TRAINING STRUCTURE, NAMING AND TYPICALITY EFFECTS IN EQUIVALENCE CLASS FORMATION

Jeanette E. Wilson

A Thesis Submitted to the
University of North Carolina Wilmington in Partial Fulfillment
Of the Requirements for the Degree of
Master of Arts

Department of Psychology

University of North Carolina Wilmington

2006

Approved by

Advisory Committee

____________________ ____________________

____
Chair

Accepted by

____
Dean, Graduate School

TABLE OF CONTENTS

ABSTRACT

While equivalence is a well-documented phenomenon, its basis is of considerable debate. The current experiment looked at the effects of training structure and naming in the acquisition of conditional discriminations and equivalence-class formation. The experiment also looked to see if typicality effects would occur in children's equivalence classes and if so, whether it would be impacted differently by the different training structures. Four groups of children learned conditional discriminations using a match-to sample (MTS) procedure. The stimuli used were trigrams and one-to four-feature stimuli. Three different training structures were employed in the training, the one-to-many training structure using the trigram as the node, the many-to-one training structure using the trigram as the node and the many-to-one training structure using a two-feature stimuli as the node. Results showed that children learned the conditional discriminations more quickly in the many-to-one, two-feature-as-node training structure. The results also showed the formation of equivalence classes with different training structures. An analysis of typicality effects was also formulated.

ACKNOWLEDGEMENTS

I would like to thank my parents, Terry and Linda Wilson who have always believed in me. I would also like to thank Dr. Pilgrim who has always pushed me to do my best.

LIST OF TABLES

Table Page

LIST OF FIGURES

INTRODUCTION

Equivalence and Equivalence Research

In the applied behavioral world, many concepts have emerged over the years as being important, fundamental concepts of behavior. One of these is the concept of stimulus equivalence. Stimulus equivalence and stimulus relations are important to study for many reasons. One of the main reasons that they are so important is their relation to verbal and written language. All words, whether we are the speaker or the listener, have effects according to the environment or context in which they are spoken or written. However, words are not physically “the same as” their environmental referents. For example, a boat is not physically the same as the spoken word "boat", the written word 'boat' or even a picture of a boat. All four are completely different from each other and yet they function equivalently, or the same, in our behavioral response to them. According to Sidman (1994), “we often react to words and other symbols as if they are the things or events that they refer to.” In noting that equivalence is found inherently in language, it is important to realize how equivalence forms and what leads to the quickest acquisition of equivalence.

The Basis for Equivalence

Stimulus equivalence begins with conditional discrimination training. Catania (1992) states that a conditional discrimination forms when there is a choice between at least two stimuli and that choice depends upon the presence of another stimulus. A match-to-sample (MTS) procedure is often used to train conditional discriminations. For example, one would teach a child that when the spoken word ‘boat’ (A1) was presented as the sample, choosing (B1) the written word ‘boat’ would be reinforced, while choosing the written word ‘tree’ (B2) would not be reinforced. When the spoken word “tree” (A2) is presented as the sample, choosing (B2) would be reinforced. This is one arbitrary conditional discrimination. Before assessing equivalence, one must teach at least two inter-related arbitrary conditional discriminations (i.e., AB as above and AC). According to Pilgrim, Chambers and Galizio (1995), these trained “baseline relations are commonly held to be the basis for equivalence-class performance” (p. 239).

The Three Defining Requirements of Equivalence

Once the baseline conditional discriminations are acquired one can test for equivalence. The behavioral tests for equivalence are based on the defining properties of equivalence in mathematics. Three types of tests are needed to assess equivalence; all three test for relations that have not been directly taught. The first, reflexivity, is sometimes called identity matching (i.e., A=A). The second is symmetry. In this case if A=B, then B=A. The third test is for transitivity; if A=B and A=C, then B=C. These properties of equivalence are tested using the MTS arrangement. The first letter shown in the property definitions above would be presented as the sample and the second would be one of the comparisons. When the subject demonstrates the ability to perform all three tests without explicit reinforcement for doing so, then equivalence is said to have emerged. The relations that emerge are “emergent in the sense that we do not explicitly teach them in the baseline” (Sidman, 2000, p. 130). An example of the trained and emergent relations is shown in Figure 1. (Figure 1 does not include the operation of reflexivity, which would show that A = A, B = B, and C = C.) In the example described above, the subject will have demonstrated two equivalence classes. Class 1 would include stimuli A1, B1 and C1, or in the example above, the spoken word "boat", the written

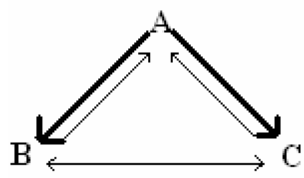


Figure 1. The basic equivalence model, with all trained and emergent relations (Sidman & Tailby, 1982). The bold lines are the relations that are explicitly or directly taught. The lighter lines are the emergent relations.

word “boat”, and the picture of a boat. The second class would contain A2, B2, and C2, or the spoken word, written word, and picture of a tree.

Sidman and Cresson (1973) demonstrated the formation of equivalence classes using two institutionalized teenage boys with severe retardation. For baseline training, the subject faced a panel that had nine circular openings, one in the center with eight surrounding it in the shape of a circle. These openings were like windows on which visual stimuli could be projected. The sample stimulus appeared in the center. The subject had to first touch the center sample stimulus and when he did the comparison stimuli appeared in the other openings. Correct responses were reinforced with a penny and ringing chimes. Incorrect responses produced no programmed consequences. There was also an oral naming test in which the subject had to name each word or picture when presented in isolation. The discriminations that were explicitly taught were: C (printed word) to C (printed word) with twenty different words, which is identity matching; A (auditory word) to B (picture) with all twenty words; and A (auditory word) to C (printed word). In the baseline A (auditory word) to C (printed word) training, the experimenters first taught nine words, then fourteen, and finally twenty. At each of the steps along the way, the experimenters also tested for symmetry (B pictures to A auditory words, C printed words to A auditory words); and transitivity (B pictures to C printed words, C printed words to B pictures). Through teaching the baseline relations, many others emerged that had not been explicitly taught. After teaching all twenty A (auditory word) to C (printed word) relations, all of the relations that were not taught emerged. This experiment shows the exponential effects of the stimulus equivalence model.

Animal Equivalence Studies

Researchers have also been interested in the possibility that nonhuman animals would show the formation of equivalence classes. Many experiments have shown category-like performances (Zentell, 1996; Urcuioli, 2003) in animals. Category formation in animals has sometimes been studied by testing for functional classes. Functional stimulus classes are "sets of discriminative stimuli that control the same behavior" (Kastak, Schusterman, & Kastak, 2001). Herrnstein and Loveland (1964) showed that pigeons could demonstrate categories or functional classes. They showed pigeons photographs. Some of the pictures included humans or human body parts. The pigeon's responses were reinforced when a human or human part was in the picture but not when pictures appeared without humans. When later tested with novel probes, which were new pictures that did or did not contain humans, the pigeons showed (by correctly responding when a human was pictured) that slides showing humans had become a functional class. This experiment was repeated using other stimuli such as flowers and cats and the pigeons showed the same results (Herrnstein & Loveland, 1964). These findings show generalization to novel stimuli with pigeons. This is an important finding and has been discussed in terms of categorization and functional class formation based on reinforcement contingencies.

Vaughan (1988) performed an experiment on the formation of equivalence sets using pigeons. In this experiment he used six experimentally naive pigeons. The stimuli used were forty pictures of trees. These pictures were divided into two main sets, the positive set, Set 1, and the negative set, Set 2. Each picture was shown twice during each session. Pecking a key in the presence of a Set 1 picture resulted in food reinforcement.

Responses to Set 2 stimuli were unreinforced.

In Session 15, the contingencies were reversed. All of the Set 1 stimuli became negative (responses were unreinforced), and all of the Set 2 stimuli became positive (responses were reinforced). The contingencies were reversed frequently after this first reversal, at first after every seven sessions, then after every four, then five, then after five, six or seven sessions.

For data analysis, Vaughan used the measure of rho, which is “the probability of ranking a positive over a negative (p. 38).” At the beginning of baseline acquisition, rho was at 0.5, which was chance performance. As the pigeons learned the contingencies, rho approached 1, which was correct performance. At the first reversal, the pigeons returned to a rho value of about 0.5 and then began to rise to higher values. The main finding in this experiment was "a gradual increase in the value of rho as a function of the number of reversals (p. 39).” In other words, the more reversals the pigeon was exposed to, the better he was at detecting the reversal quickly and at reversing responses to all stimuli in the sets after exposure to only a few. According to the author the pigeons, who had no language skills, showed arbitrary stimulus equivalence.

The concept of functional classes and equivalence in nonhuman animals was also described in Kastak et al. (2001). They trained two female sea lions two functional classes using arbitrary stimuli, letters and numbers, and a two-choice simple discrimination task. Correct responses were reinforced with a tone and a piece of fish. Incorrect responses were not reinforced and produced a vocal signal “no.” Within each session one class of arbitrary stimuli (either letters or numbers) was assigned the positive role, which signified that responding to members of that class resulted in reinforcement.

Upon mastering a series of letter-number discrimination, the contingencies were reversed such that responses to the class that was previously negative were now reinforced, and responses to the other class produced no reinforcement. Functional classes were demonstrated when experience with a few members of the class following a contingency reversal changed the responding to the other members of that class. By the end of this phase of the experiment, both of the animals were correctly responding on the reversal after only a few trials. The sea lions had demonstrated functional classes.

The second experiment by Kastak et al. (2001) asked whether the functional classes would serve as a basis for conditional discriminations. The same apparatus was used, this time in a MTS format. The sample was presented in the middle of the two comparisons. The first step in this experiment was to demonstrate maintenance of previously trained MTS relations with different stimuli. A subset of the stimuli from the previous experiment were then assigned to three four-member groups A, B, and C. These twelve stimuli were combined to test for six new conditional discriminations. Each stimulus appeared in different trials as a sample, a positive comparison, or a negative comparison. Throughout this experiment the reversal that was established in Experiment 1 was maintained. If the functional classes had also established a basis for emergent conditional discriminations then, when a number was the sample, the animals should pick a number as the correct comparison, and the same with the letters. The animals' responding was significantly better than chance.

The third experiment (Kastak et al., 2001) tested to see if the conditional discriminations shown in the second experiment would generate equivalence classes. There were two stages in this experiment. The first involved training the animals to match new stimuli to an established class member via a MTS procedure. The second involved testing these novel stimuli in the simple discrimination task of Experiment 1 to see if they reversed with all the other letters or numbers. “Untrained relations emerged between the new stimuli and the remaining members of each functional class” (Kastak et al., 2001, p. 150). The results of this experiment showed that two California sea lions formed equivalence classes.

One reason that animal studies are important is because they suggest that naming is not a requirement for equivalence. The issue of naming is described in detail later in this paper.

Different Theories of Equivalence

While equivalence is a well-documented phenomenon, its basis is of considerable debate. There are three main competing theories of equivalence. The first is Murray Sidman’s view (e.g., 1994, 2000), the second is relational frame theory (e. g., Hayes, Fox, Gifford, Wilson, Barnes-Holmes, & Healy, 2001; Lipkens, Hayes & Hayes, 1993), and the third is the naming theory (Dugdale & Lowe, 1990; Horne & Lowe, 1996).

Sidman’s Theory of Equivalence

According to Sidman (2000), “equivalence is a direct outcome of reinforcement contingencies” (p. 127-128). The equivalence relation describes observable behavior. In Sidman’s view, there are two outcomes of any reinforcement contingency; equivalence relations and analytical units. A contingency implies a dependency, for example, between a response and a reinforcer delivery. A simple three-term contingency is the basis for discrimination learning. A four-term contingency results in a conditional discrimination. An analytic unit may have two, three, four, five or greater term contingencies. According to Sidman equivalence would result from any of these arrangements. Equivalence relations consist of “ordered pairs of all positive elements that participate in the contingency” (p. 131). In a four-term contingency, these positive elements include the conditional stimulus, the discriminative stimulus, the response, and the reinforcer. Equivalence is defined in mathematical terms, but that definition alone says nothing about where equivalence comes from (Sidman, 2000).

Because the reinforcer is part of the equivalence relation, if a single reinforcer is used for conditional discrimination training, then all of the stimuli involved could become equivalent by virtue of their relation to that reinforcer. In other words, if correct selections of any stimulus were followed by the visual presentation of stars, all of the stimuli may become equivalent because the stars are related to every stimulus. This could cause what is known as class collapse (Sidman, 2000). The same is true when a single response is used throughout conditional discrimination training. The best way to avoid this issue is to employ class-specific reinforcers and responses.

Sidman’s theory is that equivalence is a fundamental process. Equivalence relations are the outcome of reinforcement contingencies.

Relational Frame Theory (RFT)

Relational frame theorists (RFT) do not agree with Sidman’s suggestion that equivalence is a direct outcome of reinforcement contingencies. They argue instead that equivalence is learned behavior. The responding must be taught (Hayes, S. et al. 2001) Lipkens, Hayes and Hayes (1993) suggest that the necessary history includes “training that leads to the development of generalized arbitrarily applicable relational responding ” (p. 203). The theory is that direct training with many different exemplars allows relations to be abstracted. The direct training is what allows these abstracted relations to be applied to new exemplars. Equivalence is but one example of a relation that can be learned.

Relational frame theory also emphasizes two additional points. The first is that data have shown that humans and nonhumans can respond to non-arbitrary relations between stimuli (e.g., this object is smaller than that object), and that responding may then come under contextual control. From that point, only the contextual cues would be needed to allow abstracted responding to be arbitrarily applied to any event. Second, “organisms can learn overarching behavioral classes containing virtually unlimited numbers of members” (Lipkens et al., p. 203). Relational frame theory in this sense is much broader than the Sidman view. Relational frame theory includes all relational responding Hayes, S. et al. (2001). An example of an overarching operant class would be generalized imitation.

Naming Theory

Proponents of the naming theory believe that naming is both necessary and sufficient for equivalence to occur. According to Dugdale and Lowe (1990), “Stimuli can not become equivalent unless the subject names them” (p. 117). Naming is considered to be an arbitrary verbal response. Horne and Lowe (1996) believed that the learning of listener behavior is a crucial precursor to the development of linguistic behavior. Children learn listener behavior mainly from their caregivers. Naming comes from this listener behavior. According to Horne and Lowe (1996), naming involves “the establishment of bi-directional or closed loop relations between a class of objects and events and the speaker-listener behavior they occasion” (p. 200). Thus, it is through the verbal community that the child learns to assign common names to groups of objects.

These groups become functional or equivalent stimulus classes. The name itself is thought to link the objects together.

Horne and Lowe (1996) describe their idea of how equivalence occurs through naming; in their words, ‘naming is classifying’. According to these authors when we name a stimulus, we are in fact indicating that the stimulus is part of a class. Emergent behavior is not trained according to these authors; it is just the consequence of the different stimuli being within the same name relation. Equivalence by naming can occur in several different ways. The first is common naming. This is where the individual learns a common name for several different things. The common name then results in the individual treating the different things as interchangeable. A second way would be intraverbal naming. Intraverbal naming is where an individual uses the naming process to define a relationship between words. For example if a child were to learn to say to the dog, “good dog”, when prompted with “good…” the child would respond “dog.” This is because the child learned a set relation between the word "good" and the word "dog". To the child, "good" and "dog" go together. The only way in which to have equivalence or functional classes is to have naming occur.

