

## **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.**

**ProQuest Information and Learning  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
800-521-0600**

**UMI<sup>®</sup>**



**ESTABLISHMENT OF TRANSITIVITY-INDICATIVE PERFORMANCES VIA  
MULTIPLE EXEMPLAR TRAINING AND STAGED TESTING**

by

**ANTONIOS VARELAS**

**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**

**2002**

UMI Number: 3063892

Copyright 2002 by  
Varelas, Antonios

All rights reserved.

**UMI<sup>®</sup>**

---

UMI Microform 3063892

Copyright 2002 by ProQuest Information and Learning Company.  
All rights reserved. This microform edition is protected against  
unauthorized copying under Title 17, United States Code.

---

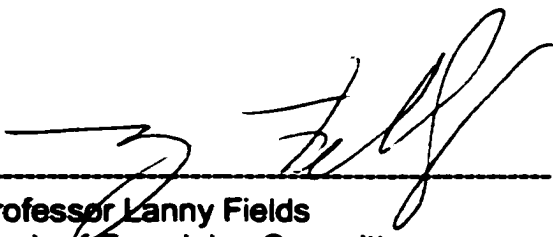
ProQuest Information and Learning Company  
300 North Zeeb Road  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

**© Copyright 2002**  
**ANTONIOS VARELAS**  
**All Rights Reserved**

**Approval Page**

**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.**

8/24/02  
Date

  
\_\_\_\_\_  
**Professor Lanny Fields**  
**Chair of Examining Committee**

8/24/02  
Date

  
\_\_\_\_\_  
**Professor Joseph Glick**  
**Executive Officer**

**Professor Bruce L. Brown**

**Professor Nancy Hemmes**

**Professor Peter Balsam**

**Professor Philip N. Himeline**

\_\_\_\_\_  
**Supervisory Committee**

**THE CITY UNIVERSITY OF NEW YORK**

**Abstract****ESTABLISHMENT OF TRANSITIVITY-INDICATIVE PERFORMANCES VIA  
MULTIPLE EXEMPLAR TRAINING AND STAGED TESTING**

by

**Antonios Varelas****Adviser: Professor Lanny Fields**

**The differential emergence of transitivity-indicative performances was observed following exposure to multiple-exemplar transitivity training (MET) and/or staged testing (ST). Twenty-four college students who failed to demonstrate transitivity-indicative performances in an initial test served as experimental participants. The initial test included the presentation of six stimulus sets drawn from three visual domains (abstract 'glyphs', satellite images and nonsense syllables) in one test block (concurrent test). Participants were then randomly assigned to one of four groups. Participants in Group 1 did not receive any intervention. Participants in Group 2 (MET) received transitivity training with 20 new glyph sets. Participants in Group 3 (ST) were re-exposed to the same stimuli that were presented in the initial test, but in a series of test blocks. Each test block, contained stimuli from one domain (staged test). Participants in Group 4 (MET & ST) received both transitivity training and the staged test. Participants in all groups were then re-exposed the initial concurrent test without further training with these stimuli. When performances during the initial and final test were compared, participants in Group 1 showed no change in the number of sets that showed transitivity-indicative performances. Participants in Groups 2 and 3 showed moderate**

increases and participants in Group 4 showed much larger increases. The improvement shown by Groups 2 and 3, however, varied with stimulus domain. The glyph domain showed the lowest level of emergence, the satellite domain showed moderate levels of emergence and the nonsense syllable domain showed the highest level of emergence. These results suggest that the emergence of transitivity-indicative performances was related to the intervention used and the nameability or perceptual organization of the stimuli in each of the stimulus domains.

## **Acknowledgements**

**I am eternally indebted to Dr. Lanny Fields for his endless patience, support and guidance throughout the course of this research and my entire academic life at The City University of New York. Gratitude is also extended to Drs. Bruce Brown and Nancy Hemmes for their valuable guidance during that time and to Drs. Peter Balsam and Philip Hine for their insightful comments and observations on my dissertation. I would also like to thank fellow Learning Labbies Jim, Ken, Devorah and Priya, for putting up with me over the years and helping make coming to work a blast.**

**Most importantly I would like to thank my wife Carol. Without her unwavering love and support, this endeavor would not have been possible.**

**Sagapo Koukla Mou!**

**Table of Contents**

<b>Approval Page.....</b>	<b>iii</b>
<b>Abstract.....</b>	<b>iv</b>
<b>Acknowledgments.....</b>	<b>vi</b>
<b>Table of Contents.....</b>	<b>vii</b>
<b>List of Tables.....</b>	<b>viii</b>
<b>List of Figures.....</b>	<b>ix</b>
<b>Introduction.....</b>	<b>1</b>
<b>Method.....</b>	<b>10</b>
<b>Results.....</b>	<b>25</b>
<b>Discussion.....</b>	<b>50</b>
<b>Tables.....</b>	<b>60</b>
<b>Figures.....</b>	<b>71</b>
<b>Bibliography.....</b>	<b>80</b>

### List of Tables

<b>Table 1.</b>	<b>Design of Experiment.....</b>	<b>60</b>
<b>Table 2.</b>	<b>Stimuli Used During Keyboard Familiarization (Phase 1).....</b>	<b>61</b>
<b>Table 3.</b>	<b>Sequence of Training and Test Trials for Preliminary Testing (Phase 2).....</b>	<b>62</b>
<b>Table 4.</b>	<b>Sequence of Training and Test Trials for Transitivity Training (Phase 3).....</b>	<b>63</b>
<b>Table 5.</b>	<b>Sequence of Training and Test Trials for Staged Test and Posttest (Phase 4) .....</b>	<b>67</b>
<b>Table 6.</b>	<b>Sequence of Training and Test Trials for No Staged Test and Posttest (Phase 5).....</b>	<b>68</b>
<b>Table 7.</b>	<b>Preliminary Test and Posttest Performances for All Participants.. .....</b>	<b>69</b>
<b>Table 8.</b>	<b>Number of Transitivity Failures During Preliminary Testing.....</b>	<b>70</b>

### List of Figures

<b>Figure 1.</b>	<b>Stimuli Used in Preliminary Testing (Phase 2).....</b>	<b>71</b>
<b>Figure 2.</b>	<b>Stimuli Used During Transitivity Training (Phase 3).....</b>	<b>72</b>
<b>Figure 3.</b>	<b>Mean Emergence of Transitivity-Indicative Performances per Group.....</b>	<b>73</b>
<b>Figure 4.</b>	<b>Mean Difference on Baseline Trials and Transitivity Probes per Group.....</b>	<b>74</b>
<b>Figure 5.</b>	<b>Emergence of Transitivity for Individual Participants.....</b>	<b>75</b>
<b>Figure 6.</b>	<b>Mean Difference on Baseline Trials and Transitivity Probes for Individual Participants.....</b>	<b>76</b>
<b>Figure 7.</b>	<b>Emergence of Transitivity-Indicative Performances Per Domain.. .....</b>	<b>77</b>
<b>Figure 8.</b>	<b>Individual Participant Performances During Glyph Training (Phase 4).....</b>	<b>78</b>
<b>Figure 9.</b>	<b>Individual Participant Performances During Staged Testing (Phase 5) .....</b>	<b>79</b>

**ESTABLISHMENT OF TRANSITIVITY-INDICATIVE PERFORMANCES VIA  
MULTIPLE EXEMPLAR TRAINING AND STAGED TESTING**

by

**ANTONIOS VARELAS**

**Equivalence Classes. An equivalence class is a finite group of physically disparate stimuli, all of which become interchangeable after the establishment of a small set of conditional discriminations (Fields & Verhave, 1987; Sidman & Tailby, 1982). For example, with stimuli represented by the letters A, B and C, a matching-to-sample procedure can be used to train baseline conditional discriminations such as AB, and BC. Thereafter, the emergence of an equivalence class is assessed by the presentation of the remaining set of untrained stimulus pairs that can be derived from the trained relations. These probes assess the properties of symmetry, transitivity and the combined properties of symmetry and transitivity and are collectively called emergent-relations probes. Symmetry probes are generated by reversing the sample and comparison functions of the stimuli in a specific baseline relation. In this case, the symmetry probes would be BA and CB. Transitivity probes are generated by taking the sample from one baseline relation and comparison stimuli from a different baseline relation. The stimuli are related to each other by linkage to a common stimulus that served as a comparison and sample in training. Thus, the transitivity probe in this example would be AC. Finally equivalence probes assess the properties of symmetry and transitivity in combination. Thus, the equivalence probe in this example is CA. If class-consistent responding was**

occasioned by all of the emergent relations probes, the stimuli are functioning interchangeably, and are members of an equivalence class.

Equivalence classes have been formed in laboratory settings using a variety of training and testing protocols. Some procedures involve the serial introduction of the baseline conditional discriminations and emergent-relations probes. One such procedure, called the simple-to-complex protocol is maximally programmed. It involves the training of each baseline relation separately and the presentation of each emergent-relations probe as soon as all of its prerequisites have been demonstrated. The simple-to-complex protocol leads to the reliable emergence of equivalence classes. Across experiments, classes emerge between 93-100% of participants (Adams, Fields & Verhave, 1993; Arntzen & Holth, 2000; Fields, Reeve, Adams & Verhave, 1991; Lynch & Cuvo, 1995; Smeets, Leader & Barnes, 1997).

Protocols on the other end of the procedural continuum involve the presentation of all baseline conditional relations in a single block of randomly presented trials until all of the baseline conditional relations are learned. This is then followed by the random presentation of all emergent relations probes in a single test block. This procedure, called the simultaneous protocol, is relatively unlikely to produce equivalence classes. Only 58% of college students show the emergence of 1-node 3-member classes (Buffington, Fields and Adams, 1997) and approximately 20% of college students show the emergence of 3-node 5-member classes (Fields, Landon-Jimenez, Buffington & Adams, 1995; Fields,

Reeve, Rosen, Varelas, Adams, Belanich & Hobbie, 1997; Smeets et al., 1997) under the simultaneous protocol.

In addition to the actual training and testing protocol, the likelihood of class formation is influenced by a variety of preliminary training procedures that precede equivalence-class training. For example, Adams, Fields and Verhave (1999) found that the prior formation of unrelated conditional discriminations interfered with the subsequent formation of new equivalence classes under the simple-to-complex protocol. Wulfert, Dougher and Greenway (1991) however, found that teaching participants to respond to relations between the perceptual features of stimulus arrays enhanced the subsequent formation of new equivalence classes also under the simple-to-complex protocol.

Buffington et al. (1997) and Fields et al. (1997) showed that formation of equivalence classes under the simple-to-complex protocol enhanced the subsequent formation of new equivalence classes under the simultaneous protocol. This enhancement effect observed during the simultaneous protocol was a direct function of the size and number of nodes in the previously established equivalence classes under the simple-to-complex protocol.

Transitivity and Enhanced Equivalence Class Formation. A recent study by Fields, Varelas, Reeve, Belanich, Wadhwa, DeRosse and Rosen (2000) showed that the enhancement effects observed in the Buffington et al. (1997) and Fields et al. (1997) studies were not produced by the establishment of equivalence classes during the preliminary training. Rather, Fields et al. (2000) demonstrated that the subsequent formation of new equivalence classes under

the simultaneous protocol was due to exposure to components of the simple-to-complex procedure during preliminary training. They found that a very low percentage of participants showed the emergence of new equivalence classes under the simultaneous protocol when no preliminary training was conducted or when preliminary training consisted of the establishment of baseline conditional discriminations alone. Preliminary training that included symmetry testing alone resulted in only a small increase in the percentage of participants who showed the emergence of equivalence classes under the simultaneous protocol. The largest increase in the percentage of participants who showed the emergence of new equivalence classes under the simultaneous protocol followed the preliminary training that included either (a) transitivity testing alone, (b) transitivity and symmetry testing and (c) transitivity, symmetry and equivalence testing. The results of the Fields et al. (2000) study showed that transitivity was the factor most responsible for maximizing the emergence of new equivalence classes under the simultaneous protocol.

The importance of transitivity for equivalence class formation can also be seen by studying the performances responsible for the failure of equivalence class formation. Work by Fields et al. (2000) and Fields, Adams, Newman and Verhave (1992), showed that participants who did not form equivalence classes under the simple-to-complex protocol, failed to demonstrate transitivity even though they had established the requisite baseline relations and demonstrated symmetry.

The studies cited have shown that the probability of establishing new equivalence classes was enhanced if a transitivity-indicative repertoire was present prior to equivalence class formation. Furthermore, the failure to establish new equivalence classes occurs if this transitivity-indicative repertoire was absent prior to the establishment of those classes. Thus, the presence of a transitivity-indicative repertoire prior to equivalence class formation was clearly important for the establishment of equivalence classes. No studies, however, have examined the parameters that influence the establishment of a transitivity-indicative repertoire.

Establishment of Transitivity Repertoire. Transitivity-indicative performances are demonstrated when a stimulus from a conditional discrimination is selected in the presence of another stimulus from another conditional discrimination, even though they have never before been paired together. The stimuli in a transitivity probe are related to each other by linkage to a common nodal stimulus that served two functions in training. The nodal stimulus served as a comparison in one baseline relation and a sample in another baseline relation (Fields & Verhave, 1987). To test for transitivity-indicative performances using a matching-to-sample procedure, in the presence of stimulus A serving as a sample, a reinforcer is provided for the selection of comparison stimulus B. Furthermore, in the presence of stimulus B serving as a sample, a reinforcer is provided for the selection of comparison stimulus C. Transitivity is demonstrated if the presence of stimulus A serving as the sample, occasions the selection of stimulus C, serving as a comparison.

In the above example, a participant may be successful in demonstrating transitivity-indicative performances with one stimulus set. To say that a participant has established a generalized transitivity repertoire, however, the participant must respond in accordance with transitivity in a spontaneous fashion in the presence of many novel stimulus sets that are discriminable from those used in training. For example, after learning one set of linked baseline conditional discriminations (AB and BC), a participant may respond correctly to the corresponding transitivity probe (AC). Demonstrating transitivity with this one set, however, does not imply that transitivity-indicative performances will be occasioned by a new set of linked conditional discriminations. For example, after training linked baseline conditional discriminations for the novel stimulus set XY and YZ, will the participant respond correctly to the new transitivity probe, XZ? To date, no study has identified the training conditions that will establish generalized transitivity-indicative performances. Thus, the primary purpose of the present study was to identify the variables that would induce a generalized transitivity repertoire.

Transitivity and Stimulus Class Formation. A potential answer to this question can be adduced by viewing a generalized transitivity repertoire from the vantage point of stimulus class formation. A stimulus class is a set of stimuli, all of which evoke a particular response while stimuli in other sets evoke alternative responses (Adams et al., 1993; Herrnstein, 1979, 1990; Herrnstein & deVilliers, 1980, Keller & Schoenfeld, 1950; Lea, 1984, Malott & Siddall, 1972). Furthermore, a response trained to occur in the presence of one or a few stimuli

in the set will be evoked by all other members of that same set and not by the stimuli in other sets. In addition, if stimulus classes are being studied in the context of a matching-to-sample format, each stimulus within a set will occasion the selection of all the other members within the same set and not occasion the selection of stimuli from alternative sets (Adams et al., 1993; Sidman, 1994; Sidman & Tailby, 1982). Members of the same class that have never before been presented together will also occasion these performances (Barnes-Holmes, Barnes-Holmes, Roche & Smeets, 2001a, 2001b; Boelens & Van Den Broek, 2000; Dymond & Barnes, 1995, 1996; Fields & Reeve, 2000; Hayes & Hayes, 1989; Premack, 1971, 1976; Smoke, 1932; Wright, Cook & Kendrick, 1989).

Following this analysis, it is plausible to consider a generalized transitivity repertoire as a type of response class that comes under the control of a stimulus class that consists of any pair of linked conditional discriminations. To illustrate, the sets of linked baseline conditional discriminations (e.g. AB-BC, XY-YZ) would be functioning as members of a class of transitive relations provided a participant responds correctly to the corresponding transitivity probes (AC and XZ respectively). Thus, each set of linked baseline conditional discriminations would be functioning as a member of the class of transitive relations. Furthermore, this class of transitive relations could be demonstrated by an infinite number of sets of linked conditional discriminations.

