the two sets (i.e. cats and flowers) half the slides served as S+s to peck a particular key. The other half served as S+s for pecks to another key location. The partitioning of the two categories corresponded to the 10 slides used in training and the 10 slides used in transfer testing from the Bhatt et al. (1988) study. For example, cat slides 1-10 would correspond to key 1, cat slides 11-20 would correspond to key 3, flower slides 1-10 would correspond to key 2, and flower slides 11-20 would correspond to key 4.

Given the four response keys, responding by chance would have resulted in about 25% of responses being correct. Correct responding by the pigeons, however, reached an asymptote of about 70% correct, which was significantly higher than chance. This demonstrated that the pigeons had (a) learned the discrimination between the two fuzzy categories and (b) learned a discrimination between the two subcategories within each fuzzy class. These results provided indirect evidence that the stimuli used for training and testing in the previous Bhatt et al. (1988) study were discriminable from each other. This demonstrated that control of responding transferred to these new stimuli even though they were discriminable from the training stimuli. Although this evidence of discriminability was between-subject in nature, Wasserman et al. (1988) inferred that pigeon performances in the earlier study by Bhatt et al. (1988) demonstrated the emergence of four fuzzy classes.

Because complete discrimination was not demonstrated by the pigeons in the Wasserman et al. (1988) study, the type of "errors" made as discrimination training progressed could be analyzed. At the start of training, about 75% of the responses were incorrect as chance would predict. As training progressed, the number of overall errors decreased until a floor of 30% incorrect responses was reached. The errors made by the pigeons consisted of two different types. One error was a failure to discriminate between the experimenter-defined fuzzy categories. For example, a pigeon might make an error by pecking a key assigned to a cat slide when a flower slide was presented. The other error type was a failure to discriminate between the subcategories within the fuzzy class. For example, a pigeon might peck a key corresponding to flower slides 11-20 when flower slide 7 was presented. The latter error type, which is an incorrect response to the key associated with the wrong sub-category of a larger class, is called a conceptual error (Wasserman et al., 1988). The results showed that as overall errors decreased, however, the number of conceptual errors remained constant. Thus, by the end of training, the majority of remaining errors were conceptual errors.

Wasserman et al. (1988) concluded that if the pigeons had simply memorized which individual slides were associated with a particular response, the type of errors made would not have differed. Specifically, of the three incorrect keys a bird could peck on any trial, one would be indicative of a conceptual error while the other two would be failures to discriminate between categories. Chance would thus predict that only 33% of all errors would be conceptual. Because the results indicated that conceptual errors accounted for the majority of all errors, Wasserman et al. concluded that the pigeon's behavior suggests more within-category similarity than between-category similarity.

The third measure of discriminability involves the establishment of "quasi-" or "pseudo-categories" (Cook et al., 1990; Edwards & Honig, 1987; Herrnstein & de Villiers, 1980; Honig & Stewart, 1988; Wasserman et al., 1988). Unlike perceptual classes, stimuli in such sets are arbitrarily assigned "membership" to a class. Typically, pseudo-categories are comprised of an equal number of stimuli drawn from different experimenter-defined perceptual classes. For example, two different pseudo-categories could each contain an equal number of "fish" and "non-fish" exemplars (Herrnstein & De Villiers, 1980). This arrangement results in as much within-set similarity as there is between-set (Vaughan, 1988). Thus, membership within such a set is not based on any physical characteristics specific only to that set (McIlvane et al., 1993; Stromer & Mackay, 1997).

In a study by Honig and Stewart (1988, Exp. 1), a simple successive discrimination training procedure was used to compare performances between two groups of pigeons trained with either two “true” perceptual classes or two “pseudo-categories.” The stimuli consisted of photographic representations of two different locations on a college campus. For pigeons trained with the true perceptual classes, pecks emitted in the presence of views of one location were reinforced while pecks emitted in the presence of views of another location were not reinforced. For pigeons trained with the pseudo-categories, pecks to half of the views of both locations were reinforced; the other half of the views of both locations were not followed by reinforced. Thus, each pseudo-category contained an equal number of views of both locations.

The results of training showed that all birds in the “true” group learned the discrimination between the two locations. In contrast, only half the birds learned the discrimination between the two pseudo-categories. In addition, for those birds that did learn this discrimination, it took twice as many sessions to do so than it did for pigeons in the “true” class group. This suggests that it was more difficult to learn to respond on an arbitrary basis than it was to respond on the basis of similarity across stimuli.

Once this discrimination training was completed, pigeons in the true group were presented with novel instances of the same two locations to determine the extent to which responding would transfer to these new instances. The results showed that some loss of control occurred but that correct responding to novel instances was still above chance levels. Although such performances suggest that the pigeons in the true group had formed perceptual classes, the low levels of discrimination exhibited by two of the birds in the pseudo group seem to suggest that many of the instances of a given location were not discriminable from one another. Although such a comparison was between-subject in nature, if this is the case, then we cannot infer that perceptual classes had been established. Responding may have simply been maintained by stimuli that were indistinguishable from those used in training. The opposite argument, however, could also be made because two of the birds did show discrimination among at least some of the stimuli comprising a particular location. Based on these data, it would appear that the pigeons in the true group had indeed formed perceptual classes.

Factors responsible for fuzzy class formation. In a fuzzy class, no one exemplar stimulus contains all the necessary and sufficient features that define membership. In addition, no one stimulus feature is either necessary or sufficient for membership. Rather, membership of a stimulus in a potential fuzzy class requires the presence of a combination of some of many features (Herrnstein, 1990; Jitsumori, 1993, 1996; Lea & Ryan, 1984; Medin & Smith, 1984; Rosch & Mervis, 1975; Wasserman et al., 1988). Each feature occurs with different probabilities across the stimulus members.

By necessity, participants must be exposed to more than one class exemplar during discrimination training between exemplars drawn from different classes for control of responding to be exerted by these combinations of features. As the number of exemplars used during discrimination from a given class increases, there is a greater probability that participants will be exposed to some of the many stimulus features that define class membership. When responding in the presence of each successive presentation of these different exemplar S+s is reinforced, some of the many features that define membership in that class become correlated with reinforcement for responding (Williams, 1984). Thus, these particular features become relevant for class membership.

In contrast, there are certain features that may also appear in stimuli, but they do not define membership in the class (Bourne et al., 1979). Because these features are found with an equal probability in stimuli that do (S+) and do not (S-) define a particular class, they are irrelevant for class membership. During discrimination training, this would result in the irrelevant features being equally correlated with reinforcement and extinction for responding. As a result, responding in the presence of the some of many features that define a particular fuzzy class is reinforced with a greater relative frequency than responding in the presence of irrelevant features. It is these differential reinforcement histories that select the combinations of features differentiating positive and negative exemplars of a class (Balsam, 1988). Only the presentation of multiple exemplars of the potential fuzzy class during training makes it likely that the class-defining combinations of many features, albeit ambiguously specified, will come to control responding (Becker, 1971; Cook et al., 1990; Haring et al., 1989; Homa & Little, 1985; Lea & Ryan, 1990).

When responding has come under the control of the unspecified "some" features defining the class, the presentation of new stimuli that share some of the many features will also occasion responding to the same degree as the training stimuli. This would be evidenced by class-consistent performances during a transfer test. Responding would thus generalize to new instances of the potential fuzzy class.

When only few or single exemplars are used during training, however, there is a decreased probability that a sufficient number of class-defining features will be correlated with reinforcement. The presentation of a single exemplar provides an equal likelihood that both relevant and irrelevant features will be correlated with reinforcement for responding. This would be evident in non-differential responding to both positive and negative instances of a class during a transfer test. Such performances indicate that responding during training had simply come under the control of individual stimuli rather than by some of the many features that define a fuzzy class (Cook et al., 1990).

Factors responsible for dimensional class formation. Although the establishment of fuzzy classes may require the use of multiple exemplars during training, some dimensional classes have been established with single exemplar training using simple discrimination procedures (Hrycenko & Harwood, 1980). In contrast, single exemplar training using a conditional discrimination procedure may or may not be sufficient to establish dimensional classes. In the study by Fields, Reeve, Adams, Brown, and Verhave (1997), single exemplar discrimination training was followed by the presentation of many test stimuli during a test for generalization. This initial test occurred under a forced-choice testing format with no reinforcement for responding. Following this, three-choice generalization testing revealed the formation of two independent dimensional classes. Thus, it is possible that single exemplar identity training alone, and/or exposure to forced-choice test trials, may have been responsible for the formation of the classes in the Fields, Reeve, Adams, Brown, and Verhave (1997) study.

If the use of single exemplar identity training alone is sufficient to induce the formation of dimensional classes, inclusion of the forced-choice generalization test would not influence three-choice generalization test performances. Three-choice testing alone, immediately following single exemplar identity training, would have resulted in a demonstration of the formation of two independent dimensional classes. Because the forced-choice generalization test was always presented between single exemplar identity training and three-choice generalization testing in the Fields, Reeve, Adams, Brown, and Verhave (1997) study, it was not possible to isolate the effects of single exemplar identity training on dimensional class formation.

It is also possible that exposure to forced-choice generalization testing was responsible for the formation of the dimensional classes in the Fields, Reeve, Adams, Brown, and Verhave (1997) study because all students were exposed to the test. If this analysis is confirmed, attributing dimensional class formation to single exemplar training under a conditional discrimination would be an inappropriate inference.

The present study addressed these possibilities parametrically by exposing some students to varying numbers of forced-choice generalization test trials following single exemplar identity training. Other students received no forced-choice generalization testing and only single exemplar identity training. All students were then presented with the same number of three-choice generalization test trials that included the response option, Neither. Performances obtained from the three-choice generalization testing determined (a) the range of stimuli that functioned as members of perceptual classes, and (b) whether any differences in class formation and class size occurred as a function of prior exposure to varying numbers of forced-choice generalization test trials.

Experiment 1

Method

Participants

Twenty undergraduate students at Queens College / CUNY were recruited from introductory psychology classes. Each student was randomly assigned to one of four experimental groups. No student was familiar with the research area. Students received partial course credit upon completion of the experiment, but credit did not depend upon performance. All participants were required to read and sign a statement of informed consent. The experiment was completed in a single session that lasted approximately 2-3 h.

Apparatus and Stimuli

The experiment was conducted with a personal computer that displayed all stimuli on a black-and-white VGA monitor in 320 x 240 dpi graphics mode. Responses consisted of touching specific keys on a standard keyboard. The experiment was controlled by custom software that programmed all stimulus presentations and recorded all keyboard responses.

Figure 1 depicts five of 19 stimuli used in Experiment 1. Each consisted of a 2 x 2 in. square region without a border. Each square contained a different percentage of white pixels on a black background. The percentage of white pixels defined each square's "fill" value, which ranged from 23-77% (Fill23-Fill77) in 3% increments. The pixels were large enough that the squares did not appear as different shades of gray. The pattern of white pixels for a given fill percentage value was randomly generated and remained fixed throughout the experiment. The stimuli were generated using custom designed software.

The Fill23 and Fill77 squares depicted in the top and bottom of Figure 1 defined the endpoints of the dimension of fill percentage. These two squares were used as comparison stimuli on all matching to sample trials during training and testing. The other stimuli depicted in Figure 1 represent the midpoint of the dimension (Fill50) along with two intermediate fill values (Fill35 and Fill62) on either side of the midpoint. The remaining 14 squares are not illustrated. During primary generalization tests, each of the 19 squares was used as a sample.

Procedure

Trial format, contingencies, and responses within a trial. All training and testing trials were conducted using a matching-to-sample format (Cumming & Berryman, 1965). Each trial began when "Press ENTER" appeared on the screen. Pressing the Enter key on the keyboard cleared the screen and displayed a sample stimulus at the top center of the monitor. When the space bar was pressed, the sample remained on the screen and the comparison stimuli were displayed at the bottom left and right corners. During trials in which the third comparison (the words "If NEITHER press 4") was programmed to appear, pressing the space bar concurrently displayed this message between the Fill23 and Fill77 comparison squares.

During a trial, one of three different keys was used to select a comparison. Pressing the number 1 or 2 key selected the comparison on the left or right, respectively. Pressing the number 4 key selected the Neither comparison during trials in which three comparisons were present. The selection of a comparison cleared the screen, and was followed by the presentation of a feedback message centered on the screen.

During training trials in which informative feedback was scheduled, the message that appeared depended on the accuracy of the comparison selection. On all training trials, only one comparison was correct; this was called the positive comparison. The remaining comparison(s) were incorrect, each of which was called a negative comparison. If the positive comparison was chosen, "RIGHT" appeared on the monitor and remained there until the student pressed the R key. If a negative comparison was chosen, the message "WRONG" appeared and remained there until the student pressed the W key. When non-informative feedback was scheduled, a dashed line (----) was presented following a comparison selection. This served to signal the end of a trial. The dashed line remained on screen until the student pressed the "E" key. The E key was used as an observing response (Wyckoff, 1952) to the non-informative feedback because it is located between the R and W keys on the keyboard. After the appropriate observing response (R, W, or E) was made, the screen was cleared and the next trial began (Fields, Landon-Jimenez, Buffington, & Adams, 1995).

Trial block structure and feedback contingencies. Each phase of training and testing was conducted in blocks of trials. Within each block, the trials were presented in random order without replacement. During all test blocks, each comparison selection was followed by noninformative feedback only.

At the start of training, a block of trials was presented repeatedly until all trials occasioned 100% correct responding (the mastery criterion). During these blocks, informative feedback was provided after each trial. Thereafter, the percentage of trials in a training block that occasioned informative feedback was reduced, in order, to 75%, 25%, and finally to 0% over successive blocks as long as performance within a block was maintained at 100% accuracy. During feedback reduction, the trials that were followed by informative feedback were randomly determined. If the mastery criterion was not achieved within three blocks at a given feedback level during training, the student was returned to the previous feedback level for that particular block.

Phase 1: Keyboard familiarization. Prior to beginning the experiment, students were presented with the following instructions on the computer monitor:

"Thank you for volunteering to participate in this experiment.
PLEASE DO NOT TOUCH ANY KEYS ON THE KEYBOARD YET!
In this experiment you will be presented with many trials.
Each trial contains three or four CUES that are
shapes, symbols, or common words.
YOUR TASK IS TO DISCOVER HOW TO RESPOND
CORRECTLY TO THE CUES.
Initially, there will also be INSTRUCTIONS that tell you how
to respond to the cues, and LABELS that will help you to
identify the cues on the screen. The labels and the instructions
that tell you which KEYS to press will slowly disappear.
Your task will be to RESPOND CORRECTLY to the CUES and the

INSTRUCTIONS by pressing certain keys on the computer's keyboard.

The experiment is conducted in phases. When each phase ends,

the computer will sometimes tell you how you did.

If you want to take a break at any time, please call the experimenter.

Thank you for your cooperation! Press the space bar to continue."

Following the instructions, students then learned to emit the appropriate keyboard responses in the presence of each cue used within a trial. To accomplish this, each trial in Phase 1 contained three English words. The semantic relation between the sample word and one of the comparisons was used to prompt the selection of the correct comparison. The words used in each trial in Phase 1 are listed in Table 1. Each trial was presented once per block. Informative feedback was scheduled during all trials.

Correct responding to the stimuli in a trial was also facilitated by the presence of instructional prompts, which were deleted in a serial manner across trials in Phase 1 (Fields, 1980; Fields, Adams, Verhave, & Newman, 1990; Fields, Reeve, Varelas, Rosen, Adams, Belanich, & Hobbie, 1997). This is illustrated in Figure 2. Phase 1 ended once the stimuli were presented without prompts and performance exceeded 85% accuracy (14/16 correct trials) during a single block. For the remainder of the experiment, if a non-experimentally defined key was pressed during a trial, the instruction used to prompt the appropriate key press during Keyboard Familiarization reappeared on the screen for that trial in addition to the next two trials.

<u>Phase 2: Two-choice identity conditional discrimination training</u>. After Keyboard Familiarization was completed, students received identity conditional discrimination training using the Fill23 and Fill77 squares. During each trial in a block, either Fill23 or Fill77 was presented as a sample in a random order without replacement. The two comparisons were Fill23 and Fill77. Each of the two comparisons appeared equally

often on the left and the right of the screen within a block. A correct response was the selection of the comparison that was identical to the sample.

When 100% feedback was scheduled, each block contained 32 trials with each stimulus triad presented 16 times per block. When 75, 25, or 0% informative feedback was scheduled, each block contained 16 trials with each triad presented eight times per block. As long as 100% response accuracy was maintained, informative feedback was reduced as described in Trial Block Structure and Feedback Contingencies. Phase 2 ended once a student responded with no errors during a block with 0% informative feedback.

Phase 3: Two-choice primary generalization testing. Once the identity conditional discriminations were established, a primary generalization test for fill percentage was conducted. In the test block, each of the 19 squares that varied in fill value from 23-77% was presented as a sample on two trials; this totaled 38 trials in a block. The Fill23 and Fill77 squares served as the two comparisons on all trials. In the presence of each of the 19 samples, each comparison was presented once on the left and once on the right of the computer screen in random order. During this test, all comparison selections were followed by non-informative feedback.

Students were exposed to different numbers of two-choice primary generalization test trials across the four groups. The four values of this independent variable were used to assess the potential effects of the number of two-choice generalization test trials on subsequent three-choice generalization test performances. The design of Experiment 1 is depicted in Table 2. Students in one group were presented with four test blocks (152 two-choice generalization trials with eight presentations of each of the 19 sample stimuli). Twelve blocks were presented to the students in a second group (456 two-choice trials with 24 presentations of each of the

19 sample stimuli). Twenty blocks were presented to the students in a third group (760 trials with 40 presentations of each of the 19 sample stimuli). A fourth group of students received no two-choice generalization trials.