Sidman and Tailby (1982) reported results indicating that naming may not be sufficient for the formation of equivalence. After baseline training, one subject (JO) correctly labeled the stimuli from each class with a common name (i.e., the spoken word that was trained as the sample stimulus). However, he never showed the formation of equivalence classes as judged by performance on equivalence probes. This would indicate that naming is not sufficient for equivalence to emerge. Another subject (EW) showed the formation of equivalence classes as judged by performance on equivalence testing, but never consistently labeled the stimuli from each class with a common name. This would indicate that naming is not a necessary prerequisite for the formation of equivalence classes.

Carr, Wilkinson., Blackman, and McIlvane, (2000) conducted two experiments to assess naming. In Study 1, the subjects were three individuals with severe mental retardation. None of these participants had any significant oral naming skills. The subjects also had no reading skills.

All of the subjects received training on a series of discrimination skills involving stimuli that were different from those used in the experimental training and testing. Experimental baseline training was conducted using a match-to-sample procedure. Subjects were taught AB, CB and DB conditional discriminations. Initial tests for equivalence included tests for BC and BD symmetry and AC, AD, CD, and DC combined tests for equivalence. All subjects acquired the baseline at 95% accuracy and maintained this during probe tests. All three subjects demonstrated equivalence class formation.

Study 2 had two subjects. The first was 13 years old and had no speaking repertoire (BN); the second was 14 years old and had a speaking repertoire of 3 years and one month (HF). There were two sets of stimuli, A and B. The stimuli were 6 physically different figures that were drawn on cards. Both subjects had a programmed training sequence that established AB and AC conditional discriminations.

Both subjects acquired the baseline relations. BN had near perfect symmetry performance, but demonstrated the gradual emergence of equivalence classes. Probe data for HF showed no evidence of emergent matching relations. HF's baseline performances also deteriorated during the probe sessions.

These two studies demonstrated equivalence in subjects who lacked well-developed verbal repertoires. This finding counters the naming theory position that naming is required for the emergence of equivalence. The authors speculated that a good reason for this might be a difference in training procedures. Unfortunately in order to test the naming hypothesis definitively, participants would have to be those who possessed no skills that even resembled communication. This could be very difficult to do with humans.

Training Structures

Three main training structures have been used to establish conditional discriminations in studies of stimulus equivalence. There are the sequential or linear structure, the one-to-many or sample-as-node structure, and the many-to-one or comparison-as-node structure. A node is a particular stimulus that is related to more than one other stimulus in equivalence baseline training (Saunders, R. & Green, 1999). The difference in training structure is interesting because there has been great debate over which structure produces the best conditional discrimination acquisition and equivalence class performance.

Linear Series (Sequential)

An example of linear series training would be to train the following conditional discriminations; AB, BC, and CD. In this example the B and C stimuli function as both samples and comparisons in different conditional discriminations, while the A stimuli are presented only as samples within the AB training, and the D stimuli are presented only as comparisons in the CD training. According to Saunders and Green (1999), linear series training results “in a higher probability of failure on equivalence tests than training with comparison-as-node structures” (p. 125).

To illustrate work involving linear series training, Sigurdardottir, Green and Saunders, R. (1990) published two experiments that showed equivalence formation using the linear-series training structure. In the first experiment three adult females were the subjects. There were six phases within the experiment. Specific instructions preceded each phase. Phase 1 involved pre-training in which subjects produced sequences by putting letters in alphabetical order.

In Phase 2, sequence trials began with five stimuli displayed in eight possible locations on the upper part of the computer screen. There were four different sets of stimuli. The first three were the position stimuli and the fourth was a distracter set. The distracters were never correct. When a subject touched a stimulus, it immediately moved to the first available position on the bottom of the screen. This procedure was continued until all three spaces were filled on the bottom of the screen. The stimuli had to be arranged in a particular order to produce reinforcement. Correct sequences were followed by a jingle, and a buzzer followed incorrect responses.

Phase 3 involved testing for ordinal classes by using a match-to-sample procedure. Each trial contained a sample and two comparisons. A correct response showed the “conditional relations predicted by class formation based on ordinal position” (Sigurdardottir et al., 1990, p. 52). No consequences followed responses. All subjects performed almost without error on reflexivity tests. All of the subjects had trouble demonstrating classes among the third-position and distracter stimuli. All were given a review of the sequence training and then they displayed the remaining emergent MTS relations. This showed that the classes were functional classes, but it did not yet show that they were equivalence classes.

Phase 4 sought to expand the classes using a match-to-sample procedure. The experimenter trained the subject to match the D stimuli to new E stimuli using the MTS procedure. After acquisition, reinforcement was reduced to 20%. Phase 5 was designed to test for expanded equivalence classes. The tests were combined tests for symmetry and transitivity. Results showed that all three ordinal classes and the distracter class had become equivalence classes.

Phase 6 tested whether the subjects would place the E stimuli in the sequence consistent with the existing ordinal functions of their classes. Results in this phase indicated that the ordinal functions transferred to the new class members. Equivalence classes can form and novel class members can be added to the class using the linear series training.

Many-to-one (Comparison-as-node)

In the many-to-one (MTO), or comparison-as-node training structure, multiple sample stimuli are trained to a single comparison set. Examples of this in the literature are found in Saunders, K., Saunders, R., Williams, D. and Spradlin, J. (1993); and in Carr, Wilkinson, Blackman and McIlvane (2000). An example of a MTO training design might involve training BA, CA, and DA conditional discriminations. Many believe that MTO training should be the preferred way to study equivalence and that equivalence emerges much more quickly with this form of training structure (Saunders & Green, 1999). The reason for this, according to Saunders and Green, is that only the MTO structure provides all of the simple discriminations within the training that are needed for consistently positive outcomes on all of the tests for equivalence.

Saunders et al. (1993) sought to determine whether or not the training structure played a role in performance outcome. They performed three experiments to address this question. All three experiments used the same general method. All of the subjects were adults and adolescents with mild mental retardation. Each individual discrimination (BA, CA, and DA) was trained until a mastery criterion was met. Then sessions that contained all three discriminations were presented. Test trials were then presented. The test trials were tests of combined symmetry and transitivity, called equivalence trials.

Study 1 looked at the effects of instructions. When the comparison-as-node (MTO) training structure was employed the question was, whether equivalence classes would form using the comparison-as-node structure without the instructions. To answer this question, six subjects were given instructions and five were not. The results showed that five out of the six subjects who received instructions formed equivalence classes, and only one of five who did not receive instructions showed class consistent responding on the equivalence test trials. The conclusion was that instruction facilitated performance in the comparison-as-node training structure.

Study 2 was a follow up of Study 1. This experiment asked whether subjects who had not formed equivalence in the first study would demonstrate the classes if instructions were given. These subjects were given new stimulus sets and instructions. The hypothesis was that the instructions had facilitated equivalence in Experiment 1 by occasioning comparison naming. Thus, one subject was taught, as a correction procedure, to say the comparison names, but had no feedback on the correctness of his naming. After this correction procedure, he showed almost perfect test accuracy. Another subject who had the same procedure showed facilitation with the comparison naming. One subject showed 0% accuracy, which was not chance performance; he might have been responding by using arbitrary oddity. Once classes were formed using instructions, new classes that had not been instructed were also demonstrated. However, if a certain stimulus set had been tested previously, performance did not change after naming.

Study 3 was designed to test stimulus effects. The authors wanted to make sure that the failure to acquire equivalence in the first study was not because of the stimuli used. Three individuals with mild mental retardation were tested using a comparison-as-node procedure without instructions. The conclusions drawn were that instructions may only be needed in certain stimuli sets.

Saunders et al. (1993) stated that they had formed three main conceptual conclusions based on the results from these experiments. " (1) Mildly retarded subjects are more likely to demonstrate equivalence classes under a comparison-as-node procedure than under a sample-as-node procedure, (2) This difference as a function of training procedure may be less likely with normal older children or adults than with persons with retardation, and (3) With certain sets of stimuli, subjects with retardation are more likely to show equivalence if they receive verbal instructions than if they do not when comparison-as-node training procedures are used" (p. 737-738). The instructions may have done one of two different things. They may have encouraged the subject to give a common name to all of the stimuli, thereby putting them all in one functional class, or they may have just established how the task worked.

One-to-many (Sample-as-node)

In the training structure known as one-to-many (OTM), or sample-as-node, a single set of stimuli serves as the sample stimuli in each of the trained conditional discriminations, while the comparison sets differ. An example of this would be to train the following conditional discriminations; AB, AC, and AD. Many studies in the equivalence literature have employed this training structure including Pilgrim, Jackson and Galizio (2000), Pilgrim et al. (1995), Lipkens et al. (1993), Pilgrim and Galizio (1990), and Sidman and Cresson (1971).

As noted above, much of the current literature states that the probability of equivalence is higher following a comparison-as-node training structure than a sample-as-node training structure. One study that seems to refute the current literature on the effects of training design is the study by Arntzen and Holth (1997). This experiment showed the opposite results.

This study involved 40 experimentally naive college students, who learned three classes of Greek letters. The students were given specific instructions. In conditional discrimination training they were also presented with the correct comparison only for the first nine trials. These students were divided into four groups. Group1 received the linear series training (A:B, B:C). Group 2 received the many-to-one or comparison-as-node training, (A:B, C:B). Group 3 received one-to-many or sample-as-node training, (B:A, B:C). Group 4 received the linear series as did Group 1, but they were presented with symmetry trials before the equivalence probes.

The experimenters looked at reaction time (RT), number of errors in training, and also the differences in equivalence outcomes following the different training structures. The equivalence tests consisted of 24 trials. These were divided into two halves. After the first twelve equivalence trials, equivalence was shown in none of the Group 1 subjects, five in Group 2, ten in Group 3, and one in Group 4. In the second twelve test trials the subjects showing equivalence were three, seven, ten, and three, respectively. In the first test half, there was a significantly higher probability of equivalence following one-to-many than there was following many-to-one training. There was also a significant difference between the many-to-one and the linear series. In all groups there were longer RTs during the first than in the final phase of testing.

"The main finding was that the OTM *(one-to-many)* procedure was significantly more effective than MTO *(many-to-one)* and that LS *(linear series)* was the least effective procedure in creating the emergent relations indicative of equivalence classes" (Arntzen & Holth (1997). p. 317.)

To summarize, the effects on equivalence performance and baseline acquisition have been mixed in regard to training structure. No single training structure seems to always lead to the best results. Another thought that needs consideration is whether equivalence performance and baseline acquisition are affected in the same manner with regard to the training structure employed. One structure could be best for acquisition of the baseline relations, while another training structure could obtain the best results on equivalence performance.

Prototype Studies

Categorization is a critical behavioral function often associated with language. Categorization and categories are well studied phenomena; however, how an individual comes to form categories is of debate (Rosch & Mervis, 1975; Lakoff, 1987; Margolis, 1994; and Wittgenstein, 1953). Most traditional theories have treated category membership as "a digital, all-or none phenomenon" (Rosch & Mervis, 1975, p. 573). In this approach every member within the category must have every defining attribute in common, and every member of the category has equal membership or representativeness (Lakoff, 1987).

In contrast to this classical view, literature has shown that the best examples or members of a category are the ones that are the most representative of their group and the least representative of other groups (Rosch & Mervis, 1975). An important implication of this view is that categorization of new stimuli may occur based on initial exposure to category members. The stimuli that are the most representative of the categories would be learned more rapidly than members that are not good examples. Another implication is that when learning category membership, one should learn the members that are the best examples of the category before learning those that are the worst examples of the categories.

To test these implications Mervis and Pani (1980) designed two experiments. In these experiments the stimuli were arbitrary objects designed by the experimenters. The stimuli were 24 three-dimensional objects that fit into six categories with four members in each category. Each member of the category had a different level of goodness-of-example. Good examples had little in common with the other categories, while the poor examples could hold many attributes in common with the other categories. These stimuli were designed to be toy-like and varied to keep the attention of the young children.

In Experiment 1, adults and 5-year olds were taught the name of one of the objects from each of the six categories. There were two different conditions. In the first condition, the GE group, the subjects were taught to name the object that was considered to have the best goodness-of-example rating. The other group, the PE group, was taught a name for the object that had the worst goodness-of-example rating. There were four hypotheses in this study. The first was that the subjects in the GE group would be more likely to generalize the category names correctly than those in the PE group. The second hypothesis was that if a misassignment occurred, it would be more likely to occur for an object with a poor goodness-of-example rating. The third hypothesis was that if and when overgeneralization occurred, it would occur with a member of the category within the same contrast set, those with similar overall shapes. The final hypothesis was that generalization of the name would be the same for comprehension and production. Production was defined as naming and comprehension was defined as choosing the correct object upon prompting. All four of these hypotheses were supported by the data collected in this experiment. Many of the children, however, never met criterion for production.

Experiment 2 was focused on the development of generalization, from the initial attempt to correct generalization. The main hypothesis was that the manner "in which a person is introduced to a category...would influence how easy it was for him to correctly generalize the category name." It may be easier for a child to learn the name for one member of a category and then generalize than it would be to learn names for all of the members. The authors expected that overgeneralizations would not extend beyond the contrast set boundaries and that goodness-of-example would be based on category structure, rather than on the first-labeled exemplar. In this experiment there were three different conditions, GE, PE and ALL. In the ALL conditions, subjects were taught names for all of the members of the categories. Subjects were asked to do a post-sort and rating task. All of the hypotheses were supported by the experimental results. As far as overgeneralization was concerned, significantly more children in the ALL condition made errors than children in the GE and PE conditions. For the adults, significantly fewer subjects in the GE condition made overgeneralization errors than did subjects in either the ALL or PE conditions. Many of the children refused to rank order the members of the categories. With the adults all ranking was as predicted with four exceptions. All of these exceptions involved switching the middle two exemplars.

Through these experiments, the authors found that naming the best exemplar was often more effective than naming all of the exemplars within the category. They also found that the relationship between comprehension and production was especially strong.

Rosch and Mervis (1975) have suggested that instead of the classical or traditional theory, categories can be structured in such a way as to allow for a prototype of that category. A prototype is the clearest or best example of the category (Rosch & Mervis, 1975). These prototypes can be formed through learning. They are not in place automatically. For example, when an individual is asked to say the first word that comes to mind when the experimenter says “bird,” the individual will say “robin” or “cardinal” more often than they will say “penguin” and “ostrich“. According to Rosch and Mervis, the robin is a “better” example of a bird than a penguin. The robin would be considered to be the prototype for the category “birds.” Wittgenstein (1953) proposed that categories involve a network of features (attributes) that occur commonly, but that no feature must appear in every member. He drew an analogy to a family. The father and two brothers may have brown hair, while the mother and sister have blonde hair. The brown hair is a frequent feature of that family, even though not all family members have brown hair and all are considered to have equal membership. This theory is called the family resemblance categorization model. The more family attributes a member has, the more it will be judged prototypical.