The most common method used to establish stimulus classes involves exposing participants to many positive and negative exemplars during training. A number of studies show that presentation of many positive and negative

exemplars during training increased the likelihood establishing of perceptual classes (e.g. Bhatt, Wasserman, Reynolds & Knauss, 1988; Boelens & Van Den Broek, 2000; Cook, Wright & Kendrick, 1990; Homa & Chambliss, 1975; Homa, Cross, Cornell, Goldman & Schwartz, 1973; Homa & Little, 1985; Homa, Sterling & Treple, 1981). When training is conducted with many variations of S+'s and S-'s, the concept-defining features are present in the S+'s and are not present in the S-'s. Additionally, irrelevant features appear with equal frequency in both the S+'s and S-'s. Because these irrelevant features are not predictive of reinforcement, the likelihood of responding in their presence decreases. Furthermore, the likelihood of responding increases in the presence of the stimulus containing the concept-defining features because they are solely predictive of reinforcement (Fields & Reeve, 2000; Homa, 1978; Homa et al., 1973; Homa & Vosburgh, 1976; Lea, 1984). Therefore, exposing participants to many positive and negative exemplars during training allows responding to come under the control of the relational and not the absolute properties of the stimuli. Exposure to many stimuli during training also decreases the likelihood that any particular irrelevant feature (or features) that is present in one exemplar will control behavior.

Multiple-exemplar training results in the emergence of stimulus classes. Thus, The first purpose of the current experiment was to determine whether transitivity-indicative performances could be established with multiple-exemplar training. That is, would multiple-exemplar training of a transitivity repertoire, result in the emergence of transitivity-indicative performances in the presence of

many new sets of linked conditional discriminations? This was accomplished by exposing half of the participants to multiple-exemplar training of the transitivity repertoire and the other participants to no training of the repertoire. Differences in performance between the groups would index the effect of multiple-exemplar training on the emergence of transitivity-indicative repertoire.

Another variable that increases the likelihood of stimulus class formation is the repetition of test trials. A number of studies show that the emergence of equivalence classes is a direct function of the repeated presentations of emergent relations test trials (e.g. Devaney, Hayes & Nelson, 1986; Fields, Adams, Newman & Verhave; 1990; Fields et al., 1992; Holth & Arntzen, 1998; Kennedy & Laitinen, 1988; Saunders, Wachter & Spradlin, 1988; Sidman, Kirk & Willson-Morris, 1985; Sidman, Willson-Morris & Kirk, 1986; Sigurardottir, Green & Saunders, 1990; Spradlin, Cotter & Baxley, 1973; Yamamoto, 1986). In addition, Reeve and Fields (2001) demonstrated that the likelihood of inducing perceptual classes was a direct function of the number of forced-choice generalization tests trials that preceded the assessment of class formation. All of these experiments demonstrate that test repetition is implicated in the emergence of classes, regardless of class type.

Repeated testing results in the emergence of stimulus classes. Thus, the second goal of the current experiment was to determine whether repetition of the transitivity test trials would facilitate the establishment of transitivity-indicative performances. This was accomplished by exposing half of the participants to a series of transitivity-test blocks, each of which contained stimuli from one

stimulus domain (staged testing). Following staged testing, these participants were then exposed to a block of transitivity-tests that contained stimuli from the three domains (concurrent testing). The other participants were exposed to concurrent testing only. Difference in performance between the participants would index the effect of repeated testing on the emergence of transitivity-indicative repertoire.

### Method

#### Participants

One hundred-one undergraduate students from Queens College, CUNY were recruited from introductory psychology classes to participate. The participants had no prior experience with the research area. They received partial course credit upon completion of the experiment, regardless of performance. Each participant participated in one to three experimental sessions over the course of 10 days, with each session lasting no more than 2.5 hours.

All participants were exposed to a preliminary testing condition that assessed the presence of transitivity-indicative performances with the stimulus sets used in the current experiment. Participants who did not show transitivity-indicative performances in at least two of the stimulus domains were randomly assigned to one of the four groups. All other participants were excused from the study. This procedure is described in detail in Phase 2 of the procedure.

Seventy-seven participants demonstrated transitivity-indicative performances with the stimulus sets used in preliminary training. The remaining twenty-four participants who failed to demonstrate transitivity-indicative

performances were randomly assigned to one of four groups that differed in terms of training and/or testing.

### **Apparatus and Stimuli**

The experimental stimuli were presented to the participants using an IBM-compatible computer with a color VGA monitor. Each participant was seated alone in a cubicle at a table facing the computer. Responses were made by touching specific keys on a computer keyboard. The recording of responses and the presentation of the stimuli were controlled by software designed specifically for the current experiment.

The stimuli used during the preliminary test and posttest are presented in Figure 1 with their corresponding symbolic representation. Six sets, designated 1 through 6, each contained three stimuli designated A, B, and C. The number immediately to the right of each letter identified class membership. The particular stimuli that corresponded to a given symbolic designation did not vary across conditions and did not change from preliminary test to posttest. The stimuli in sets 1 and 2 were comprised of 'glyphs' that were abstract black and white block drawings from Microsoft Word 97. The stimuli in sets 3 and 4 were color satellite images of various 10 kilometer by 10 kilometer regions of the world provided by the Army Research Institute Topographical Engineering Center. The stimuli in sets 5 and 6 were three-letter nonsense syllables that consisted of consonant-vowel-consonant letter strings.

The stimuli used during training are presented in Figure 2 with their corresponding symbolic representations. Training was conducted with 20 three-

member sets designated 7 through 26, each of which contained stimuli designated A, B, and C. The particular stimuli that corresponded to a given symbolic designation did not vary across conditions. Only glyph stimuli were used for transitivity training.

All stimuli were presented on the screen in 5 x 5-cm square region. The glyphs and nonsense syllables appeared black on a white background, and the satellite images appeared in color with a thin white frame surrounding the stimulus.

### Procedure

Trial structure, contingencies, and responses with a trial. All training and testing trials were conducted using a matching-to-sample format (Berryman, Cumming, Cohen & Johnson, 1965; Cumming and Berryman, 1961). Each trial began when the words 'press ENTER to continue the experiment' appeared centered on the screen. After the participant pressed the ENTER key, this message was removed and the sample stimulus was displayed in the upper portion of the screen. The participant was then required to press the space bar to display the two comparison stimuli along with the sample on the screen. The stimuli were displayed in a triangular pattern, with the sample stimulus at the vertex of the triangle and each of the two comparison stimuli at the corners of the base of the triangle. During each trial, the sample stimulus and the positive comparison stimulus (Co+) were from the same class, whereas the negative comparison stimulus (Co-) was from the other class.

The participant selected the comparison on the left by pressing the 1 key and selected the comparison on the right by pressing the 2 key. During training trials in which informative feedback was scheduled, after the participant made a response, a feedback message was displayed centered on the screen. If the participant selected the Co+, the word 'RIGHT' appeared on the screen until the participant pressed the 'R' key. If the participant selected the Co-, the word 'WRONG' appeared on the screen until the participant pressed the 'W' key. On trials when non-informative feedback was scheduled, the letter E appeared on the screen after the participants' response, regardless of comparison selection. It remained on the screen until the participant pressed the 'E' key. The 'E' key was selected as an observing response, and because it was between the 'R' and 'W' keys on a standard QWERTY keyboard. After pressing the 'R', 'W', or 'E' key, the screen cleared and the next trial began (Fields et al., 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 the training of the baseline conditional discriminations, a block of trials was presented repeatedly until all trials occasioned 100% correct responding. During these blocks, informative feedback was provided after each trial. The percentage of trials during each baseline-conditional-discrimination-training block that occasioned informative feedback was then reduced over successive blocks as long as performance within a block was maintained at 100% accuracy (mastery criterion). Trials that were followed by informative

feedback were randomly determined. If the mastery criterion was not met within three blocks at a given feedback level, the participant was returned to the previous feedback level.

### Overall Design

Participants who did not pass the preliminary test were randomly assigned to one of four experimental conditions. The four conditions and their corresponding combination of the two experimental variables, transitivity training and staged testing, are presented in Table 1. In Phase 1, all participants were exposed to keyboard familiarization. In Phase 2, all participants were given a preliminary test of transitivity-indicative performances under the glyph, satellite and nonsense syllable stimulus domains. After preliminary testing, in Phase 3, half of the continuing participants received transitivity training with glyph sets and half were not exposed to transitivity training. In Phase 4, half of the participants were then exposed to the staged testing protocol and half were not. Finally, in Phase 5, all participants were exposed to the posttest, which was the same test that participants received at the end of the preliminary testing.

In general, differences between performances on the preliminary test and posttest indicated how transitivity training and/or staged testing influenced the emergence of transitivity-indicative performances across stimulus domains. It is possible, however, that there were potential additive effects between the two independent variables on the induction of transitivity-indicative performances during the posttest. Thus, each of the four experimental conditions provided specific information on the influence of the independent variables, alone or in

combination, on the emergence of transitivity-indicative performances during the posttest.

Group I examined how repeated testing alone influenced the induction of transitivity-indicative performances during the posttest. Comparing the performances of participants in Group I to Groups II and III showed how the emergence of a transitivity-indicative repertoire was influenced by transitivity training and staged testing, respectively. Comparing the performances of the participants in Group I to Group IV showed how the emergence of a transitivity-indicative repertoire was influenced by exposure to transitivity training and staged testing. Furthermore, comparing the performances from participants in Groups II and III to Group IV provided additional information on the additive effects of transitivity training and staged testing in the fourth Group.

Phase 1: Keyboard familiarization. The purpose of this phase was to provide a description of performances expected during the experiment. When a participant entered the experimental cubicle, the following instructions were on the computer screen:

Thank you for volunteering to be a participant in this  
experiment.  
PLEASE DO NOT TOUCH ANY OF THE KEYS ON THE KEYBOARD  
YET! In this experiment you will be presented with many trials.  
Each contains three CUES that will be common words or pictures.

\*\*\*

YOUR TASK IS TO DISCOVER WHICH CUES GO  
TOGETHER.

\*\*\*

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 will tell you which KEYS to press. These labels

will slowly disappear. Your task is to **RESPOND CORRECTLY** to the **CUES** and the **INSTRUCTIONS** by pressing the appropriate **keys** on the computer keyboard.

\*\*\*

The experiment is conducted in blocks. When each block ends, the screen will tell you how you did. If you need to take a break after any of these blocks, please call the experimenter.

Press = = = > SPACE BAR < = = = to continue.

Following the instructions, participants were taught to emit the appropriate keyboard responses required, in the presence of each cue, to progress through each trial in the experiment. During these trials, semantically related English words were used as samples and comparisons with five instructional prompts. The samples and their corresponding positive and negative comparison for each trial in Phase 1 are listed in Table 2. Each trial was presented once per block with informative feedback scheduled following all trials. The Co+ was the semantically related word and the Co- was the semantically unrelated word.

The instructional prompts were deleted in a serial manner across trials in Phase 1 (Fields et al., 1990; Fields et al., 1997). Phase 1 ended once the stimuli were presented without prompts and performance exceeded 85% accuracy 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 this phase reappeared on three successive trials.

Phase 2: Preliminary Testing. The purpose of this phase was to test for transitivity-indicative performances with six sets (1-6) across three stimulus domains. Participants who did not show transitivity-indicative performances in at

least two of the stimulus domains were randomly assigned to one of the four experimental conditions. All other participants were excused from the study. The sequence of training and testing trials and the number of trials in each block is presented in Table 3.

Baseline conditional discriminations (AB and BC) were first established with stimuli in each of the domains. Transitivity (AC) was then assessed in a 'mixed' test block that also contained the baseline conditional discriminations. A mixed test is a block of trials that contains stimuli from more than one stimulus domain.

The AB conditional discrimination for each stimulus domain was established in separate blocks at the 100% feedback level. The AB conditional discrimination was established first with glyphs then with satellite images and finally with nonsense syllables. Once the 100% correct criterion had been met under each stimulus domain, the AB baseline conditional discriminations were collapsed into a mixed AB block that contained stimuli from all domains. Once the 100% correct criterion had been met in the mixed AB block, the BC conditional discriminations were established following the same protocol. Following the completion of the mixed BC block, the mixed AB block was repeated followed by another mixed BC block. There were an equal number of trials from each of the stimulus domains in the mixed blocks.

Following the second presentation of the mixed BC block, the AB and BC conditional discriminations were combined into a mixed AB/BC block. Feedback

in the mixed AB/BC block was systematically reduced from 100-75-25-0% as stipulated above.

Following completion of the mixed AB/BC block, transitivity-indicative performances were tested in a mixed test block under extinction conditions. The mixed test block contained the baseline conditional discriminations (AB and BC) and the transitivity probes (AC) for all stimulus domains. There were an equal number of baseline conditional discriminations and transitivity probes in the mixed test blocks. The criterion for transitivity-indicative performances was at least 95% correct across the 3 mixed preliminary test blocks in any two of the stimulus domains.

Phase 3: Transitivity Training. The purpose of this phase was to directly train transitivity-indicative performances with 20 glyph sets. This was done for half of the participants (Groups II and IV). The participants that did not receive transitivity training were immediately exposed to Phase 4 (Staged Testing).

The participants who were assigned to glyph-training established baseline conditional discriminations (AB and BC) with twenty 3-member glyph sets, designated 7-8, 9-10, 11-12, 13-14, 15-16, 17-18, 19-20, 21-22, 23-24, and 25-26. Transitivity probes (AC) followed the baseline conditional discriminations. If a participant failed to show transitivity-indicative responding, the repertoire was directly trained. The sequence of training and testing trials and the number of trials in each block is presented in Table 4.

Glyph training began with the establishment of the AB baseline conditional discrimination for glyph sets 7 and 8 at 100% feedback. Once criterion was met,

the BC baseline conditional discriminations for glyph sets 7 and 8 were trained at 100% feedback. Once criterion was met on the BC conditional discriminations, the AB and BC conditional discriminations were presented in a single block. Feedback during these blocks was systematically reduced from 100-75-25-0% as stipulated above. Once criterion was met at the 0% feedback level, AC transitivity probes were presented with the baseline conditional discriminations in 2 test blocks under extinction conditions. There were an equal number of baseline conditional discriminations and transitivity probes in each of the test blocks. If the participant demonstrated transitivity-indicative performances with this glyph set, the next set of baseline conditional discriminations were presented. Transitivity-indicative performances were defined as 100% set consistent responding in either of the test blocks or 98% set consistent responding (one incorrect trial) across both test blocks.

If the participant failed to demonstrate transitivity-indicative performances during the test, AC transitivity was directly trained in a block of trials that contained their corresponding baseline conditional discriminations. Feedback on the transitivity training trials was systematically reduced from 100-50-0% as stipulated above. Following the direct training of the transitivity discriminations, the next set of baseline conditional discriminations were presented. The protocol for the glyph sets 9 and 10 were the same as glyph sets 7 and 8.

Following transitivity training with the glyph sets 9 and 10, the AB baseline conditional discriminations for glyph sets 11 and 12 were trained at 100% feedback. Once criterion was met, the AB baseline conditional discriminations

for glyph set 13 and 14 were trained at 100% feedback. Once criterion was met for glyph set 13 and 14, the AB baseline conditional discriminations for all sets were collapsed into a single mixed AB training block at 100% feedback. Mixed blocks indicate those blocks that contained stimuli from more than two glyph sets. Once criterion was met, the BC baseline conditional discriminations for glyph sets 11 and 12 were trained at 100% feedback followed by the BC baseline conditional discriminations for glyph set 13 and 14. Once criterion was met for glyph sets 13 and 14, the BC baseline conditional discriminations for all sets were presented in a single mixed BC training block at 100% feedback. Once criterion was met, the AB and the BC baseline conditional discriminations for all sets were presented in a mixed AB/BC training block at 100% feedback. Feedback on these baseline-training trials was systematically reduced from 100-75-25-0% as stipulated above.

Once criterion was met at the 0% feedback level, AC transitivity probes with the baseline conditional discriminations for glyph sets 11 and 12 was presented under extinction conditions. If the participant demonstrated transitivity-indicative performances with these glyph sets, a transitivity test was conducted with glyph sets 13 and 14. If the participant failed to demonstrate transitivity-indicative performances during the AC tests with glyph sets 11 and 12, AC transitivity was directly trained following the protocol described above. Transitivity tests were then conducted with glyph set 13 and 14. If the participant failed to demonstrate transitivity-indicative performances during the AC tests with glyph sets 13 and 14, AC transitivity was directly trained following the protocol

described above. When the participant demonstrated transitivity-indicative performances with glyph sets 13 and 14, an equal number of AC transitivity probes and their corresponding baseline conditional discriminations from glyph sets 11-12 and 13-14 were presented in a single mixed test block. If the participant demonstrated transitivity-indicative performances under those conditions, the next set of baseline conditional discriminations was presented.

If the participant failed to demonstrate transitivity-indicative performances during the mixed AC tests, transitivity was directly trained by presenting the transitivity probes and the baseline conditional discriminations for glyph sets 11-12 and 13-14 in a single mixed training block. Feedback on the mixed transitivity training trials was systematically reduced from 100-50-0% as described above. Following mixed transitivity training, the next set of baseline conditional discriminations was presented.