Phase 4: Training the use of the Neither comparison. Because only two comparisons were provided during the Phase 3 generalization test, students were required to assign each test sample to one fill comparison or the other. Thus, the generalization gradients obtained for selection of the Fill23 and Fill77 comparisons in Phase 3 were, by necessity, complements of each other. If a square with an intermediate level of fill did not appear to be perceptually similar to either comparison, students were still required to assign that stimulus to one of the available fill comparisons (Fields et al., 1993; Fields et al., 1996; Fields, Reeve, Adams, Brown, & Verhave, 1997). In Phase 4, an additional response option, called the Neither comparison, was provided to allow students to assign such stimuli to neither comparison. Training the appropriate use of the Neither comparison began with the presentation of the following instructions on the computer monitor:

> "You will now learn how to use the number 4 key to select a third choice.
> Please select the comparison stimulus that is related to the sample.
> If no comparison is related to the sample, press the number 4 key.
> Thank you for your cooperation!"

The trials used in Phase 4 are listed in Table 2. Each trial was presented once per block. On all trials, the sample and two of the comparisons were English words. The Neither comparison was also available on all trials as a third response option. On some trials, one of the word comparisons was semantically related to the sample; its selection resulted in the display of the feedback message RIGHT. On other trials, neither of the two comparison words was related to the sample, in which case the selection of the

Neither comparison displayed the feedback message RIGHT (McIlvane & Stoddard, 1981; Urcuioli & Nevin, 1975). As long as responding remained accurate, feedback was reduced as indicated in the Trial Block Structure and Feedback Contingencies section. Phase 4 ended once a student responded with no errors during a block with 0% informative feedback. After the completion of Phase 4, the Neither comparison was included on all subsequent phases in Experiment 1.

Phase 5: Three-choice primary generalization testing. Primary generalization gradients were obtained by presenting the same generalization test block used in Phase 3 with one addition: the Neither comparison was available as a third response option on all trials. This test allowed for the measurement of the range of low and high fill stimuli that occasioned the exclusive selection of the Fill23 and Fill77 comparisons, respectively. The test block was presented four times for a total of 152 trials for students in all four groups. The availability of the Neither comparison during the test allowed students to select one of the fill comparisons (Fill23 or Fill77) in the presence of a given fill sample without forcing a complementary change in the likelihood of selecting the other fill comparison. The availability of the Neither comparison, therefore, potentially made the gradients measured with the Fill23 and Fill77 squares functionally independent of each other.

Phase 6: Discriminability training and testing. To infer that a range of stimuli are functioning as members of a perceptual class, it must be demonstrated that stimuli within the set are discriminable from each other. Phase 6 assessed the discriminability of the Fill23 and Fill77 squares used in training from those used during generalization testing. Students were presented with the same trial block used in the primary generalization test in Phase 5. At the start of Phase 6, however, all comparison selections were followed by informative feedback (RIGHT / WRONG). Selection of the

Fill23 square was followed by RIGHT when the sample was Fill23; the selection of the Fill77 square was followed by RIGHT when the sample was Fill77. The selection of the Neither comparison was followed by RIGHT when the remaining 17 intermediate fill squares (26-74%) were presented as samples. Any other selection was followed by the feedback message WRONG. These contingencies maximized the discrimination between the intermediate fill squares and the Fill23 and Fill77 squares used in training (Blough & Blough, 1977; Hamilton & Coleman, 1933; Wright, 1972; Wright & Cumming, 1971).

The block was repeated with 100% informative feedback until 100% correct responding was obtained. Feedback was then reduced to 75, 25, then 0% of the trials as previously described. No informative feedback was presented for the last four blocks in Phase 6. Performances during these last four blocks were used to determine the degree of discriminability between those squares used in training and testing.

Results

The results focus on five outcomes: (a) performances obtained during the identity conditional discrimination training for fill values, (b) the area of the generalization gradients obtained during two-choice generalization testing, (c) the differential effects of the number of previously presented two-choice generalization trials on the area of the generalization gradients obtained during subsequent three-choice generalization testing, (d) the range of test stimuli that occasioned exclusive selection of either the Fill23 or Fill77 comparison squares during three-choice generalization testing, and (e) performances obtained during the discriminability training and testing procedure for fill values.

Identity conditional discrimination performances. During Phase 2, all students learned the Fill23-Fill23 and Fill77-Fill77 identity conditional discriminations. Across all students, a mean of 5.2 blocks was needed to reach criterion. There was no difference across groups in the number of blocks needed to learn the identity conditional discriminations, F(3,16) = .42, n. s.

Two-choice primary generalization test performances. Table 3 shows the likelihood of selecting the Fill23 and Fill77 comparisons for individual students exposed to 152, 456, and 760 trials during two-choice generalization. There are no data for students in the 0-trial group because they were not exposed to two-choice generalization testing. No systematic differences were observed in student performances either across groups or across comparison selections. Therefore, group functions were computed by averaging the percentage of trials in which each fill pattern occasioned the selection of Fill23 and Fill77 across students. These data are depicted in Figure 3.

Figure 3 shows the average performances obtained during two-choice generalization testing for students exposed to 152, 456, and 760 trials. The graph on the left shows the likelihood of selecting the Fill23 comparison square in the presence of the 19 fill samples. The graph on the right of the figure shows the likelihood of selecting the Fill77 comparison. These data are the complement of those presented in the left of the figure.

As seen in the left of the figure, the gradients used to measure selection of Fill23 overlapped with one another. Similarly, the gradients for selection of Fill77 overlapped. The likelihood of a given fill pattern occasioning selection of Fill23 or Fill77 did not differ systematically with changes in the number of two-choice generalization test trials.

When generalization gradients are used to assess the extent of stimulus control exerted by stimuli along a dimension, one characteristic of the gradient that can be used is area (Honig & Urcuioli, 1981). The area under a gradient was quantified by summing the percentage of trials that resulted in selection of one comparison type for all test stimuli. These data were then subjected to a 3 x 2 (Number of Two-choice Trials x Comparison Type) ANOVA.

A comparison of the area under the gradient showed that there was no difference with changes in number of test trials (*F*(1,12) = 2.35, *n. s.*) across the groups. A comparison of the area under the gradients within each group showed that the area did not change regardless of selection of Fill23 or Fill77 (*F*(2,12) = 2.23, *n. s.*). In addition, the number of two-choice trials used for generalization testing did not interact with the comparison selection used to measure generalization (*F*(1,12) = .73, *n. s.*).

Effect of two-choice generalization testing on later three-choice generalization test performances: Individual performances. Figure 4 shows the likelihood of selecting the Fill77 comparison during three-choice generalization testing plotted as a function of the number of prior two-choice generalization test trials. From top to bottom, each row represents individual student performances for 760, 456, 152, and 0 two-choice generalization test trials, respectively. The area under the gradients for individual students was used to rank order the graphs within each group. The gradients with the greatest area are depicted on the left and decrease in area moving to the right. By comparing performances for students in the same column, this arrangement shows how the number of prior two-choice generalization trials affected the area of the generalization gradients for students with comparable ranking in each group.

When students in the left-most column are considered, the student exposed to 760 two-choice trials showed the greatest area under the three-choice gradients. There was a systematic reduction in the area under the gradients for the students exposed to 456, 152, and 0 two-choice trials, respectively. The same systematic effect of prior two-choice trials can also be observed for students in each of the remaining four columns of the figure. Thus, the area under the gradients for selection of the Fill77 comparison during three-choice generalization testing was a direct function of the number of prior two-choice generalization test trials.

In addition to affecting the area under the gradients, the number of prior two-choice generalization test trials also influenced the range of fill values that occasioned the near exclusive selection of Fill23 and Fill77. The two ranges were obtained in the following manner. For Fill77, the value of the lowest fill percentage that occasioned selection of Fill77 on at least 87.5% (7/8) of the trials was subtracted from the endpoint value (77%). For example, if samples Fill77 through Fill68 all occasioned selection of the Fill77 comparison on at least 87.5% of the trials, the resulting range would be 9% (77 - 68 = 9). For Fill23, the value of the highest fill percentage that occasioned selection of Fill23 was subtracted from the endpoint value (23%) that occasioned selection of Fill23. If only the endpoint values occasioned selection of themselves on at least 87.5% of the trials, the range would be 0%.

In Figure 4, each gradient in the far left column indicates that some range of high fill values almost exclusively (87.5% or greater) occasioned selection of Fill77. This range was a direct function of the number of prior two-choice trials. This effect can also be seen in the remaining four columns. In general, then, the range was a direct function of the number of prior two-choice generalization test trials.

Looking across the top row in Figure 4, the performances of all five students indicate that a large range of high-fill stimuli all occasioned the near exclusive selection of Fill77 following the presentation of 760 two-choice generalization test trials. In the second row, the gradients of all five students exposed to 456 two-choice generalization test trials also showed the same effect, albeit each ranges was smaller in breadth than for students in the 760-trial group. Four of the five students in the 152-trial group show a range of high-fill stimuli exceeding 0% that occasioned the near exclusive selection of Fill77. Each of these four ranges was smaller in breadth than for those obtained for students in the 456-trial group. Lastly, the gradient for only one of the students that were exposed to 0 two-choice generalization test trials has a range exceeding 0%.

Figure 5 shows the likelihood of selecting the Fill23 comparison during three-choice generalization testing plotted as a function of the number of prior two-choice generalization test trials. As in the previous figure, each row represents individual performances for students exposed to 760, 456, 152, and 0 two-choice generalization test trials, from top to bottom. The graphs were rank ordered by area as described for Figure 4.

When considering the area under the gradients, in the left column, the student exposed to 760 two-choice trials showed the greatest area under the three-choice gradient. There was a systematic reduction in the area under the gradients for the student exposed to 456 two-choice trials. This difference can also be seen in the remaining four columns for students exposed to 760 or 456 two-choice trials.

A comparison of performances in the first and second columns shows that the area under the gradients for students in the 456 group was somewhat greater than that for students in the 152 group. In contrast, columns 3-5 show that there was no systematic difference between the area under the gradients for students in the 152 and 456 groups. Finally, the smallest area under the gradients was observed for students exposed to 0 two-choice test trials group. The one exception to this was the gradient in the first column which did not differ systematically from the gradient for the student exposed to 152 two-choice trials.

These data show that reductions in the number of prior two-choice generalization test trials resulted in a systematic decrease in the area of the generalization gradients obtained during three-choice generalization testing for selection of the Fill23 comparison. Although not as robust, this replicated the effect seen for selection of the Fill77 comparison.

In Figure 5, each gradient in the far left column indicates that some range of low-fill values almost exclusively (87.5% or greater) occasioned selection of Fill23. This range was a direct function of the number of prior two-choice trials. This effect can also be seen in the remaining four columns.

Looking across the top row in Figure 5, the gradients of all five students in the 760-trial group show that a large range of low-fill stimuli occasioned selection of Fill23. In the second row, four of five students exposed to 456 two-choice generalization test trials showed a range greater than 0%. In the third row, four of the five students in the 152-trial group showed a range of low-fill stimuli exceeding 0% that occasioned the near exclusive selection of Fill77. Lastly, the gradients of only two of the students that were exposed to 0 two-choice generalization test trials have a range exceeding 0%.

Area under three-choice generalization gradients. In addition to visual inspection of individual performances, group data were also computed to quantify the effect of prior exposure to different numbers of two-choice trials on the area under the generalization gradients obtained during the three-choice generalization test.

Figure 6 shows the mean area under the generalization gradients obtained for three-choice generalization testing plotted as a function of the number of prior two-choice generalization trials. The area under the gradients during three-choice testing was a direct function of the number of previously presented two-choice generalization test trials. These differences were statistically significant as measured by a 4 x 2 (Number of Two-choice Trials x Comparison Type) ANOVA, $F(3, 16) = 12.05$, $p = .0002$). Across groups, the area under the gradients that measured the likelihood of selection of Fill23 and Fill77 were nearly identical, F(1, 16) = .001, n. s. Thus, the number of prior two-choice trials had a similar effect on the area under the gradients obtained by measuring selection of Fill23 and Fill77. In addition, no significant interaction was found between the number of two-choice generalization trials and comparison type, F(1, 16) = 3.13, n. s.).

Post-hoc pair-wise comparisons using Fisher's TSD were also calculated and are depicted in Table 4. These showed that the number of previously presented two-choice trials resulted in a significantly larger area under the gradient for selection of the Fill77 square for the 760-trial group as compared to the area for the 456-,152-, and 0-trial group. In addition, the area under the gradient for selection of the Fill77 square was significantly greater for the 456-trial group than the 0-trial group, but was not greater than the area under the gradient for the 152-trial group. The number of prior two-choice trials had no significant effect on the area under the gradient for the 152-trial group as compared to the area under the gradient for the 0-trial group.

The number of prior two-choice trials resulted in a significantly larger area under the gradient for selection of the Fill23 square for the 760-trial group as compared to the 0-trial group. There was no significant difference, however, in the area of the gradients between the 456-trial and the 0-trial groups, the 456- and 152-trial groups, and the 0- and 152-trial groups.

Range of stimuli that occasioned exclusive selection of Fill23 or Fill77. The effects of prior two-choice generalization testing on the range of sample fill patterns that occasioned the selection of the Fill23 or Fill77 comparisons on at least 87.5% (7/8) of the three-choice generalization trials was also quantified. The two ranges were obtained in the following manner. For Fill77, the value of the lowest fill percentage that occasioned selection of Fill77 was subtracted from the endpoint value (77%). For example, if samples Fill77 through Fill68 all occasioned selection of the Fill77 comparison on at least 87.5% of the trials, the resulting range would be 9% (77 - 68 = 9). For Fill23, the value of the highest fill percentage that occasioned selection of Fill23 was subtracted from the endpoint value (23%) that occasioned selection of Fill23. If only the endpoint values occasioned selection of themselves on at least 87.5% of the trials, the range would be 0%.

Figure 7 shows that the range of low-fill squares that occasioned selection of Fill23, and the range of high-fill squares that occasioned selection of Fill77 during three-choice generalization testing, were both direct functions of the number of prior two-choice trials. A 4 x 2 (Number of Prior Two-Choice Trials x Comparison) mixed factorial ANOVA showed that these differences were significant, $\underline{F}(3, 16) = 18.04$, $\underline{p} < .0001$. Within each group, however, a comparable range of fill patterns occasioned selection of the Fill23 or Fill77 comparisons on at least 87.5% of the test trials, $\underline{F}(1, 16) = .60$, n. s.. In addition, no interaction effects were found between the number of two-choice generalization test trials and comparison type, $\underline{F}(3, 16) = 2.56$, n. s.

Post-hoc pair-wise comparisons using Fisher's TSD were also calculated for the different ranges and are depicted in Table 5. When considering the range of high-fill stimuli that occasioned the exclusive selection of the Fill77 comparison, the range for the 760-trial group was significantly larger than the range for the 456-, 152-, and 0-trial group. In addition, the range was significantly larger for the 456-trial group as compared to the 0-trial group, but was not greater than the range for the 152-trial group. Finally, the range for the 152-trial group was not significantly larger than the range for the 0-trial group.

When considering the range of low-fill stimuli that occasioned the exclusive selection of the Fill23 comparison, the range for the 760-trial group was significantly larger than that for the 456-, 152-, and 0-trial group. There was no significant difference in the size of the range between the 456-trial group, and the 0- and 152-trial groups. In addition, there was no significant difference in the size of the range for the 0- and 152-trial groups.

In summary, during three-choice primary generalization testing, a range of high-fill squares occasioned the near exclusive selection of Fill77. In addition, a range of low-fill squares occasioned exclusive selection of Fill23. In both cases, these ranges were direct functions of the number of prior two-choice generalization trials. The ranges, however, were similar in size when measuring selection of either Fill23 or Fill77. Thus, the number of prior two-choice generalization test trials had the same effect on two separate ranges of fill pattern. These provide a within-subject replication of the effect of prior two-choice testing on the range of fill patterns that occasioned selection of either endpoint of the fill dimension.

Discriminability performances. During discriminability training and testing, the likelihood of selecting the Fill23, Fill77, and Neither comparisons was measured in the presence of each of the 19 sample fill patterns. High levels of stimulus control developed during the first few training trials, and were maintained during all subsequent training and testing blocks. A mean of 8.3 blocks was needed to complete the discriminability training (feedback levels of 100, 75, and 25%). There was no difference across student performances in the different groups, $F(3,16)=.91$, n. s.

During the last four blocks of discriminability testing, student performances were very similar, both within and across groups. Therefore, data for each fill value were averaged and are depicted in Figure 8. This averaged function is representative of individual performances.

Figure 8 shows the likelihood of selecting the Fill23 and Fill77 comparisons for all

19 samples during the last four blocks of discriminability testing. The left function was obtained for selection of the Fill23 comparison. When Fill23 was the sample, the Fill23 comparison was selected nearly exclusively on all trials. The Fill23 comparison was rarely selected in the presence of Fill26 and was almost never selected with increasingly higher sample fill percentages. Although not depicted, these intermediate fill values occasioned selection of the Neither comparison nearly exclusively. Thus, Fill26 through Fill77 were almost completely discriminable from the Fill23 square.

The right discriminability function was obtained for the selection of the Fill77 square. The shape of the function mirrors that obtained for selection of Fill23. It shows that squares Fill74 through Fill23 were nearly completely discriminable from Fill77. Taken together, these data show that Fill23 and Fill77 were nearly completely discriminable from all other fill values.

Discussion

Five major outcomes will be discussed: (a) identity conditional discrimination learning (b) the width of the generalization gradients obtained during two-choice generalization testing, (c) the differential effects of the number of two-choice generalization trials on the width and area of the generalization gradients later obtained during three-choice generalization testing, (d) the range of test stimuli that occasioned exclusive selection of either the Fill23 or Fill77 comparison squares during three-choice generalization testing, and (e) performances obtained during the discriminability training and testing procedure for fill values.