Rosch and Mervis (1975) performed a series of experiments to look at questions concerning prototypes and the family resemblance model. Experiment 1 used 400 introductory psychology students and six common categories of nouns. The typicality of the items within the categories were already rated. Subjects were asked to list attributes for the words that they received. All of the attributes were collected and analyzed by the experimenters. This experiment showed that “the more an item has attributes in common with other members of the category, the more it will be considered a good and representative member of the category” (Rosch & Mervis, 1975, p. 582). Experiment 5 looked at artificial categories in which the members differed only in the degree of family resemblance, or the amount of overlap. Overlap is the degree to which a member shared characteristics with other categories. The authors looked at three different dependent variables: rate of learning, reaction time, and typicality rating. All of the stimuli were six letter strings that differed in respect to family resemblance and overlap of letters. The subjects were instructed that they would be presented with an item from one of two categories. Subjects pressed one key if the item belonged to Group1 and another key if the item belonged to Group 2. They were further instructed when they were right or wrong. After they had achieved mastery of these stimuli, they were asked to rank order the stimuli; this provided a prototypicality rating. Items that had a greater degree of family resemblance were learned more quickly, had faster reaction times, and were judged to be more prototypical than the others.

The research of Stewart (1999) looked at typicality effects in contingency-shaped stimulus classes. She looked at possible similarities between laboratory-produced equivalence classes and natural language classes. She did this by testing for typicality effects. She taught eight arbitrary condition discriminations to college students using a match to sample (MTS) procedure. The stimuli used were three trigrams (WUG, JOM, and NIZ) and 24 abstract stimuli. These stimuli had from one to four class-defining features; fill, insert, appendage, and base, as well as up to three different irrelevant features.

For MTS training subjects were instructed simply to pick the one that "goes with" the sample. Correct responses (which were defined by the presence of any class-defining features) were marked by music and the presentation of one point. Incorrect responses were marked by a buzzer and the subtraction of one point. Subjects were told to get as many points as possible.

Phase 1 involved baseline training. Stewart (1999) used a one-to-many training design in blocks of 24 trials. Samples for the MTS training were one of the three trigrams. Phase 2 involved novel tests and probe tests. The novel tests presented stimuli that had combinations of the relevant and irrelevant features different from those used during the training. In these trials, three-choice arrays of novel stimuli were intermixed within the baseline trials. Probe tests included symmetry and equivalence (symmetry and transitivity) trials. Responses on novel and probe trials were never reinforced. Phase 3 was a sorting and rating task. Subjects were given sheets of paper that had the trigrams on them and cards with all of the stimuli. They were told to sort the cards in the three categories. The subjects then had to rate the stimuli within each category from the "best" to the "worst" example of the category.

The results were analyzed on several different levels. Acquisition data were analyzed. An analysis of errors as a function of number of features was computed by taking the number of errors for each feature number summed together and dividing that by the total number of trials for that number of features. The means and standard error (SE) were computed. This showed evidence for typicality effects because there was a higher percentage of errors for the one and two feature types than there was for the three and four feature types. A within-subject ANOVA and Tukey post hoc tests showed that these typicality effects were significant.

Reaction times (RT) during acquisition were analyzed for correct responses in all trial blocks excluding the first. RTs were compared across trial types at each level of feature inclusion. Mean latencies were alsocomputed. A 2 (session) X 4 (feature) within-subject ANOVA showed that there were differences between latencies for one and two feature versus three and four feature stimuli, but that there was no significant difference between the one and two feature or between three and four feature stimuli. When the mean latencies and speed scores were analyzed, typicality effects were confirmed. There was a main effect for number of defining features.

On the probe trials the subjects responded class consistently. This showed that the subjects demonstrated the formation of equivalence classes. On the novel probes the subjects were accurate on 98-100% of the trials. This showed that equivalence classes are open-ended in the same way as naturally occurring language classes and that new exemplars could be added to the class. An analysis of latencies on novel trials was analyzed in the same way as above and also showed strong evidence for typicality effects. Novel trials that contained more features had faster response times, and those with fewer features had slower response times, thus showing typicality effect. On the rating task subjects gave higher typicality ratings for stimuli with the greater number of features.

Madden (2001) performed an experiment that was in many ways like the Stewart experiment, with the exception that children aged 5-10 years served as subjects. All training was in the MTS format with trigrams serving as the samples and abstract stimuli serving as comparisons. Training was broken down into blocks of 24 MTS trials.

The experiment was divided into four different phases. Phase 1 was pre-training. The pre-training assessed whether or not the child could adequately use a computer. Pre-training began with identity matching with familiar stimuli; hearts, clouds and squiggles. The next phase of pre-training presented familiar words (e.g. “cow”) as samples and pictures of the objects as comparisons. Mastery criteria were met when a child scored over 90% on one 36-trial block in both phases. Phase 2 was baseline training. The comparison stimuli used in the baseline training were abstract stimuli that had two irrelevant features, (i.e., shape and color) and four relevant stimuli, (i.e., appendage, fill, base, and insert). There were also three trigrams, WUG, NIZ and JOM, as the samples. Mastery was met when a subject scored 90% correct on two consecutive blocks. If the child did not meet criterion within 12 blocks of training, a shaping procedure was used.

Phase 3 was the testing phase. The test trials were intermixed with baseline trials. The subjects were tested for symmetry and transitivity. There were also tests to see if the subjects would respond the same way with novel feature combinations as they did with the trained stimuli and to see if the novel stimuli would become class members. Phase 4 was a sorting and rating task.

Out of the twelve subjects who were exposed to the baseline training, six failed to master the conditional discriminations after twelve 24-trial blocks and therefore received the shaping procedure. The four oldest subjects met criterion without the shaping procedure. All subjects eventually mastered the baseline conditional discriminations. Typicality effects were determined by looking at the errors per opportunity. Most of the older subjects showed typicality effects, while most of the younger subjects did not show these patterns. In other words, the older children had fewer errors with the stimuli that had more features; for the younger children, errors were not a function of the number of features. In looking at speed of responding for both the younger and older children, there were no clear "differences between feature levels in speed scores" (Madden, 2001, p. 63). On probe trials, all seven of the older children showed equivalence and all but one of the younger children also showed equivalence class formation. Four or five of the older children also showed generalization to novel stimuli, while only one of the younger subjects showed generalized equivalence to novel class members.

All but two subjects who were given the sorting and rating task sorted correctly according to the baseline training. Typicality effects were demonstrated when a child rated a stimulus with more features as a more representative of the group. The older children showed evidence of typicality, while the younger children did not show this effect.

Purpose of this Study

The purpose of the current study was to extend previous work on equivalence and prototypes to assess different aspects of the current literature debate. The first question was whether the children would acquire the baseline conditional discriminations regardless of the training structure employed. Previous research has shown that it can be difficult for children of different ages to master the baseline conditional discriminations (e.g., Madden, 2001; Pilgrim, Jackson, & Galizio, 2000).

Some authors believe that equivalence can be demonstrated more readily by employing certain training structures over others (e.g., Saunders et al., 1993; Arntzen & Holth, 1997). Therefore, another measure to assess was equivalence performance. Would each training structure produce strong evidence of equivalence-class formation, as demonstrated by a high percentage of class-consistent responding on the equivalence probes?

In considering class performances, it is also important to ask if equivalence is really a good approximation of natural language classes. Natural language classes are elastic in that they allow stimuli that were not originally part of the class to become class members based on common features. Would children in both training conditions show generalized equivalence? Stewart (1999) and Madden (2001) both found generalized equivalence performance that allowed for novel stimuli to become part of the class. Madden (2001), however, found that the younger children were not likely to show this generalized equivalence performance. Would this be the case regardless of the training structure used?

Another issue is that of naming. Is it necessary for equivalence class formation? Many studies have addressed this issue (e.g., Horne & Lowe, 1996; Sidman & Tailby, 1982; and Carr et al., 2000). The present work addressed naming by manipulating the nature of the training node. Would children show acquisition of the baseline conditional discriminations, regardless of the stimuli presented as node?

These questions were addressed in the following experiments. The present study consists of two different, yet related, experiments. Experiments 1 and 2 used the arbitrary trigrams (i.e., WUG, JOM and NIZ) and twenty-four abstract stimuli. One group of children began by learning baseline conditional discriminations in the sample-as-node training structure with the trigrams as the sample stimuli. Once training and equivalence testing were completed, a new stimulus set was introduced, and new conditional discriminations were trained in the comparison-as-node training structure with trigrams as nodes. A second group of children began baseline conditional discrimination training with the comparison-as-node training structure and the trigram as comparison. The subsequent condition included using the sample-as-node with a new stimulus set.

In Experiment 2 one group of children began with the comparison-as-node training structure using the trigram as the node, while another group began with the comparison-as-node training structure using the two-feature stimuli as the node. This dealt directly with naming questions. Would both the groups demonstrate equivalence performance?

METHOD

Participants

The participants in this study were twenty-six 4-10 year olds in two after-school care programs in Wilmington, NC (see Table 1). Some of the subjects had prior MTS training experience and others were experimentally naive. Permission slips signed by the custodial guardian were obtained for all participants. Participants were randomly assigned to one of four groups.

Table 1

Demographic data by subject

Experiment 1				
Subject	Age at 1st Sessions	Gender	Ethnicity	Group
YC1	4yr 2mo	M	Caucasian	2
YC2	4yr 6mo	F	Caucasian	2
YC3	5yr 0mo	M	African American	1
YC4	5yr 1mo	M	Caucasian	2
YC5	5yr 3mo	F	Caucasian	1
YC6	5yr 3mo	M	Caucasian	1
YC7	5yr 4mo	F	Caucasian	2
YC8	5yr 6mo	M	Caucasian	2
YC9	5yr 7mo	M	Caucasian	2
YC10	5yr 9mo	M	Caucasian	1
YC11	5yr 10mo	F	African American	1
YC12	6yr 4mo	M	African American	2
YC13	6yr 10mo	M	Caucasian	1

Table 1 (continued)

Experiment 2

Subject	Age at 1st Sessions	Gender	Ethnicity	Group
OC1	7yr 5mo	F	African American	4
OC2	7yr 6mo	F	African American	3
OC3	7yr 7mo	F	Caucasian	4
OC4	8yr 2mon	F	Caucasian	4
OC5	8yr 3mo	M	African American	4
OC6	8yr 5mo	M	African American	3
OC7	8yr 9mo	M	African American	3
OC8	8yr 11mo	F	African American	4
OC9	9yr 3mo	F	Caucasian	4
OC10	9yr 5mo	M	African American	3
OC11	9yr 8mo	M	Caucasian	3
OC12	10yr 8mo	M	African American	3
OC13	10yr 9 mo	M	African American	4
OC14	10yr 11mo	F	Caucasian	4

Each participant was given three pieces of candy or stickers upon completion of each day's session. They were also given a sticker to place on their prize chart. The candy and the sticker were dependent upon participation, not performance. Every fifth sticker resulted in the participant being able to choose an age-appropriate prize from the prize box. All prizes were priced at approximately $0.50-$1.00.

Apparatus

The participants were tested four to five days a week in a quiet environment. Testing occurred in the cafeteria, library, or classroom at the school the participant attended. All subjects were tested on a Macintosh Apple desktop or laptop computer. All computers were equipped with a 3 1/2 inch floppy drive, a CD ROM drive and a mouse. The children tested with the laptop used the built-in mouse pad. The participant manipulated the mouse or his finger to click on the appropriate location on the monitor. All of the children were tested using the MTS software designed by Dube (1991).

Stimuli

Three different sets of stimuli were used in the baseline training for the two experiments. Each set contained three trigrams and 24 abstract stimuli that had from one to four class-relevant features. There were also four irrelevant features. The stimuli for each condition were the same for all groups. Each new condition used new arbitrary stimuli and trigrams. The first stimulus set was the same one used by Madden (2001) as seen in Figure 2. The trigrams in this set were "WUG", "JOM", and "NIZ". The relevant features in this set were the appendage, base, insert and fill (see Table 2). The appendages were the attachments to the outside of the main stimulus. The base was the single black object protruding from the edge of the stimulus. The fill was the shading of the inner

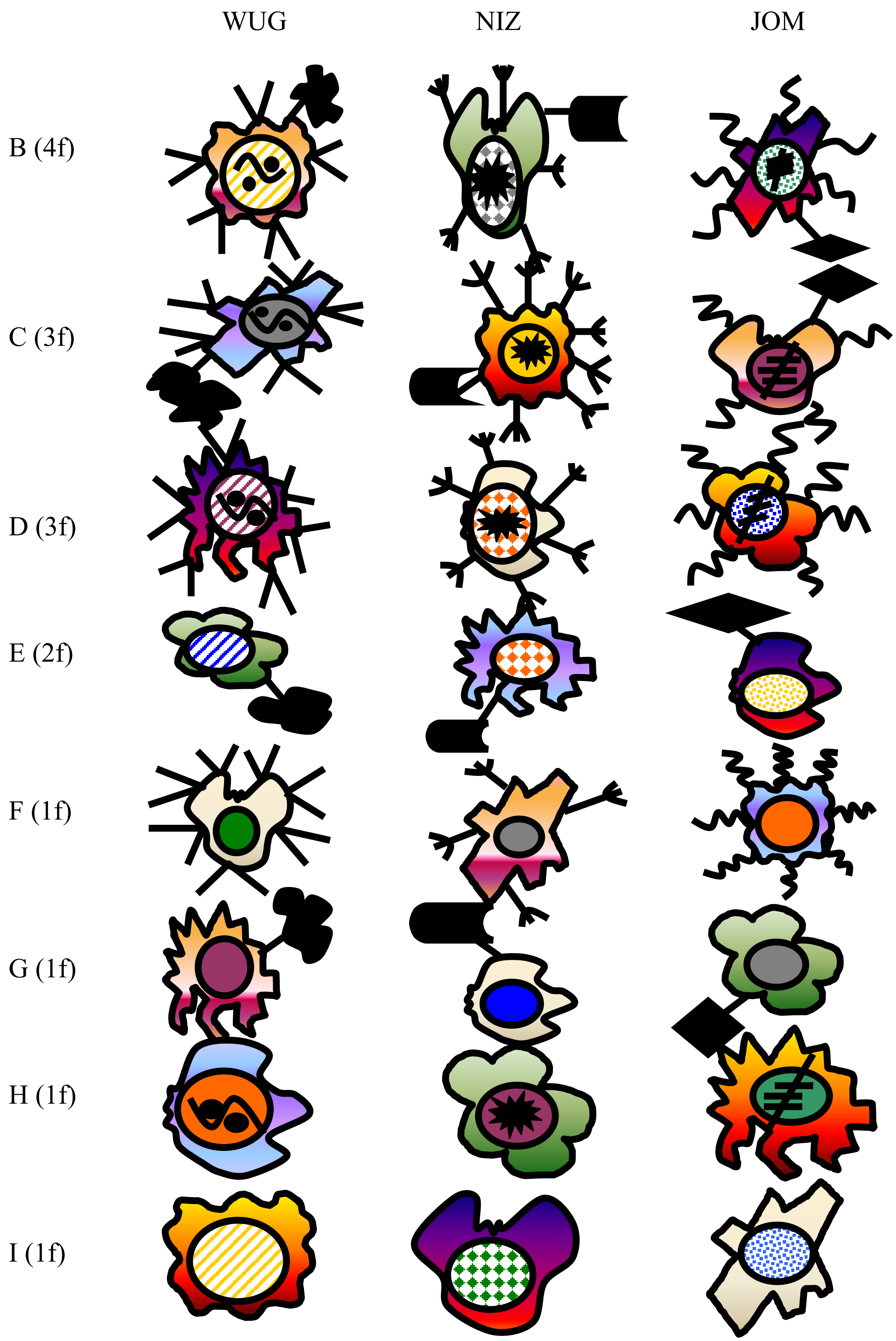


Figure 2. Baseline color stimulus Set 1 and trigrams used in the training phase

circle of the stimulus. The insert was the design inside of the inner circle. In this set the class-irrelevant features were; the color of the inner circle, color of the outer shape, shape, and positioning, as described in Madden's (2001) color set.