Following transitivity training and testing with glyph sets 11-12 and 13-14, the AB baseline conditional discriminations for glyph sets 15 and 16 were trained at 100% feedback. Once criterion was met, the AB baseline conditional discriminations for glyph set 17 and 18 were trained at 100% feedback. Once criterion was met, the AB baseline conditional discriminations for glyph sets 19 and 20 were trained at 100% feedback. Once criterion was met for glyph sets 19 and 20, the AB baseline conditional discriminations for all three sets were presented in a single mixed AB training block at 100% feedback. Once criterion was met, the BC baseline conditional discriminations for glyph sets 15 and 16 were trained at 100% feedback followed by the BC baseline conditional

**discriminations for glyph sets 17 and 18. Once criterion was met, the BC baseline conditional discriminations for glyph set 19 and 20 were trained at 100% feedback. Once criterion was met for glyph sets 19 and 20, the BC baseline conditional discriminations for sets 15-16, 17-18 and 19-20 were presented in a single mixed BC training block at 100% feedback. Once criterion was met for the mixed BC training block, the AB and the BC baseline conditional discriminations for sets 15-16, 17-18 and 19-20 were presented in a mixed AB/BC training block at 100% feedback. Feedback on these baseline-training trials was systematically reduced from 100-75-25-0% as stipulated above.**

**Once criterion was met at the 0% feedback level, an equal number of AC transitivity probes and the baseline conditional discriminations for glyph sets 15 and 16 were presented under extinction conditions. If the participant demonstrated transitivity-indicative performances with this glyph set, the same transitivity test was conducted with glyph set 17 and 18. If the participant failed to demonstrate transitivity-indicative performances during the AC tests with these glyph sets, AC transitivity was directly trained following the protocol described above. Transitivity tests were then conducted with glyph sets 17 and 18. If the participant demonstrated transitivity-indicative performances with these glyph sets, the same transitivity test was conducted with glyph set 19 and 20. If the participant failed to demonstrate transitivity-indicative performances during the AC tests with glyph sets 17 and 18, AC transitivity was directly trained following the protocol described above. Transitivity tests were then conducted with glyph set 19 and 20. If the participant demonstrated transitivity-indicative performances**

with this glyph set, AC transitivity probes and their corresponding baseline conditional discriminations from glyph sets 15-16, 17-18 and 19-20 were presented in a single mixed test block. If the participant failed to demonstrate transitivity-indicative performances during the AC tests with glyph sets 19 and 20, AC transitivity was directly trained following the protocol described above. Mixed transitivity tests were then presented with glyph sets 15-16, 17-18 and 19-20. If the participant demonstrated transitivity-indicative performances during the mixed AC tests, the next set of baseline conditional discriminations were presented.

If the participant failed to demonstrate transitivity-indicative performances during the mixed AC tests, transitivity was directly trained by presenting the transitivity probes and the baseline conditional discriminations for glyph sets 15-16, 17-18 and 19-20 in a single mixed training block. Feedback on the mixed transitivity training trials was systematically reduced from 100-50-0% as stipulated above. Following mixed transitivity training, the next set of baseline conditional discriminations was presented. The protocol for glyph sets 21-22, 23-24 and 25-26 were the same as glyph sets 15-16, 17-18 and 19-20.

**Phase 4: Staged Testing.** The purpose of this phase was to test for transitivity-indicative performances in each domain. Tests were conducted in separate test blocks, each of which contained stimuli in a given domain (Staged testing). This was done for half of the participants (groups III and IV). The participants that did not receive staged testing were immediately exposed to

**Phase 5 (Post Testing). The sequence of training and testing and the number of trials in each block for Phase 4 are presented in Table 5.**

**Mixed training blocks were used to reestablish the AB baseline conditional discriminations with the same glyph, satellite and nonsense syllable stimuli used in preliminary testing (sets 1-6) at 100% feedback for all participants. Once criterion was met on the AB baseline conditional discriminations, the BC baseline conditional discriminations for sets 1 through 6 were retrained at 100% feedback. Once criterion was met on the BC conditional discriminations, the AB and the BC baseline conditional discriminations for sets 1 through 6 were combined in a single mixed training block. Feedback on the mixed AB/BC baseline conditional discrimination training was reduced from 100-75-25-0% as stipulated above. Once criterion was met at the 0% feedback level, participants were then assigned the no-staged or staged-testing testing condition.**

**Participants who were not assigned to the staged-testing condition went directly to the mixed AC test (Phase 5). The remaining participants, who were assigned to the staged-testing condition, were then tested on AC transitivity probes and their corresponding baseline conditional discriminations in a staged manner. That is, transitivity was first tested with the glyph stimulus domain, then the satellite images and finally the nonsense syllables under extinction conditions. Up to three blocks of tests for each stimulus domain were presented. An equal number of baseline conditional discriminations and transitivity probes were presented. If a participant failed to show transitivity-indicative responding**

under any of the stimulus domains however, the repertoire was not directly trained.

**Phase 5: Post Testing.** The purpose of this phase was to test for the emergence of transitivity-indicative performances with stimuli in each of the stimulus domains. Re-exposing participants to the mixed test blocks used in Phase 2 accomplished this. Those mixed tests included stimuli from sets 1-6. The sequence of training and testing and the number of trials in each block for Phase 5 are presented in Table 6. Note that subjects who were exposed to staged testing (Phase 4) were not retrained on the baseline conditional discriminations for sets 1 – 6. They were instead put directly into the mixed test.

### **Results**

**Dependent Variables.** The effect of explicit transitivity training with glyphs (glyph training) and staged-testing was indexed with three dependent variables. The first dependent variable measured the percentage of sets that occasioned transitivity-indicative performances. This value was computed by averaging the percentage of sets that occasioned transitivity-indicative performances across all participants in a condition and subtracting preliminary test from posttest performances. Data for these computations are listed on Table 7.

The second dependent variable measured the number of participants in a condition who satisfied a standardized performance criterion. The criterion was an increase of a specific size of the number of sets that occasioned transitivity-indicative performances between the preliminary test and the posttest.

The final dependent variable measured the percentage of participants who showed specific increases in transitivity-indicative performances under each stimulus domain. This value was computed by counting the number of participants who showed transitivity-indicative performances under a given stimulus domain and subtracting preliminary test from posttest performances for each stimulus domain under each condition.

### Difference Scores

Figure 3 illustrates the average change and standard error of the percentage of sets that occasioned transitivity-indicative responding between the preliminary test and the posttest. The data are partitioned into the groups that the participants had been assigned. Data represented by the white bars indicate the conditions in which glyph training was not conducted and data represented by the gray bars indicate those conditions in which glyph training was conducted. The data represented by the pair of bars on the left represent performances of participants who were not exposed to staged testing whereas the data represented by the pair of bars on the right represent performances of participants who were exposed to staged testing.

Glyph Training and Staged Testing. The effect of glyph training can be seen by comparing the two white bars (no glyph training) with the two gray bars (glyph training). The percentage of sets that occasioned transitivity-indicative performances was 11.08% for the no-glyph-training conditions and was 55.62% for the glyph training conditions. Thus, the presence of glyph training led to a

larger percentage of sets that occasioned transitivity-indicative performances when compared to no glyph training.

The effect of staged testing can be seen by comparing the two bars on the left (no staged testing) with the two bars on the right (staged testing). The percentage of sets that occasioned transitivity-indicative performances was 19.48% for the no staged testing conditions and was 47.22% for the staged testing conditions. Thus, the presence of staged testing led to a larger percentage of sets that occasioned transitivity-indicative performances when compared to no staged testing.

A two-factor analysis of variance (Glyph Training x Staged Testing) showed a significant effect of glyph training  $F(1,20) = 19.15, p < .05$ , and a significant effect of staged testing  $F(1,20) = 7.42, p < .05$ , but no interaction between glyph training and staged testing  $F(1,20) = 0.84, p > .05$ .

Effect of Glyph Training When Not Followed by Staged Testing. The data represented by the pair of bars on the left show the effect of glyph training alone. When no glyph training was scheduled (bar on the left), there was virtually no change in the average number sets that occasioned transitive indicative performances. Therefore, repetition of mixed test did not induce a transitivity repertoire when compared to an expected outcome of zero,  $t(5) = 0.28, p > .05$ . Glyph training (bar on right), resulted in a relatively large increase in the average number of sets that occasioned transitivity-indicative performances in the posttest. That increase was significant when compared to an expected outcome

of zero,  $t(5) = 2.60$ ,  $p < .05$ . Thus, glyph training induced a transitivity repertoire in the absence of staged testing.

The relative effects of glyph training versus no glyph training when not followed by staged testing can be seen by a comparison of the data represented by the two bars on the left. Glyph training resulted in a significant increase in the average number of sets that occasioned transitivity-indicative performances in the posttest when compared to the no glyph training condition,  $LSD = -2.44$ ,  $p < .05$ . Thus, glyph training was more effective at inducing transitivity-indicative performances than mere repetition of the mixed test block.

Effect of Glyph Training When Followed by Staged Testing. The data represented by the pair of bars on the right shows the effect of glyph training when followed by staged testing. When no glyph training (bar on left) was scheduled, there was a small but significant increase in the average number sets that occasioned transitivity-indicative performances when compared to an expected outcome of zero,  $t(5) = 3.05$ ,  $p < .05$ . Thus, staged testing induced transitivity-indicative performances. Glyph training (bar on right) also resulted in a significant increase in the average number of sets that occasioned transitivity-indicative performances in the posttest when compared to an expected outcome of zero,  $t(5) = 6.72$ ,  $p < .05$ .

The relative effect of glyph training when followed by staged testing was shown by a comparison the data represented in the two bars on the right. Glyph training resulted in a significant increase in the average number of sets that occasioned transitivity-indicative performances when compared to the no-glyph-

training condition,  $LSD = -3.74, p < .05$ . Therefore, while staged testing alone induced transitivity-indicative performances, glyph training followed by staged testing was more effective at inducing transitivity-indicative performances.

#### **Effect of Staged Testing When Not Preceded by Glyph Training.** A

comparison of the data represented by the white bars show how the average number of sets that occasioned transitivity-indicative performances was influenced by the presence of staged testing that was not preceded by glyph training. Staged testing alone produced a significant increase in the average number of sets that occasioned transitivity-indicative performances in the posttest when compared to the no-glyph-training, no-staged-testing condition,  $LSD = -1.24, p < .05$ . Thus, staged testing induced a transitivity repertoire in the absence of glyph training.

#### **Effect of Staged Testing When Preceded by Glyph Training.** A

comparison of the data represented by the gray bars show how the average number of sets that occasioned transitivity-indicative performances was influenced by the presence of staged testing that was preceded by glyph training. The introduction of staged testing after glyph training resulted in a significant increase on the average number of sets that emerged as transitivity classes when compared to no staged testing,  $LSD = -2.58, p < .05$ . Thus, a transitive repertoire emerged when staged testing followed glyph training.

#### **Posttest Performances Relative to Preliminary Test Performances**

The purpose of this analysis was to determine whether the performances shown in the previous figure reflect a breakdown in performance on the baseline

conditional-discriminations or the performances occasioned by the transitivity probes. The top panel of Figure 4 illustrates the average change and standard error in performance on the baseline trials from the preliminary test to the posttest.

During the preliminary mixed test, the baseline conditional discriminations occasioned 89.87% correct responding across the four groups. The average change in performance on the baseline trials was 4.73% across groups. A two-factor analysis of variance (Glyph Training x Staged Testing) however, showed a significant effect of glyph training  $F(1,12) = 5.22, p < .05$ , but not for staged testing  $F(1,12) = 0.04, p > .05$  and no interaction between glyph training and staged testing  $F(1,12) = 4.61, p > .05$ . The reason for these differences is due to the inclusion of a participant's data from Group 3 (No glyph training – staged testing condition). This participant showed a difference of 15.3%, which was responsible for the statistically significant finding. This participant's datum certainly appears to be an outlier and thus, the statistically significant difference is not functionally significant.

The bottom panel of Figure 4 illustrates the average change and standard error in performance on the transitivity probes from the preliminary test to the posttest. During the preliminary mixed test, the transitivity probes occasioned 54.86% correct responding across the four groups. The average percentage of sets that showed transitivity-indicative performances during the posttest, however, was lowest for the no-glyph-training, no-staged-testing condition. Intermediate levels of performance were observed for the glyph training alone

and the staged testing alone conditions and the highest levels of performance was observed for the glyph-training, staged-testing condition. A two-factor analysis of variance (Glyph Training x Staged Testing) however, showed a significant effect of glyph training  $F(1,12) = 56.83, p < .05$ , but not for staged testing  $F(1,12) = 2.04, p > .05$  and no interaction  $F(1,12) = .37, p > .05$ .

The levels of performance on the transitivity probes mirrored percentage of sets that showed transitivity-indicative performances presented in Figure 3. Thus, the performances observed in the posttest reflect differences in responding on the transitivity probes and were not due to a breakdown in the baseline conditional discriminations.

#### Single-Participant Performances

The purpose of this analysis was to determine whether the data shown in Figure 3 reflect the performances of individual participants in each group or whether those performances are an artifact of averaging extreme scores across participants in a group. Figure 5 illustrates the difference in percentage of sets that occasioned transitivity-indicative performances that emerged across all stimulus domains between all the preliminary test blocks and the final posttest block, for individual participants in each condition. The graphs in the bottom row depict data for participants who were not exposed to staged testing before the posttest. The graphs in the top row depict data for participants who were exposed to staged testing before the posttest. The graphs in the left column depict participants who were not exposed to any glyph training. The graphs in the right column depict participants who were exposed to glyph training.

There was a discrepancy in the overall level of performance across the staged and the no staged testing conditions. In general, the staged testing conditions produced higher levels of transitivity-indicative performances than the no staged testing conditions. Additionally, the glyph training conditions produced higher levels of performance than the no glyph training conditions. To make a meaningful comparison of the number of participants who showed a particular level of transitivity-indicative responding, the data are analyzed using two performance criteria. For the no staged testing conditions, the criterion was at least a 20% increase in the number of sets that occasioned transitivity-indicative performances. For the staged testing conditions, the criterion was at least a 65% increase in the number of sets that occasioned transitivity-indicative performances. For the no glyph training conditions, a criterion of at least a 20% increase in the number of sets that occasioned transitivity-indicative performances was used. In the glyph training conditions, a criterion of at least 65% increase in the number of sets that occasioned transitivity-indicative performances was used.

Effect of Glyph Training Not Followed by Staged Testing. The lower left-hand graph in Figure 5 illustrates the effects of repetition of the mixed test block from the preliminary test and the posttest with no intervening glyph training as a parameter. One of six participants showed at least a 20% increase in the number of sets that occasioned transitivity-indicative performances. Thus, repetition of the mixed test in the absence of glyph training did not appear to induce an

increase in the number of participants who showed at least a 20% increase in the number of sets that occasioned transitivity-indicative performances.

The lower right hand graph of Figure 5 shows the effects of the glyph training alone on the emergence of transitivity-indicative performances. Four of the six participants also showed at least a 20% increase in the number of sets that occasioned transitivity-indicative performances. Thus, glyph training induced an increase in the number of participants who showed at least a 20% increase in the number of sets that occasioned transitivity-indicative performances.

The relative effect of glyph training in the absence of staged testing can be seen by a direct comparison of the data represented by the two graphs on the bottom of Figure 5. When compared to the no glyph training condition, glyph training resulted in a significant increase in the average number of participants who showed at least a 20% increase in the number sets that occasioned transitivity-indicative performances,  $\chi^2(1, n = 6) = 3.09, p < .05$ .

Effect of Glyph Training When Followed by Staged Testing. The data represented by the upper left-hand graph in Figure 5 illustrates the effect of staged testing when not preceded by glyph training. When no glyph training was conducted, no participants showed at least a 65% increase in the number of sets that occasioned transitivity-indicative performances.

The upper right-hand graph of Figure 5 shows the effect of staged testing when preceded by glyph training. Five of the six participants showed at least a 65% increase in the number of sets that occasioned transitivity-indicative performances. Thus, when staged testing was scheduled after glyph training,

there was a large increase in the number of participants that showed at least a 65% increase in the number of sets that occasioned transitive indicative performances.

The relative effect of glyph training in the presence of staged testing can be seen by a direct comparison of the data represented by the two graphs on the top of Figure 5. Glyph training resulted in a significant increase in the number of participants that showed at least a 65% increase in the number of transitivity classes that emerged when compared to the no glyph training condition,  $\chi^2(1, n = 6) = 8.57, p < .05$ . Thus, there was a significant increase in the number of participants who showed at least a 65% increase in the number of sets that occasioned transitivity-indicative performances when staged testing followed glyph training.