Identity conditional discrimination training. All students learned the identity conditional discriminations. There were no differences across groups in the number of blocks needed to learn the discriminations. Based on these data, differences in performances obtained during subsequent 3-choice generalization testing cannot be attributable to performance differences during identity conditional discrimination training.

Two-choice primary generalization test performances. During two-choice primary generalization testing, no systematic differences in the gradients were observed, either across student groups or across comparison selections. Repeated testing, therefore, had no effect on the width or shape of the two-choice primary generalization gradients. Based on these data, differences in subsequent three-choice generalization performances could not be due to differences in performances during two-choice generalization testing.

The size and shape of the gradients for the selection of the Fill23 and Fill77 stimuli were complements of each other. Thus, low-fill stimuli that ranged from Fill23 to an intermediate point along the fill dimension occasioned selection of Fill23 at least some of the time. High-fill stimuli from Fill77 to an intermediate point along the fill dimension occasioned selection of Fill77 at least some of the time. The two gradients crossed at the midpoint of the fill dimension (Fill50). These performances demonstrate that students partitioned the fill dimension into two roughly equal regions corresponding to the actual physical midpoint of the dimension.

Effect of Neither comparison on generalization testing. During two-choice generalization testing, intermediate fill stimuli occasioned the selection of Fill23 during some of the trials and Fill77 during other trials on a complementary basis. When the Neither comparison was introduced, these intermediate stimuli sometimes occasioned either (a) the exclusive selection of the Neither comparison, (b) the complementary selection of the Neither comparison and Fill23, but not Fill77, or (c) the complementary selection of the Neither comparison and Fill77, but not Fill23. Because the likelihood of selection of Fill23 in the presence of the test fill samples was no longer the complement of the selection of Fill77, the inclusion of the Neither comparison led to the functional separation of the fill dimension into two separate regions. Thus, the removal of the forced-choice allowed for the independent measurement of the range of stimuli that occasioned the near exclusive selection of Fill23 and Fill77. These results replicate those previously reported by Fields et al. (1993, Exp. 2), and by Fields, Reeve, Adams, Brown, and Verhave (1997).

Use of Neither comparison to disclose the formation of perceptual classes. During two-choice generalization testing, two different non-overlapping ranges of fill patterns occasioned the near exclusive selection of the Fill23 or Fill77 comparisons without direct training. Thus, stimuli in each of these ranges may have been functioning as members of two different perceptual classes: one consisting of low-fill stimuli and the other consisting of high-fill stimuli. Based on these performances alone, however, it cannot be concluded that all of the stimuli in each range were members of a perceptual class.

While some fill stimuli always occasioned the selection of Fill23 or Fill77 because they were members of a perceptual class, the selection response always occasioned by other fill stimuli may have been forced by the available response options during the two-choice generalization test. These latter fill values occasioned selection of Fill23 or Fill77 even though they were not members of a perceptual class (Fields et al., 1991; Fields et al., 1993; Fields, Reeve, Adams, Brown, & Verhave, 1997). Thus, the range of stimuli that occasioned selection of Fill23 or Fill77 during two-choice generalization testing overestimated the breadth of the perceptual classes. The range of stimuli that was indicative of perceptual classes, however, could not be identified during two-choice generalization testing.

The range of fill values that were members of each perceptual class, however, was identified by the introduction of the Neither comparison during three-choice generalization testing. This option permitted students to select the neither comparison if a test fill stimulus was judged to be perceptually dissimilar from both the Fill23 and Fill77 comparisons. Thus, the remaining fill values that continued to occasion the selection of Fill23 or Fill77 were functioning as members of each perceptual class.

Effects of number of two-choice trials on three-choice generalization performances. The removal of the forced-choice disclosed the effect of number of prior two-choice trials on the size of perceptual classes. During three-choice generalization testing, some range of low fill stimuli occasioned the exclusive selection of Fill23, and some range of high fill stimuli occasioned the exclusive selection of the Fill77 comparison. The size of each range was a direct function of the number of prior two-choice generalization test trials. Within each group, however, there was no difference between the size of the low- and high-fill classes. Thus, the number of prior two-choice trials had the same effect on the size of the low- and high-fill classes for each student. This replication across two distinct perceptual classes demonstrates the generality of the effect of the number of two-choice trials on the size of perceptual classes.

Discriminability of stimuli within perceptual classes. Performances during three-choice generalization testing showed that some range of fill values occasioned selection of Fill23 and Fill77. There are two possible accounts for such performances. One is that the test fill values that occasioned a common comparison selection were not discriminable from one another. In that case, the common selection response would reflect a student's failure to discriminate the test fills from the fill endpoint used in training. This argument is similar to that proposed by Lashley and Wade (1946) to account for constancy of responding across stimuli presented in primary generalization tests. Such an outcome would preclude referring to the range of stimuli as a class (Fields, Reeve, Adams, Brown, & Verhave, 1997; Lea, 1984; Wasserman et al., 1988).

Another account is that the test squares that occasioned the same selection response during three-choice generalization testing were discriminable from each other. If so, each of these squares occasioned the same selection responses, even though the responses had been trained to occur in the presence of only one of the squares. Therefore, the set of discriminable squares would be functioning as members of a perceptual class (Bourne et al., 1979; Cook et al., 1990; Fields et al., 1996; Fields, Reeve, Adams, Brown, & Verhave, 1997; Lea & Ryan, 1984).

Identification of the appropriate account was determined by obtaining an independent measure of the discriminability of the squares that occasioned the same selection response during three-choice generalization testing. Discriminability testing demonstrated that the range of squares that occasioned selection of Fill23 were discriminable from Fill23. Likewise, the squares that occasioned selection of Fill77 were discriminable from Fill77. Thus, the test performances obtained during the three-choice generalization test did not reflect a failure to discriminate among the fill patterns. Because each of the fill values in a given range evoked the same selection response, each range of discriminable squares functioned as a perceptual class (Bourne et al., 1979; Cook et al., 1990; Fields et al., 1996; Fields, Reeve, Adams, Brown, & Verhave, 1997; Lea & Ryan, 1984).

<u>Factors responsible for establishment of perceptual classes</u>. During discriminability testing, none of the test fill stimuli occasioned selection of either Fill23 or Fill77. In contrast, during three-choice generalization testing, different ranges of test fill stimuli occasioned selection of either Fill23 or Fill77. In Figures 4 and 5, such ranges were seen for 5 of 5 students in the 760-trial group, 4 of 5 students in the 456-trial group, and 4 of 5 in the 152-trial group. Thus, the majority of the students in these groups showed the emergence of perceptual classes. In contrast, only 1 of 5 students

showed the emergence of classes, albeit of minimal size, in the 0-trial group.

These results raise questions about the role of reinforcement in the formation of perceptual classes established along the same dimension. Explicitly programmed reinforcement was used with all students to establish identity conditional discriminations using the endpoints of the fill dimension as stimuli. The use of reinforcement with identity training alone, however, was not sufficient to reliably establish perceptual classes. This was evidenced by three-choice generalization test performances for students in the 0-trial group: only one of five student showed the emergence of classes.

Once identity training was completed, students in the 152-, 456-, and 760-trial groups were exposed to forced-choice generalization testing that did not include explicit reinforcement. Following this, three-choice generalization testing showed the reliable emergence of perceptual classes, the size of which was a direct function of the number of prior two-choice generalization trials. These results support the view that perceptual classes were established by an experimental operation that did not include concurrently programmed reinforcement.

This analysis also suggests that the explicitly programmed reinforcement during identity conditional discrimination had a remote effect on the formation of perceptual classes. The intradimensional nature (Balsam, 1988) of the identity conditional discrimination training used in the present experiment was most likely responsible for the establishment of control of responding by the dimension of fill. Such dimensional control by fill was evidenced by two-choice generalization test performances for students in the 152-, 456-, and 760-trial groups.

Once established, dimensional control by fill influenced the conditional selections emitted by students during forced-choice generalization testing. Specifically, students were predisposed to select Fill23 in the presence of the lower fill values, and to select

Fill77 in the presence of the higher fill values. In turn, these unreinforced conditional selections (Saunders et al., 1988) may have established one set of relations among the low-fill stimuli, and one set among the high-fill stimuli, resulting in the formation of two perceptual classes. Thus, while identity training with reinforcement was not sufficient to reliably establish perceptual classes by itself, its use was of historical necessity for the emergence of the perceptual classes. Without the use of explicitly programmed reinforcement during the identity intradimensional discrimination training, it is unlikely that subsequent forced-choice generalization testing would have resulted in the formation of perceptual classes.

Summary. In Experiment 1, both the area under the generalization gradients, and the range of fill values that occasioned the exclusive selection of Fill23 or Fill77, were direct functions of the number of prior two-choice generalization trials. A discriminability test also revealed that Fill23 and Fill77 were discriminable from each other and the other fill values. Thus, a range of low-fill and high-fill squares were functioning as members of perceptual classes that were defined along the dimension of fill. The size of the classes was a direct function of exposure to the number of forced-choice generalization test trials. This latter effect occurred without the use of direct reinforcement. In addition, the reliable establishment of two distinct dimensionally based perceptual classes required exposure to some two-choice generalization test trials.

Experiment 2

In Experiment 1, the presentation of many fill stimuli in the two-choice primary generalization test was the apparent operation responsible for the establishment of the perceptual classes. One requirement to infer that perceptual classes have been established is that some behavior established in the presence of a subset of stimuli transfers to new stimuli not used in training (Adams et al., 1993; Fields et al., 1996;

Dinsmoor, 1995b; Keller & Schoenfeld, 1950; Lea, 1984; Herrnstein, 1990; Wasserman & DeVolder, 1993). Because all of the fill values used to assess class formation in the three-choice tests were also presented in the two-choice generalization tests, the restriction of novel stimulus presentation was not met. This renders these stimuli as potentially unusable as a transfer test for class formation. Thus, there is a possibility that the ranges of stimuli referred to as perceptual classes in Experiment 1 may not have functioned in that capacity.

This potential problem can be resolved by assessing another property of perceptual classes. Specifically, members of a class should function as a transfer network. That is, a behavioral function acquired by one member of the presumed class should spontaneously transfer to the remaining stimuli in the set without direct training (Dougher & Markham, 1994; Dougher, Augustson, Markham, Greenway, & Wulfert, 1994; Fields et al., 1996; Hayes, 1991; Sidman, Wynne, Maguire, & Barnes, 1989). Indeed, various researchers have suggested that a test other than the one used to define the emergence of the class should be conducted to definitively conclude that classes have been established (Fields et al., 1993; Hayes, 1991; Lea, 1984; Sidman, 1994). If it could be shown that the stimulus sets established by forced-choice testing in Experiment 1 functioned as transfer networks for a new response, the conclusion that the stimulus sets were functioning as perceptual classes can be made.

In Experiment 2, this possibility was assessed in the following manner. First, the 760 two-choice generalization test trial condition in Experiment 1 was replicated with five new college students. Next, students were trained to select a different glyph in the presence of Fill23 and Fill77. Following this, three-choice generalization testing in which all 19 fill values were presented as samples along with the two glyphs and Neither as comparisons was then conducted. This determined whether the same range of fill stimuli that occasioned selection of Fill23 or Fill77 would also occasion selection of the glyph related to that fill value through conditional discrimination training. The outcome of this test determined whether perceptual classes established through forced-choice generalization testing functioned as effective transfer networks for a new response.

In Experiment 1, discriminability functions showed that Fill23 and Fill77 were discriminable from all other fill values that comprised each low- and high-fill perceptual class. No discriminability measures resulted for the intermediate fill values. To infer that perceptual classes have been established, however, many of the stimuli in the set must be discriminable from one another (Cook et al., 1990; Fields, Reeve, Adams, Brown, & Verhave, 1997; Wasserman et al., 1988). If only the endpoint values were discriminable from other stimuli comprising each class, the perceptual classes that were established in Experiment 1 would be of limited scope, regardless of apparent size. Thus, to make a more precise inference as to the breadth of the perceptual classes that are established, it must be shown that some stimuli throughout the range of values that are functioning as a class are discriminable from one another.

In Experiment 2, a more comprehensive discriminability test than that used in Experiment 1 was conducted. It began with the same discriminability test used in Experiment 1 in which selection of the endpoint values (Fill23 and Fill77) was reinforced in the presence of themselves, and selection of the Neither comparison was reinforced in the presence of all intermediate fill values (Fills 26-74%). Following this, the fill dimension was successively reduced in size. Thus, identity training was conducted with the endpoints of a dimension that became increasingly smaller in size. In addition, selection of Neither was reinforced for all intermediate fill values regardless of the width of the dimension. This allowed a more accurate determination as to the breadth of the perceptual classes that had been established.

Method

Participants

Five undergraduate students at Queens College/CUNY were recruited from introductory psychology classes. None of the students was familiar with the research area. Students received partial course credit upon completion of the experiment but credit did not depend upon performance. All participants were required to read and sign a statement of informed consent. A single experimental session lasted approximately 2-3 h.

Apparatus and Stimuli

The apparatus used was the same as that in Experiment 1. The 19 filled squares that were used in Experiment 1 were also used in Experiment 2. These 19 squares appeared as samples during all training and generalization testing. In addition, two different glyphs served as comparisons during part of Experiment 2: Glyph1 = Φ, Glyph2 = £.

Procedure

The trial format, block structure, response contingencies, feedback reduction, and responses within a trial were the same as those used in Experiment 1.

Phase 1: Keyboard familiarization. The same training procedure from Experiment 1 was used to train students to emit the appropriate keyboard responses within a trial.

Phase 2: Two-choice identity conditional discrimination training. After Keyboard familiarization was completed, students received identity conditional discrimination training using the Fill23 and Fill77 squares. This procedure was the same as that used in Experiment 1.

Phase 3: Two-choice primary generalization testing. Once the identity conditional discriminations were established, the same two-choice primary generalization test for fill percentage used in Experiment 1 was conducted. Students were presented with 20 blocks test blocks consisting of 760 two-choice generalization trials. The purpose of this was to replicate the effect of prior two-choice generalization testing on later three-choice generalization performances.

Phase 4: Training the use of the Neither comparison. Following two-choice generalization testing, students were trained on the use of the Neither comparison using the same procedure as in Experiment 1.

Phase 5: Three-choice primary generalization testing. After Neither comparison training, primary generalization gradients were obtained by presenting the same three-choice generalization test block used in Phase 5 in Experiment 1. The block was presented four times for a total of 152 trials.

Phase 6: Two-choice fill-glyph conditional discrimination training. After three-choice generalization testing, students received conditional discrimination training in which either the Fill23 or Fill77 squares served as samples. Two different glyphs served as comparisons on all trials. Each glyph comparison appeared equally often on the left and the right within a block. When Fill23 was the sample, the correct response was the selection of Glyph1. When Fill77 was the sample, the correct response was the selection of Glyph2. The Neither comparison was not available during Phase 6 conditional discrimination training.

When 100% feedback was scheduled, each block contained 32 trials with each triad presented 16 times per block. When 75, 25, or 0% informative feedback was scheduled, each block contained 16 trials with each triad presented eight times per block. As long as 100% response accuracy was maintained, feedback reduction was scheduled as outlined in Trial Block Structure and Feedback Contingencies in

Experiment 1.

Phase 7: Three-choice fill-glyph generalization testing. Once the fill-glyph conditional discriminations were established, a generalization test was presented in which each of the 19 squares (Fill23 through Fill77) was used a sample. Glyphs 1 and 2 served as comparisons on all trials in addition to the Neither comparison. Four blocks of 38 trials were presented to all students. In each block, each glyph comparison was presented once on the left and once on the right of the computer screen. During this test, all comparison selections were followed by non-informative feedback only. This block determined the extent to which generalization performances obtained during fill-fill generalization testing were maintained during fill-glyph generalization testing.

Phases 8-12: Discriminability training and testing. Experiment 1 assessed the discriminability of the fill patterns used in training from those used in testing. In Experiment 2, a more comprehensive procedure was used to assess the discriminability of many of the squares along the fill dimension. During each of the five phases of discriminability testing, a new pair of fill values served as comparisons along with the Neither comparison. Once the four discriminability test blocks were completed in each phase, discriminability training and testing was conducted with the next pair of fill stimuli.

At the beginning of each phase of discriminability testing, all comparison selections were followed by informative feedback (RIGHT / WRONG). Selection of a given fill comparison was reinforced only if it was identical to the sample. Selection of the Neither comparison was reinforced only when the sample presented was not identical to either of the comparison fill values. These contingencies maximized the discrimination between the intermediate fill values and the fill values at the endpoints of the dimension (Blough & Blough, 1977; Hamilton & Coleman, 1933; Wright, 1972; Wright & Cumming, 1971). In each phase, the training block was repeated with 100%

informative feedback until 100% correct responding was obtained. Feedback was then reduced as indicated in Trial Block Structure and Feedback Contingencies.

No informative feedback was presented for the last four blocks during Phases 8-12. In each phase, performances during these last four blocks determined the degree of discriminability between those fill values used as comparisons from all other test fill values.

In Phase 8, students received the same discriminability training and testing used in Experiment 1 (Phase 6). All nineteen fill values were presented as samples along with Fill23, Fill77, and Neither as comparisons. Phase 8 assessed the discriminability of Fill23 and Fill77 from those intermediate fill values used during three-choice generalization testing.

In Phase 9, students were presented with the same trial block used in the Phase 8 with the following exceptions. Instead of 19 fill stimuli, 17 fill values ranging from Fill26 to Fill74 were used as samples during the training and testing of this block; the Fill23 and Fill77 squares were no longer presented as either samples or comparisons. The comparisons used in Phase 9 were Fill26 and Fill74 along with the Neither comparison.