The 24 pictorial stimuli were divided into three different classes, each including one of the trigrams. The classes were defined by the four relevant features. In each class there was one stimulus with four relevant features, two stimuli with three relevant features, one stimulus with two relevant features, and four stimuli with one relevant feature each. Each class included eight abstract pictorial stimuli and one of the trigrams. As seen in Figure 2, the A stimuli were the trigrams. The B stimuli were designed to be the prototypes because they included all four of the relevant features. The C and D stimuli included three relevant features (base, appendage, and insert for C; appendage, fill and insert for D). The two relevant features used in set E were fill and base. Stimuli F, G, H and I all included only one of the relevant features (appendage, base, insert, or fill respectively). Each of the relevant features was used four times within each stimulus class. The irrelevant features were also balanced across the baseline training trials.

Each trial block included one exposure to each of the trial types illustrated in Figure 2. In order to decrease the chance that the children would learn the order of the trials, there were six different versions of each trial block. The different versions were presented consecutively so that the child was exposed to all six of the versions before repeating one.

The second stimulus set also consisted of three trigrams, "RUP", "LOY" and "KIF", and 24 abstract pictorial stimuli (see Figure 3). These were specifically designed for this experiment. The irrelevant features in this set were body shape, mouth, placement

Table 2

Stimulus patterns for all color experimental stimuli Set 1

Feature Type	Appendages	Base	Insert	Fill
Training And Novel Features				
B, J (4-features)	X	X	X	X
C, K (3-features)	X	X	X	
D, L (3-features)	X		X	X
E, M (2-features)		X		X
F, N (1-feature)	X			
G, O (1-feature)		X		
H, P (1-feature)			X	
I, Q (1-feature)				X
Novel Combinations				
R (3-features)	X	X		X
S (3-features)		X	X	X
T (2-features)	X	X		
U (2-features)	X		X	
V (2-features)			X	X

of hand, and background color. The relevant features were type of hat, shape of toy in the hand, type of neckwear, and type of shoes. The 24 stimuli were once again divided into three different classes, each including one of the trigrams. The classes were defined by the four relevant features. As was the case for Stimulus Set 1, each class included one stimulus with four relevant features, two stimuli with three relevant features, one stimulus with two relevant features, and four stimuli with one relevant feature each (see Table 3). Each class included eight abstract stimuli and one of the trigrams. An important difference between this stimulus set and Set 1 was the coloration. Each of the relevant features in these abstract stimuli was color coded. The hat was always red, the toy was blue, the neckwear was green and the shoes were brown

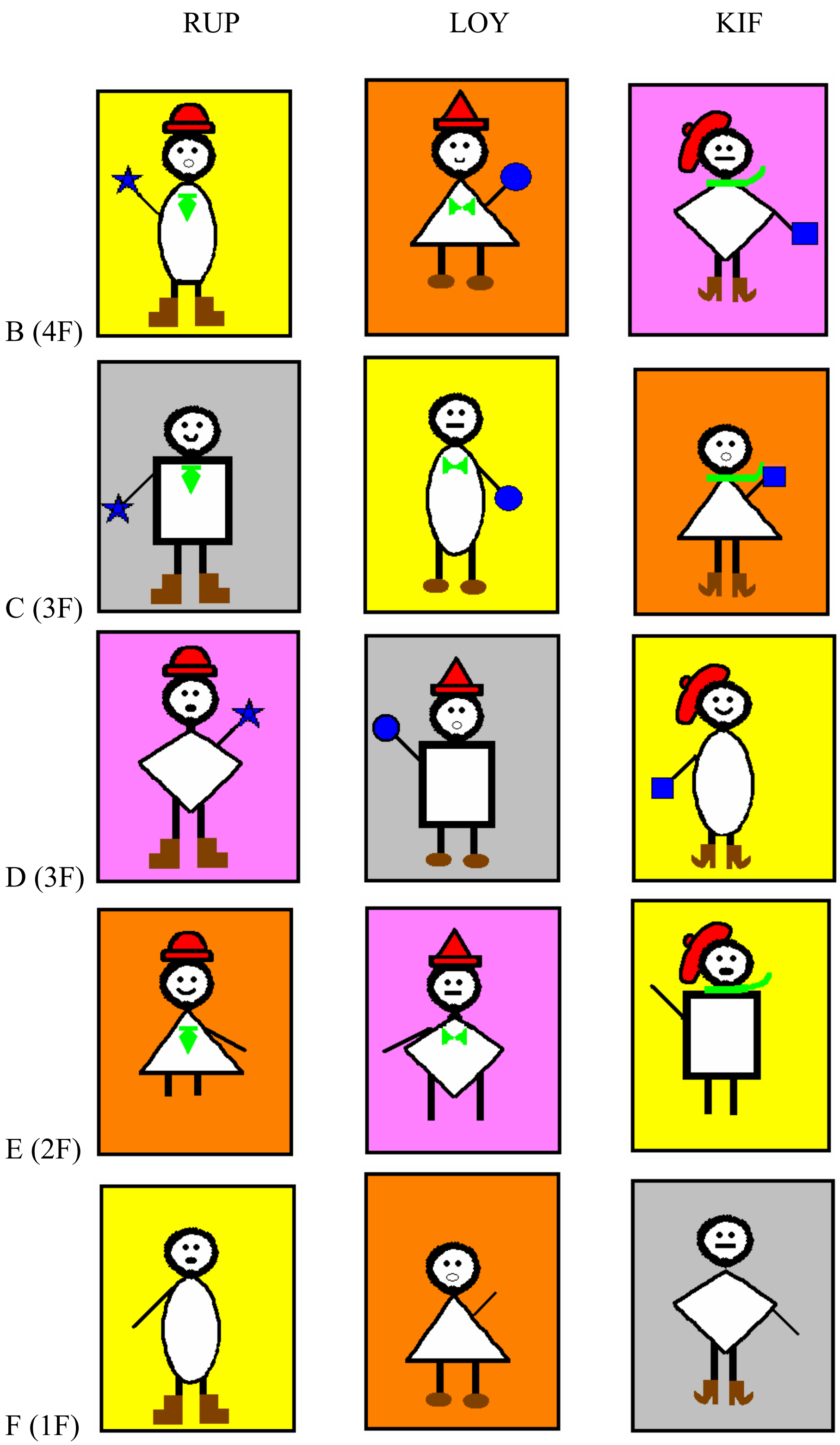
Experimental Overview

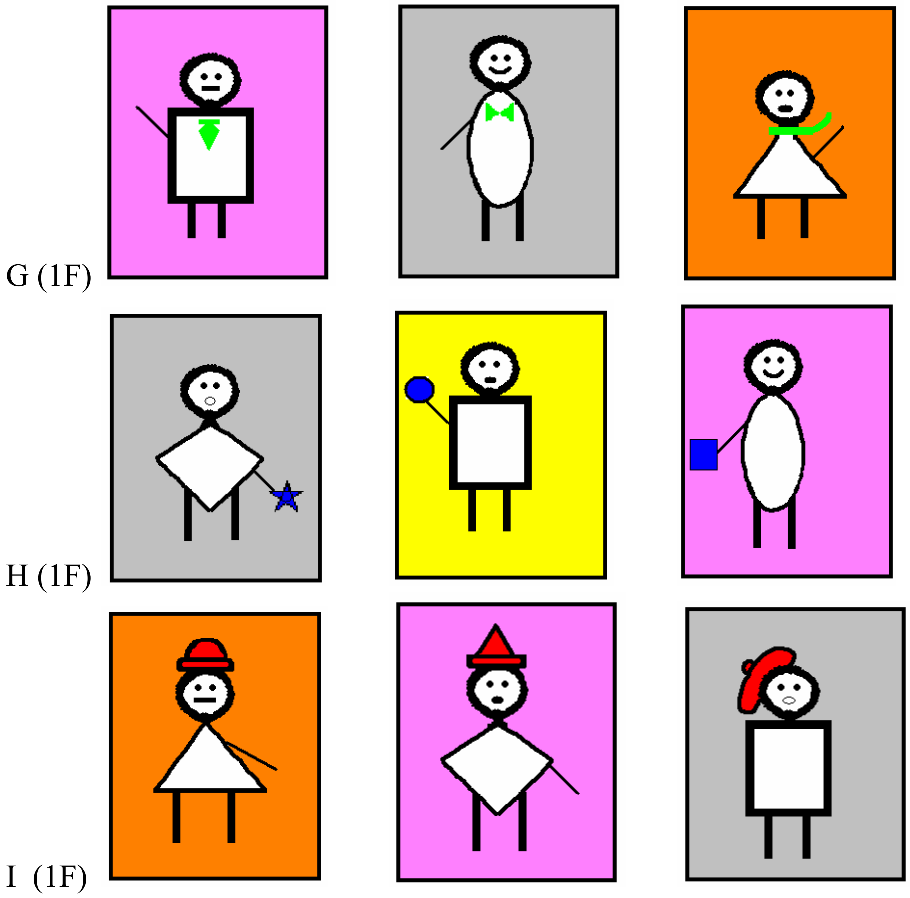
Within this study, there were two separate experiments. Each experiment was designed to be a complete study within itself, and each addressed a different experimental question. Both experiments used the same two stimulus sets. The training and testing for all children consisted of five phases (see Table 2), a modification of the procedure used by Madden (2001). The phases were completed in order, and no child moved on to the next phase until mastery or stability criteria were met.

PROCEDURE

Phase 1: Pre-Training

In order to familiarize participants with the mouse and the MTS format, many experiments begin with pre-training techniques. The first pre-training exercise used in this experiment was called Identity matching. In this exercise, a picture of a familiar


RUP
LOY
KIF
B (4F)
C (3F)
D (3F)
E (2F)
F (1F)



<u>Figure 3.</u> Baseline color stimulus Set 2 and trigrams used in the training phase

Table 3

Stimulus patterns for all color experimental stimuli Set 2

Feature Type	Hat	Toy	Neckwear	Shoes
Training and Novel Features				
B, J (4-features)	X	X	X	X
C, K (3-features)		X	X	X
D, L (3-features)	X	X		X
E, M (2-features)	X		X	
F, N (1-feature)				X
G, O (1-feature)			X	
H, P (1-feature)		X		
I, Q (1-feature)	X			
Novel Combinations				
R (3-features)	X	X		X
S (3-features)		X	X	X
T (2-features)	X	X		
U (2-features)	X		X	
V (2-features)			X	X

object (i.e., baby, flower, or tree) appeared as the sample stimulus in the middle of the screen. When the child manipulated the mouse to “click” on the sample, three comparison stimuli appeared in three of the four corners of the screen. The instructions to the child were simply to “Click the middle picture, now pick one.” If the child picked the identical picture from the comparisons, then colorful stars appeared and a jingle sounded. If the child picked one of the other comparisons, a buzzer sounded. There were 24 trials in each block, eight trials with each of the stimuli (baby, flower, and tree) as sample within each block. The child began the next pre-training exercise when a criterion of 90% correct was met on two consecutive blocks of trials.

The next exercise in the pre-training phase was a program that used familiar words and pictures. Here a written word, such as “cow”, “boat”, or “tree” appeared as a sample. All of the children were familiar with the spoken and pictorial counterparts of these written words; some of the children were familiar with the written word. When the child clicked on the sample, three comparison pictures appeared in the corners of the screen. Choosing the picture of the cow was reinforced when the sample was the word “cow.” The same pattern was followed with the boat and tree stimuli. A buzzer sounded upon clicking on an “incorrect” choice. In this exercise there were 24 trials per block, eight trials involving each of the written words.

Phase 2: Pre-Sort

Upon completion of Phase 1, the child sat at a table facing 24 cards. The cards were laminated 2 X 2 ½ in. pictures of the stimuli to be used in the experiment. The child was told to look at the cards and to put them in piles that “go together.” No other instructions were given and the child was allowed to put the cards in as many or few

Table 4

Session Sequences for Phases 1-4

Phase 1- Pre-training
Identity MTS
Word MTS
Phase 2- Pre-sort
Phase 3 Baseline Training and Reinforcement Reduction
Baseline (until 90% on 2 consecutive blocks)
Mixed Reinforcement 75% (until 90% on 2 consecutive blocks)
Mixed Reinforcement 50% (until 90% on 2 consecutive blocks)
Phase 4-Probe Sessions Sequences
Reflexivity A
Reflexivity B
Reflexivity C
Symmetry A
Symmetry B
Symmetry C
Transitivity A
Transitivity B
Transitivity C
Novel Features A
Novel Features B
Novel Combinations A

categories or piles, as they chose. The child was encouraged to “keep trying, you are doing fine,” when they said that they did not know which ones to put together. This phase determined whether the stimuli would be sorted according to experimentally defined relevant features prior to the baseline training. If any child did sort according to experimentally defined relevant features, they were to be excluded from the experiment; this never occurred. The pre-sort was completed one time by each child. They then continued on to Phase 3.

Phase 3: Baseline Training and Reinforcement Density Reduction

In the baseline training there were 24 trials per block. Eight conditional discriminations were taught and each trial type was presented once in each block (see Tables 5 and 6). Each trial consisted of a sample and three comparisons, in the same format used during Pre-training. The sample appeared in the middle of the screen and when clicked, comparison stimuli appeared in three of the four corners of the screen. Trials were balanced within the blocks so that no sample appeared more than twice in a row, no comparison appeared in the same corner more than twice in a row, and no corner was the correct one for more than two trials in a row.

The comparison choices within a trial all had the same relevant feature combination (see Figure 2). Choosing the comparison stimulus that belonged to the same class as the sample produced a consequence that included an audio and a visual component. All reinforcement was class-specific (CSR). That is, correct choices on trials involving a “WUG” sample, for example, had a specific visual consequence and a specific audio jingle. Only correct choices on the “WUG” trials produced this particular combination of visual and auditory elements. The same pattern held true for trials with the other samples. Class-specific reinforcers were used because many believe that these procedures facilitate acquisition (Sidman, 2000).