Effect of Staged Testing When Not Preceded by Glyph Training. The effect of staged testing with no prior glyph training as a parameter is shown by a comparison of the data represented by the graphs in the left-hand column of Figure 5. Staged testing had a significant effect on the number of participants that showed at least a 20% increase in the number sets that occasioned transitivity-indicative performances,  $\chi^2(1, n = 6) = 3.09, p < .05$ , but no effect on the number of participants that showed at least a 65% increase.

Effect of Staged Testing When Preceded by Glyph Training. The effect of staged testing with prior glyph training as a parameter is shown by a comparison of the data represented by the graphs in the right-hand column of Figure 5. Staged testing had a significant effect on the number of participants that showed

at least a 65% increase in the number sets that occasioned transitivity-indicative performances,  $\chi^2(1, n = 6) = 3.09, p < .05$ . Thus, staged testing following glyph training produced the largest effect on the number of participants who showed at least a 65% increase in the number of sets that occasioned transitivity-indicative performances. The combination of glyph training and staged testing maximized the induction of transitivity-indicative performances.

Individual Participant Performances on Baseline Trials. The purpose of this analysis was to determine whether the performances shown in Figure 5 reflect a posttest breakdown in performance on (A) the baseline conditional discriminations and/or (B) the transitivity probes. Figure 6 illustrates the difference in percentage of baseline trials and transitivity probes that occasioned set-consistent responding between all the preliminary test blocks and the final posttest block. The data are presented for all participants in each group. The lower left quadrant shows the data for participants in the no-glyph-training, no-staged-testing group. The lower right quadrant shows the data for participants in the glyph-training, no-staged-testing group. The upper left quadrant shows the data for participants in the no-glyph-training, staged-testing group. The upper right quadrant shows the data for the participants in the glyph-training, staged-testing group. There are two panels in each quadrant. The top panel in each quadrant shows the performances on the baseline trials and the bottom panel in each quadrant shows the performances on the transitivity probes. The order of displaying data for the participants in each condition corresponds to the order used in Figure 5.

The average change in performance on the baseline trials between the preliminary mixed test and the posttest was 4.73%. The top panel in each quadrant shows that this outcome was observed for individual participants across all conditions. There was very little change in the performance on the baseline trials from the preliminary mixed test to the posttest for individual participants.

Performances on the transitivity probes, illustrated in the lower panel in each quadrant, show differential levels that parallel the outcomes shown in Figure 5. The participants who showed the least amount of change on the transitivity probes were in the no-glyph-training, no-staged-testing group. Low levels of performance were also observed for the participants in the no-glyph-training, staged-testing group. Higher levels of performance were observed for the participants in the glyph-training, no-staged-testing group. The highest levels of performance were observed for the participants in the glyph-training, staged-testing group. Thus, the change performances observed in Figure 5 reflected changes in performance on the transitivity probes and maintenance of the baseline performances for individual participants.

#### **Generalization of Transitivity Repertoire across Stimulus Domains**

The design of the experiment permitted the measurement of the generalization of transitivity-indicative performances from the sets of glyphs used in training to novel stimuli in the same domain and to stimulus sets in two other domains. Generalization was measured by the change between the preliminary

**test and the posttest in the percentage of participants who showed transitivity-indicative performances with stimuli in each domain, in each group.**

**Transitivity-indicative performance during the preliminary test was defined as at least 95% class-consistent responding in any stimulus domain across the three mixed-test blocks. For participants who met this criterion under any stimulus domain, the effects of glyph training and staged testing could not be measured. Thus, the data for these participants were removed from the analysis. The number of participants who did not show transitivity-indicative performances under each of the stimulus domains for all groups is presented in Table 8. Overall, no more than two participants met criterion in any group under any stimulus domain.**

**Figure 7 shows the change from preliminary training to the posttest in the percentage of participants who demonstrated the emergence of transitivity-indicative performances, within each stimulus domain. Emergence of transitivity-indicative performances in the posttest was defined as 100% class consistent responding within a given domain in any one of the three mixed blocks. For all graphs, the pair of bars on the left indicated the performances observed in the no-staged-testing conditions and the pair of bars on the right indicated the performances observed in the staged-testing conditions. The shaded bars indicated the performances observed in the glyph training conditions whereas the white bars indicate the performances observed in the no glyph training conditions. The rightmost column in each panel is a stacked bar which will be used for a theoretical analysis presented in the discussion.**

### **Glyph Training Effects When Not Followed By Staged Testing:**

**Glyph Stimuli.** The data represented by the pair of bars on the left of the top panel in Figure 7 illustrate the effect of glyph training on the emergence of transitivity-indicative performances in the presence of glyph stimuli with no staged testing as a parameter. When no glyph training or when glyph training was scheduled (columns 1 and 2), there was no increase in the percentage of participants who showed the emergence of transitivity. Thus, neither the repetition of the mixed test block nor glyph training induced transitivity-indicative performances with the glyph sets used in testing.

The relative effect of glyph training can be seen by comparing the data represented by the bars on the left of the top panel in Figure 7. Glyph training did not result in a significant increase in the number of participants who showed the emergence of transitivity with the glyph sets when compared to the no glyph training condition,  $\chi^2(1, n = 11) = 0.02, p > .05$ . Thus, glyph training that was not followed by staged testing did not induce transitivity-indicative performances for the glyph sets.

**Satellite Stimuli.** The data represented by the pair of bars on the left of the middle panel in Figure 7 illustrate the effect of glyph training on the emergence of transitivity-indicative performances in the presence of satellite stimuli with no staged testing as a parameter. When no glyph training was scheduled (column 1), there was no increase in the percentage of participants who showed the emergence of transitivity-indicative performances.

When glyph training was scheduled (column 2), there was an increase in the percentage of participants who showed the emergence of transitivity-indicative performances. Thus, glyph training induced transitivity-indicative performances to the sets of satellite images used in testing but repetition of the mixed test block did not.

The relative effect of glyph training can be seen by comparing the data represented by the bars on the left of the middle panel in Figure 7. Glyph training resulted in a significant increase in the number of participants who demonstrated the emergence of transitivity-indicative performances with sets of satellite images when compared to the no glyph training condition,  $\chi^2(1, n = 9) = 0.9, p < .05$ . Thus, glyph training that was not followed by staged testing induced transitivity-indicative performances for the sets of satellite images used in testing.

Nonsense Syllable Stimuli. The data represented by the pair of bars on the left of the bottom panel in Figure 7 illustrate the effect of glyph training on the emergence of transitivity-indicative performances in the presence of nonsense syllables with no staged testing as a parameter. When no glyph training was scheduled (column 1), there was an increase in the percentage of participants who showed the emergence of transitivity-indicative performances. Thus, repetition of the mixed test block did not induce transitivity-indicative performances the sets of nonsense syllables used in testing.

When glyph training was scheduled however (column 2), there was an increase in the percentage of participants who showed the emergence of

transitivity-indicative performances. Thus, glyph training induced transitivity-indicative performances with the sets of nonsense syllables used in testing.

The relative effect of glyph training can be seen by comparing the data represented by the bars on the left of the bottom panel in Figure 7. Glyph training resulted in a significant increase in the number of participants who demonstrated the emergence of transitivity-indicative performances with sets of nonsense syllables when compared to the no glyph training condition,  $\chi^2(1, n = 10) = 4.28$ ,  $p < .05$ . Thus, glyph training that was not followed by staged testing induced transitivity-indicative performances for the sets of nonsense syllables.

#### Glyph Training Effects When Followed By Staged Testing:

Glyph Stimuli. The data represented by the pair of bars on the right of the top panel in Figure 7 illustrate the effect of glyph training on the emergence of transitivity-indicative performances in the presence of novel glyph sets with staged testing as a parameter. When no glyph training was scheduled (column 3), there was no increase in the percentage of participants who demonstrated the emergence of transitivity-indicative performances with novel glyph sets. Thus, induction of transitivity-indicative performances to the same stimulus domain was not observed when staged testing was conducted with no glyph training. When glyph training was scheduled (column 4), however, there was an increase in the percentage of participants who showed the emergence of transitivity-indicative performances. Thus, glyph training when followed by staged testing induced transitivity-indicative performances to the glyph sets used in testing.

The relative effect of glyph training can be seen by comparing the data represented by the bars on the right of the top panel in Figure 7. Glyph training followed by staged testing resulted in a significant increase in the number of participants who showed the emergence of transitivity-indicative performances with the glyph sets when compared to the no-glyph-training condition,  $\chi^2(1, n = 11) = 7.64, p < .05$ . Thus, glyph training that was followed by staged testing induced transitivity-indicative performances with novel sets of glyph stimuli.

Satellite Stimuli. The data represented by the pair of bars on the right of the middle panel in Figure 7 illustrate the effect of glyph training on the emergence of transitivity-indicative performances in the presence of sets of satellite images with staged testing as a parameter. When no glyph training was scheduled (column 3), there was no increase in the percentage of participants who demonstrated the emergence of transitivity-indicative performances. Thus, induction of transitivity-indicative performances to sets of satellite images was not observed when staged testing was conducted with no glyph training. When glyph training was scheduled (column 4), however, there was an increase in the percentage of participants who showed the emergence of transitivity with the sets of satellite images. Thus, glyph training that was followed by staged testing induced transitivity-indicative performances to the satellite stimulus domain.

The relative effect of glyph training can be seen by comparing the data represented by the pair of bars on the right of the middle panel in Figure 7. Glyph training resulted in a significant increase in the number of participants who showed the emergence of transitivity with the sets of satellite images used in

testing when compared to the no glyph training condition,  $\chi^2(1, n = 9) = 3.6, p < .05$ . Thus, glyph training that was followed by staged testing induced transitivity-indicative performances with sets satellite images.

**Nonsense Syllable Stimuli.** The data represented by the pair of bars on the right of the bottom panel in Figure 7 illustrate the effect of glyph training on the emergence of transitivity-indicative performances in the presence of sets of nonsense syllables with staged testing as a parameter. When no glyph training was scheduled (column 3), there was an increase in the percentage of participants who showed the emergence of transitivity with the sets of nonsense syllables. Thus, staged testing that was not preceded by glyph training induced transitivity-indicative performances to sets of nonsense syllables used in testing.

When glyph training was scheduled (column 4), there was also an increase in the percentage of participants who demonstrated the emergence of transitivity with the nonsense syllables sets used in testing. Thus, induction of transitivity-indicative performances to sets of nonsense syllables was also observed when glyph training preceded staged testing.

The relative effect of glyph training can be seen by comparing the data represented by the bars on the right of the bottom panel in Figure 7. Glyph training followed by staged testing resulted in no significant increase in the percentage of participants who demonstrated the emergence of transitivity with the sets of nonsense syllables when compared to the no glyph training-staged testing condition,  $\chi^2(1, n = 11) = 7.64, p < .05$ . Thus, glyph training that was

followed by staged testing did not induce transitivity-indicative performances with sets of nonsense syllables used in testing.

#### Intervention Effects Across Stimulus Domains.

There was a differential effect of stimulus domain on the percentage of participants who showed the emergence of transitivity-indicative performances when glyph training was conducted in the absence of staged testing (column 2). Under the glyph stimulus domain, there was a small increase in the percentage of participants who showed the emergence of transitivity. Relative to the glyph stimulus domain, the satellite images showed a larger increase in the percentage of participants who showed the emergence of transitivity-indicative performances. Relative to the glyph stimulus domain, the sets of nonsense syllables showed an even larger increase in the percentage of participants who showed the emergence of transitivity-indicative performances. While these differences illustrate a trend, the failure to show statistical significance is due to the small number of participants in each condition.

There was also a differential effect in the percentage of participants who showed the emergence of transitivity-indicative performances across stimulus domains when staged testing was not preceded by glyph training (column 3). Under the glyph stimulus domain, there was no increase in the percentage of participants who showed the emergence of transitivity. Relative to the glyph stimulus domain, the satellite images resulted in a small increase in the percentage of participants who showed the emergence of transitivity-indicative performances that approached significance,  $\chi^2(1, n = 10) = 2.50, p = .057$ . With

the sets of nonsense syllables, there was a very large increase in the percentage of participants who showed the emergence of transitivity-indicative performances. This difference is significant when compared to the glyph stimulus domain,  $\chi^2(1, n = 11) = 5.24, p < .05$  but is not significant when compared to the satellite images stimulus domain,  $\chi^2(1, n = 11) = 0.78, p > .05$ .

#### Staged Testing Effects When Not Preceded by Glyph Training

Comparison of the data represented by the white bars in each panel showed how staged testing influenced the emergence of transitivity-indicative performances in each stimulus domain with no-glyph training as a parameter. Comparison of the data represented by the gray bars in each panel showed how staged testing influenced the emergence of transitivity-indicative performances in each stimulus domain with glyph training as a parameter.

Glyph Stimuli. Comparison of the data represented by the white bars in the top panel of Figure 7 (columns 1 & 3) show, for the glyph stimulus domain, no significant difference between the staged and no-staged testing conditions when it was not preceded by glyph training,  $\chi^2(1, n = 10) = 1.11, p > .05$ . Thus, staged testing not preceded by glyph training did not induce transitivity-indicative performances with novel sets of glyph stimuli.

Comparison of the data represented by the gray bars in the top panel of Figure 7 (columns 2 & 4) show a significant difference between the staged and no-staged-testing conditions when it was preceded by glyph training,  $\chi^2(1, n = 12) = 5.33, p < .05$ . Thus, staged testing that followed glyph training induced transitivity-indicative performances with novel sets of glyph stimuli.

**Satellite Images.** Comparison of the data represented by the white bars in the middle panel of Figure 7 (columns 1 & 3) show, for the satellite images, no significant difference between the staged and no staged testing conditions when it was not preceded by glyph training,  $\chi^2(1, n = 10) = 0.48, p > .05$ . Thus, staged testing that was not preceded by glyph training did not induce transitivity-indicative performances with sets of satellite images.

Comparison of the data represented by the gray bars in the middle panel of Figure 7 (columns 2 & 4) showed, for the satellite images, the difference between the staged and no staged testing conditions when it was preceded by glyph training approached significance,  $\chi^2(1, n = 8) = 2.67, p = .051$ . Thus, staged testing that followed glyph training induced transitivity-indicative performances with sets of satellite images.

**Nonsense Syllables.** Comparison of the data represented by the white bars in the bottom panel of Figure 7 (columns 1 & 3) show, for the sets of nonsense syllables, a significant difference between the staged and no staged testing conditions when they were not preceded by glyph training,  $\chi^2(1, n = 11) = 5.24, p < .05$ . Thus, staged testing not preceded by glyph training induced transitivity-indicative performances with sets of nonsense syllables.

Comparison of the gray bars in the bottom panel of Figure 7 (columns 2 & 4) show, for the sets of nonsense syllables, no significant difference between the staged and no staged testing conditions when they were preceded by glyph training,  $\chi^2(1, n = 10) = 0.48, p > .05$ . Thus, staged testing that followed glyph

training failed to induce transitivity-indicative performances with sets of nonsense syllables.

### Glyph Training Performances

Figure 8 summarizes individual participant performances during glyph training. The bars within each panel identify the glyph sets that required transitivity training for each participant. The bars illustrate the number of transitivity training blocks required at each feedback level. Three blocks was the minimum number of blocks required to pass the transitivity training. The arrows identify the mixed-test blocks.

Participants from both glyph training groups were rank ordered according to the number of transitivity training blocks required at the 100% feedback level. This was done because there were no differences in the number of sets that required transitivity training between the staged-testing and the no-staged-testing conditions (58 and 60 sets respectively),  $t(10) = 0.12$ ,  $p > .05$ , two-tailed. Additionally, there was no difference in the number of participants who required training during the mixed-test blocks between each condition,  $\chi^2(1, n = 12) = 1.5$ ,  $p > .05$ . Thus, there are no differences between the participants who were subsequently exposed to the staged-testing or the no-staged-testing protocol.

When considering the demonstration of transitivity-indicative performances, only three of the 12 participants (HL7, MY5, IH3) required direct transitivity training on fewer than six glyph sets. Nine of the 12 participants required direct transitivity training on at least eight glyph sets. For participants who required training on at least eight glyph sets, there was no difference in the

number of sets that required transitivity training between the conditions,  $t(9) = 1.32$ ,  $p > .05$ , two-tailed. Thus, students showed difficulty in demonstrating transitivity-indicative performances with these glyph sets.

Additionally, performances observed during glyph training illustrate the failure of transitivity-indicative performances to emerge over the course of training. Only three participants (HL7, MY5, IH3) showed transitivity-indicative performances without direct training on the final 10 glyph sets. Only two other participants (SS9, JH5) did not require training on the final four glyph-sets and the corresponding mixed-test. Thus, 9 of the 12 participants required some level of transitivity training after the 10<sup>th</sup> glyph set.