In Phase 10, 13 squares ranging from Fill32 to Fill68 were used as samples. Fill20 and Fill68 served as comparisons along with the Neither comparison. In Phase 11, nine squares were used as samples. They ranged from Fill38 to Fill62. Fill38 and Fill62 served as comparisons along with the Neither comparison. In the final phase of discriminability training and testing (Phase 12) , only five squares were used as samples: Fill44, Fill47, Fill50, Fill53, and Fill56. The comparisons were Fill44, Fill56, and the Neither comparison.

Results

The results focus on five outcomes: (a) performances obtained during the identity conditional discrimination training for fill values, and the Fill-Glyph conditional discrimination training, (b) the area of the generalization gradients obtained during two-choice generalization testing, (c) a comparison of the areas under the gradients obtained for selection of Fill23 and Glyph1, or Fill77 and Glyph2 during subsequent three-choice generalization testing, (d) a comparison of the range of test stimuli that occasioned exclusive selection of either Fill23 and Glyph1 or Fill77 and Glyph2 during three-choice generalization testing, and (e) performances obtained during the discriminability training and testing procedure for fill values.

Fill-Fill and Fill-Glyph conditional discrimination performances. During Phase 2, all students learned the Fill23-Fill23 and Fill77-Fill77 identity conditional discriminations. Across all students, a mean of 4.3 blocks was needed to reach criterion. During Phase 6, all students learned the Fill23-Glyph1 and Fill77-Glyph2 conditional discriminations. A mean of 11 blocks was needed to reach criterion.

Two-choice primary generalization test performances. Table 6 shows individual performances obtained during two-choice generalization testing. Because no systematic differences were observed in the area under the gradients during two-choice generalization testing, averaged functions were computed and are depicted in Figure 9. The area was quantified as described in Experiment 1.

Figure 9 shows the average performances obtained during two-choice generalization testing. The left graph shows the likelihood of selecting the Fill23 comparison square in the presence of the 19 fill samples. The right graph shows the likelihood of selecting the Fill77 comparison. These data are the complement of those presented in the left of the figure. As can be seen in the figure, the area under the gradients used to measure selection of Fill23 did not differ from that used to measure selection of Fill77, (*F*(1, 4) = 3.41, *n. s.*).

A comparison of the areas under the gradients obtained for selection of Fill23 and Glyph1, or Fill77 and Glyph2 during three-choice generalization testing. Table 7 shows individual performances obtained during the Phase 5 fill-fill and Phase 7 fill-glyph three-choice generalization testing. Because no systematic differences were observed in the area under the gradients used to measure selection of Fill23 and Fill77 or Glyph1 and Glyph2, averaged functions were computed and are depicted in Figure 10.

Figure 10 shows the mean area under the generalization gradients obtained for selection of Fill23, Fill77, and Glyph1 and Glyph2. The stimuli along the fill dimension had similar likelihoods of occasioning selection of different comparisons. Regardless of comparison selection, there were no systematic differences for the area under the four gradients. In the left graph, a comparison of performances shows that the area under the gradients for selection of Fill23 and Glyph1 were not different. In addition, the area under the gradients for selection of Fill77 and Glyph2 were not different. Thus, these gradients overlapped.

A 2 x 2 (Fill Comparison Type x Glyph Comparison Type) ANOVA confirmed that there was no significant difference in the area under the gradients across fill comparisons, F(1,8) = 1.52, n. s., or across glyph comparisons, F(1,8) = .86, n. s. In addition, no significant interaction between the two comparison types was observed, F(1,8) = .86, n. s..

Range of stimuli that occasioned exclusive selection of Fill23, Fill77, Glyph1, and Glyph2. The range of sample fill patterns that occasioned the selection of the Fill23, Fill77, Glyph1 and Glyph2 comparisons on at least 87.5% (7/8) of the three-choice generalization trials was also quantified as described in Experiment 1.

Figure 11 depicts the range for each individual student. The top graph shows that a similar range of low-fill samples occasioned selection of Fill23 and Glyph1 on at least 87.5 % (7/8) of the trials for a given student. The size of the range, however, varied across students. The bottom graph shows similar results. That is, a similar range of high-fill samples occasioned selection of Fill77 and Glyph2. In this case, however, the size of the range was identical across all five students. Thus, the two ranges of stimuli that occasioned the selection of Fill23 and Fill77 in the presence of all the fill stimuli during three-choice generalization testing were highly predictive of the two ranges of stimuli that occasioned the selection of Glyph1 and Glyph2 during the fill-glyph generalization test.

A 2 x 2 (Fill Comparison x Glyph Comparison) ANOVA showed that there was no significant difference in the ranges used to measure selection of a fill comparison, F(1,8) = .83, n. s. or a glyph comparison, F(1,8) = .02, n. s. In addition, there was no significant interaction between these two variables on the size of the range, F(1,8) = .02, n. s.

Discriminability performances. During first part of discriminability training and testing, the likelihood of selecting the Fill23, Fill77, and Neither comparisons was measured in the presence of each of the 19 sample fill patterns. High levels of stimulus control developed during the first few training trials, and were maintained during all subsequent training blocks when the fill dimension was reduced in width.

Because individual performances for all five students were very similar during the last four blocks of each of the five discriminability tests, data for each fill value were averaged and are depicted in Figure 12. This averaged function is representative of individual performances.

The top of Figure 12 shows the likelihood of selecting the Fill23 and Fill77 comparisons for all 19 samples during the last four blocks of Phase 8 discriminability testing. The left function was obtained for selection of the Fill23 comparison. When Fill23 was the sample, the Fill23 comparison was selected nearly exclusively on all trials. The Fill23 comparison was rarely selected in the presence of Fill26 and was almost never selected with increasingly higher sample fill percentages. Although not depicted, these intermediate fill values occasioned the near exclusive selection of the Neither comparison. Thus, Fill26 through Fill77 were almost completely discriminable from the Fill23 square. The right discriminability function was obtained for the selection of the Fill77 square. The shape of the function mirrors that obtained for selection of Fill23. It shows that squares Fill74 through Fill23 were nearly completely discriminable from Fill77. These results replicate those from Experiment 1.

The second row of Figure 12 shows the likelihood of selecting the Fill26 and Fill74 comparisons for all 17 samples during the last four blocks of Phase 9 discriminability testing. When Fill26 was the sample, the Fill26 comparison was selected nearly exclusively on all trials. The Fill26 comparison was rarely selected in the presence of Fill29 and was almost never selected with increasingly higher sample fill percentages. Rather, these intermediate fill values occasioned the near exclusive selection of the Neither comparison. Thus, Fill29 through Fill74 were almost completely discriminable from Fill26 square. The right discriminability function was obtained for the selection of the Fill74 square. The shape of the function mirrors that obtained for selection of Fill26. It shows that squares Fill71 through Fill26 were nearly completely discriminable from Fill74.

The third row of Figure 12 shows the likelihood of selecting the Fill32 and Fill68 comparisons for all 13 samples during the last four blocks of Phase 10 discriminability testing. The functions show that Fill32 was nearly completely discriminable from all other fill values. Likewise, Fill68 was nearly completely discriminable from all other fill values.

Similar results can be observed in the fourth row of Figure 12 which depicts the likelihood of selecting the Fill38 and Fill62 comparisons for all 9 samples during the last four blocks of Phase 11 discriminability testing. Fill38 was nearly completely discriminable from all other fill values. Likewise, Fill62 was nearly completely discriminable from all other fill values.

The last row of Figure 12 shows the likelihood of selecting the Fill44 and Fill56 comparisons for all 5 samples during the last four blocks of Phase 12 discriminability testing. Fill44 was nearly completely discriminable from all other fill values. Likewise, Fill56 was nearly completely discriminable from all other fill values.

Discussion

Three major outcomes will be discussed: (a) the width of the generalization gradients obtained during two-choice generalization testing, (b) performances obtained during the discriminability training and testing procedure, and (c) a comparison of generalization test performances obtained during three-choice fill-fill and fill-glyph generalization testing.

Two-choice primary generalization test performances. During two-choice generalization testing, no systematic differences were observed in the gradients either across individual students or across comparison selections. Thus, the size and shape of the gradients for the selection of the Fill23 and Fill77 stimuli were complements of each other. In addition, the fill dimension was partitioned into two roughly equal regions corresponding to the actual physical midpoint of the dimension. These findings replicate those seen in Experiment 1.

Discriminability of stimuli within perceptual classes. Discriminability testing demonstrated that adjacent fill values were discriminable from each other along the entire width of the fill dimension. The range of squares that occasioned selection of Fill23 were discriminable from Fill23 and from each other. Likewise, the squares that occasioned selection of Fill77 were discriminable from Fill77 and from each other. Thus, the test performances obtained during the three-choice generalization test did not reflect a failure to discriminate among the fill patterns. Because each of the fill values in a given range evoked the same selection response, each range of discriminable squares functioned as a perceptual class (Bourne et al., 1979; Cook et al., 1990; Fields et al., 1996; Fields, Reeve, Adams, Brown, & Verhave, 1997; Lea & Ryan, 1984). Unlike in Experiment 1, this more comprehensive measure of discriminability along the fill dimension indicates that the apparent breadth of the perceptual classes defined by the range measure was accurate and was not due to an inability to discriminate some of the stimuli in the range that were functioning as perceptual class members.

A comparison of Fill-Fill and Fill-Glyph performances during three-choice generalization testing. During three-choice testing, a given stimulus along the fill dimension occasioned a similar likelihood of selecting the Fill23 and Glyph1 comparisons. This held for each test fill value. Thus, performances for selection of Fill23 rather precisely predicted the degree of generalization observed for selection of Glyph1. Similar results were obtained for selection of the Fill77 and Glyph2 comparisons. These performances show that the degree of stimulus control exerted by the test fill stimuli was the same when the comparison was a Fill stimulus that was physically similar to the sample fill value, or when the comparison was a glyph that bore no resemblance to the sample fill value.

The range of low-fill stimuli that occasioned the near exclusive selection of Fill23 also occasioned the near exclusive selection of Glyph1. Likewise, the same range of high-fill stimuli that occasioned the near exclusive selection of Fill77 also occasioned the near exclusive selection of Glyph2. Thus, the same ranges of sample fill values occasioned the selection of Fill23 and Glyph1 or Fill77 and Glyph2 even though the glyphs were never paired with any of the test fill values.

One property of a stimulus class is that it should function as a transfer network (Dougher et al., 1994; Dougher & Markham, 1994; Fields et al., 1996; Hayes, 1991; Sidman et al., 1989). In Experiment 2, when a new selection-based response was trained in the presence of one stimulus member in each range of low- or high-fill stimuli, that same selection-based response was also occasioned by the remaining class members without benefit of direct training. Thus, the ranges of stimuli that were referred to as perceptual classes in Experiments 1 and 2 functioned as transfer networks. These findings confirm that the ranges were in fact also functioning as perceptual classes (Dougher et al., 1994; Dougher & Markham, 1994; Fields et al., 1996; Hayes, 1991; Sidman et al., 1989).

General Discussion

Extension of perceptual classes. The training of the conditional discrimination between a given fill value and a glyph, followed by fill-glyph generalization testing, confirmed that perceptual classes were established. By definition, the performances occasioned by stimuli that function as members of a class remain stable both over time (Spradlin, Saunders, & Saunders, 1992; Saunders, Saunders, Kirby, & Spradlin, 1990) and across contexts (Bush, Sidman, & de Rose, 1988; Lynch & Green, 1991; Meehan & Fields, 1995). In Experiment 2, two different testing contexts were used. One consisted of fill comparisons while another consisted of glyph comparisons. Regardless of comparison, the same range of stimuli occasioned selection of a given fill comparison or its corresponding glyph. Thus, the requirement that stimuli in a class remain stable over different contexts was met.

Because the selection of a particular glyph was occasioned by all of the fill values in a given range, all of these stimuli functioned as a stimulus class. Some stimuli in the class were physically similar (a range of fill stimuli), while one stimulus (a glyph) was physically disparate from all the other class members. Therefore, the fill-glyph test performances demonstrated that each perceptual class extended beyond its initial defining domain of fill percentage. This extended class bears formal similarity to other complex stimulus classes such as superordinate semantic categories (Medin & Smith, 1984; Rosch & Mervis, 1975), natural kinds (Gelman, 1988a; Gelman, 1988b; Gelman & Markman, 1986; Gelman & Markman, 1987) and generalized equivalence classes (Barnes & Keenan, 1993; Bush, 1993; DeGrandpre, Bickel, & Higgins, 1992; Fields et al., 1991; Fields et al., 1993; Fields et al., 1996; Haring et al.,1989) because each consists of some stimuli that are perceptually similar and some stimuli that are perceptually disparate.

The establishment of the extended fill-glyph classes in Experiment 2 may be accounted for in terms of the operations needed to establish associative transitivity between stimuli (Fields & Verhave, 1987; Sidman & Tailby, 1982). This can be illustrated by considering the following example for the low-fill class. A relation was established between each test fill value (called Fillx) and Fill23 when Fill23 was selected during the forced-choice generalization testing. This occurred as a result of unreinforced conditional selections of comparison stimuli (Saunders, Saunders, Kirby, & Spradlin, 1988). These conditional relations can be referred to as Fillx∩Fill23. Next, the Fill23∩Glyph1 relation was established through direct conditional discrimination training. To test for transitivity, the likelihood of selecting Glyph1 is measured in the presence of the test fill stimuli (Fillx). In Experiment 2, the likelihood of the selection of Glyph1 was very high for a range of low-fill stimuli, thereby demonstrating a transitive relation between the test fill values and Glyph1. A similar argument can also be made to explain associative transitivity among stimuli in the high-fill and Glyph2 extended class.

It is conceivable that the classes established in Experiment 2 were generalized equivalence classes. Although a demonstration of transitivity is necessary to make this inference, additional tests must also be conducted to infer that these stimuli are functioning as a generalized equivalence class (Fields, Reeve, Adams, Brown, & Verhave, 1997; Fields & Verhave, 1987; Sidman & Tailby, 1982). Because these other tests were not conducted, however, the inference that generalized equivalence classes were established in Experiment 2 is equivocal.

Establishment of dimensionally based perceptual classes. The traditional method used to establish perceptual classes has involved direct discrimination training with multiple exemplars (Bhatt et al., 1988; Cook et al., 1990; Herrnstein, 1990; Herrnstein & de Villiers, 1980; Herrnstein & Loveland, 1964; Herrnstein, Loveland, & Cable, 1976; Honig & Stewart, 1988; Lubow, 1974; Malott & Siddall, 1972; Porter & Neuringer, 1985; Wasserman et al., 1988). Although the use of direct reinforcement contingencies may facilitate class formation, the results of Experiments 1 and 2 demonstrate that perceptual classes can be established without the use of direct reinforcement when forced-choice generalization testing is presented. Forced-choice testing may be sufficient by itself to establish perceptual classes, particularly those that are dimensionally based. Thus, the results of Experiments 1 and 2 expand the range of operations that can be used to establish perceptual classes.

Tests for the emergence of such classes might include the presentation of novel stimuli that are physically similar to stimuli used in training. In addition, a second transfer test must also be used to demonstrate that such classes also function as effective transfer networks for a new behavior or stimulus function. If this second proviso becomes another defining characteristic of perceptual classes, then following the emergence of such classes, a new behavior must be trained to one class member to determine whether this response transfers to the other members.

Table 1

Stimuli and Trial Configurations Used During Keyboard Familiarization (Phase 1).

Sample	Positive Comparison	Negative Comparison
ALCOHOL	DRUNK	MOUSE
ANT	BEE	COW
CANARY	SPARROW	STARS
CAT	MOUSE	DRAGONS
COMETS	STARS	FATHER
DOG	WOLF	DARK
DUNGEONS	CHAINS	PENCIL
EGGS	BACON	SPARROW
KINGS	QUEENS	CAMELS
LIGHT	DARK	SOCK
MOTHER	FATHER	BACON
MUD	PIG	HAT
PAPER	WRITE	OCEAN
RED	COLOR	PEAR
SOAP	WATER	THAT
THIS	THAT	KINGS

Table 2

Stimuli and Trial Configurations Used During Neither Comparison Training (Phase 4).

Sample	(+) Comparison	(-) Comparison	(-) Comparison
SOAP	NC	COMPUTER	TRASH
PAPER	NC	COFFEE	OCEAN
CANARY	NC	FIRE	STARS
KINGS	NC	TRUCK	ROCK
DOG	WOLF	DARK	NC
RED	COLOR	PEAR	NC
MUD	PIG	HAT	NC
MOTHER	FATHER	BACON	NC
ANT	BEE	COW	NC
DUNGEONS	CHAINS	PENCIL	NC
SHOE	SOCK	DARK	NC
COMETS	STARS	FATHER	NC

Note. NC = The Neither Comparison.

Table 3

Design of Experiment 1

Condition	Single Exemplar Identity Training	Two-choice Generalization Testing	Three-choice Generalization Testing
0 Group	YES	NONE	152 trials
152 Group	YES	152 trials	152 trials
456 Group	YES	456 trials	152 trials
760 Group	YES	760 trials	152 trials

Table 4

Individual Performances In Experiment 1 For Selection of Fill23 During Two-Choice Primary Generalization Testing.