After the child completed two consecutive trial blocks at or above 90% accuracy, reinforcement density was reduced to 75% of the trials. The reinforcement density was reduced by programming six of the baseline trials to include no reinforcement. The six trials without reinforcers were balanced over sessions. Once the mastery criterion was met (two blocks at or above 90%), reinforcement was reduced once again, this time to 50% of the trials. The trials programmed with no consequence were balanced across sessions. When the child completed two consecutive blocks at or above 90% accuracy with reinforcement programmed on 50% of the trials, he or she proceeded to Phase 4.

Phase 4: Probe Session Sequence

After the baseline conditional discriminations were mastered, the children were presented with probe trials. As described above, the reinforcement density was reduced during the baseline training to prepare for this testing. Probe trials were intermixed with baseline trials. Responses on probe trials were never reinforced. There were four different types of probe trials. These were Reflexivity, Symmetry, Transitivity and Novel.

For the reflexivity trials, the participant was presented with a sample. When clicked, three comparisons appeared. One of these was identical to the sample. In each block of probe trials there were twelve reflexivity probes intermixed with twenty-four baseline trials. These test trials assessed whether the participant would choose the comparison that was identical to the sample (e.g., A=A).

On symmetry probes, the stimuli that were presented as samples in baseline training were presented as comparisons, and the previous comparisons were presented as

Table 5

Baseline conditional discriminations trained in the first and second experimental conditions for Experiment 1

Group 1		Group 2	
Condition 1	Condition 2	Condition 1	Condition 2
AB	BA	BA	AB
AC	CA	CA	AC
AD	DA	DA	AD
AE	EA	EA	AE
AF	FA	FA	AF
AG	GA	GA	AG
AH	HA	HA	AH
AI	IA	IA	AI

Table 6

Baseline conditional discriminations trained in the first and second experimental conditions for Experiment 2

Group 3		Group 4	
Condition 1	Condition 2	Condition 1	Condition 2
AE	BA	BA	AE
BE	CA	CA	BE
CE	DA	DA	CE
DE	EA	EA	DE
FE	FA	FA	FE
GE	GA	GA	GE
HE	HA	HA	HE
IE	IA	IA	IE

samples. Each block of symmetry trials included 12 symmetry probes intermixed with twenty-four baseline trials. For children in Group 1, the baseline relations trained were AB, AC, AD, AE, AF, AG, AH, and AI (Table 5). Thus, symmetry probes tested for BA, CA, DA, EA, FA, GA, HA, and IA relations. Children in Groups 2 and 4 were taught BA, CA, DA, EA, FA, GA, HA, and IA as baseline conditional discriminations (Tables 5 and 6); thus, symmetry tests assessed AB, AC, AD, AE, AF, AG, AH, and AI relations. Children in Group 3 learned baseline conditional discriminations AE, BE, CE, DE, FE, GE, HE, and IE (Table 6), so symmetry tests in this case included EA, EB, EC, ED, EF, EG, EH, and EI.

Transitivity probe trials assessed relations between class members that were never presented together in training trials. The specific transitivity probes that were presented were selected from the many that could have been used so as to include only trials in which there was no relevant feature overlap between the sample and the correct comparison choice. In other words, the participant could not feature match in making the "correct" choice. The transitivity probes are presented in Table 7.

On novel probe trials the trigrams from baseline training were used as sample stimuli. The comparison stimuli included new combinations of relevant features with different combinations of irrelevant features (see Tables 2 & 3). These novel probes assessed class membership of untrained stimuli sharing critical features. A number of different arrangements were possible for novel trial testing. The present arrangement was selected to be standard across all testing conditions and because trials with a trigram as sample are most relevant to the issue of naming. To complete the probe phase, the participant needed to be stable on all probe performances and maintain baseline accuracy.

Stability was assessed by the following formula: $| (\sum X_4\text{-}X_6) - (\sum X_3\text{-}X_1) | \leq (\sum X_1\text{-}X_6)$, for each trial type.

Specialized Training Conditions

Each child was individually monitored for acquisition. Children progressed though the phases as described above. If the child did not show any trend toward acquisition within the first 18 sessions, they proceeded to the second condition using Stimulus Set 2. If the child did not meet the mastery accuracy within the first 18 sessions, but showed a trend toward acquisition by scoring at least 75% correct by the eighteenth session, they were allowed to continue without intervention.

Those children who did not meet this criterion received a series of specialized training procedures in an attempt to facilitate acquisition of the baseline conditional discriminations. A variety of specialized training procedures were used. For the children who received specialized training, the specific procedures and their order of presentation were designed to address each individual's pattern of errors. This was determined by doing an error analysis of responses made on the baseline training sessions. Table 8 presents the specialized training procedures used with each individual child and their order of presentation. In some of the specialized training procedures, children learned only one conditional discrimination at a time; sessions included eight trials of each type, 24 trials in all. Some of the children received training sessions that presented two conditional discriminations. There were four trials of each type per session, 24 trials in all.

A delay procedure was used with some participants. After the comparisons were presented and after a predetermined time, all of the incorrect comparisons disappeared

Table 7

Transitivity relations tested in probes

Set 1
I-H
H-D
E-I
G-F
I-F
D-H
F-E
G-I
F-H
H-G

and only the correct stimulus remained. Every correct response lengthened the delay between the presentation of the comparisons and the removal of the incorrect ones. Some of the children in both experiments were given the delay procedure in conjunction with other specialized training procedures, such as a session with only one conditional discrimination.

A correction procedure was also utilized for some participants. In this procedure, an incorrect choice resulted in the trial being repeated until a correct choice was made. This was also used in conjunction with other training procedures.

Some of the children were also given instructions, “Which one goes with this one (sample)?” Several children in Experiment 1 (see Table 8) were given the explicit instructions “When you see this one (sample), pick this one (correct comparison).”

EXPERIMENT 1

Procedure

All 13 children in Experiment 1 received training in two different conditions, the one-to-many, sample-as-node training structure and the many-to-one, comparison-as-node, training structure. Table 1 displays all of the children in Experiment 1, their age and group affiliation. Children in Group 1 were taught first with the one-to-many training condition, in which the sample was a trigram, while the comparisons were abstract stimuli. Children in Group 2 were first taught with the many-to-one training structure, in which the samples were the abstract stimuli, and the comparisons were the three trigrams. Experiment 1 examined the effectiveness of both the one-to-many training structure and the many-to-one training structure for generating conditional discrimination acquisition,

Table 8

Specialized training for children in Experiment 1. The first number is the order in which each condition was presented; the second number represents the number of sessions spent in that training condition.

	Children				
	YC5	YC6	YC10	YC11	YC13
			Group1		
BL Delay			1-4		
Alt BL & BL Delay			2-2		
FA	5-3*	2-13+*	4-5	2-23+*	3-2*
	7-5	8-6*			
FA Delay	4-5•+*		3-9+		2-17*
Alt FA & FA Delay			5-3		
HA		9-4		3-6*	4-5*
FA & HA	1-11	1-12+		1-12+	1-5
	6-3	3-11*			6-9*
		7-7			
FA & HA Delay	2-6				

Table 8 continued

	Children				
	YC5	YC6	YC10	YC11	YC13
FA & HA Block					5-9*
Alt FA, HA & Delay	3-5				
GA		5-6* 7-3*			7-4*
FA, HA & GA Block					8-8*
GA & IA		4-14 6-18			

Table 8 continued

	Children		
	YC2	YC7	YC12
		Group 2	
AF	3-3	2-6*	2-10*
	8-9*		
AF Delay	4-18		1-4
	5-4Δ		
Alt AF & AF Delay	6-1	1-13	
	7-4		
AH	9-9Δ	3-22Δ*	3-13*
AF & AH		5-2*	5-2*
AF & AH Delay	1-9		
AF & AH Block		4-5*	4-15*
Alt AF, AH & Delay	2-1		
AG		6-2	6-3*
AF, AH, AG Block			7-5*

Note.

Δ-- Explicit Instructions

•-- Tracing

+-- Instructions

*--Mastery

equivalence-class formation, and generalized equivalence.

The conditional discriminations targeted in the one-to-many training were AB, AC, AD, AE, AG, AH and AI (see Table 3). For children in Group 1, when mastery of the conditional discriminations was achieved and equivalence classes were tested, or after 18 sessions if there was no trend towards mastery, the second experimental condition was introduced with a new set of stimuli. Training with this second set was conducted with the many-to-one training structure. The targeted conditional discriminations were BA, CA, DA, EA, FA, GA, HA, and IA (see Table 3). If there was still no mastery, the children were exposed to specialized training.

Children in Group 2 were exposed to the same stimuli and trigrams as children in Group 1, but had a different order of experimental conditions (many-to-one, followed by one-to-many). Thus, the conditional discriminations targeted in the first condition were BA, CA, DA, EA, FA, GA, HA, and IA (see Table 3). In the second experimental condition, AB, AC, AD, AE, AG, AH and AI (see Table 3) were targeted with a new stimulus set.

The questions to be answered here were as follows: Would children in this experiment show acquisition of the baseline conditional discriminations regardless of training structure? Would each training structure produce strong equivalence class formation as demonstrated by a high percentage of class-consistent responses on equivalence probes? Would generalized equivalence result from each of the training structures?

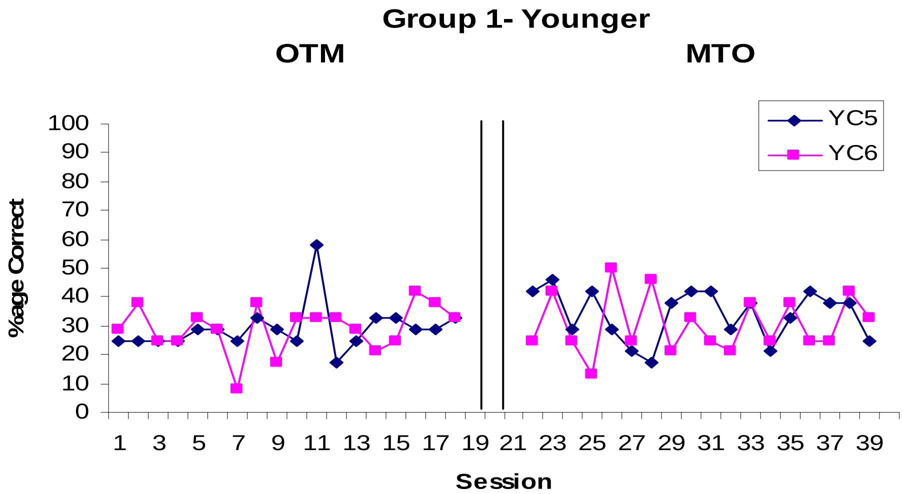
Results

Acquisition

Experiment 1 involved investigation of the one-to-many and the many-to-one training structures with a nonsense syllable as the node. In Experiment 1, none of the13 children acquired the baseline conditional discriminations within the allotted 18 sessions in the first training condition. Four children (YC3, YC4, YC8, and YC9) left the after-school program before completing 18 sessions in the first training condition. Of the nine remaining children, four (YC1, YC2, YC7, and YC12) began with the many-to-one training structure, and five (YC5, YC6, YC10, YC11, and YC13) began with the one-to-many training structure. Figures 4 and 5 show the percentage of trials on which correct selections were made on the initial baseline sessions for each child during their first and second training condition. The first panel of Figure 4 shows the younger children in Group 1. The second panel shows the older children in Group 1. Figure 5 shows the children in Group 2 and the percentage of trials on which correct selections were made on the initial baseline sessions for the first and second training conditions.

All nine children failed to master the conditional discriminations in the first training condition, and there were no trends towards acquisition for any of the children. Each child then received training with a second set of stimuli and a different training structure. One child (YC1), who began with the many-to-one training structure, left the school before completing the first 18 sessions in this second condition. Of the remaining eight participants, all failed to acquire the conditional discriminations within the 18 sessions of their second training condition (Figures 4 & 5).

Each of the eight remaining participants then received specialized training, based



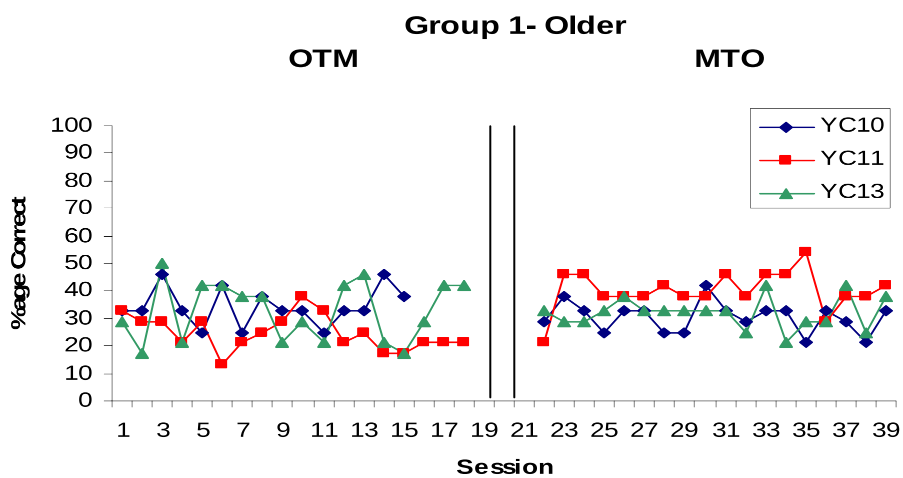


Figure 4. The percentage of trials on which correct selections were made on the baseline sessions for each child in Group 1 during their first and second training conditions. The individual scores for the two youngest children are displayed in the top panel. The bottom panel shows data for the older three children in Group 1.

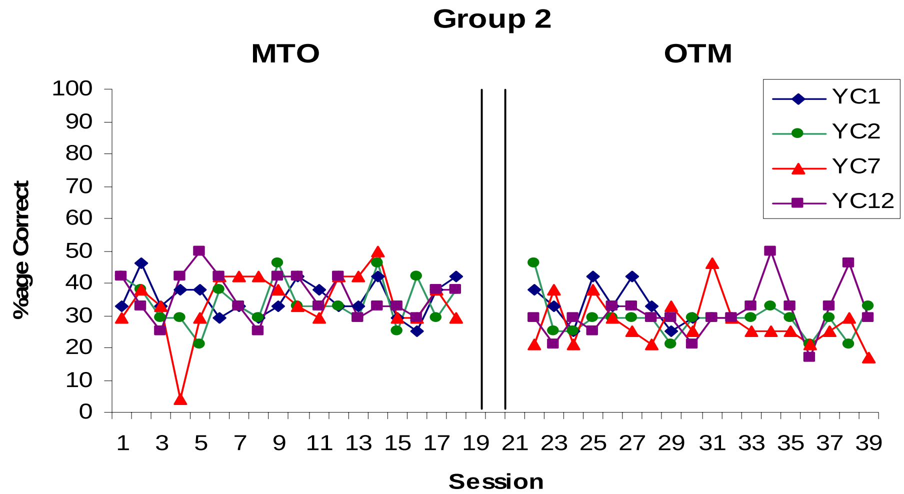


Figure 5. The percentage of trials on which correct selections were made on the baseline sessions for each child in Group 2 during their first and second training conditions.

on their individual error patterns. Table 8 shows the specialized training procedures provided for each child in Experiment 1. For each participant, Table 8 shows the order of each condition presented during the specialized training , as well as the number of sessions spent in that training condition, and whether or not mastery criteria were met. Consider participant YC5 as an example (see Table 8). YC5 began her training with the one-to-many training structure. After failing to master the baseline conditional discriminations in the allotted 18 sessions, a second stimulus set and the many-to-one training structure were introduced. YC5 failed to acquire the baseline conditional relations in the specified number of sessions, and so the specialized training procedures were introduced. First, the number of conditional discriminations presented in the training sessions was reduced from eight to just two conditional discriminations, FA and HA. YC5 completed 11 sessions in this condition with no trends towards acquisition.