When transitivity training was conducted however, criterion levels of performance were met quickly. Transitivity training was completed in fewer than ten training blocks in 44 of the 65 occurrences.

There was no correlation between the number of transitivity training blocks required during glyph training and the percentage of sets that showed an emergence of transitivity-indicative performances during the posttest,  $r = -0.38$ ,  $n = 12$ ,  $p > .05$ , two-tailed. Furthermore, performances during glyph training did not predict the emergence of transitivity-indicative performances to the sets of stimuli used in the posttest per domain. For the glyph domain,  $r = -0.31$ ,  $n = 12$ ,  $p > .05$ , two tails, the satellite domain,  $r = -0.53$ ,  $n = 8$ ,  $p > .05$ , two-tailed and for the nonsense syllable stimulus domain,  $r = 0.29$ ,  $n = 10$ ,  $p > .05$ , two-tailed. Therefore, performances observed during glyph training did not predict the later

emergence of transitivity-indicative performances to the sets of stimuli used in the posttest.

### Staged Testing Performances

Figure 9 showed the percentage of trials that occasioned set consistent responding for the participants exposed to staged testing. Participants' performance under each stimulus domain is shown. The data are presented for participants who had not shown transitivity-indicative performances during preliminary testing. The lines connected with squares show the percentage of baseline trials that occasioned set consistent responding and the lines with connected triangles show the percentage of transitivity trials that occasioned transitivity-indicative responding.

In general, there was maintenance of the baseline conditional discriminations across the staged testing blocks for both conditions. When staged testing followed glyph training, 94% (30 of 32) of staged testing blocks occasioned baseline performances that were 75% correct or greater. When staged testing did not follow glyph training, 84.2% (32 of 38) of the staged testing blocks occasioned baseline performances that were 75% correct or greater. Thus, there was no difference in the percentage of baseline trials that occasioned sets consistent responding between the staged-testing-no-glyph-training condition and the staged-testing-glyph-training condition,  $t(70) = 1.24$ ,  $p > .05$ , two-tailed.

In general, the baseline conditional discriminations resulted in a higher percentage of set-consistent responding than the transitivity trials. This is

observed when excluding blocks that occasioned 100% set consistent performances with both the baseline and transitivity trials. When staged testing occurred after glyph training, 90.5% (19 of 21) of the staged testing blocks occasioned baseline performances that were greater than the performances occasioned by the transitivity trials. When staged testing occurred after no glyph training, 71.9% (23 of 32) of the staged testing blocks occasioned baseline performances that were greater than the performances occasioned by the transitivity trials. Additionally, in six of those nine blocks in which the performance on the baseline trials was not greater than the performance on the transitivity trials, the discrepancy was 12.5% or less.

Overall, there is a significant difference in the number of sets that showed transitivity-indicative performances when the no glyph-training, staged-testing condition was compared to the glyph-training, staged-testing condition,  $\chi^2(1, n = 31) = 4.01, p < .05$ . When staged testing was not preceded by glyph training, 38% of the sets (6 of 16) occasioned transitivity-indicative performances. In contrast, when staged testing was conducted after glyph training, 73% of the sets (11 of 15) occasioned transitive repertoire. Thus, history with glyph training did induce the delayed emergence of transitivity-indicative performances during staged testing at a greater level than staged testing that did not follow glyph training.

There is, however, no difference in the number of participants who showed transitivity-indicative performances on the first staged test block (immediate emergence) when the no-glyph-training, staged-testing condition is

compared to the glyph-training, staged-testing condition,  $\chi^2(1, n = 31) = 0.80, p > .05$ . When staged testing followed glyph training, six of the 15 participants showed immediate emergence of the transitivity repertoire (JH5 – Glyph domain; CY6, AT9 HL7 – Satellite domain; HL7, EG0 – Nonsense Syllable domain). When staged testing did not follow glyph training, four of the 16 participants showed immediate emergence of the transitivity repertoire (JW3 – Satellite domain; DL8, LM7 JE2 – Nonsense Syllable domain). Thus, history with glyph training did not induce the immediate emergence of transitivity-indicative performances during staged testing at a greater level than staged testing that did not follow glyph training.

There was, however, a significant difference in the number of participants who showed the emergence of transitivity-indicative performances after the first staged test block (delayed emergence) across the conditions,  $\chi^2(1, n = 21) = 3.5, p < .05$ . When staged testing followed glyph training, five of the nine participants showed delayed emergence of the transitivity repertoire (EG0, AB6 – Glyph domain; AT9, JH5, AB6 – Nonsense Syllable domain). When staged testing did not follow glyph training, only two of the 12 participants showed delayed emergence of the transitivity repertoire (SS7, AV0 – Satellite domain). Thus, history with glyph training did induce the delayed emergence of transitivity-indicative performances during staged testing at a greater level than staged testing that did not follow glyph training.

### Discussion

The current experiment showed the effects of transitivity training and/or

staged testing on the emergence of transitivity-indicative performances under three stimulus domains. The emergence of transitivity-indicative performances was measured with three dependent variables: (a) the change in the number of test trials that occasioned transitivity-indicative performances across all stimulus domains for each condition, (b) the emergence of transitivity-indicative performances for individual participants averaged across stimulus domains and (c) the emergence of transitivity-indicative repertoires under each of the stimulus domains.

Effect of No Glyph Training and No Staged Testing. The repetition of the mixed test block with no intervening glyph training or staged testing did not lead to the emergence of transitivity-indicative performances with stimuli in any of the stimulus domains. This outcome was observed with all three dependent variables. Thus, repetition of the mixed test block in the absence of either of the experimental variables did not induce transitivity-indicative performances with any of the stimuli used in testing.

Effects of Glyph Training When Not Followed by Staged Testing and Staged Testing When Not Preceded by Glyph Training. The effect of glyph training that did not follow staged testing was the same for the first two dependent variables. Glyph training led to a significant increase in the emergence of transitivity-indicative performances across all stimulus domains when compared to the no-glyph-training no-staged-testing condition. When glyph training was not followed by staged testing, an intermediate number of participants showed the emergence of a high level ( $\geq 65\%$ ) of transitivity-

indicative responding.

The effect of staged testing not preceded by glyph training was the same for the first two dependent variables. Staged testing alone led to a significant increase in the emergence of transitivity-indicative performances across all stimulus domains when compared to the no-glyph-training no-staged-testing condition. When staged testing did not follow glyph training, a moderate number of participants show a modest increase ( $\geq 20\%$ ) in transitivity-indicative responding. Thus, staged testing, regardless of glyph training, induced moderate increases in transitivity-indicative performances with the stimuli used in testing.

Effects of Stimulus Domains. Glyph training that was not followed by staged testing and staged testing that did not follow glyph training both resulted in the differential emergence of transitivity-indicative performances across stimulus domains. Each condition did not induce transitivity-indicative performances for the glyph sets, but did induce transitivity-indicative performances to a moderate degree for the satellite stimuli and to a high degree for the nonsense syllable stimuli.

The failure of glyph training to generalize to stimuli within the same domain was an unexpected outcome. There are a number of possible explanations for the differential induction of transitivity-indicative performances across stimulus domains. One might involve an interaction with the 'nameability' of the stimuli in each domain. Research has shown that the formation of equivalence classes can be enhanced when verbal labels are trained to one or more members of the set or when members of the sets are 'readily nameable.'

(Eikeseth & Smith, 1992; Randall & Remington, 1999; Smith, Dickins & Bentall, 1996). The results of the current experiment suggest that the properties of the stimuli used in testing correlate with nameability and thus influenced the likelihood of demonstrating transitivity-indicative performances in the context of one experimental variable.

The nonsense syllables, which yielded the highest degree of transitivity-indicative performances, were of consonant-vowel-consonant construction and thus, pronounceable. The content of these stimuli then, may have prompted their naming and may have facilitated the induction of transitivity-indicative performances. The satellite images, which were color pictorial stimuli of various landmasses, yielded a moderate degree of transitivity-indicative performances. The presence of color and contour in each of the images may have made labeling possible but not as easy as with the nonsense syllables. Intermediate levels of nameability would therefore account for the moderate level of transitivity-indicative performances. The glyph images yielded no induction of transitivity-indicative performances under testing. These stimuli were black and white images and were less organized in terms of colors and/or contours. The lack of perceptual organization in the glyph images may correspond to difficulty in naming and thus account for the failure of induction of transitivity-indicative performances under this domain.

Another possible explanation for the differential induction of transitivity-indicative performances across stimulus domains might be governed by the prevalence of the experimental stimuli in each domain in natural settings.

Subjects' differential exposure, on a daily basis, to stimuli like those used in the current experiment may have played a part in the observed outcome. Research has shown that the recognition and recall of English words is a function of the frequency of occurrence in the culture (e.g., McGinnies, Lacey & Comer, 1952).

Typically developing adults are exposed to massive amounts of vocabulary, both verbal and visual, on a daily basis. Exposure to this large number of exemplars then should maximally facilitate the establishment of new discriminations among stimuli that are similarly constructed, like the consonant-vowel-consonant nonsense syllables. Typically developing adults are also exposed to large amounts of visual topographical information on a daily basis, but less so than words. The presence of color and distinct contours in each of the satellite images may parallel the information participants are exposed to and thus account for the moderate level of transitivity-indicative performances. Finally, typically developing adults are exposed to relatively small amounts of abstract black and white shapes like the glyph images. Thus, the lack of history with stimuli that are not perceptually organized may account for the failure of induction of transitivity-indicative performances under the glyph domain.

Single and Combined Effects of Glyph Training and Staged Testing. The bars in the fourth column of Figure 7 show the combined effects of glyph training and staged testing. The combination of both experimental variables resulted in similar high levels of transitivity-indicative performances across stimulus domains. The relative effects of each of these two variables on the induction of transitivity-indicative performances, however, varied with stimulus domain, as

illustrated by the stacked bars in the fifth column of Figure 7. The stacked bars combine the outcomes of glyph training alone (column 2) and staged testing alone (column 3). A comparison of this stacked bar with the bar in fourth column illustrates the extent to which the column four outcome is predicted by the summated effects of glyph training and staged testing.

Glyph training alone and staged testing alone produced little to no induction of the generalized transitivity repertoire with the stimuli in the glyph domain. Thus, the high level of transitivity-indicative performances observed when the two experimental variables were combined (column 4) cannot be accounted for by the effect of glyph training alone combined with the effect of staged testing alone (column 5). Hence, the combined effects of both experimental variables (column 4) surpass the predicted outcome (column 5) by 67%.

The effect of glyph training alone combined with the effect of staged testing alone (column 5) would predict a moderate induction of the transitive repertoire with the stimuli in the satellite domain. This prediction did not equal the actual level of transitivity-indicative performances observed when the two experimental variables were combined (column 4). Rather, the combined effects of both experimental variables (column 4) are greater than the predicted outcome (column 5) by 16%. Furthermore, the disparity between the predicted and the observed outcomes was smaller for the satellite than the glyph domain.

Glyph training alone and staged testing alone each produced high levels of the transitivity-indicative responding repertoire with the stimuli in the nonsense

syllable domain. The high level of transitivity-indicative performances observed when the two experimental variables were combined (column 4) does not meet the predicted outcome when the effect of glyph training alone combined with the effect of staged testing alone (column 5). The predicted outcome exceeds the observed outcome by 20%.

Glyph Training. Another unexpected outcome was observed with participants exposed to glyph training. In general, the transitivity repertoire failed to emerge during glyph training. One reason for this failure may have been the concurrent presentation of baseline conditional discriminations and transitivity test probes from more than two glyph sets. This multiple-glyph-set training and testing occurred after training was completed with sets 9 and 10. After transitivity training was completed with sets 9 and 10, transitivity training and testing was then conducted with more than one pair of glyph sets at once (i.e. 11-12 and 13-14). Hence, the training and testing of for than one pair of glyph sets concurrently may have inhibited the acquisition and subsequent demonstration of the transitive repertoire. Research that examined the formation of equivalence classes has shown that procedures that include the concurrent presentation of training and testing trials are relatively unlikely to produce equivalence classes (Buffington et al., 1997; Fields et al., 1995; Fields et al., 1997; Smeets et al., 1997). To explore this possibility with the induction of transitivity-indicative performances, future studies should limit transitivity training and testing to no more than a single pair of glyph sets at a given time.

Another possible reason for the failure of the transitivity repertoire to

emerge during glyph training may have been that too few glyph sets were presented during training. Many experiments have showed how the number of training exemplars has influenced the formation of stimulus sets (e.g., Bhatt et al., 1988; Homa et al., 1973; Omohundro, 1981). Future studies should explore whether the inclusion of many more glyph sets during transitivity training would lead to the emergence of transitivity-indicative performances.

Staged Testing. When glyph training preceded staged testing, a significant number of sets showed transitivity-indicative performances during the subsequent staged testing. This did not occur when glyph training did not precede staged testing. Thus, glyph training would appear to be needed for the induction of transitivity-indicative performances during staged testing.

Glyph training involved the presentation of two linked baseline conditional discriminations and a transitivity probe for 20 glyph sets. It is possible that transitivity-indicative performances during staged testing could be induced by presentation of multiple conditional discriminations, even if they are unlinked. Varelas, Fields, Rosen, Reeve and Belanich (1999) however, reported that presentation of unlinked conditional discriminations did not lead to the later emergence of equivalence classes. Therefore, it is unlikely that exposure to unlinked conditional discriminations would induce transitivity-indicative performances during staged testing. Further research is required to evaluate this prediction.

Generalized Transitivity Repertoire and Relational Frame Theory. In the current study, transitivity-indicative performances were observed in the presence

of sets of novel stimuli following multiple-exemplar training and/or staged testing. These results were akin to other studies that have induced generalized repertoires. These include learning set (Harlow, 1949), identity matching (Brown et al., 1995; Wright, Cook, Rivera, Sands & Delius, 1988) mirror-image symmetry (Delius & Haber, 1978), associative symmetry (Boelens & Van Den Broek, 2000) and imitation (Young, Krantz, McClannahan, & Poulson, 1994).

The generalized repertoire of associative transitivity observed in the current study may also be characterized as a relational frame (Hayes & Barnes, 1997; O'Hora, Roche, Barnes-Holmes & Smeets, 2002; Steele & Hayes, 1991). Two defining characteristics of relational frames, such as greater than, less than, opposite and different, include (a) the occurrence of a given response in the presence of novel stimuli, following training with a number of exemplars and (b) the stimuli in the frame contain a necessary relation among its elements that control responding. Indeed, the performances in the current study meet each of these criteria. First, transitivity-indicative performances are demonstrated when a stimulus from a conditional discrimination is selected in the presence of another stimulus from another conditional discrimination, even though they have never before been paired together. Second, multiple exemplar training induced transitivity performances in the presence of novel stimuli.

The current study however, also identified staged testing as a variable that induced transitivity indicative performances. To date, research examining the establishment of relational frames has only considered multiple exemplar training and has not considered any other variables that may induce the

repertoire in the presence of novel stimuli. If a generalized transitivity repertoire is a relational frame, it is possible that staged testing might also result in the formation of relational frames. Thus, the current study has the potential of expanding our understanding of the variables that are responsible for the establishment of relational frames.