		Fill Values																		
Grp.	Student	23	26	29	32	35	38	41	44	47	50	53	56	59	62	65	68	71	74	77
	EK	100	100	100	88	75	100	88	75	38	25						13			
	IM	100	100	100	100	100	88	88	100	50	25	13	13						13	
152	GL	100	100	100	88	100	100	100	100	100	100	75	50	25	38	13				13
	EC	100	100	88	100	100	88	75	100	75	63	25	13				13			
	JU	100	100	100	100	100	63	88	75	75	75	38					13		13	
	MEAN	100	100	98	95	95	88	88	90	75	58	30	15	5	8	3	8	0	5	3
	PF	100	100	96	96	96	96	96	75	67	67	21	13	4		4				
	MD	96	96	96	100	100	96	75	92	42	13					4				4
456	IS	92	100	100	96	100	100	100	88	75	63	29	4	8				4	4	
	DV	100	100	100	100	100	100	100	88	75	68					4		4		
	DL	100	100	100	100	100	88	88	71	54	25	8		4						
	MEAN	98	99	98	98	99	96	92	83	63	45	12	3	3	0	3	0	2	1	1
	KS	100	100	100	100	95	93	78	53	23	35	8	3	3						
	PP	100	100	95	98	100	93	95	90	75	85	40	13	8	3	3	13	13	3	5
760	DJ	100	95	100	100	98	100	100	90	58	60	25	10	8			5			3
	AT	98	95	95	97	95	85	80	60	15	23	5	3				5		3	
	NR	95	93	93	93	93	90	78	63	35	35	15	5				3	3		3
	MEAN	99	97	97	97	96	92	86	71	41	48	19	7	4	1	1	5	3	1	2

Note. Percentages are rounded to the nearest whole number. Empty Slots = 0.

Table 5

Quantification of Area Differences During Three-choice Generalization Testing in Experiment 1 Using Fisher's Protected T-Tests For Pair-Wise Comparisons.

	152-HIGH	456-HIGH	760-HIGH
0-HIGH	-1.40	-2.98**	-7.25**
152-HIGH	—	-1.58	-5.85**
456-HIGH	—	—	-4.27**

	152-LOW	456-LOW	760-LOW
0-LOW	-1.48	-.79	-3.71**
152-LOW	—	.69	-2.22*
456-LOW	—	—	-5.03**

	0-LOW	152-LOW	456-LOW	760-LOW
0-HIGH	1.43	—	—	—
152-HIGH	—	1.51	—	—
456-HIGH	—	—	-.76	—
760-HIGH	—	—	—	-2.12

Note. LOW and HIGH indicate area under the gradients for selection of Fill23 and Fill77, respectively. ** p<.01, * p<.05

Table 6

Quantification of Range Differences During Three-choice Generalization Testing in Experiment 1 Using Fisher's Protected T-Tests For Pair-Wise Comparisons.

	152-HIGH	456-HIGH	760-HIGH
0-HIGH	-1.55	-3.32**	-7.53**
152-HIGH	—	-1.77	-5.98**
456-HIGH	—	—	-4.21**

	152-LOW	456-LOW	760-LOW
0-LOW	-1.55	-1.55	-4.21**
152-LOW	—	0.00	-2.66*
456-LOW	—	—	-2.66*

	0-LOW	152-LOW	456-LOW	760-LOW
0-HIGH	.88	- - -	- - -	- - -
152-HIGH	- - -	.89	- - -	- - -
456-HIGH	- - -	- - -	-.89	- - -
760-HIGH	- - -	- - -	- - -	-2.44*

Note. LOW and HIGH indicate range for selection of Fill23 and Fill77, respectively.
** p<.01, * p<.05

Table 7

Individual Performances in Experiment 2 for Selection of Fill23 During Two-Choice Primary Generalization Testing.

Student	Fill Values																	
	23	26	29	32	35	38	41	44	47	50	53	56	59	62	65	68	71	74	77
ET	98	100	98	100	98	98	95	78	25	28	5						3		
JC	100	100	100	95	98	95	100	95	70	83	10	8				3			3
RA	100	100	98	100	100	95	98	83	70	68	20	5				3			
MP	100	95	100	100	100	100	93	90	30	40		3	3		5				3
CF	98	95	98	100	93	98	95	85	67	53	28	5	3	3			3	3	
MEAN	99	98	99	99	98	97	96	86	53	54	13	4	1	1	1	1	1	1	1

Note. Percentages are rounded to the nearest whole number. Empty slots = 0.

Table 8

Individual Performances During Phase 5 (Fill-Fill) and Phase 7 (Fill-Glyph) Three-Choice Generalization Testing in Experiment 2.

Percent Selection of Fill23 Fill Values

Student	23	26	29	32	35	38	41	44	47	50	53	56	59	62	65	68	71	74	77
ET	100	100	100	100	75	75	13	13											
JC	100	100	100	100	100	100	100	88	75	38	13								
RA	100	100	100	100	100	100	63	50	50	50									
MP	100	100	100	100	100	88	75	63	50	38	13		13						
CF	100	100	100	100	100	100	100	100	88	75	50		38		13	13	13		
MEAN	100	100	100	100	95	93	70	63	53	40	15	0	10	0	3	3	3	0	0

Percent Selection of Glyph1 Fill Values

Student	23	26	29	32	35	38	41	44	47	50	53	56	59	62	65	68	71	74	77
ET	75	100	88	88	75	75	38	38	13										
JC	100	100	100	100	100	88	100	100	63	50	25	13	13	13	13		25		
RA	100	100	88	88	88	63	63	25								13			
MP	88	100	88	100	88	75	50	38	13	38	13					13			
CF	88	88	100	100	100	88	75	88	50	38	25	25			13				
MEAN	90	98	93	95	90	78	65	58	28	25	13	8	3	3	5	5	5	0	0

(continued on next page)

Percent Selection of Fill77 Fill Values

Student	23	26	29	32	35	38	41	44	47	50	53	56	59	62	65	68	71	74	77
ET									25	13	50	63	100	100	100	100	100	100	100
JC		13									13	13	38	88	100	100	100	100	100
RA											50	50	50	88	88	100	100	100	100
MP									13		25	38	88	100	100	100	100	100	100
CF									13			13	13	88	75	88	88	100	100
MEAN	0	3	0	0	0	0	0	0	10	3	28	35	58	93	93	98	98	100	100

Percent Selection of Glyph2 Fill Values

Student	23	26	29	32	35	38	41	44	47	50	53	56	59	62	65	68	71	74	77
ET					13			13		38	38	38	88	100	100	100	100	100	100
JC									13		13		13	75	88	88	75	100	100
RA					13	13		25			63	63	75	100	100	88	100	100	100
MP	13	13				13	25	13					13	75	75	100	100	100	100
CF	13							13	38		13	38	75	88	88	88	100	100	100
MEAN	5	3	0	0	5	5	5	13	10	8	25	28	53	88	90	93	95	100	100

Note. Percentages are rounded to the nearest whole number. Empty slots = 0.

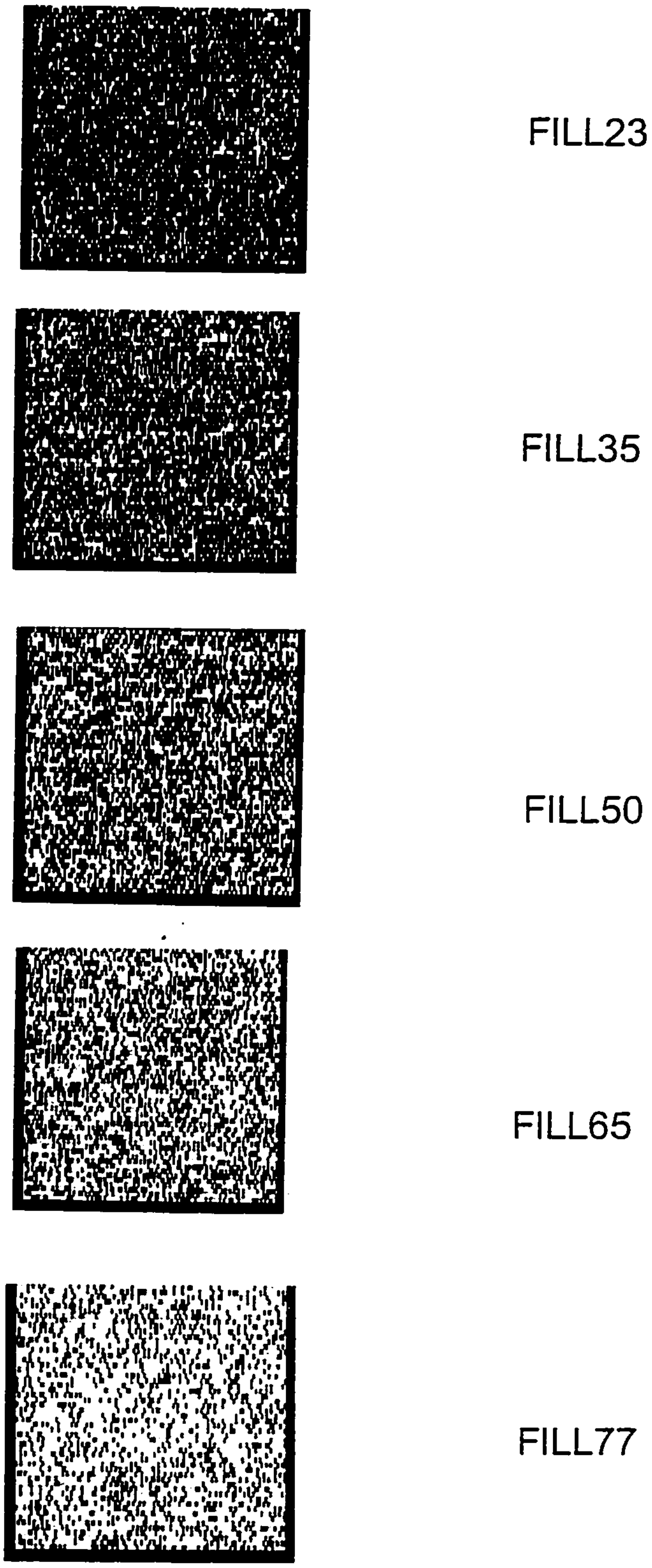


Figure 1. Five of the 19 fill stimuli used in Experiment 1. Fillx represents the percentage of pixels that were filled for a given square. Fill23 and Fill77 were the endpoints of the dimension.

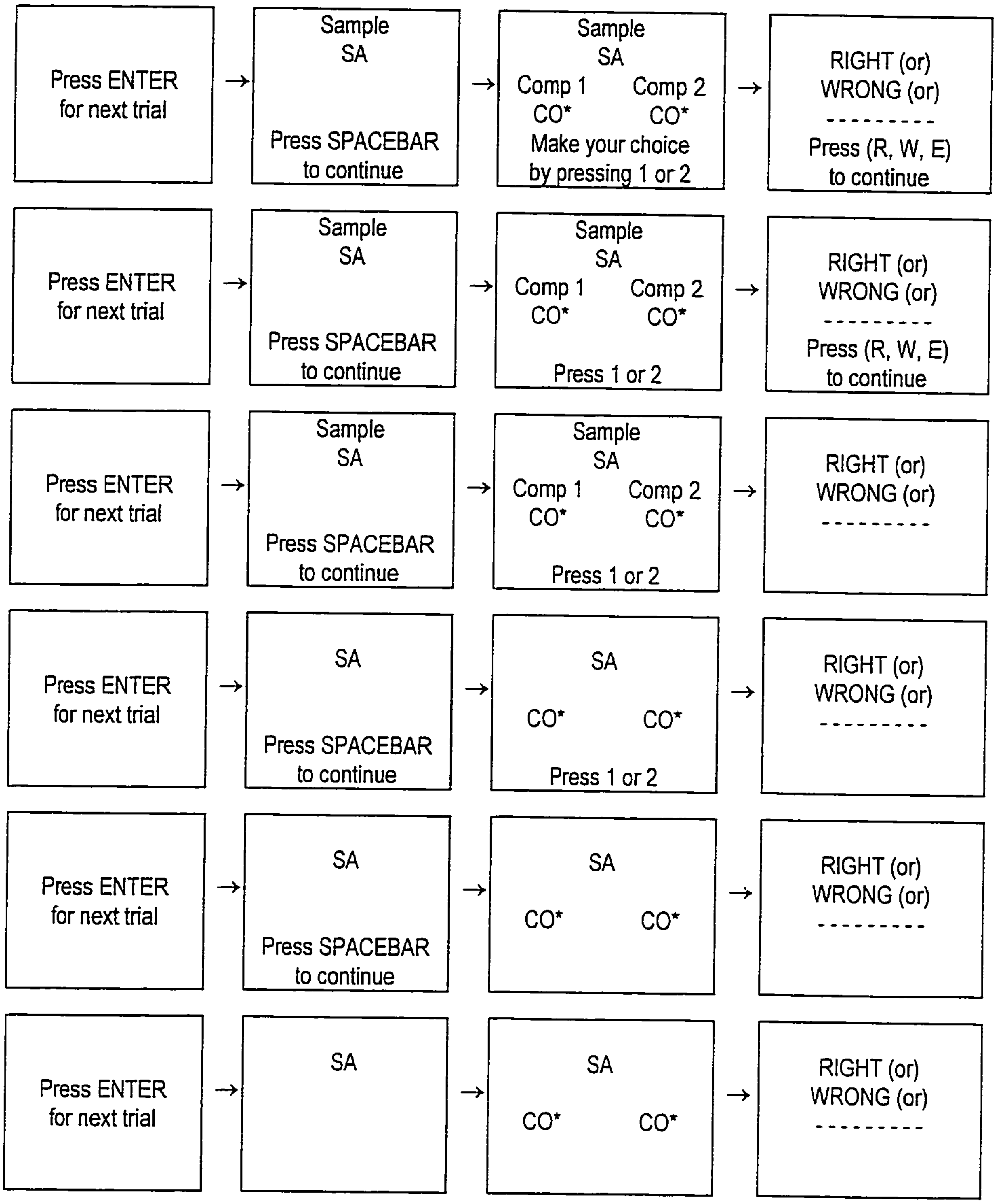


Figure 2. Sequential changes in prompts that were presented during a trial are shown across each row. "SA" represents the location of the sample stimulus; "CO*" represents the locations of the comparison stimuli.

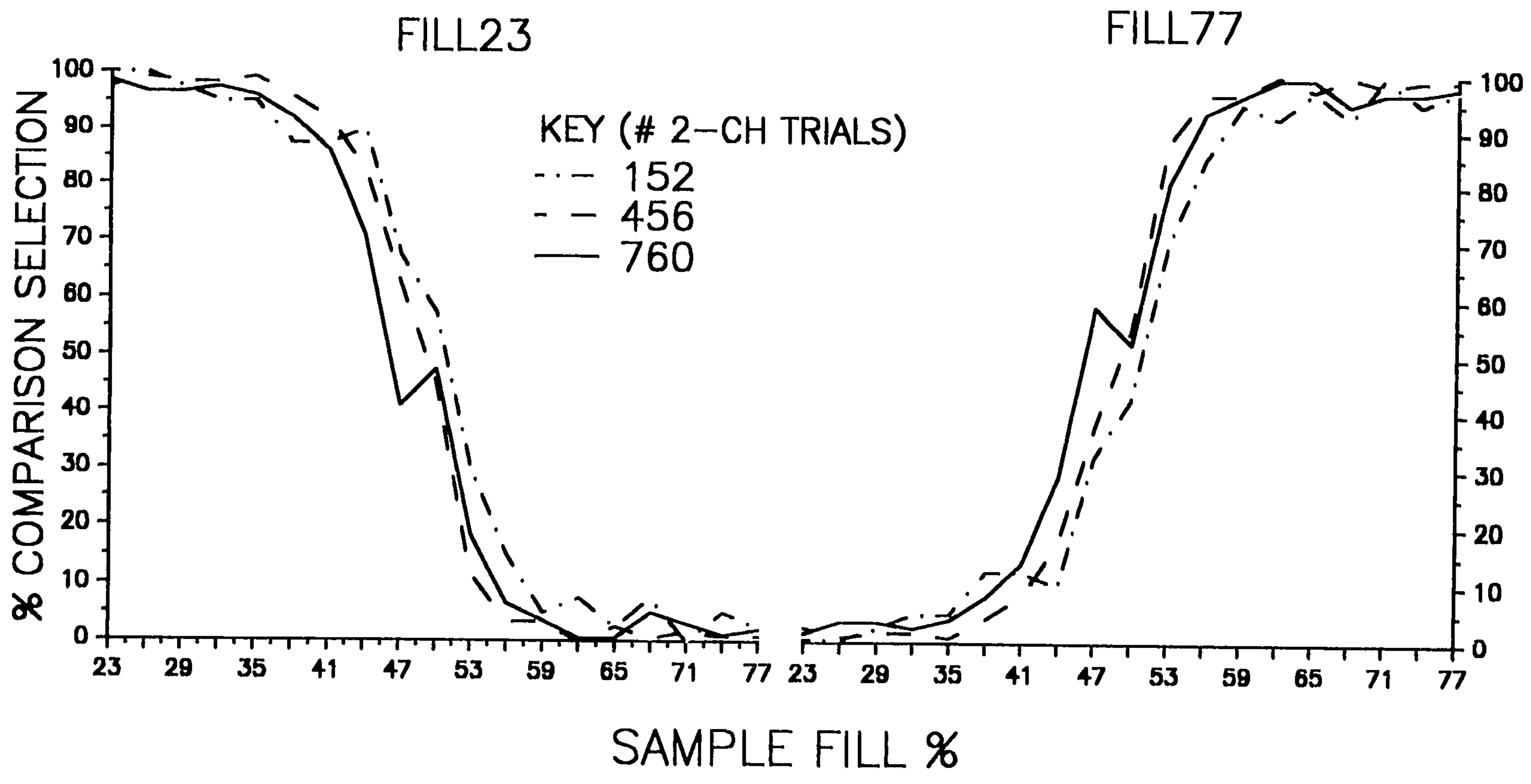


Figure 3. Two-choice primary generalization group test performances plotted as a function of selection of Fill23 and Fill77.