Next, a delay procedure was added to the previous program. She completed six sessions in this condition, again with no trend toward acquisition; she simply waited on each trial until the incorrect comparisons disappeared and then clicked on the remaining (correct) choice. She was then exposed to alternating sessions of FA and HA, with and without a delay. She completed this sequence five times with no acquisition on the non-delay programs. YC5 was then exposed to a program with just the FA condition discrimination, with a delay. In addition, she was given the general instruction, “Pick the one that goes with this one.” She was also told to place her finger on the “one in the middle” (sample) and to draw a line to her choice (comparison); the experimenter then clicked on that comparison for her (i.e., the tracing intervention). Using all of these interventions, YC5 met criterion in five sessions on the delay program with the FA conditional discrimination. She was then trained with just the FA conditional discrimination and no delay. She met criterion here in just three sessions. YC5 was then re-exposed to the FA and HA training. At this point her performance fell to chance levels for the FA conditional discrimination training, as well as for the HA conditional discrimination. YC5 was immediately re-exposed to just the FA conditional discrimination training, and she met criterion in five sessions. YC5 left the after-school program at this time.

Specialized Training for Group 1

Specialized training for children in Group 1 was conducted with the MTO training structure (following failure to meet mastery criterion in their second training condition.) During specialized training with these children, none met mastery criterion on any of the conditional discriminations until the baseline was reduced to a single conditional discrimination. Four out of the five children (YC5, YC6, YC11, and YC13) met mastery, with single conditional discrimination training while one never did, despite extensive specialized training. After learning one conditional discrimination, three of the children went on to meet mastery criteria with other conditional discriminations, but this required teaching additional conditional discriminations individually, and then in blocks. The other child (YC1) left the after-school program.

YC13 made the most progress towards mastering the baseline conditional discriminations through specialized training. Mastery criteria were first met with a single conditional discrimination, FA. He then mastered HA, as well as block and intermixed sessions with FA and HA. The GA conditional discrimination was mastered next, as were the block and intermixed sessions with all three conditional discriminations. All of the single conditional discriminations learned involved the one-feature stimuli and a nonsense syllable (see Figure 3). Upon mastering the IA conditional discrimination intermixed with the FA, HA, and GA conditional discriminations, YC13 would have been presented with the full baseline of eight conditional discriminations.

YC6 and YC11 mastered the FA conditional discrimination when presented with the delay procedure. Correct responses fell to chance levels when FA and HA trials were intermixed. The FA conditional discrimination was then re-trained, and training with the HA conditional discriminations is in progress. When this is mastered, block FA and HA sessions will be introduced.

YC10 has made the least progress in this group. He has failed to master a single conditional discrimination even with extended specialized training. He has received training for a single conditional discrimination (FA) with the delay procedure and the general instruction, “Pick the one that goes with this one.”

Specialized training for Group 2

Specialized training for children in Group 2 was conducted with the OTM training structure after failing to master the baseline conditional discriminations in the second condition. As was the case for children in Group 1, none of the children in Group 2 mastered any of the conditional discriminations until trained in isolation. After mastering the first conditional discrimination, all three children (YC2, YC7 and YC12) went on to learn additional conditional discriminations.

YC12 has made the most progress. Mastery criteria were first met with the single conditional discrimination, AF. YC12 then mastered AH in addition to block and intermixed sessions with both AF and AH. The AG conditional discrimination was mastered next, as were the block and intermixed sessions with all three conditional discriminations. All of the single conditional discriminations learned involved the one-feature stimuli and a nonsense syllable (see Figure 3). Upon mastering the AI conditional discrimination intermixed with the AF, AH, and AG conditional discriminations, YC12 will be presented with the full baseline of eight conditional discriminations.

YC7 progressed much as did YC11 and YC6. She mastered the single conditional discrimination, AF, with delay. The AH conditional discrimination was mastered next in addition to the block session of AF and AH. Upon mastery, AF and AH intermixed sessions will be introduced.

Participant YC2 took the longest time to master the first conditional discrimination. In looking at her performances, the data showed that she was making all responses based on background color. Subsequently, the background color from each of the abstract stimuli was removed. Mastery for YC2 occurred with the single conditional discrimination AF, with no background color, delay and the explicit instructions, "When you see this one, pick this one." After mastering this single conditional discrimination with the delay and instructions faded, she is currently receiving training with the AH conditional discrimination, the delay program, explicit instructions, and no background color. Upon mastery, YC2 will be presented with the block program of AF and AH, using no background color.

In looking at acquisition of the baseline conditional discriminations, there were no differences between children trained with the one-to-many training structure and the many-to-one training structure, for either the first or the second training condition (see Figures 4 & 5). No child in either group mastered any of the baseline conditional discriminations until specialized training was introduced. Because performances for all children in both groups remained at chance levels in both training conditions, no statistical analyses were conducted. Given their failure to master the full baseline training, none of the children in Experiment 1 received probe trials.

Discussion

The results of Experiment 1 indicate that presenting eight conditional discriminations simultaneously was too difficult a task for these young children. None of the children mastered any of the baseline conditional discriminations until introduction of the specialized training procedures. This is consistent with other equivalence literature where younger children have demonstrated difficulties in mastering conditional discriminations (e.g., Madden, 2001; Pilgrim, Jackson & Galizio, 2000; Lipkens et al., 1993).

It may be that age is an important indicator of conditional discrimination acquisition, at least on an MTS task like that used in the present study. This would also suggest that when working with younger children, it would be advisable to begin with as simple a procedure as possible. YC12 and YC13 mastered more conditional discriminations than any other children in Experiment 1. They are also the oldest of the children in this experiment.

In observing the children in Experiment 1, the process through which they best acquired the baseline conditional relations was through programmed instruction with simpler steps that built upon each other. Some of the specialized training procedures led to more accurate performance than others. When only one conditional discrimination was presented, more of the children began to perform accurately.

The delay procedure had mixed effectiveness. While most of the children who mastered a conditional discrimination were presented with the delay procedure, it is not clear that the delay really aided in acquisition. Many of the children exposed to this training step became dependent upon the delay. The delay procedure was designed to discourage the subject from waiting for the duration of the delay on every trial. Every correct response led to a longer delay, which should discourage the subject from simply waiting until the incorrect stimuli disappeared to make their response. This did not happen for most of the children in this study. Many of the younger children waited on every trial of the session for the correct response to be revealed, despite the fact that delays reached up to 20 seconds. While accuracy for these children was close to 100%, closer inspection of the data revealed that the children rarely responded prior to the end of delay period.

To attempt to counter this, sessions with and without the delay were alternated. This was done in hope of reducing dependence on the delay. This method appeared to facilitate acquisition. In addition, many children were also instructed that they “did not have to wait.”

Training multiple conditional discriminations simultaneously was clearly too demanding for these young children, regardless of training structure or stimulus set. Utilizing simple small steps that build upon each other appears to be a beneficial way to help young children acquire the conditional discriminations required.

It was hoped that the new stimulus set (Stimulus Set 2) would facilitate the acquisition of the baseline conditional discriminations. In this set, the relevant features were color coded and the figures were designed to be more age-appropriate (see Figure

3). These additions did not seem to aid performances. Children in Experiment 1 performed at chance accuracy with both stimulus sets until specialized training was introduced.

This study was designed to address whether children would acquire baseline conditional discriminations regardless of training structure. None of the children in Experiment 1 met mastery of the full baseline of eight conditional discriminations. Children in both training structures mastered a few of the conditional discriminations, but only with specialized training procedures. Thus, neither of the training structures facilitated greater acquisition. Questions related to the effects of training structure on equivalence performance and typicality effects could not be addressed with the present data.

EXPERIMENT 2

Procedure

All 14 subjects in Experiment 2 received training in two different conditions. Table 1 shows all of the children in Experiment 2, their age, and group affiliation. Both conditions involved the many-to-one, or comparison-as-node, training structure. The factor that differed across conditions was the nature of the stimuli acting as the node. The comparisons (nodes) in one condition were the two-feature E stimuli, and the samples were the remaining abstract stimuli and trigrams. The trained conditional discriminations were AE, BE, CE, DE, FE, GE, HE, and IE (see Table 6). In the other condition, the comparisons (nodes) were the trigrams. The conditional discriminations learned in this condition were BA, CA, DA, EA, FA, GA, HA, and IA (see Table 6). When children who began training with the 2-feature stimuli as node (Group 3) completed all training (and testing as appropriate), training began with a new stimulus set and the trigram as node. Children who began training with the trigram as node (Group 4) received the same conditions in the opposite order.

The important questions to be addressed here were: Would children show acquisition of the baseline conditional discriminations, regardless of the stimuli presented as node? Would each training arrangement produce strong evidence of equivalence-class formation, as demonstrated by a high percentage of class-consistent responding on equivalence probes? Would generalized equivalence result from each of the training arrangements?

Results

Acquisition-Group 3

Six children (OC2, OC6, OC7, OC10, OC11 and OC12) began training in the many-to-one, 2-Feature-as-node condition. Five of the six children failed to meet the mastery criterion within the allotted 18 sessions, yet four of the five (all except OC6) demonstrated at least some trend towards mastery (see Figure 6).

One child of the six, OC2, mastered the baseline conditional discriminations without the aid of specialized training procedures. However, when the reinforcement density was reduced, her percentage of correct responses began to drop (see Figure 7).

Extensive specialized training procedures, which included intermixed sessions of FA and IA with and without delay, were then implemented to facilitate accuracy (see Table 9 & Figure 7). When re-exposed to the full baseline training and the reinforcement reduction, her accuracy remained at criterion levels.

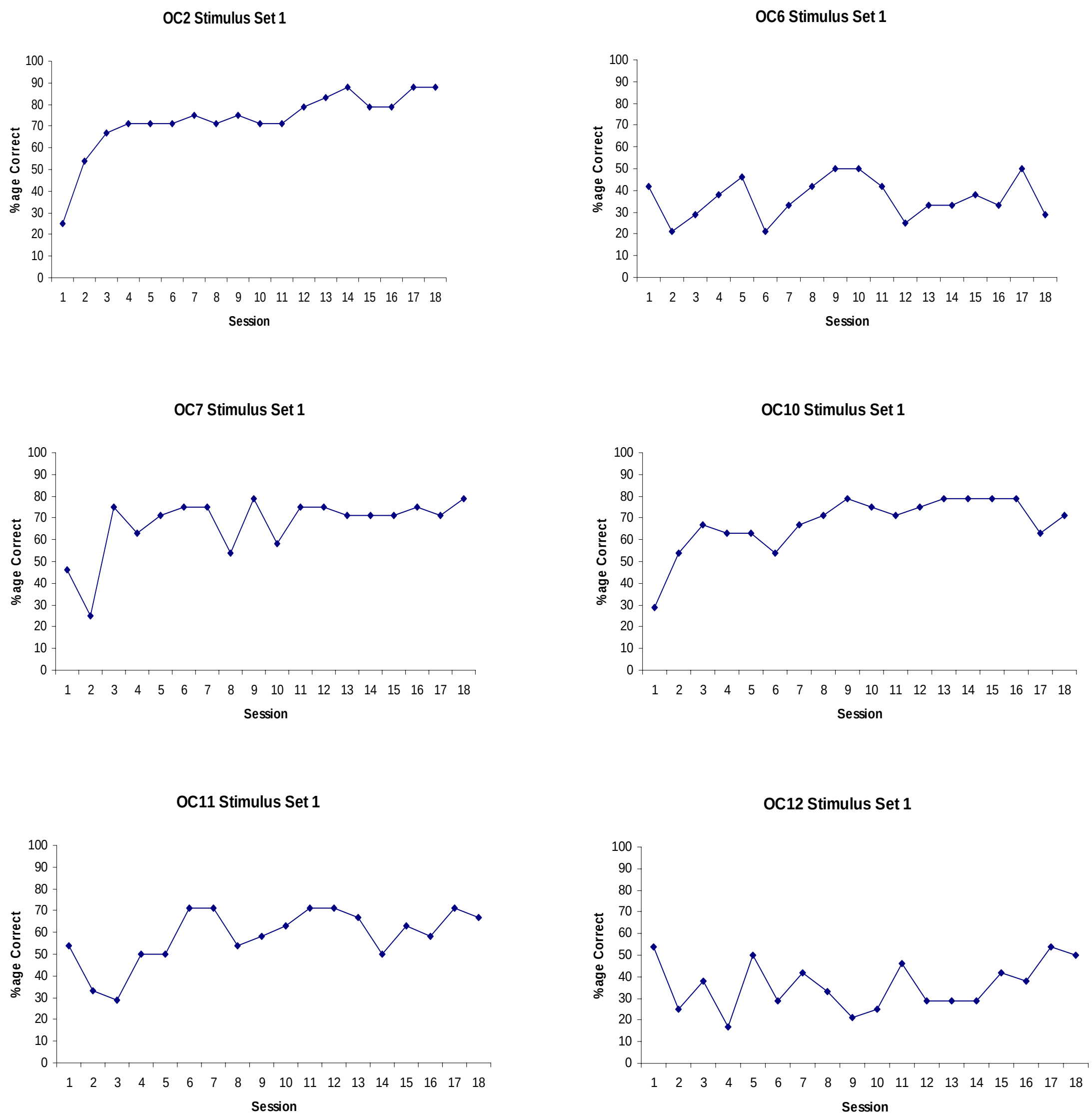


Figure 6. The percentage of trials on which correct selections were made on the baseline conditional discriminations for children in Group 3 (2F) for the first 18 sessions in the 2-feature condition.