Table 1

**Sequence of Experimental Treatments for Each Group**

Groups - Conditions	Phase				
	1	2	3	4	5
	Keyboard Familiarization	Preliminary Test	Transitivity Training	Staged Testing	Posttest
I - No Training / No Staged Testing	Y	Y	N	N	Y
II - Glyph Train / No Staged Testing	Y	Y	Y	N	Y
III - No Training / Staged Testing	Y	Y	N	Y	Y
IV - Glyph Training / Staged Testing	Y	Y	Y	Y	Y

Table 2

Stimuli Used During Keyboard Familiarization (Phase 1)

<b>Sample</b>	<b>Positive Comparison</b>	<b>Negative Comparison</b>
<b>Alcohol</b>	<b>Drunk</b>	<b>Mouse</b>
<b>Ant</b>	<b>Bee</b>	<b>Cow</b>
<b>Canary</b>	<b>Sparrow</b>	<b>Stars</b>
<b>Cat</b>	<b>Mouse</b>	<b>Dragons</b>
<b>Comets</b>	<b>Stars</b>	<b>Father</b>
<b>Dog</b>	<b>Wolf</b>	<b>Dark</b>
<b>Dungeons</b>	<b>Chains</b>	<b>Pencil</b>
<b>Eggs</b>	<b>Bacon</b>	<b>Sparrow</b>
<b>Kings</b>	<b>Queens</b>	<b>Camels</b>
<b>Light</b>	<b>Dark</b>	<b>Sock</b>
<b>Mother</b>	<b>Father</b>	<b>Bacon</b>
<b>Mud</b>	<b>Pig</b>	<b>Hat</b>
<b>Paper</b>	<b>Write</b>	<b>Ocean</b>
<b>Red</b>	<b>Color</b>	<b>Pear</b>
<b>Soap</b>	<b>Water</b>	<b>That</b>
<b>This</b>	<b>That</b>	<b>Kings</b>

Table 3

**Sequence, Number of Trials, Percent Feedback and Criterion per Block for Preliminary Test (Phase 2) (Mixed = Glyph, Satellite Images and Nonsense Syllables)**

<b>Conditional Relation</b>	<b>Stimulus Domain</b>	<b># Trials</b>	<b>% Feedback</b>	<b>Criterion - % Correct</b>
<b>AB-Train</b>	<b>Glyphs</b>	<b>8</b>	<b>100</b>	<b>100</b>
	<b>Satellite</b>	<b>8</b>	<b>100</b>	<b>100</b>
	<b>Nonsense Syl.</b>	<b>8</b>	<b>100</b>	<b>100</b>
	<b>Mixed</b>	<b>24</b>	<b>100</b>	<b>100</b>
<b>BC-Train</b>	<b>Glyphs</b>	<b>8</b>	<b>100</b>	<b>100</b>
	<b>Satellite</b>	<b>8</b>	<b>100</b>	<b>100</b>
	<b>Nonsense Syl.</b>	<b>8</b>	<b>100</b>	<b>100</b>
	<b>Mixed</b>	<b>24</b>	<b>100</b>	<b>100</b>
<b>AB-Train</b>	<b>Mixed</b>	<b>24</b>	<b>100</b>	<b>100</b>
<b>BC-Train</b>	<b>Mixed</b>	<b>24</b>	<b>100</b>	<b>100</b>
<b>AB/BC-Train</b>	<b>Mixed</b>	<b>24</b>	<b>100-75-25-0</b>	<b>100</b>
<b>AB/BC AC-Test</b>	<b>Mixed</b>	<b>48</b>	<b>0</b>	<b>100</b>

Table 4

**Sequence, Number of Trials, Percent Feedback and Criterion per Block for Transitivity Training (Phase 3)**

<b>Glyph Sets</b>	<b>Block Name</b>	<b># Trials</b>	<b>% Feedback</b>	<b>Criterion % Correct</b>	<b>Outcome</b>
7-8	AB-Train	8	100	100	
7-8	BC-Train	8	100	100	
7-8	AB/BC-Train	16	100-75-25-0	100	
7-8	AB/BC AC - Test	16	0	100	Pass->Sets 9-10 Fail ->Set 7-8 Train
7-8 Train	AB/BC AC - Train	16	100-50-0	100	
9-10	AB-Train	8	100	100	
9-10	BC-Train	8	100	100	
9-10	AB/BC-Train	16	100-75-25-0	100	
9-10	AB/BC AC - Test	16	0	100	Pass->Sets 11-12 Fail ->Set 9-10 Train
9-10 Train	AB/BC AC - Train	16	100-50-0	100	

(continued on next page)

Table 4 continued

11-12	AB-Train	8	100	100	
13-14	AB-Train	8	100	100	
11-14	AB-Train	16	100	100	
11-12	BC-Train	8	100	100	
13-14	BC-Train	8	100	100	
11-14	BC-Train	16	100	100	
11-14	AB/BC-Train	16	100-75-25-0	100	
11-12	AB/BC AC - Test	16	0	100	Pass->Sets 13-14 Fail ->Set 11-12 Train
11-12 Train	AB/BC AC - Train	16	100-50-0	100	
13-14	AB/BC AC - Test	16	0	100	Pass->Set 11-14 Fail ->Set 13-14 Train
13-14 Train	AB/BC AC - Train	16	100-50-0	100	
11-14	AB/BC AC - Test	32	0	100	Pass->Sets 15-16 Fail ->Set 11-14 Train
11-14 Train	AB/BC AC - Train	32	100-50-0	100	

(continued on next page)

Table 4 continued

15-16	AB-Train	8	100	100	
17-18	AB-Train	8	100	100	
19-20	AB-Train	8	100	100	
15-20	AB-Train	24	100	100	
15-16	BC-Train	8	100	100	
17-18	BC-Train	8	100	100	
19-20	BC-Train	8	100	100	
15-20	BC-Train	24	100	100	
15-20	AB/BC-Train	24	100-75-25-0	100	
15-16	AB/BC AC - Test	16	0	100	Pass->Set 17-18 Fail ->Set 15-16 Train
15-16 Train	AB/BC AC - Train	16	100-50-0	100	
17-18	AB/BC AC - Test	16	0	100	Pass->Set 19-20 Fail ->Set 17-18 Train
17-18 Train	AB/BC AC - Train	16	100-50-0	100	
19-20	AB/BC AC - Test	16	0	100	Pass->Sets 15-20 Fail ->Set 19-20 Train
19-20 Train	AB/BC AC - Train	16	100-50-0	100	
15-20	AB/BC AC - Test	48	0	100	Pass->Sets 21-22 Fail ->Set 15-20 Train
15-20 Train	AB/BC AC - Train	48	0	100	

(continued on next page)

Table 4 continued

21-22	AB-Train	8	100	100	
23-24	AB-Train	8	100	100	
25-26	AB-Train	8	100	100	
21-26	AB-Train	24	100	100	
21-22	BC-Train	8	100	100	
23-24	BC-Train	8	100	100	
25-26	BC-Train	8	100	100	
21-26	BC-Train	24	100	100	
21-26	AB/BC-Train	24	100-75-25-0	100	
21-22	AB/BC AC - Test	16	0	100	Pass->Sets 23-24 Fail ->Set 21-22 Train
21-22 Train	AB/BC AC - Train	16	100-50-0	100	
23-24	AB/BC AC - Test	16	0	100	Pass->Sets 25-26 Fail ->Set 23-24 Train
23-24 Train	AB/BC AC - Train	16	100-50-0	100	
25-26	AB/BC AC - Test	16	0	100	Pass->Sets 21-26 Fail ->Set 25-26 Train
25-26 Train	AB/BC AC - Train	16	100-50-0	100	
21-26	AB/BC AC - Test	48	0	100	Pass->Phase 4 Fail->Set 21-26 Train
21-26 Train	AB/BC AC - Train	48	0	100	

Table 5

**Sequence, Number of Trials, Percent Feedback and Criterion per Block for Staged Testing (Phase 4)**

<b>Conditional Relation</b>	<b>Stimulus Domain</b>	<b># Trials</b>	<b>% Feedback</b>	<b>Criterion - % Correct</b>
AB-Train	Mixed	24	100	100
BC-Train	Mixed	24	100	100
AB/BC-Train	Mixed	24	100-75-25-0	100
AB/BC AC-Test	Glyph	24	0	100
AB/BC AC-Test	Satellite	24	0	100
AB/BC AC-Test	Nonsense Syl.	24	0	100
AB/BC AC-Test	Mixed	48	0	100

Table 6

**Sequence, Number of Trials, Percent Feedback and Criterion per Block for No Staged Testing Group (Phase 5)**

<b>Conditional Relation</b>	<b>Stimulus Domain</b>	<b># Trials</b>	<b>% Feedback</b>	<b>Criterion - % Correct</b>
AB-Train	Mixed	24	100	100
BC-Train	Mixed	24	100	100
AB/BC-Train	Mixed	24	100-75-25-0	100
AB/BC AC-Test	Mixed	48	0	100

Table 7

**Percentage of Sets that Showed Transitivity-Indicative Performances in Across All Preliminary Training Mixed Test Blocks and the Final Mixed Test Block in the Posttest for All Participants**

		Glyph Training			
		No		Yes	
Staged Testing					
Participants	Preliminary Test	Posttest	Participants	Preliminary Test	Posttest
AV0	0	0	AT9	11	33
JW3	33	33	CY6	22	100
JE2	11	33	JH5	22	100
DL8	33	66	AB6	22	100
LM7	0	33	HL7	11	100
SS7	0	33	EG0	0	100

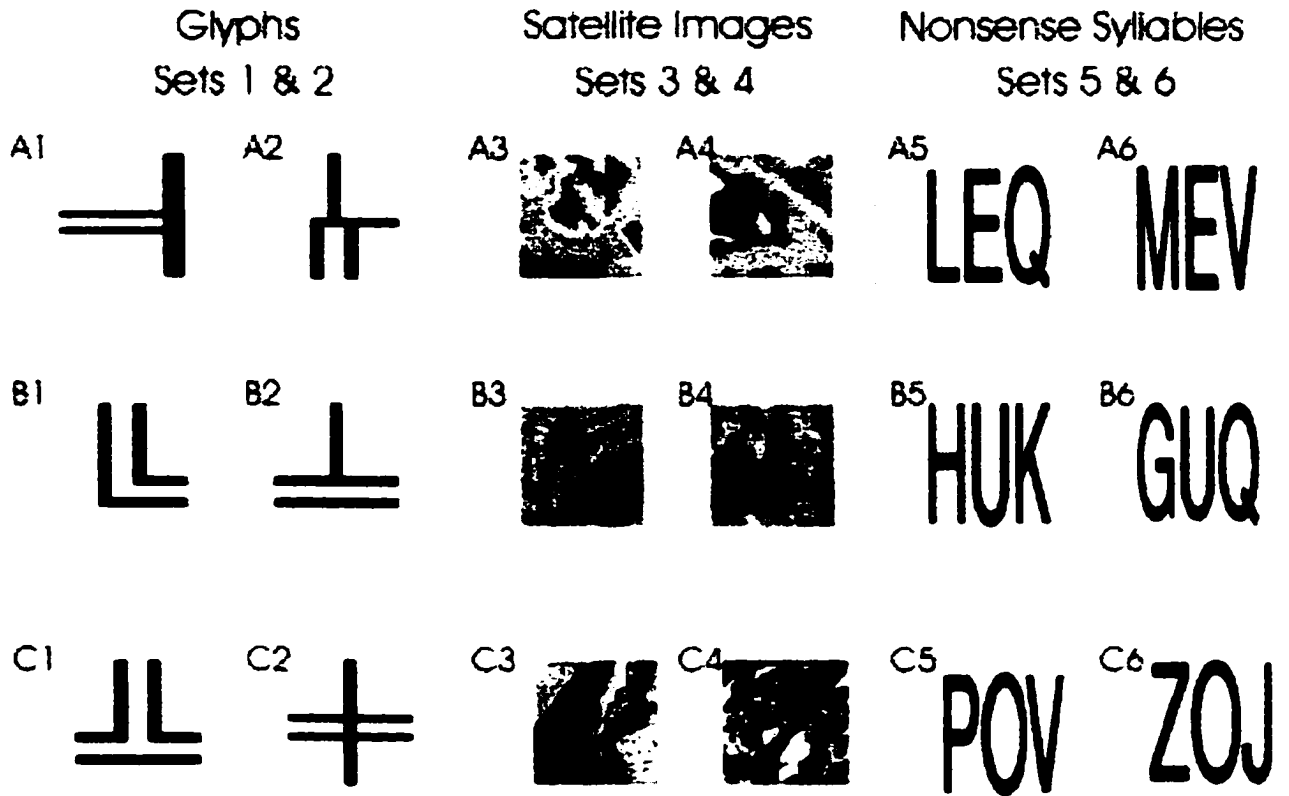
No Staged Testing					
Participants	Preliminary Test	Posttest	Participants	Preliminary Test	Posttest
SL9	11	0	YB4	11	0
NC9	11	0	SS9	22	33
YK3	0	0	EM9	11	33
AY2	33	33	HY1	11	66
CE7	33	33	IH3	0	66
IS6	0	33	MY5	22	100

Table 8

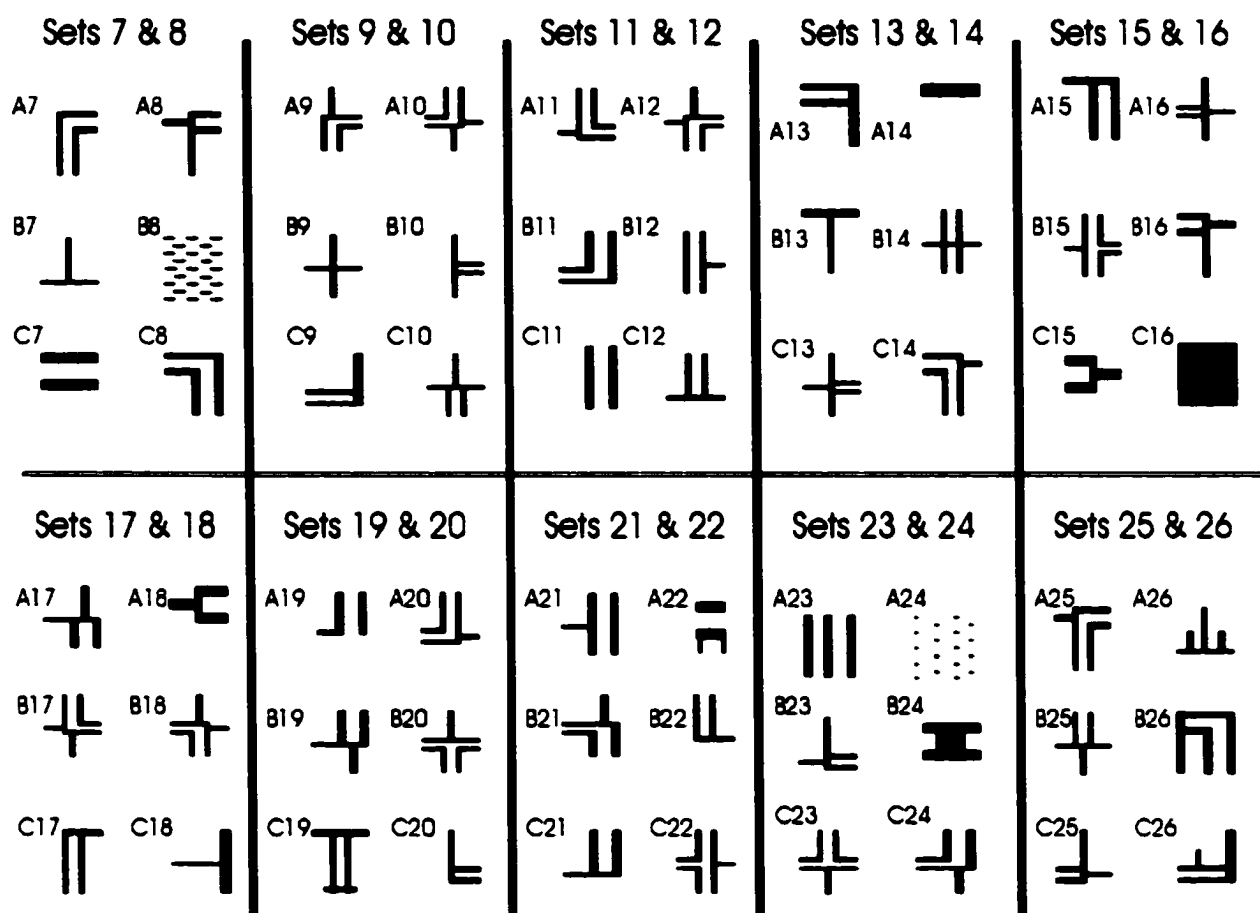
**Number of Participants in Each Group Who Failed to Show Transitivity-Indicative Performances Under Each of the Stimulus Domains During Preliminary Testing**

Group	Glyph Training	Staged Testing	Participants per domain		
			Glyph	Sat	CVC
I	N	N	5	5	5
II	Y	N	6	4	5
III	N	Y	5	5	6
IV	Y	Y	6	4	5

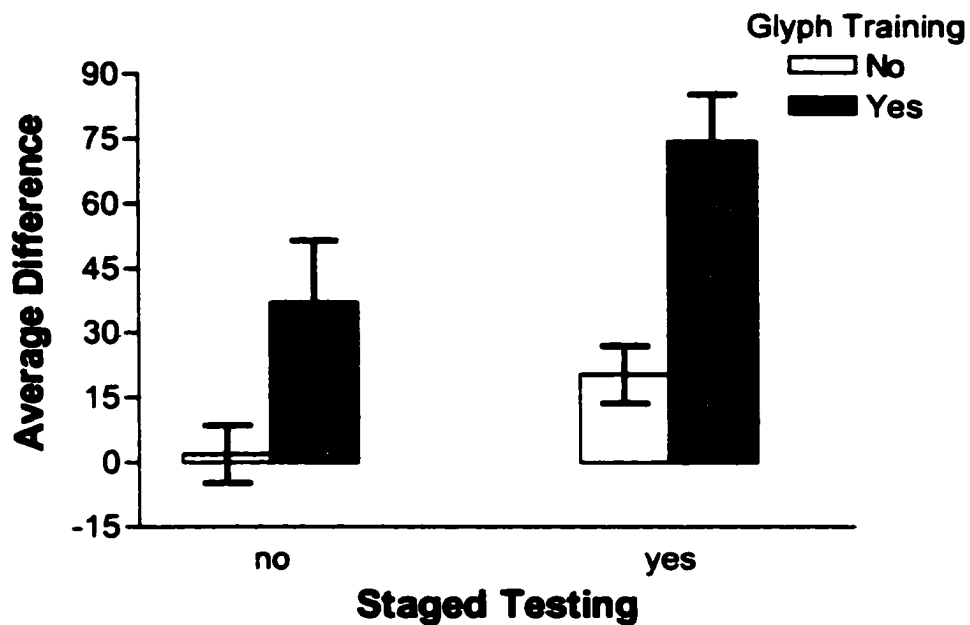
**Figure 1.** The members of the glyph, satellite and nonsense syllable stimulus sets used during preliminary training and the posttest. Stimulus position within a set is indicated by the letter (A, B or C) and set membership is represented by the number (1 - 6).