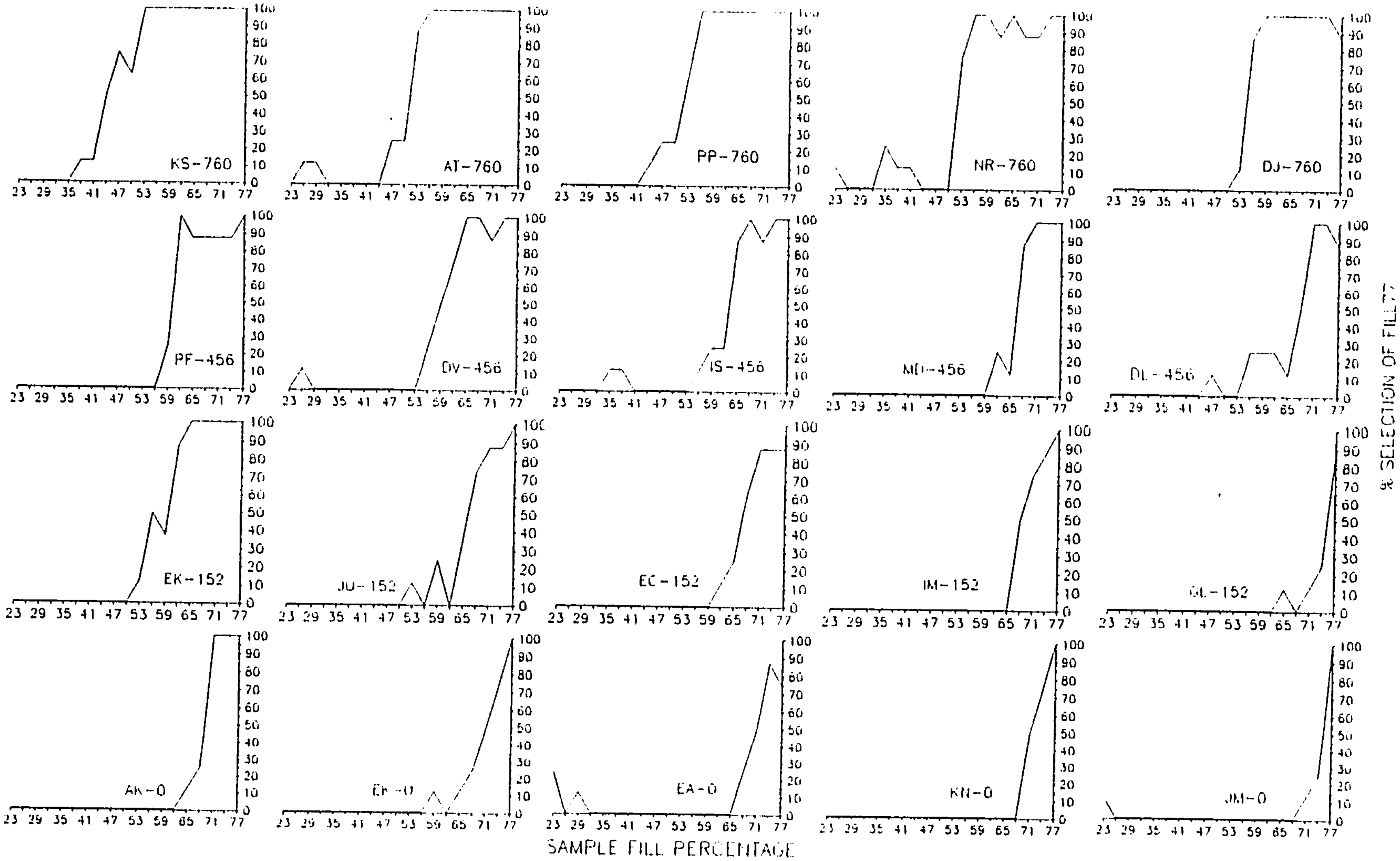


Figure 4. Individual performances for selection of Fill77 during the three-choice generalization test. Depicted from top to bottom are student performances from the 760-, 456-, 152-, and 0-trial groups, respectively.

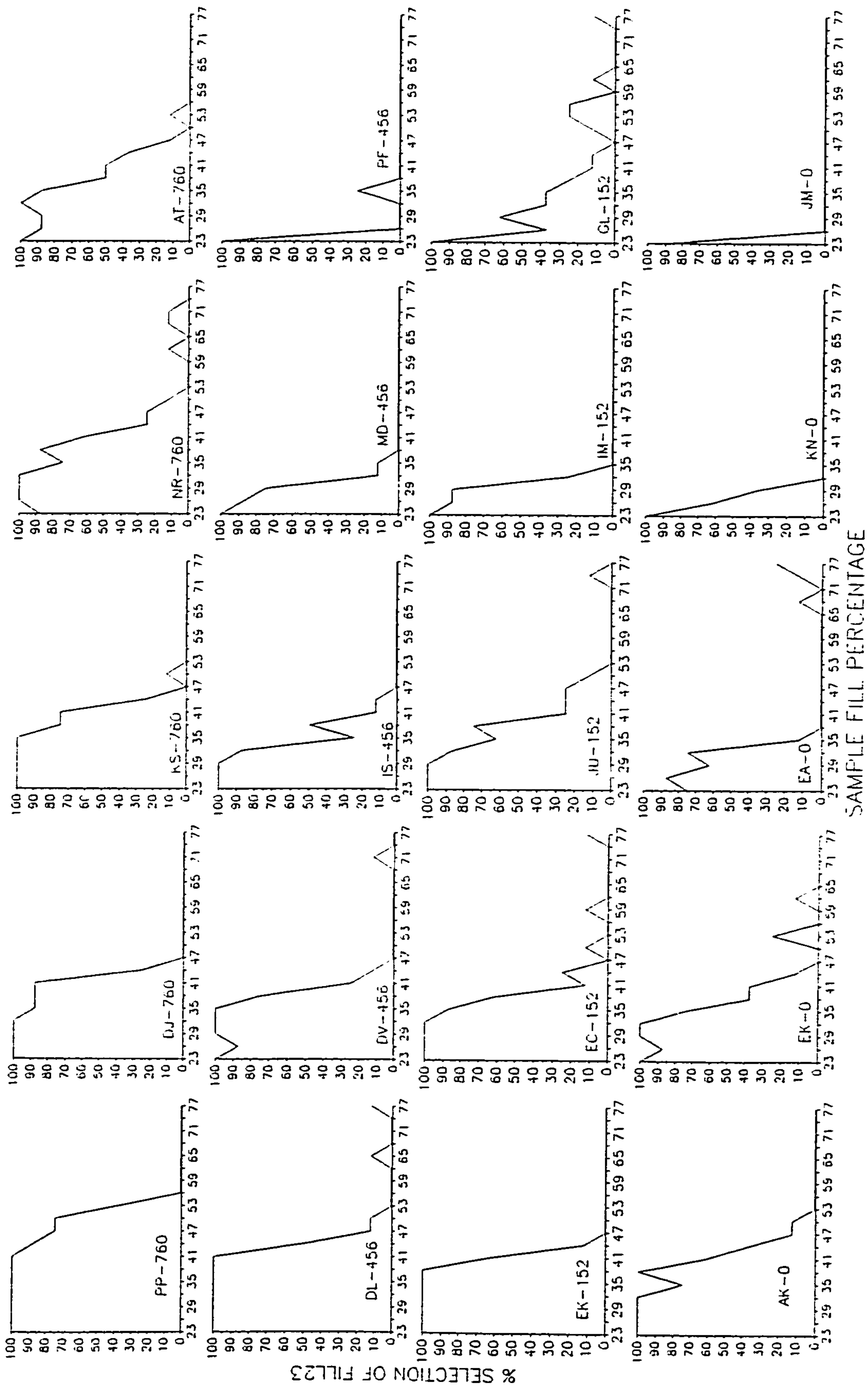


Figure 5. Individual performances for selection of Fill23 during the three-choice generalization test. Depicted from top to bottom are student performances from the 760-, 456-, 152-, and 0-trial groups, respectively.

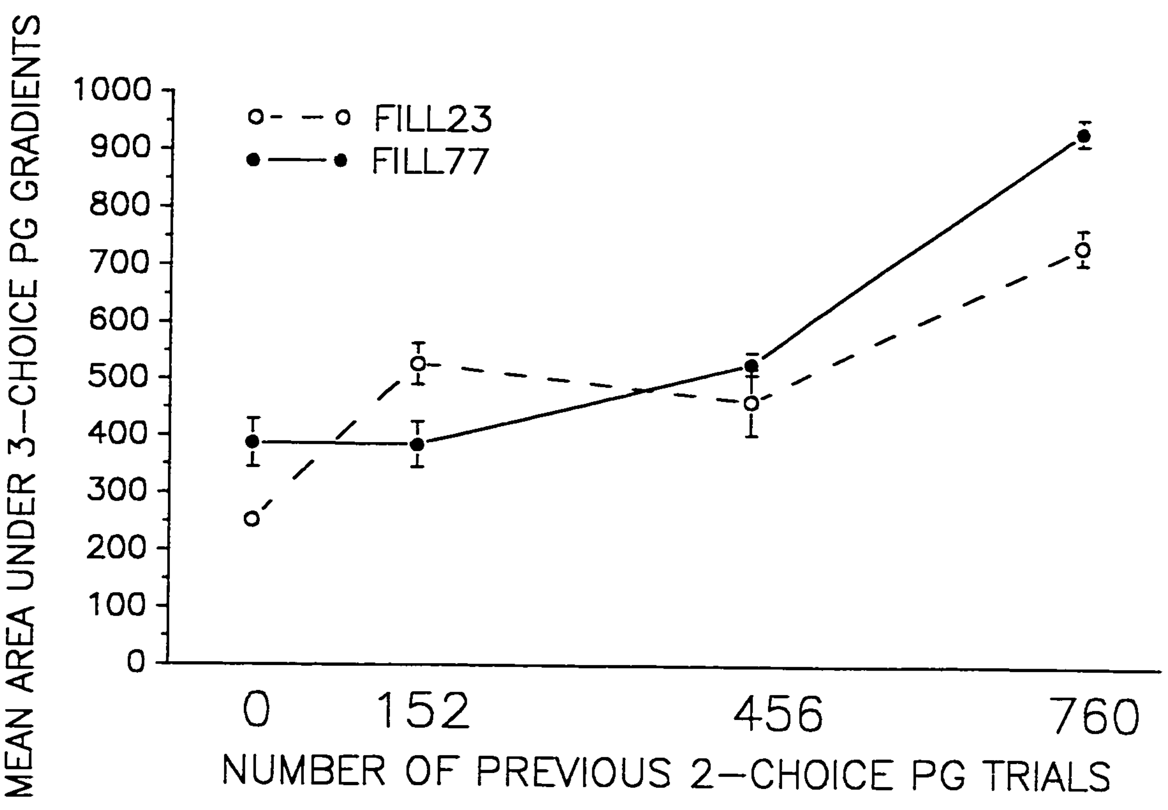


Figure 6. The mean area under the generalization gradients obtained for three-choice generalization testing plotted as a function of the number of prior two-choice generalization trials. Error bars represent plus or minus one standard error of the mean. For one data point (selection of Fill23 by students in the 0-trial group) the standard error was 0.

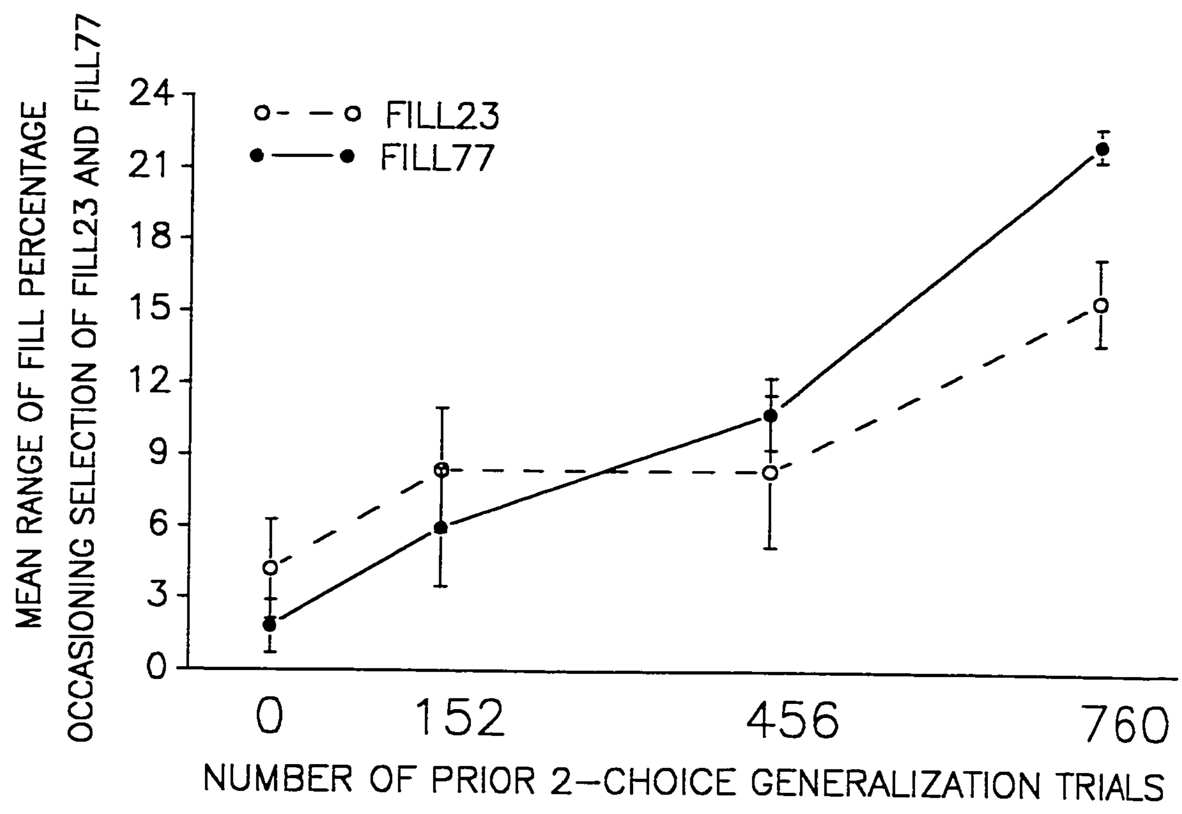


Figure 7. The mean range of fill stimuli that occasioned the near exclusive selection of Fill23 or Fill77.

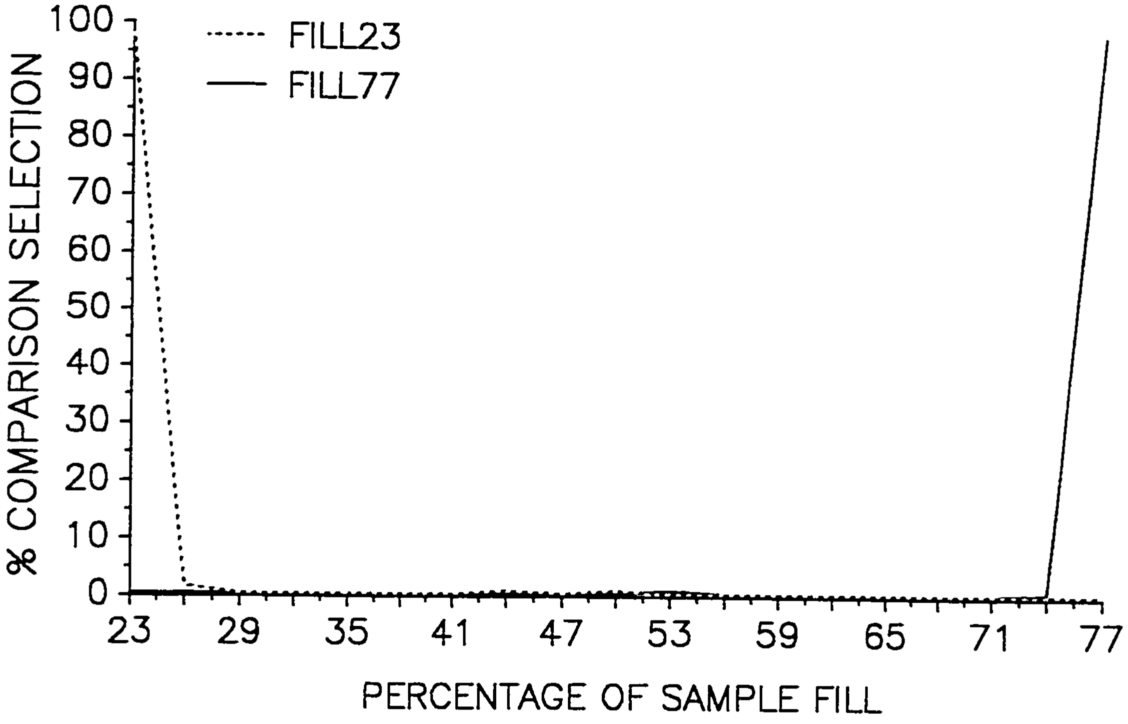


Figure 8. Mean discriminability test functions for selection of Fill23 and Fill77. The left and right functions represent the likelihood of selecting Fill23 or Fill77, respectively.