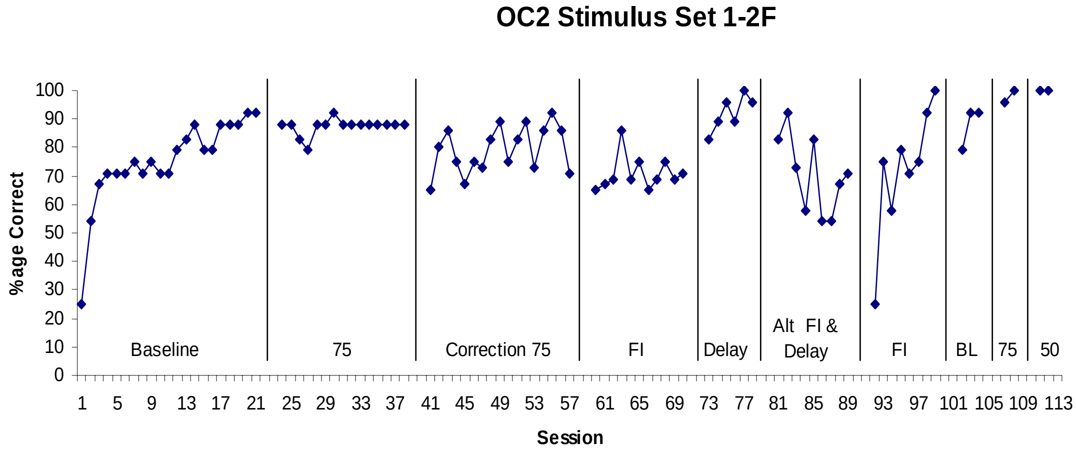
Four of the five remaining children (OC7, OC10, OC11 & OC12) received specialized training (see Table 9). OC11 mastered the baseline conditional discriminations in the 2-feature-as-node training condition after receiving specialized training with CE and HE (see Figure 7), in addition to the baseline with correction. He then proceeded to reinforcement reduction. OC11 left the after-school program immediately after this step and was never presented with equivalence probes.

The other three children (OC7, OC10 and OC12) left the after-school program before mastering the full baseline. All three had mastered six of the conditional discriminations and were receiving specialized training with the final two. OC7 and OC10 received specialized training that included the baseline with correction, intermixed FE and IE sessions, and the baseline with delay. They left the after-school program before mastering the FE and IE conditional discriminations. OC12 received specialized training with baseline correction, and intermixed DE and FE sessions. He left the after-school program before mastering the DE and FE conditional discriminations.

The final child, OC6, left the after-school program just after completing the 18 sessions in the first training condition. He received no specialized training.

Acquisition- Group 4

Eight children began training with the many-to-one training structure and the trigram as node. These children were OC1, OC3, OC4, OC5, OC8, OC9, OC13 & OC14. Figure 8 shows the percentage of trials on which accurate comparison selections were made on each session of baseline training and reinforcement reduction for OC1, OC3, OC4, OC5, OC8, OC9, OC13 and OC14. Of these eight children only one, OC4, mastered the full baseline (all eight conditional discriminations) without the assistance of



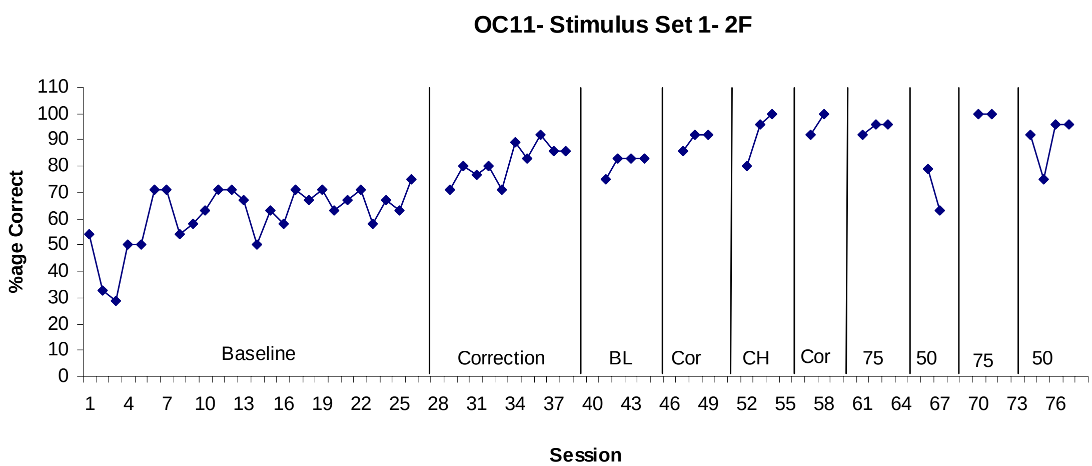


Figure 7. The percentages of trials on which correct selections were made for the two children who completed baseline training. OC2 and OC11 both required specialized training.

Table 9

Specialized training for children in Experiment 2. The first number is the order in which each condition was presented; the second number represents the number of sessions spent in that training condition.

	Children				
	OC2	OC7	OC10	OC11	OC12
			Group 3		
Return to BL	5-3*			2-4	
BL		1-5	1-10	1-10	1-9
Correction				3-3* 5-2*	
FA & IA	1-11+	2-6	2-6 3-8+		
FA & IA Delay	2-6*	4-1+	3-4		
Alt FA, IA & Delay	3-10	3-6			
DE & FE					2-3
Alt FA, IA & GA, HA				4-3*	

Table 9 Continued

	Children		
	OC1	OC3	OC9
	Group 4		
Return to	4-2	4-42	
BL	11-3		
BL	6-10+		
Delay			
GE & FE	1-18		1-6
	5-1		6-4
GE & FE	2-4		2-2
Delay			4-3
Alt GE, FE		1-19*	3-3
& Delay			5-8+
AE	9-6*		
AE	8-9*	5-9*	
Delay			
AE & HE	7-15	3-32*	
	10-8*		
CA & HA		2-17*	

Note. Δ--Explicit Instructions, •--Tracing, +--Instructions, *--Mastery specialized training. OC4 was allowed to continue in the original training condition because her data indicated trends toward mastery (see Figure 8; bottom right panel). Another child, OC3, was also allowed to continue in the original training condition. She was one of the first children tested in this study, and the 18 session cut-off had not yet been established, but only 18 sessions are presented. OC3 failed to master the full baseline in 42 sessions. Accuracy scores were between 67-79% correct for the last ten sessions. OC3 then received specialized training based on her individual pattern of errors (see Table 9). She first mastered FA and IA in intermixed sessions, followed by mastery of GA and HA in intermixed sessions. The full baseline was then re-introduced. OC3 left the after-school program at this time.

Six children in the group (OC1, OC5, OC8, OC9, OC13, and OC14) failed to master the baseline conditional discriminations within the allotted 18 sessions (see Figure 8). One child, OC8, left the after-school program just after completing the 18 sessions in the first training condition. The other five children (OC1, OC5, OC9, OC13 and OC14) were then exposed to many-to-one training with the 2-Feature stimuli as node, and a second stimulus set. OC5 quickly mastered the baseline conditional discriminations without being exposed to the specialized training. Mastery criteria were met on session 32, but his first session with a score greater than 90% was session number 17 (see Figure 9).

Of the four children remaining, OC1 and OC9 also showed improving trends within the first 18 sessions of the 2-feature-as-node condition, but failed to meet the mastery criteria for full baseline training (see Figure 9). These children then received specialized training sessions based on their individual error patterns (see Table 9). OC1

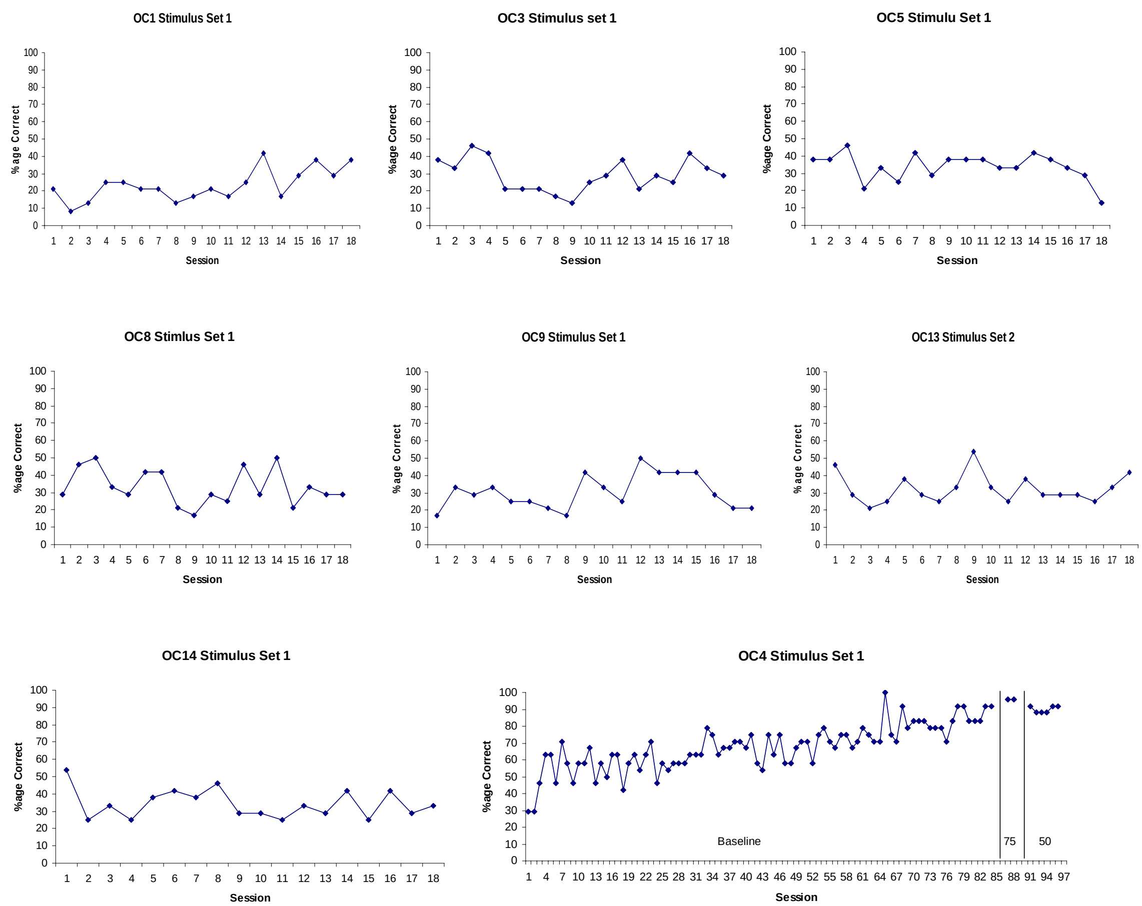


Figure 8. The percentage of trials on which correct selections were made on the baseline conditional discriminations for children in Group 4 for the first 18 sessions in the trigram-as-node condition.

received the baseline with correction. She then mastered intermixed sessions with FE and GE, with and without delay. OC1 then mastered sessions of AE, followed by intermixed sessions of AE and HE, with and without delay. The full baseline was then re-presented, alternating with sessions that included FE and HE trials with stimuli from only two of the classes. Accuracy scores for OC1 have improved to 80% on full baseline sessions. OC9 also mastered GE and FE with delay, but failed to master these conditional discriminations without the delay before leaving the after-school program.

Two of the four children (OC14 and OC13) left the after-school program after beginning the second condition, before receiving specialized training. Their individual acquisition data in this second training condition is summarized in Figure 9. OC13 had only six sessions in this condition, yet reached 83% accuracy before leaving. OC14 had more exposure to this condition, and scored as high as 88% before leaving.

The results of Experiment 2 seem to suggest that for children who were learning eight conditional discriminations simultaneously, acquisition was greatly facilitated when the 2-Feature stimulus served as the node rather than the Trigram. Figure 10 shows the last six (out of 18) training sessions in the first training condition for each child in Experiment 2. The top panel shows the percent accuracy on the last six sessions of the initial training condition for the children who began with the 2-feature-as-node training condition (Group 3). The bottom panel shows the same data for the children who began with the trigram-as-node training condition (Group 4).

Children who had the 2-Feature-as-node training condition performed at one of two different levels. OC2, OC7, OC10, and OC11, all performed at greater than chance levels, well before completion of the first 18 sessions. OC12 performed at above chance

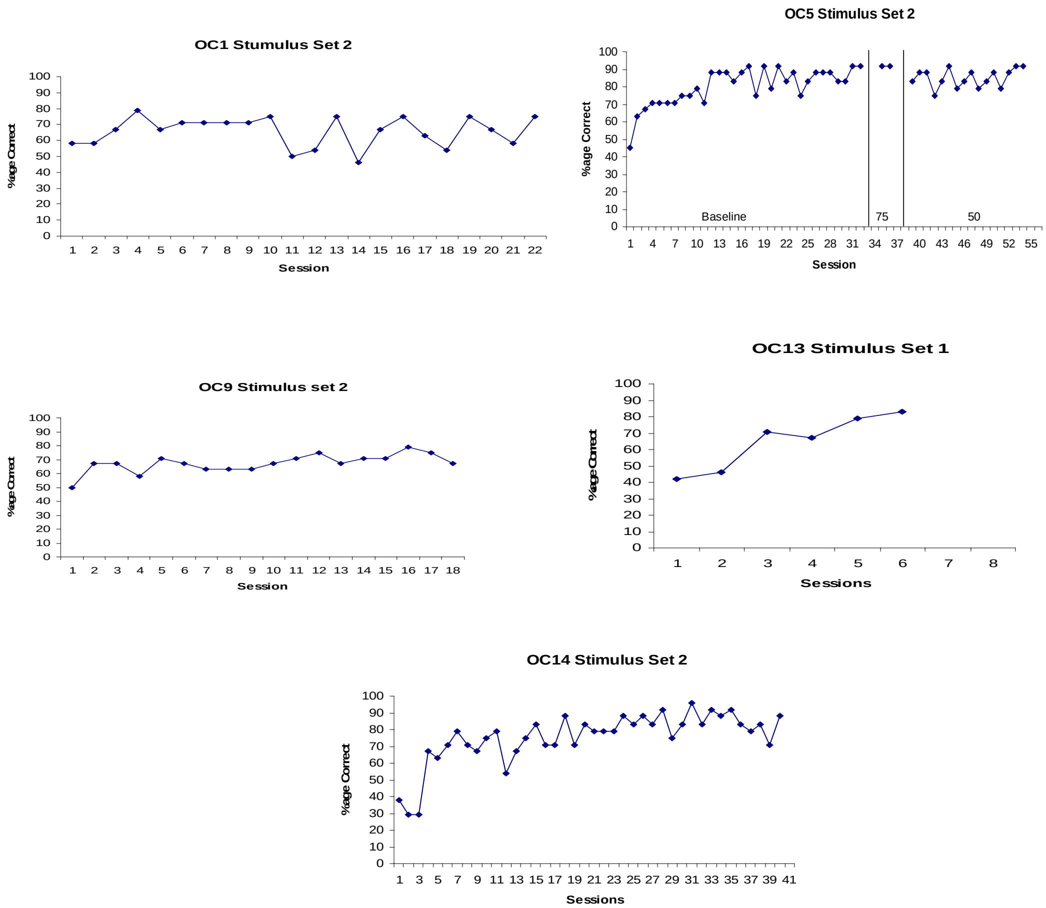
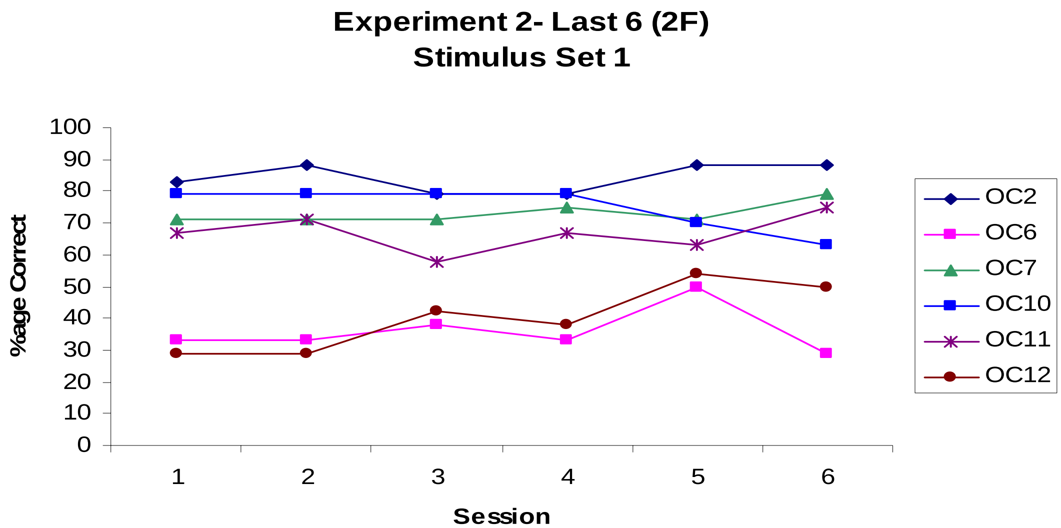


Figure 9. The percentage of trials on which correct selections were made on the baseline conditional discriminations in the second condition (2-Feature-as-node) for children in Group 4.