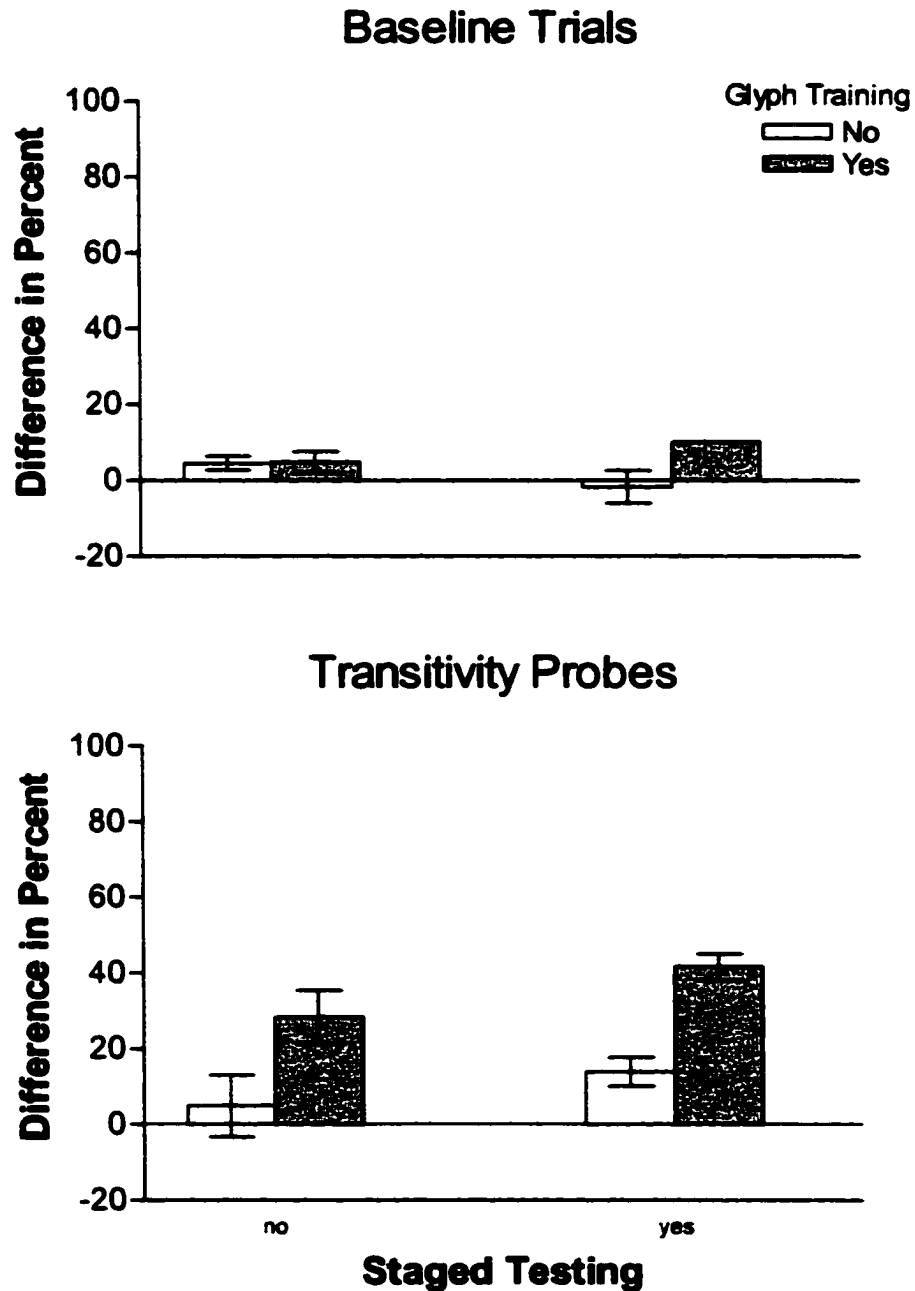
**Figure 2.** The twenty glyph sets presented during transitivity training. Stimulus position within a set is indicated by the letter (A, B or C) and set membership is represented by the number (7 - 26).



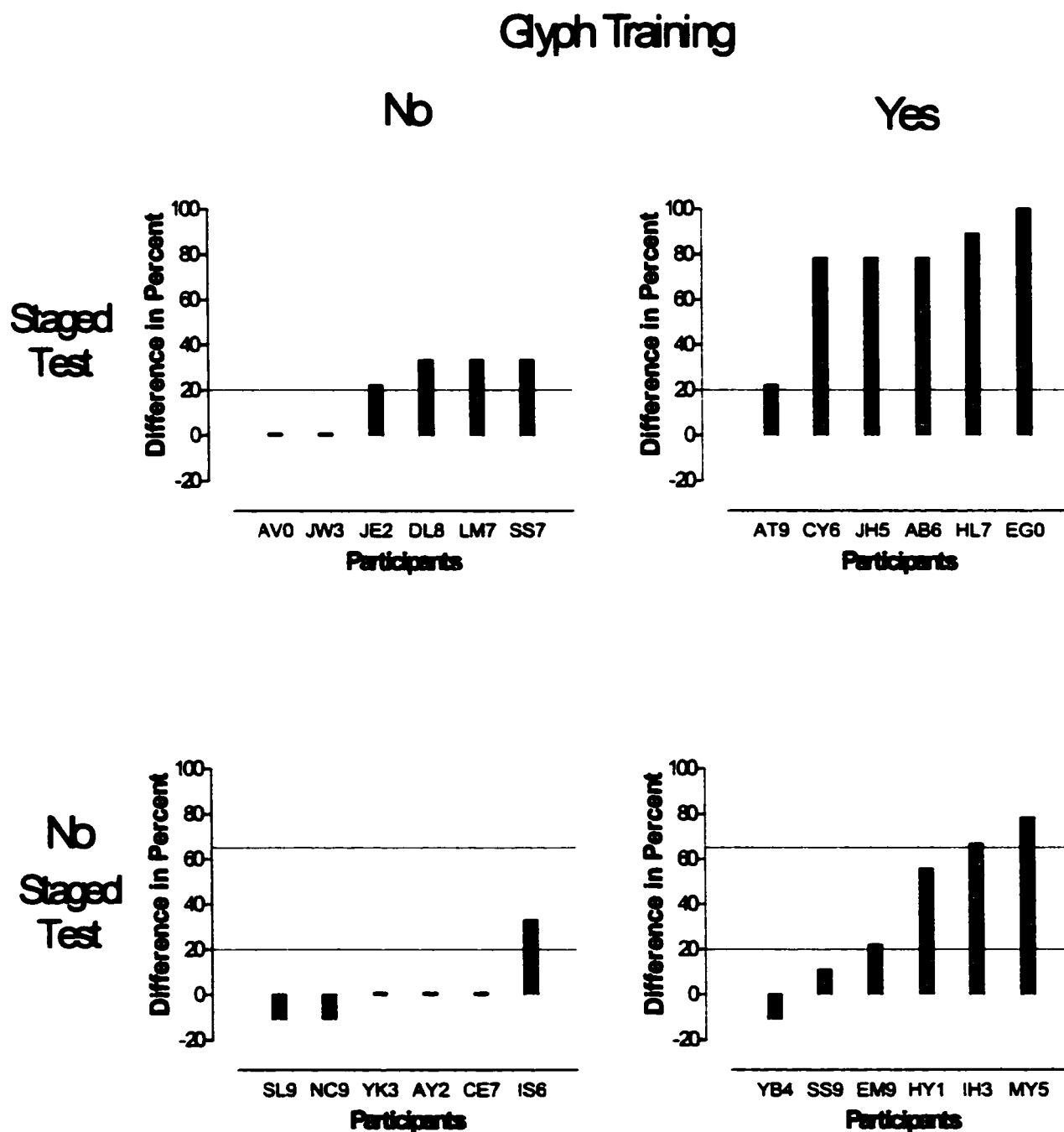
**Figure 3.** The average difference and standard error of the percentage of sets that occasioned transitivity-indicative performances between all preliminary training mixed test blocks and the final mixed test block in the posttest, for each condition.



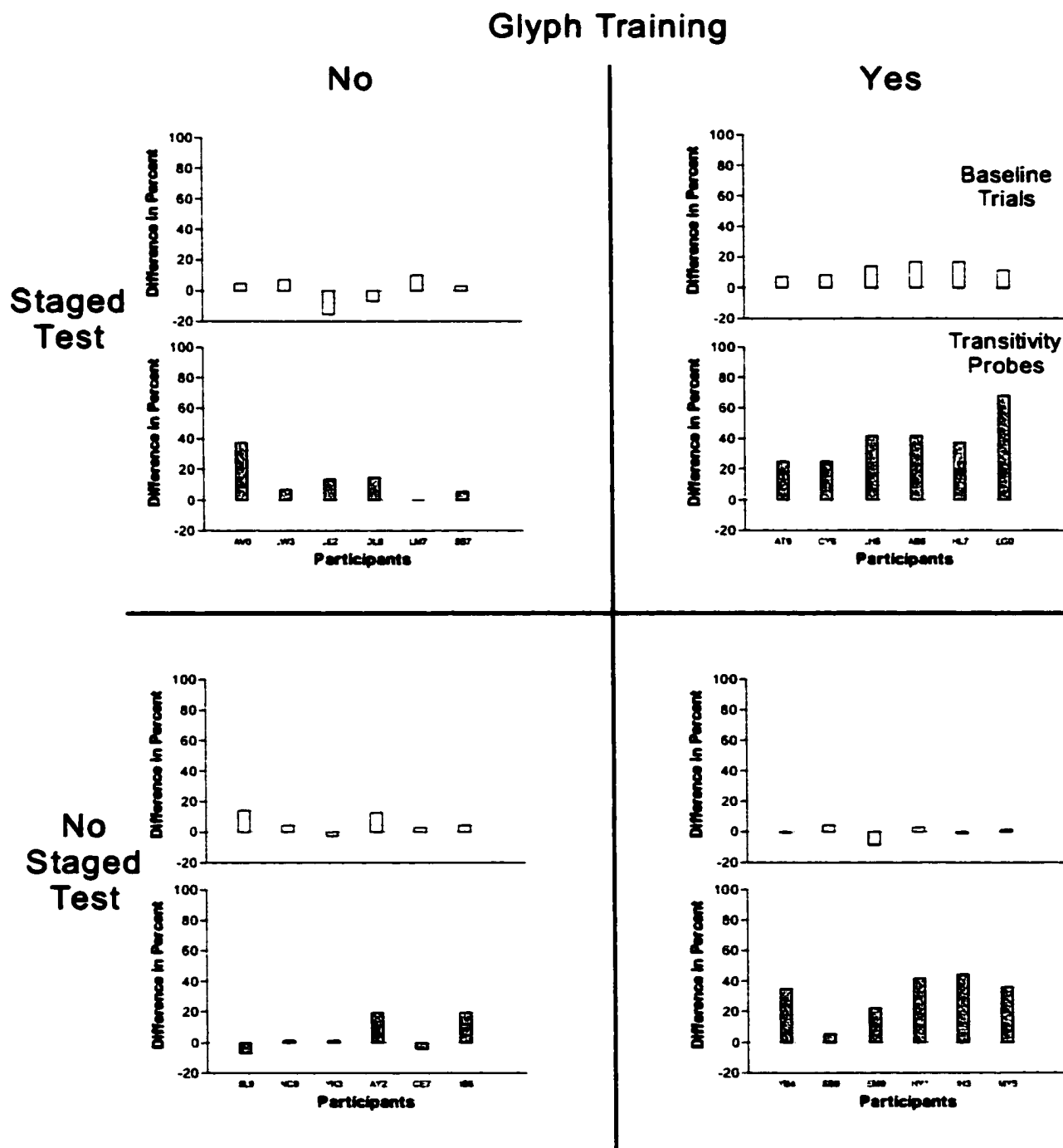
**Figure 4.** The top frame shows the average difference and standard error of the baseline conditional discrimination trials that show class-consistent responding between all preliminary training mixed test blocks and the final mixed test block during the posttest. The bottom frame shows the same for the transitivity probes. Outcomes for all groups are shown in each frame.



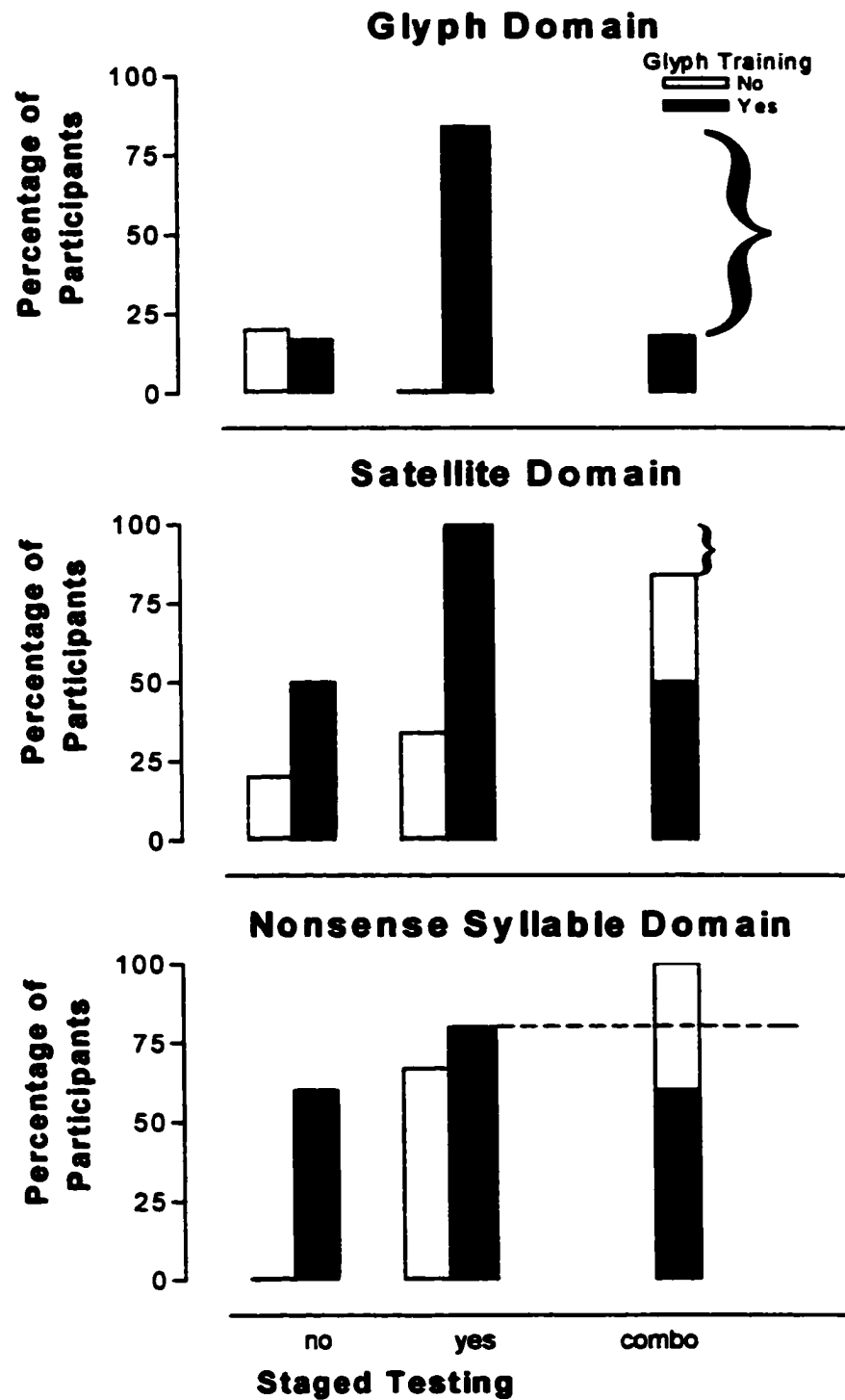
**Figure 5.** The difference in the percentage of sets that occasioned transitivity-indicative performances between all preliminary training mixed test blocks and the final mixed test block in the posttest, for each participant. The horizontal lines indicate the criterion levels of 20% and 65%. The participants in each group are arranged according to level of performance.