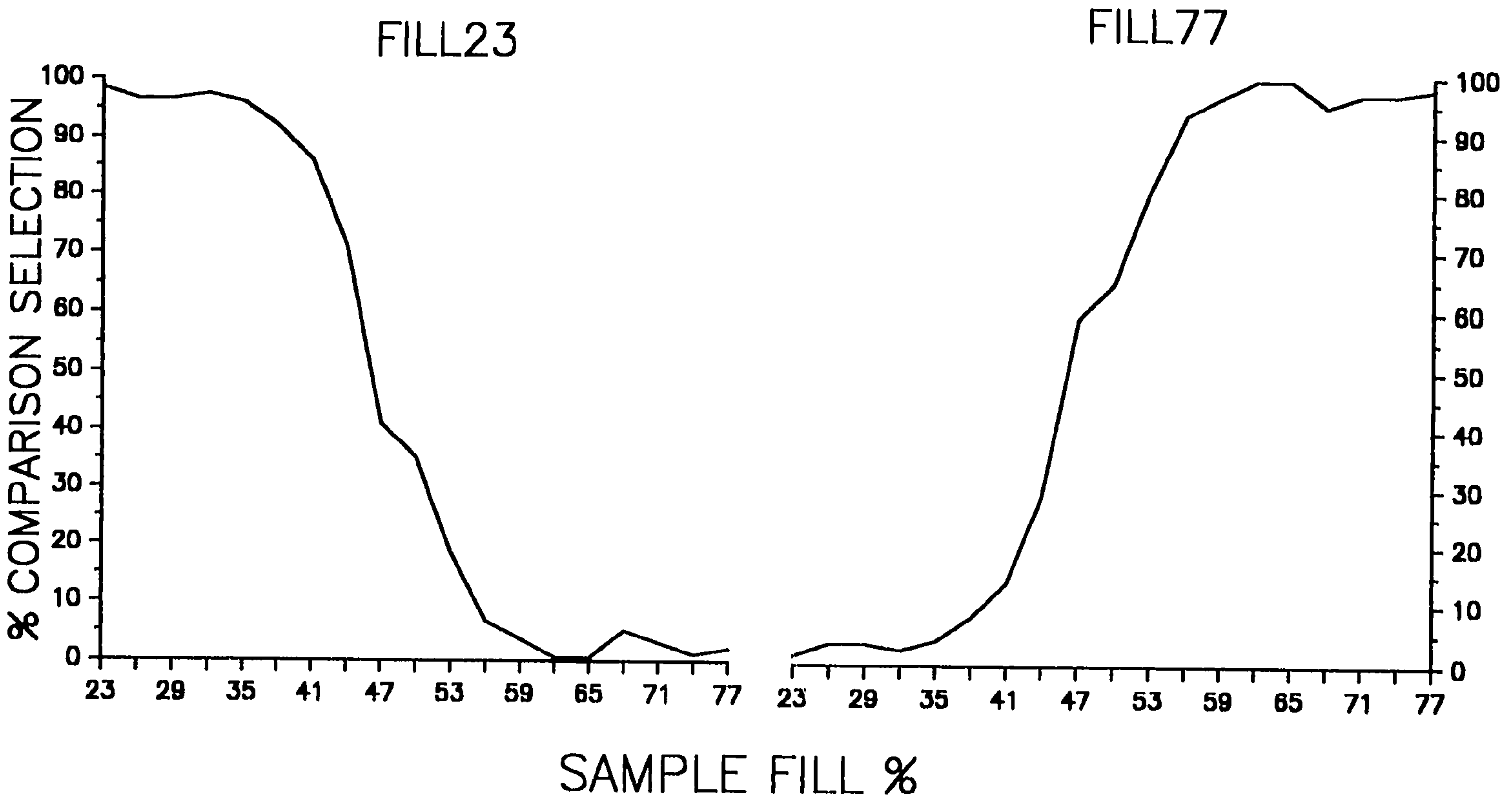


Figure 9. Mean performances obtained during two-choice generalization testing for students in Experiment 2. The left and right functions represent the likelihood of selecting Fill23 or Fill77, respectively.

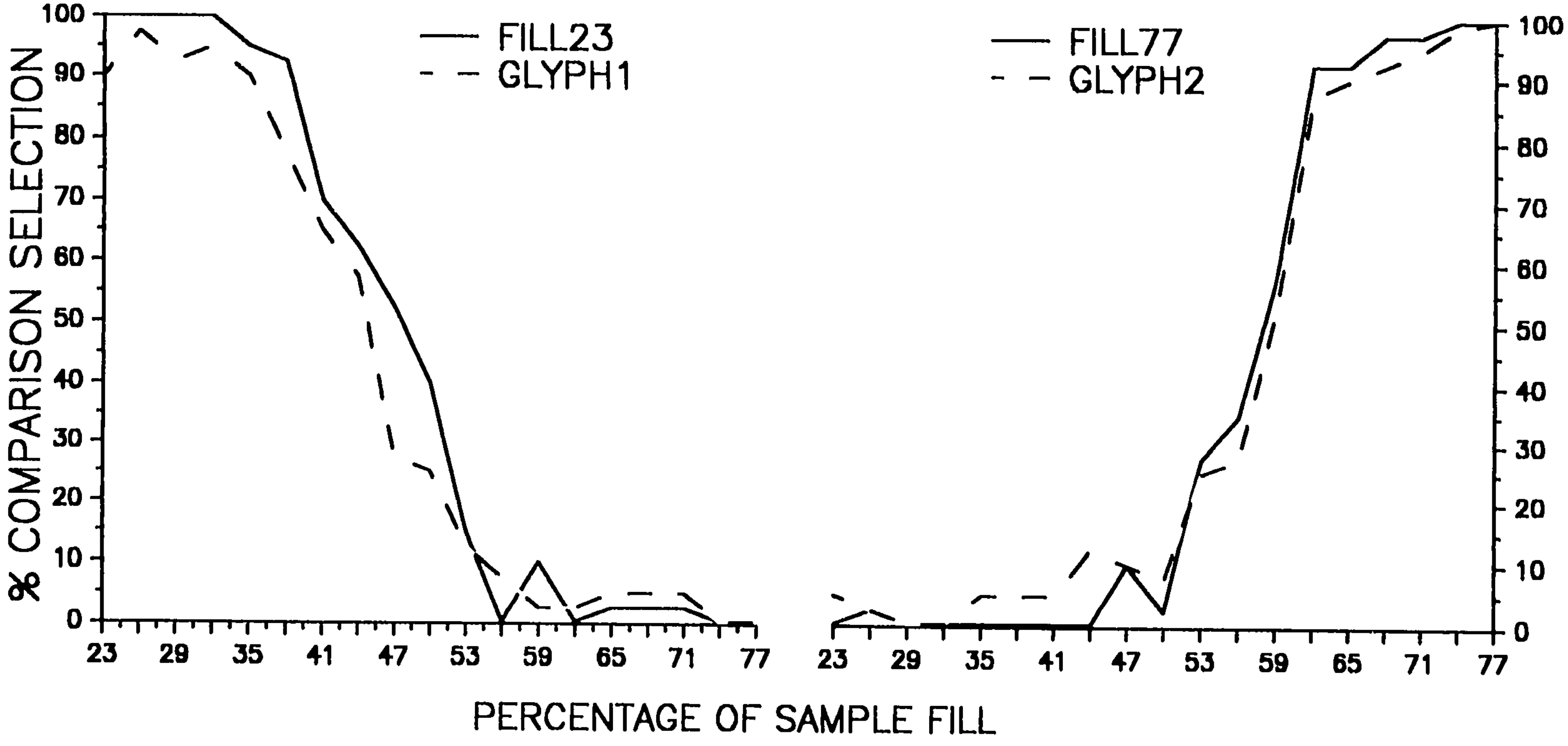


Figure 10. The area under the gradients obtained for selection of Fill23 and Glyph1 (left function), or Fill77 and Glyph2 (right function) during three-choice generalization testing. The solid lines represent selection of Fillx while the dashed lines represent selection of Glyphx.

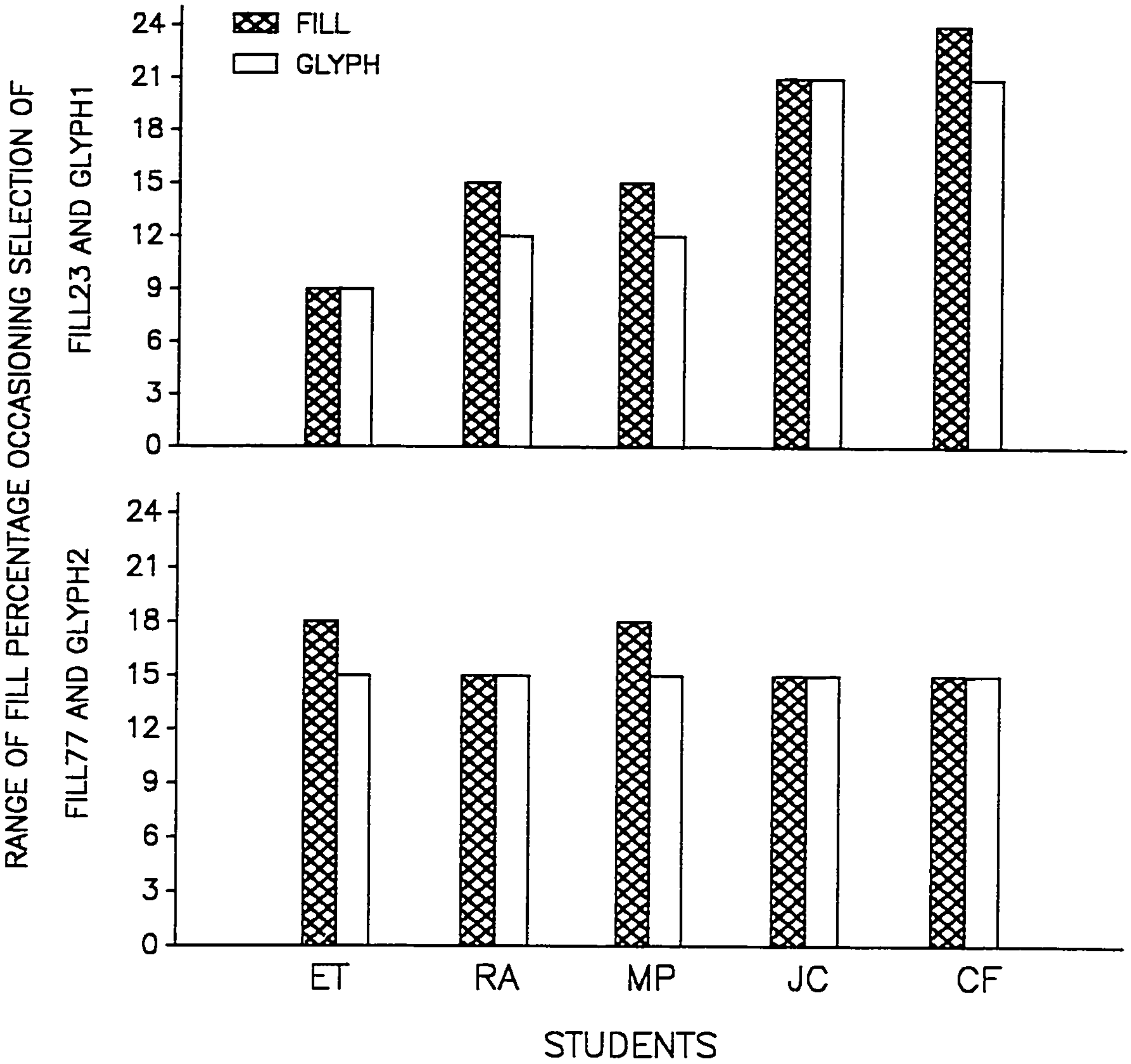


Figure 11. Range of stimuli that occasioned the near exclusive selection of Fill23, Fill77, Glyph1, and Glyph2 for each individual student.

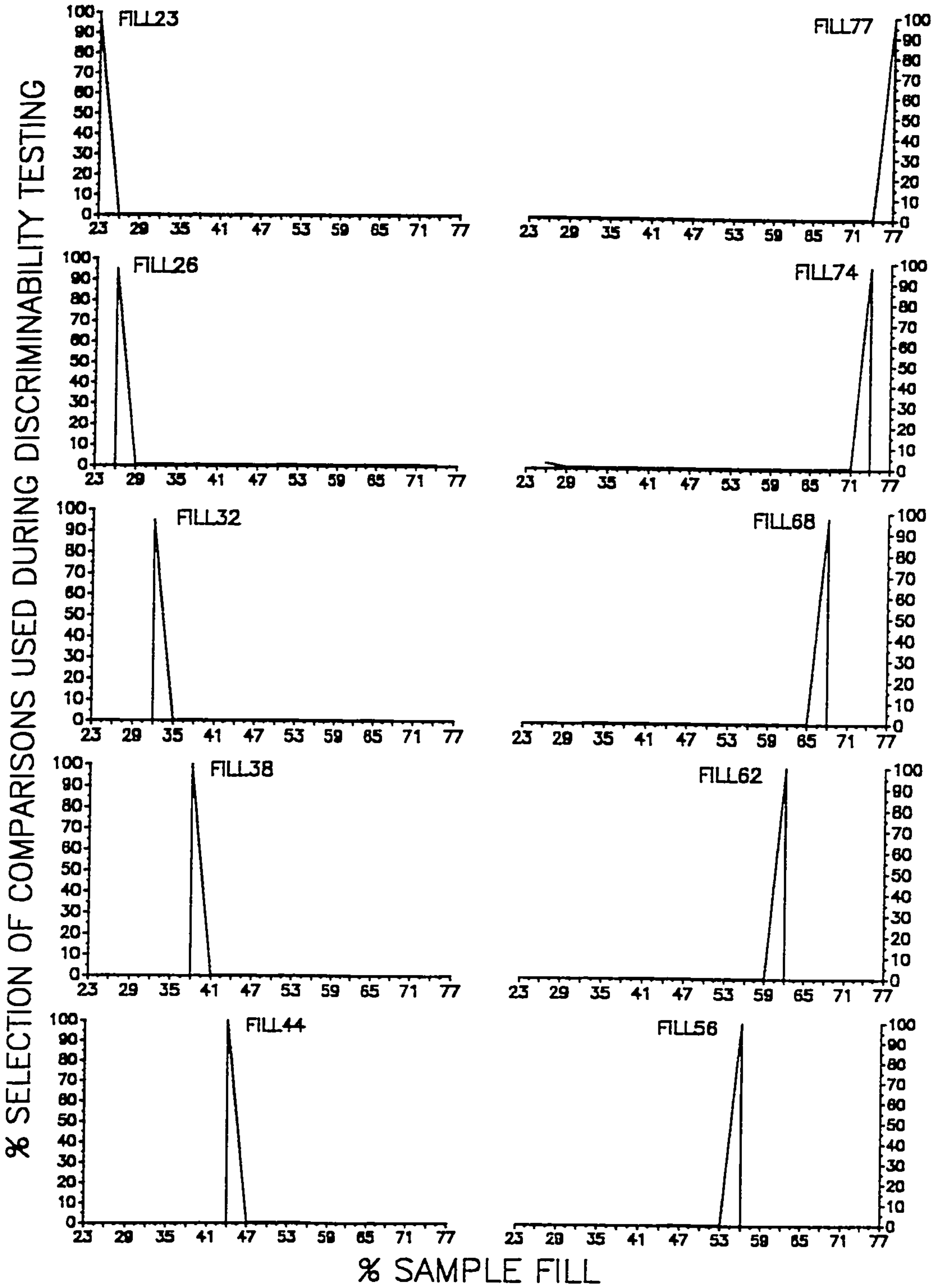


Figure 12. Mean performances for selection of each endpoint value during the last four blocks of each of the five discriminability tests. In the first row, the left and right functions represent the likelihood of selecting Fill23 or Fill77, respectively. In subsequent rows, the left and right functions represent the likelihood of selecting Fill 26 or Fill74, Fill32 or Fill68, Fill38 or Fill62, and Fill44 or Fill56, respectively.

Bibliography

Adams, B. J., Fields, L., & Verhave, T. (1993). Formation of generalized equivalence classes. The Psychological Record, 43, 553-566.

Balsam, P. D. (1988). Selection, representation, and equivalence of controlling stimuli. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.) Steven's handbook of experimental psychology, volume 2: Learning and cognition. New York: Wiley and Sons.

Barnes, D., & Keenan, M. (1993). A transfer of functions through derived arbitrary and nonarbitrary stimulus relations. Journal of the Experimental Analysis of Behavior, 59, 61-82.

Becker, W. (1971). Teaching concepts and operations, or how to make kids smart. In W. Becker (Ed.) An empirical basis for change in education. New York: Science Research Associates.

Bhatt, R. S., Wasserman, E. A., Reynolds, W. F., Jr., & Knauss, K. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. Journal of Experimental Psychology: Animal Behavior Processes, 3, 219-234.

Bhatt, R. S., & Wright, A. A. (1992). Concept learning by monkeys with video picture images and a touch screen. Journal of the Experimental Analysis of Behavior, 57, 219-226.

Blough, D. S. (1990). Form similarity and categorization in pigeon visual research. In M. L. Commons, R. J. Herrnstein, S. Kosslyn, & D. Mumford (Eds.) Models of behavior: Behavioral approaches to pattern recognition and concept formation. Hillsdale, NJ: Erlbaum.

Blough, D., & Blough, P. (1977). Animal psychophysics. In W. K. Honig & J. E. R. Staddon (Eds.), Handbook of operant behavior, (pp. 514-539). Englewood Cliffs, NJ: Prentice Hall.

Bourne, L. E., Dominowski, R. L., & Loftus, E. F. (1979). Cognitive processes. Englewood Cliffs, NJ: Prentice-Hall.

Bush, K. B. (1993). Stimulus equivalence and cross-modal transfer. The Psychological Record, 43, 567-584.

Bush, K. M., Sidman, M., & de Rose, T. (1989). Conditional control of emergent equivalence relations. Journal of the Experimental Analysis of Behavior, 38, 29-46.

Cerella, J. (1979). Visual classes and natural categories in the pigeon. Journal of Experimental Psychology: Human Perception and Performance, 5, 68-77.

Cook, R. G., Wright, A. A., & Kendrick, D. F. (1990). Visual categorization by pigeons. In M. L. Commons, R. J. Herrnstein, S. Kosslyn, & D. Mumford (Eds.) Models of behavior: Behavioral approaches to pattern recognition and concept formation. Hillsdale, NJ: Erlbaum.

Cumming, W. W., & Berryman, R. (1965). The complex discriminated operant: Studies of matching-to-sample and related problems. In D. I. Mostofsky (Ed.), Stimulus generalization (pp. 284-330). Stanford, CA: Stanford University Press.

DeGrandpre, R. J., Bickel, W. K., & Higgins, S. T. (1992). Emergent equivalence relations between interoceptive (drug) and exteroceptive (visual) stimuli. Journal of the Experimental analysis of Behavior, 58, 9-18.

Dinsmoor, J. A. (1995b). Stimulus control: Part II. The Behavior Analyst, 18, 253-270.

Domjam, M. (1998). The principles of learning and behavior. Pacific Grove, CA: Brooks / Cole Publishing.

Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of emotional respondent eliciting and extinction functions through stimulus equivalence classes. Journal of the Experimental Analysis of Behavior, 62, 331-351.

Dougher, M. J., & Markham, M. R. (1994). Stimulus equivalence, functional equivalence and the transfer of function. In S. C. Hayes, L. J. Hayes, M. Sato, & K. Ono (Eds.), Behavior analysis of language and cognition. Reno, NV: Context Press.

Edwards, C. A., & Honig, W. K. (1987). Memorization and "feature selection" in the acquisition of natural concepts in pigeons, Learning and Motivation, 18, 235-260.

Engelmann, S., & Carnine, D. (1982). Theory of instruction: Principles and applications. New York: Irvington.

Fields, L. (1980). Enhanced learning of new discriminations after stimulus fading. Bulletin of the Psychonomic Society, 15, 327-330.

Fields, L., Adams, B. J., Brown, J. L., & Verhave, T. (1993). The generalization of emergent relations in equivalence classes: Stimulus substitutability. Psychological Record, 43, 235-254.

Fields, L., Adams, B. J., Buffington, D. M., Yang, W., & Verhave, T. (1996). Response transfer between stimuli in generalized equivalence classes: A model for the establishment of natural kind and fuzzy superordinate categories. Psychological Record, 46, 665-684.

Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1990). The effects of nodality on the formation of equivalence classes. Journal of the Experimental Analysis of Behavior, 53, 345-358.

Fields, L., Landon-Jimenez, D. V., Buffington, D. M., & Adams, B. J. (1995). Maintained nodal distance effects after equivalence class formation. Journal of the Experimental Analysis of Behavior, 64, 129-146.

Fields, L., & Reeve, K. F. (in press). Synthesizing equivalence classes and natural categories from perceptual and relational classes. In J. C. Leslie & D. Blackman (Eds.). Issues in experimental and applied analyses of human behavior. Reno, NV: Context Press.

Fields, L., Reeve, K. F., Adams, B. J., Brown, J. L., & Verhave, T. (1997). Predicting the extension of equivalence classes from primary generalization gradients: The merger of equivalence classes and perceptual classes. Journal of the Experimental Analysis of Behavior, 68, 67-92.

Fields, L., Reeve, K. F., Adams, B. J., & Verhave, T. (1991). The generalization of equivalence relations. A model for natural categories. Journal of the Experimental Analysis of Behavior, 55, 305-312.

Fields, L., Reeve, K. F., Rosen, D., Varelas, A., Adams, B. J., Belanich, J., & Hobbie, S. A. (1997). Using the simultaneous protocol to study equivalence 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.

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. Journal of the Experimental Analysis of Behavior, 49, 317-332.

Gelman, S. A. (1988a). Children's expectations concerning natural kind categories. Human Development, 31, 28-34.

Gelman, S. A. (1988b). The development of induction within natural kind and artificial categories. Cognitive Psychology, 20, 65-95.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. Cognition, 23, 183-209.

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. Child Development, 58, 1532-1541.

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

Hamilton, W. F. & Coleman, T. B. (1933). Trichromatic vision in the pigeon as illustrated by the spectral hue discrimination curve. Journal of Comparative Psychology, 15, 183-191.

Haring, T. G., Breen, C. G., & Laitinen, R. E. (1989). Stimulus class formation and concept learning: Establishment of within- and between-set generalization and transitive relationships via conditional discrimination procedures. Journal of the Experimental Analysis of Behavior, 52, 13-26.

Hayes, S. C. (1991). A relational control theory of stimulus equivalence. In L. J. Hayes & P. N. Chase (Eds.), Dialogues on verbal behavior (pp. 19-40). Reno, NV: Context Press.

Herrnstein, R. J. (1984). Objects, categories, and discriminative stimuli. In H. L. Roitblatt, T. G. Bever, & H. S. Terrace (Eds.) Animal cognition (pp. 233-261). Hillsdale, NJ: Erlbaum.

Herrnstein, R. J. (1990). Levels of stimulus control: A functional approach. Cognition, 37, 133-166.

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.

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. Science, 146, 549-551.

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. Journal of Experimental Psychology: Animal Behavior Processes, 4, 285-301.

Homa, D., & Chambliss, D. (1975). The relative contributions of common and distinctive information on the abstraction from ill-defined categories. Journal of Experimental Psychology: Human Learning and Memory, 1, 351-359.

Homa, D., Cross, J., Cornell, D., Goldman, D., & Swartz, 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. (1985). The abstraction and long-term retention of ill-defined categories by children. Bulletin of the Psychonomic Society, 23, 325-328.

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

Honig, W. K., & Stewart, K. E. (1988). Pigeons can discriminate locations presented in pictures. Journal of the Experimental Analysis of Behavior, 50, 541-551.

Honig, W. K., & Urcuioli, P. (1981). The legacy of Guttman and Kalish (1956): 25 years of research on stimulus generalization. Journal of the Experimental Analysis of Behavior, 36, 405-445.

Hrycenko, O., & Harwood, D. W. (1980). Judgments of shape similarity in the Barbary dove (Streptopelia risoria). Animal Behavior, 58, 586-592.

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

Jitsumori, M. (1993). Category discrimination of artificial polymorphous stimuli based on feature learning. Journal of Experimental Psychology: Animal Behavior Processes, 19, 244-254.

Jitsumori, M. (1996). A prototype effect and categorization of artificial polymorphous stimuli in pigeons. Journal of Experimental Psychology: Animal Behavior Processes, 22, 405-419.

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

Lashley, K. S., & Wade, M. (1946). The Pavlovian theory of generalization. Psychological Review, 53, 72-87.

Lea, S. E. G. (1984). In what sense do pigeons learn concepts? In H. L. Roitblatt, T. G. Bever, & H. S. Terrace (Eds.) Animal cognition (pp. 263-276). Hillsdale, NJ: Erlbaum.

Lea, S. E. G., & Ryan, C. M. E. (1984). Feature analysis of pigeons' acquisition of concept discrimination. In H. L. Roitblatt, T. G. Bever, & H. S. Terrace (Eds.) Animal cognition (pp. 233-261). Hillsdale, NJ: Erlbaum.

Lea, S. E. G., & Ryan, C. M. E. (1990). Unnatural concepts and the theory of concept discrimination in birds. In M. L. Commons, R. J. Herrnstein, S. Kosslyn, & D. Mumford (Eds.) Models of behavior: behavioral approaches to pattern recognition and concept formation. Hillsdale, NJ: Erlbaum.

Lubow, R. E. (1974). High-order concept formation in the pigeon. Journal of the Experimental Analysis of Behavior, 21, 475-483.

Lynch, D. C., & Green, G. (1991). Development and crossmodal transfer of contextual control of emergent stimulus relations. Journal of the Experimental Analysis of Behavior, 56, 139-154.

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

McIlvane, W. J., Dube, W. V., Green, G., & Serna, R. W. (1993). Programming conceptual and communication skill development. In A. P. Kaiser & D. B. Gray (Eds.), Enhancing children's communication : Research foundations for intervention, volume 2. Baltimore MD: Brookes Publishing.

McIlvane, W. J., & Stoddard, L. T. (1981). Acquisition of matching-to-sample performance in severe retardation: Learning by exclusion. Journal of Mental Deficiency Research, 25, 33-48.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. Annual Reviews of Psychology, 35, 113-138.

Meehan, E. F., & Fields, L. (1995). Contextual control of new equivalence classes. The Psychological Record, 45, 165-182.

Migler, B., & Millenson, J. R. (1969). Analysis of response rates during stimulus generalization. Journal of the Experimental Analysis of Behavior, 12, 81-87.

Neisser, U. (1967). Cognitive psychology. New York: Appleton-Century-Crofts.

Njegovan, M., Ito, S., Mewhort, D., & Weisman, R. (1995). Classification of frequencies into ranges by songbirds and humans. Journal of Experimental Psychology: Animal Behavior Processes, 21, 33-42.

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

Porter, D., & Neuringer, A. (1985). Music discrimination by pigeons. Journal of Experimental Psychology: Animal Behavior Processes, 10, 138-148.

Risley, T. (1964). Generalization gradients following two-response discrimination training. Journal of the Experimental Analysis of Behavior, 7, 199-204.

Rosch, E. H. (1978). Principles of categorization. In E. Rosch & B. B. Lloyds (Eds.), Cognition and categorization. Hillsdale, NJ: Erlbaum.

Rosch, E. H., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.

Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Bream, P. (1976). Basic objects in natural categories. Cognitive Psychology, 8, 382-439.

Saunders, R. R., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1988). The merger and development of equivalence classes by unreinforced conditional selection of comparison stimuli. Journal of the Experimental Analysis of Behavior, 50, 145-162.

Saunders, R. R., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1990). Long term stability of equivalence relations in the absence of training or practice. American Journal of Mental Retardation, 95, 291-303.

Sidman, M. (1994). Equivalence relations and behavior: A research story. Boston, MA: Authors Cooperative, Inc.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: an expansion of the testing paradigm. Journal of the Experimental Analysis of Behavior, 37, 5-22.

Sidman, M., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. Journal of the Experimental Analysis of Behavior, 52, 261-274.

Smith, E. A. (1989). Concepts and induction. In Michael I. Posner (Ed.), Foundations of cognitive science. Cambridge, MA: MIT Press.

Spradlin, J. E., Saunders, K. J., & Saunders, R. R. (1992). The stability of equivalence classes. In S. C. Hayes & L. B. Hayes (Eds.), Understanding verbal relations (pp. 29-42). Reno NV: Context Press.

Stemmer, N. (1980). Natural concepts and generalization classes. The Behavior Analyst, 3, 41-48.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. Journal of Applied Behavior Analysis, 10, 349-367.

Stromer, R., & Mackay, H. A. (1997). Naming and the formation of stimulus classes. In T. R. Zentall & P. M. Smeets (Eds.), Stimulus class formation in humans and animals. Amsterdam, NL: North-Holland.

Urcuioli, P. J., & Nevin, J. A. (1975). Transfer of hue matching in pigeons. Journal of the Experimental Analysis of Behavior, 24, 149-155.

Vaughan, W., Jr. (1988) Formation of equivalence sets in pigeons. Journal of Experimental Psychology: Animal Behavior Processes, 14, 36-42.

Wasserman, E. A., & DeVolder, C. L. (1993). Similarity and nonsimilarity-based conceptualization in children and pigeons. The Psychological Record, 43, 779-794.

Wasserman, E. A., Gagliardi, J. L., Cook, B. R., Kirkpatrick-Steger, K., Astley, S. L., & Biederman, I. (1996). The pigeon's recognition of drawings of depth-rotated stimuli. Journal of Experimental Psychology: Animal Behavior Processes, 22 (2), 205-221.

Wasserman, E. A., Kiedinger, R. E., & Bhatt, R. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. Journal of Experimental Psychology: Animal Behavior Processes, 3, 235-246.

Williams, B. (1984). Stimulus control and associative learning. Journal of the Experimental Analysis of Behavior, 42, 469-484.

Wittgenstein, L. (1968). Philosophical investigations (3rd Edition). Translated from the German by G. E. M. Anscombe. New York: Macmillan.

Wright, A. A. (1972). The influence of ultraviolet radiation on the pigeon's color discrimination. Journal of the Experimental Analysis of Behavior, 17, 325-337.

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

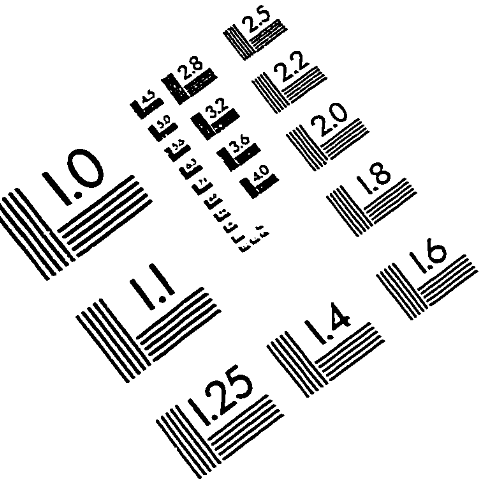
Wright, A. A. & Cumming, W. W. (1971). Color-naming functions for the pigeon. Journal of the Experimental Analysis of Behavior, 15, 7-17.

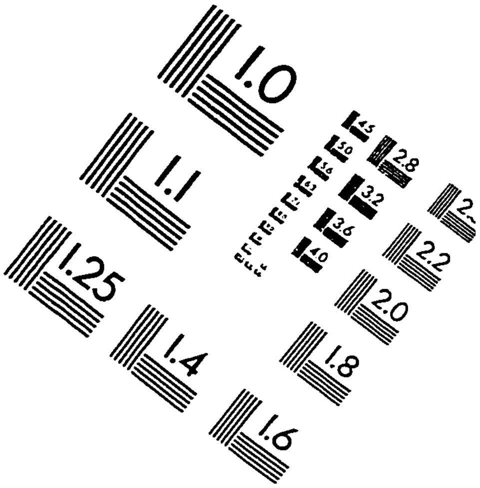
Wyckoff, L. B. (1952). The role of observing responses in discrimination learning. Part 1. Psychological Review, 66, 68-78.

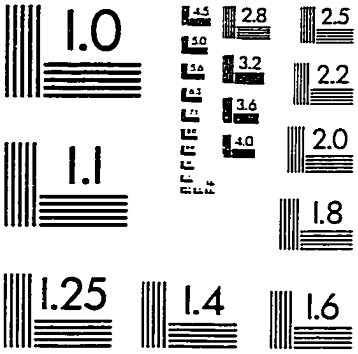
Zentall, T. R., Hogan, D. E., & Edwards, C. A. (1984). Cognitive factors in conditional learning by pigeons. In H. L. Roitblatt, T. G. Bever, & H. S. Terrace (Eds.) Animal cognition, (pp. 389-405). Hillsdale, NJ: Lawrence Earlbaum Associates.

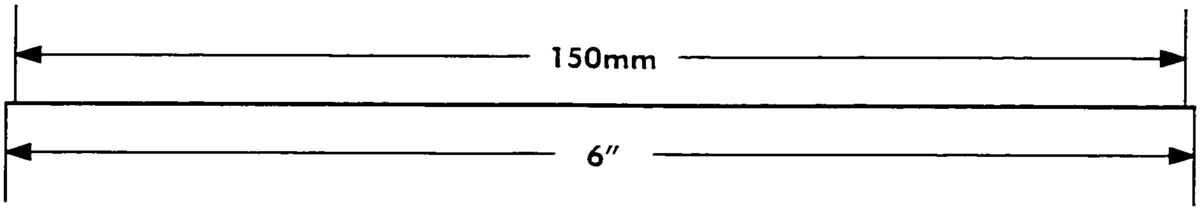
Zentall, T. R., Jackson-Smith, P., Jagielo, J. A. (1990). Categorical color and shape coding by pigeons. In M. Commons, R. J. Herrnstein, S. M. Kosslyn, & D. B. Mumford (Eds.), Quantitative analyses of behavior, vol. VIII: Behavioral approaches to pattern recognition and concept formation, (pp. 3-22). Hillsdale, NJ: Lawrence Earlbaum Associates.

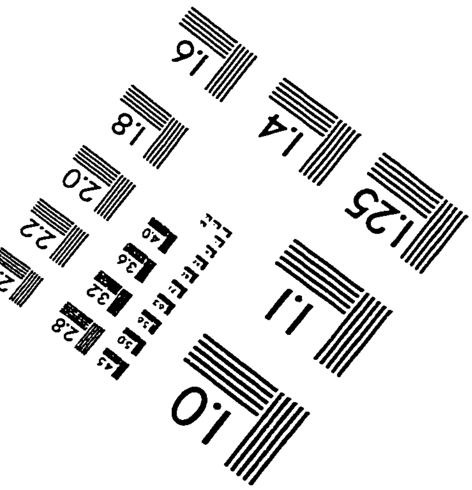
IMAGE EVALUATION TEST TARGET (QA-3)

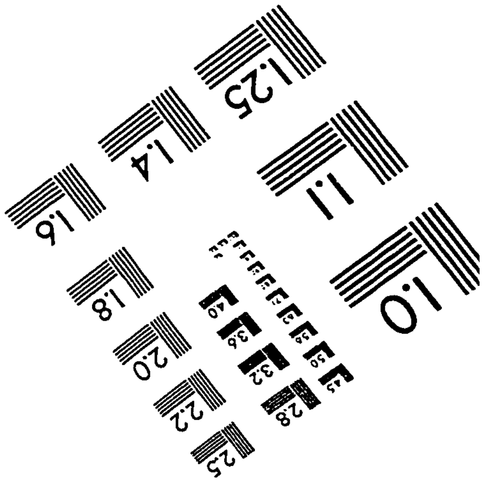












APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989