**Figure 6.** The difference in the percentage of trials that occasioned set consistent responding between all preliminary training mixed test blocks and the final mixed test block in the posttest, for each participant. The top frame in each pair shows the difference for the baseline conditional discriminations, and the bottom frame in each pair shows the difference for the transitivity probes. Participants are arranged corresponding to placement in Figure 5.

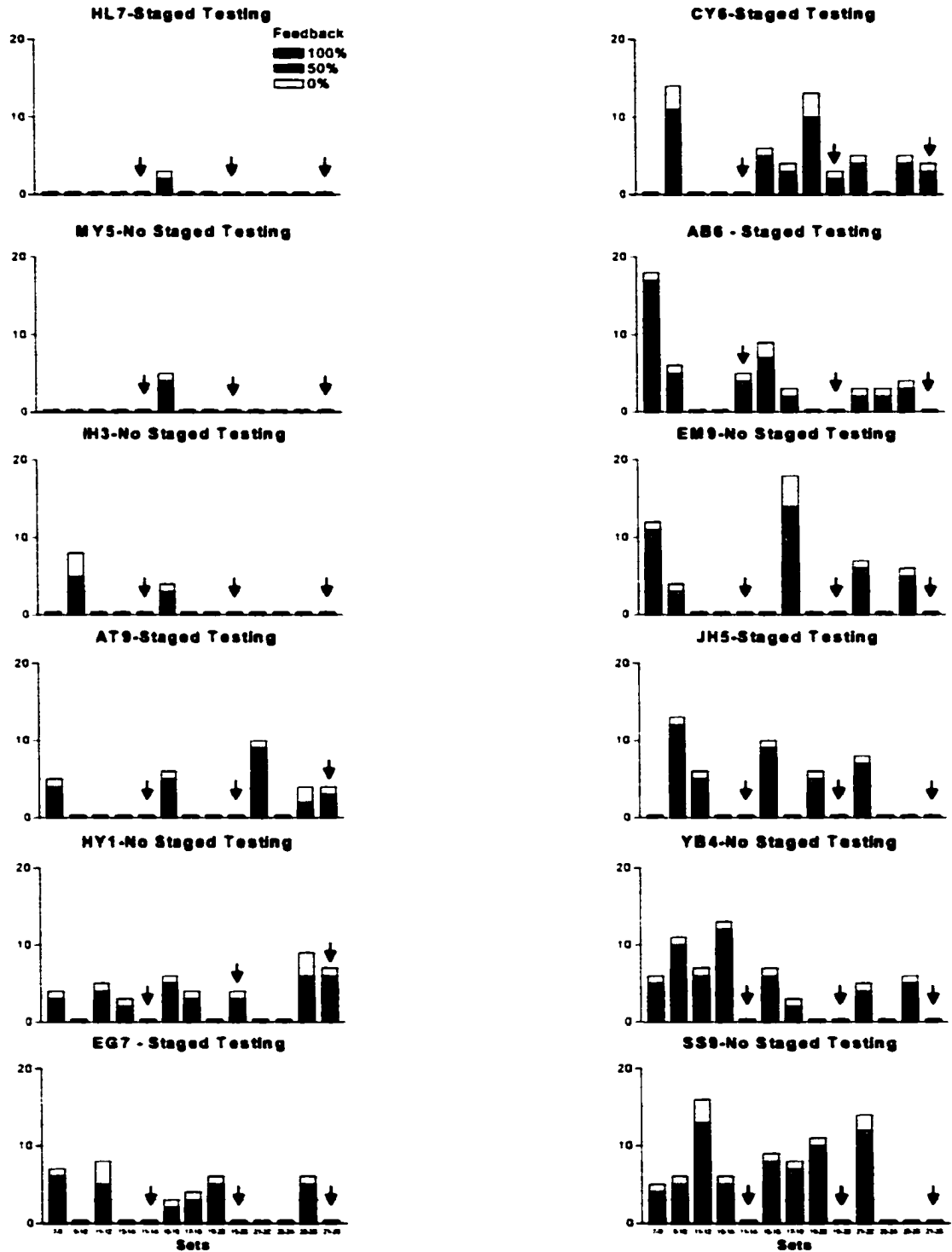


**Figure 7.** The average percentage of participants who showed an increase in sets that occasioned transitivity-indicative performances, per stimulus domain for each condition. The fifth column in each panel is a stacked bar of the data from the second and third columns respectively.

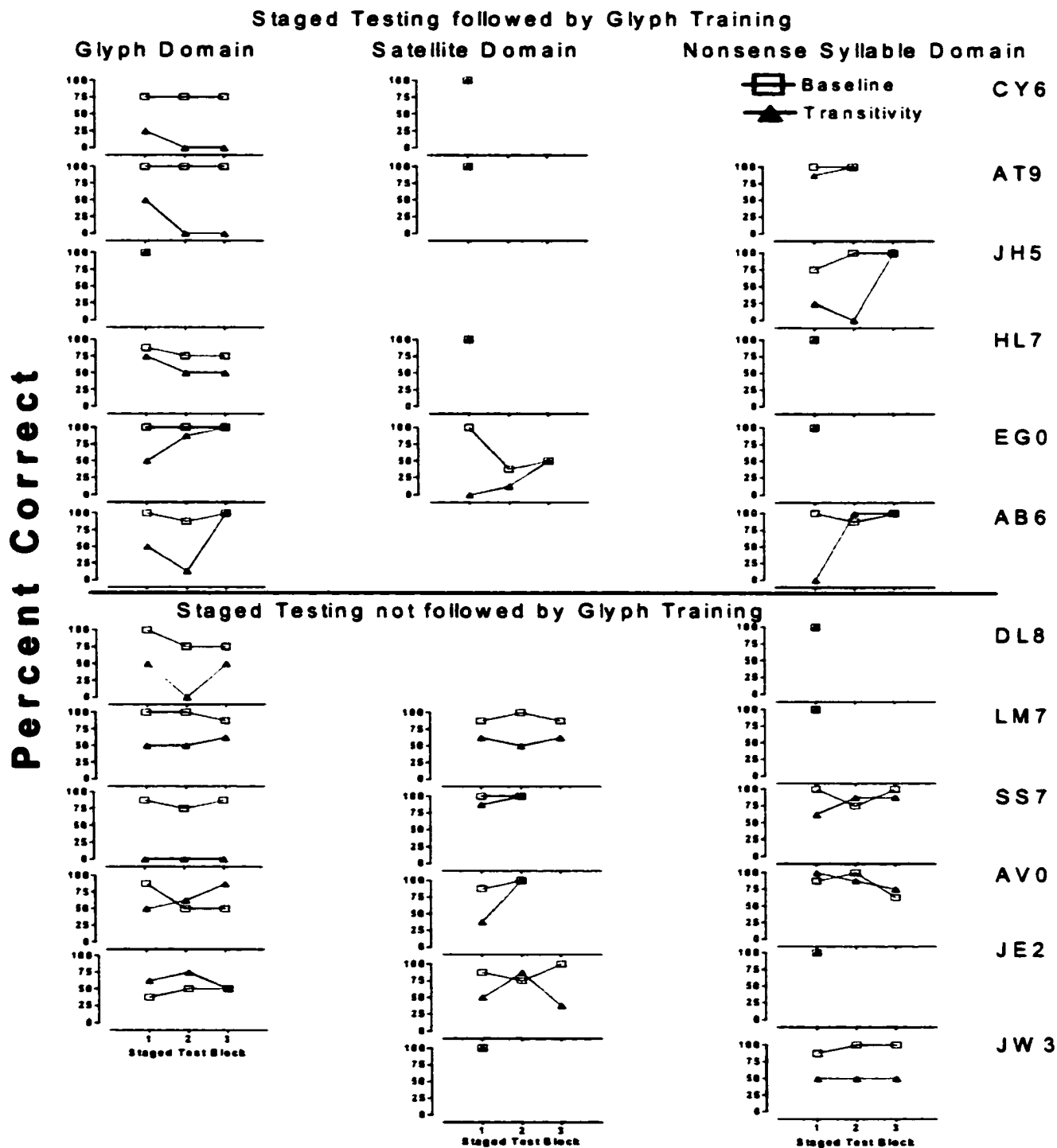


**Figure 8.** The number of training blocks, at each feedback level, from sets that did not show transitivity-indicative performances for each participant in the glyph training conditions. Participants are rank ordered according to the number of training blocks required at the 100% feedback level.

**Cumulative Number of Training Blocks**



**Figure 9.** The percentage of trials that occasioned set consistent responding for participants exposed to staged testing who failed to show transitivity-indicative performances during preliminary training under each stimulus domain. The lines connected via the squares indicated baseline performances and the lines connected by the triangles indicate performances on the transitivity trials.



## Bibliography

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

Adams, B. J., Fields, L., & Verhave, T. (1999). Effects of unreinforced conditional selection training, multiple negative comparison training, and feedback on equivalence class formation. The Psychological Record, 49, 685-702.

Arntzen, E., & Holth, P. (2000). Differential probabilities of equivalence outcome in individual participants as a function of training structure. The Psychological Record, 50, 603-628.

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2001). Exemplar training and a derived transformation of function in accordance with symmetry. The Psychological Record, 51, 287-308.

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2001). Exemplar training and a derived transformation of function in accordance with symmetry: II. The Psychological Record, 51, 589-603.

Berryman, R., Cumming, W. W., Cohen, L. R., & Johnson, D. F. (1965). Acquisition and transfer of simultaneous oddity. Psychological Reports, 17, 767-775.

Bhatt, R. S., Wasserman, E. A., Reynolds, W. F., & 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.

Boelens, H., & Van Den Broek, M. (2000). Influencing children's symmetric responding in matching-to-sample tasks. The Psychological Record, 50, 655-670.

Brown, A. K., Brown, J. L., & Poulson, C. L. (1995). Generalization of children's identity matching-to-sample performances to novel stimuli. The Psychological Record, 45, 29-43.

Buffington, D. M., Fields, L., & Adams, B. J. (1997). Enhancing equivalence class formation by pretraining of other equivalence classes. The Psychological Record, 47, 69-96.

Cook, R. G., Wright, A. A., & Kendrick, D. F. (1990). Visual categorization by pigeons. In M. L. Commons, R. J. Heermstein, S. Kosslyn, & D. Mumford (Eds.) Models of Behavior: Behavioral Approaches to Pattern Recognition and Concept Formation. Hillsdale, NJ: Erlbaum.

Cumming, W. W., & Berryman, R. (1961). Some data on matching behavior in the pigeon. Journal of the Experimental Analysis of Behavior, *29*, 565-601.

Delius, J. D., & Habers, G. (1978). Symmetry: Can pigeons conceptualize it? Behavioral Biology, *22*, 336-342.

Devaney, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. Journal of the Experimental Analysis of Behavior, *46*, 243-257.

Dymond, S., & Barnes, D. (1995). A transformation of self-discrimination response functions through the arbitrarily applicable relations of sameness, more-than, and less-than. Journal of the Experimental Analysis of Behavior, *64*, 163-184.

Dymond, S., & Barnes, D. (1996). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness and opposition. The Psychological Record, *46*, 271-300.

Eikeseth, S., & Smith, T. (1992). The development of functional and equivalence classes in high-functioning autistic children: The role of naming. Journal of the Experimental Analysis of Behavior, *58*, 123-133.

Fields, L., Adams, B. J., Newman, S., & Verhave, T. (1992). Interactions among Emergent Relations during Equivalence Class Formation. The Quarterly Journal of Experimental Psychology, *45*, 125-138.

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. (2000). 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 (pp. 59-84). Reno, NV: Context Press.

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., Varelas, A., Reeve, K. F., Belanich, J., Wadhwa, P., DeRosse, P., & Rosen, D. (2000). Effects of prior conditional discrimination training, symmetry, transitivity, and equivalence testing on the emergence of new equivalence classes. The Psychological Record, *50*, 443-466.

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

Harlow, H. F. (1949). The formation of learning sets. Psychological Review, *56*, 51-65.

Hayes, S. C., & Barnes, D. (1997). Analyzing derived stimulus relations requires more than the concept of stimulus classes. Journal of the Experimental Analysis of Behavior, *68*, 235-243.

Hayes, S. C., & Hayes, L. J. (1989). The verbal action of the listener as the basis for rule-governance. In S. C. Hayes (Ed.), Rule-governed behavior: Cognition, contingencies, and instructional control (pp. 153-190). New York: Plenum Press.

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

Herrnstein, R. J. (1990). Levels of stimulus control: A functional approach. Cognition, *37*, 133-136.

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 55 – 95. New York: Academic Press.

Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. The Psychological Record, *48*, 81-110.

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

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

Homa, D., & Little, J. (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.

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

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

Kennedy, C. H., & Laitinen, R. E. (1988). Second-order conditional control of symmetric and transitive stimulus relations: The influence of order effects. The Psychological Record, 38, 437-446.

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

Lynch, D. C., & Cuvo A. J. (1995). Stimulus equivalence instruction of fraction-decimal relations. Journal of Applied Behavior Analysis, 28, 115-126.

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

McGinnies, E., Lacey, O. L., & Comer, P. B. (1952). Visual-recognition threshold as a frequency of word length and word frequency. Journal of Experimental Psychology, 44, 65 - 69.

O'Hora, D., Roche, B., Barnes-Holmes, D., & Smeets, P. (2002). Response latencies to multiple derived relations: Testing two predictions of relational frame theory. The Psychological Record, 52, 51-76.

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

Premack, D. (1971). Language in chimpanzees? Science, 172, 808-822.

Premack, D. (1976). Intelligence in ape and man. Hillsdale, NJ: Erlbaum.

Randall, T., & Remington, B. (1999). Equivalence relations between visual stimuli: The functional role of naming. Journal of the Experimental Analysis of Behavior, 71, 395-416.

Reeve, K. F., & Fields, L. (2001). Perceptual classes established with forced-choice primary generalization tests and transfer of function. Journal of the Experimental Analysis of Behavior, 76, 95-114.

Saunders, R. R., Wachter, J. A., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member stimulus class via conditional discrimination procedures. The Journal of the Experimental Analysis of Behavior, 49, 95-115.

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

Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. Journal of Experimental Analysis of Behavior, 37, 5-22.

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., Willson-Morris, M., & Kirk, B. (1986). Matching-to-sample procedures and the development of equivalence relations: The role of naming. Analysis and Intervention in Developmental Disabilities, 6, 1-19.

Sigurdardottir, Z. G., Green, G., & Saunders, R. R. (1990). Equivalence classes generated by sequence training. Journal of the Experimental Analysis of Behavior, 53, 47-64.

Smeets, P. M., Leader, G., & Barnes, D. (1997). Establishing stimulus classes in Adults and children using a respondent-type training procedure: A follow up study. The Psychological Record, *47*, 285-308.

Smoke, K. L. (1932). An objective study in concept formation. Psychological Monographs, *42*, Whole no. 191.

Smith, A. B., Dickens, D. W., & Bentall, R. P. (1996). The role of interfering stimulus names in the emergence of equivalence relations II: The effects of interfering tasks prior to and after tests for equivalence relations. The Psychological Record, *46*, 109-130.

Spradlin, J. E., Cotter, V. W., & Baxley, N. (1973). Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. American Journal of Mental Deficiency, *77*, 556-566.

Steele, D., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. Journal of the Experimental Analysis of Behavior, *56*, 519-561.

Varelas, A, Fields, L., Rosen, D., Reeve, K. F., & Belanich, J. (May, 1997). Effects of prior establishment of pseudo-equivalence classes on the formation of new equivalence classes under the simultaneous protocol. A paper presented in the symposium, Variables that influence the emergence and extension of equivalence classes, at the Annual Meeting of the Association for Behavior Analysis. Chicago, Ill.

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

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

Wulfert, E., Dougher, M. J., & Greenway, D. E. (1991). Protocol analysis of the correspondence of verbal behavior and equivalence class formation. Journal of the Experimental Analysis of Behavior, *56*, 489-504.

Yamamoto, J. (1986). Establishing stimulus equivalences in autistic children. Japanese Journal of Behavior Analysis, *1*, 3-22.

Young, J. M., Krantz, P. J., McClannahan, L. E., & Poulson, C. L. (1994). Generalized imitation and response-class formation in children with autism. Journal of Applied Behavior Analysis, *27*, 685-697.