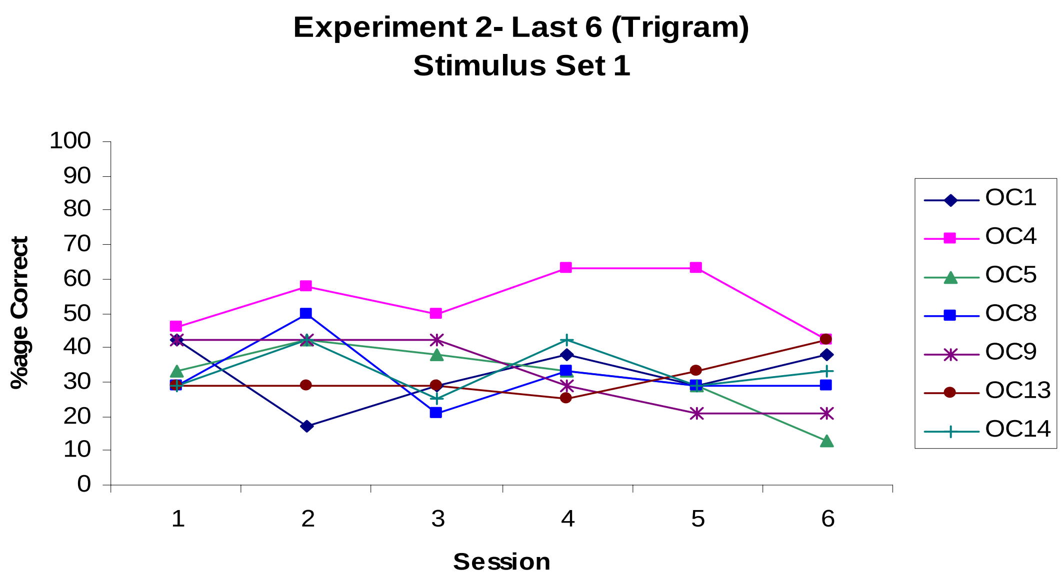


Figure 10. Each data point represents the percentage of trials on which correct selections were made on the baseline conditional discriminations for the last six sessions in the first training condition. The top panel presents data from children who had the 2-feature-as-node training condition (Group 3). The bottom panel shows the same data for the children in the trigram-as-node training condition (Group 4).

levels only on the last two sessions. OC6 had one session that was above chance performance before completing the 18 sessions. In contrast, only one child (OC4) in the trigram-as-node condition (Group 3) performed at above-chance levels of accuracy within the initial 18 sessions. For the remaining six children, accuracy scores above chance level occurred on only a single session, for a single subject (OC8). The difference between the two groups for accuracy on the last six sessions of training was significant $t(7) = 2.36$, $p < .05$.

In the second training condition, when introduced to the 2-Feature-as-node training structure, five of the children in Group 4 (OC1, OC5, OC9, OC13 and OC14) showed a trend towards acquisition within 18 sessions (see Figure 9). The other three children in Group 4 did not receive the second condition. None of the children in Group 3 required the second training condition.

Probe Performance

Group 3

After mastering the baseline conditional discriminations, OC2 received equivalence probes (see Figure 11). Figure 11 shows the percentage of class-consistent responses made on baseline and probe trials for sessions that included reflexivity, symmetry, transitivity and novel probes. On the reflexivity and symmetry probes, her percentage of class-consistent responses was near 100% throughout testing. Class-consistent responses on transitivity probes ranged between 0-50%. For this reason, she was re-exposed to the specialized training condition that included the two conditional discriminations related to the transitivity probes for which class-inconsistent responses were most common. When OC2 met mastery criterion for this specialized training condition, she was presented with alternating sessions of equivalence probes and baseline-only sessions at 50% reinforcement reduction (30 sessions in total, 15 baseline sessions and 15 probe sessions). After 15 sessions of testing for transitivity, she was exposed to novel probes. In contrast to her transitivity probe performances, class-consistent responses on novel-stimulus probe trials ranged from 67%-92% (see Figure 11).

Group 4

After mastering the baseline conditional discriminations, OC4 proceeded to equivalence and novel probe testing (see Figure 12). Figure 12 shows the percentage of class-consistent responses made on baseline and probe trials for sessions that included reflexivity, symmetry, transitivity, and novel probes. OC4 chose the class-consistent comparison on almost 100% of the trials for all reflexivity, symmetry and transitivity probes. In contrast, her responses on novel probes fluctuated between 33%-75%.

OC5 also received probe testing after mastering the baseline conditional discriminations (see Figure 12). OC5 had 100% class-consistent responding on the reflexivity probes, with the exception of the last session which was 88%. Class-consistent responding took longer to develop for the symmetry probes with the last three sessions at 100%. Responses on the transitivity probes were not class-consistent, but there was a slight trend toward greater class-consistent responding, as testing continued. Because accuracy on baseline trials during those sessions also began to drop, sessions including equivalence probes were alternated with sessions including only baseline conditional discriminations at 50% reinforcement density for the 16 sessions. Baseline accuracy remained at 88-92%. OC5 left the after-school program before receiving novel probes.

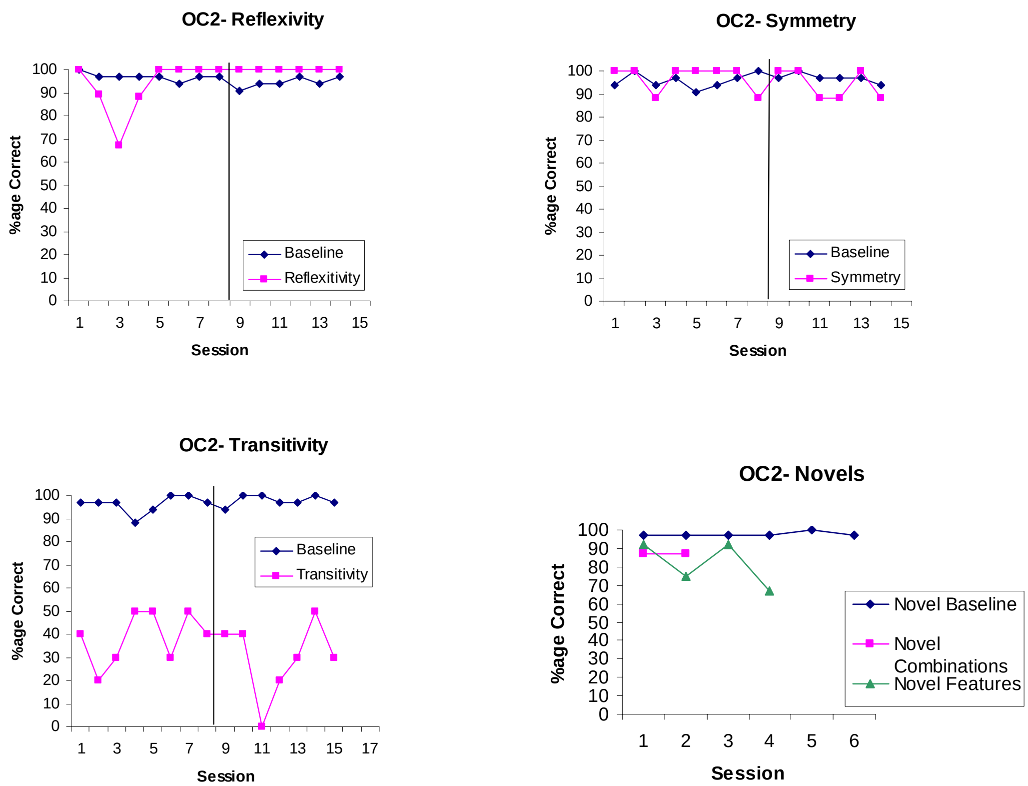


Figure 11. Equivalence and Novel probe performances for OC2. Each data point represents the percentage of class-consistent responses made on baseline and probe trials for sessions that included reflexivity, symmetry, transitivity, and novel probes. The vertical line represents the point at which alternation between probe and baseline sessions.

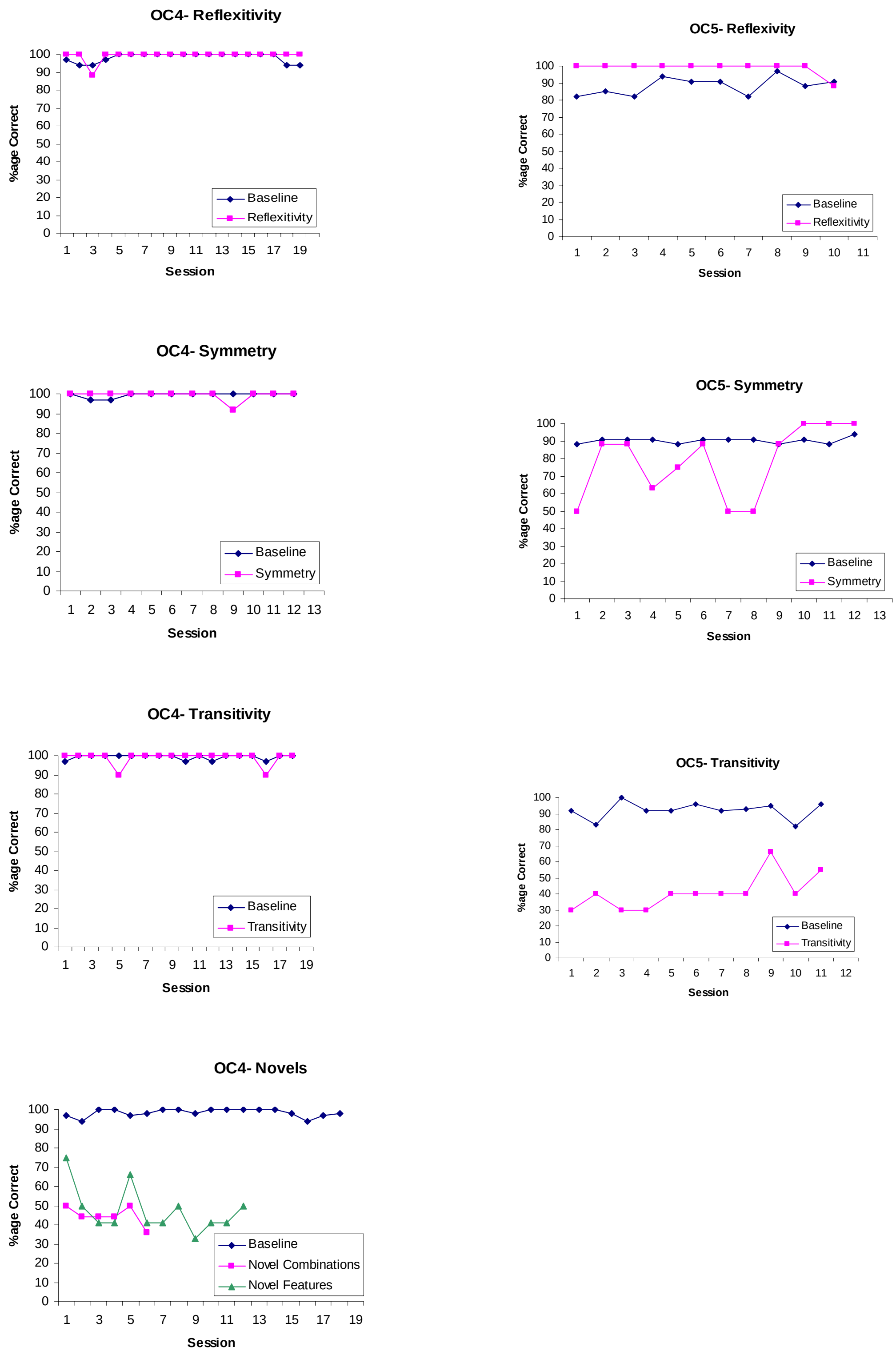


Figure 12. Each data point represents the percentage of class-consistent responses made on baseline and probe trials for OC4 and OC5.

Typicality

The possibility of typicality effects during baseline acquisition was assessed by analyzing errors as a function of number of class-consistent features. This was calculated for all nine of the subjects (OC1, OC2, OC4, OC5, OC10, OC11, OC12, OC13, and OC14) who either progressed to probe testing or achieved at least 80% accuracy on full baseline sessions (i.e., those presenting all eight conditional discriminations). The typicality data were analyzed for full baseline training sessions only, before any specialized training or reinforcement reduction was arranged (see Table 10). The number of sessions included varied greatly across subjects and may account for some of the differences in typicality effects.

The typicality data for each child are shown in Figure 13. In order to observe whether the children were feature matching (i.e., identity matching with the 2-Feature stimulus), the 1-Feature stimuli were analyzed as two different categories. The first included the 1-Feature, non-matching stimuli; these were the stimuli for which there were no relevant features in common with the sample. The second category included the 1-Feature matching stimuli; these were the stimuli for which at least one relevant feature was identical to those of the sample. The top left panel of Figure 13 presents mean accuracy for all eight of the 2-Feature-as-node subjects per feature level. The other panels present the individual means per feature level. OC4 was the only subject in the trigram-as-node training condition (see bottom, right panel).

OC11 and OC12 appeared to show some evidence of typicality. They had greater accuracy on trials with the 3- and 4-feature matched stimuli than on trials with the 1-feature, matched stimuli. Six of the children in the 2-feature-as-node group (OC1, OC2,

Table 10

Number of baseline training sessions included in each child's typicality analysis.

Name	# of Sessions
OC1	
OC2	21
OC4	86
OC5	31
OC10	23
OC11	26
OC12	31
OC13	6
OC14	40

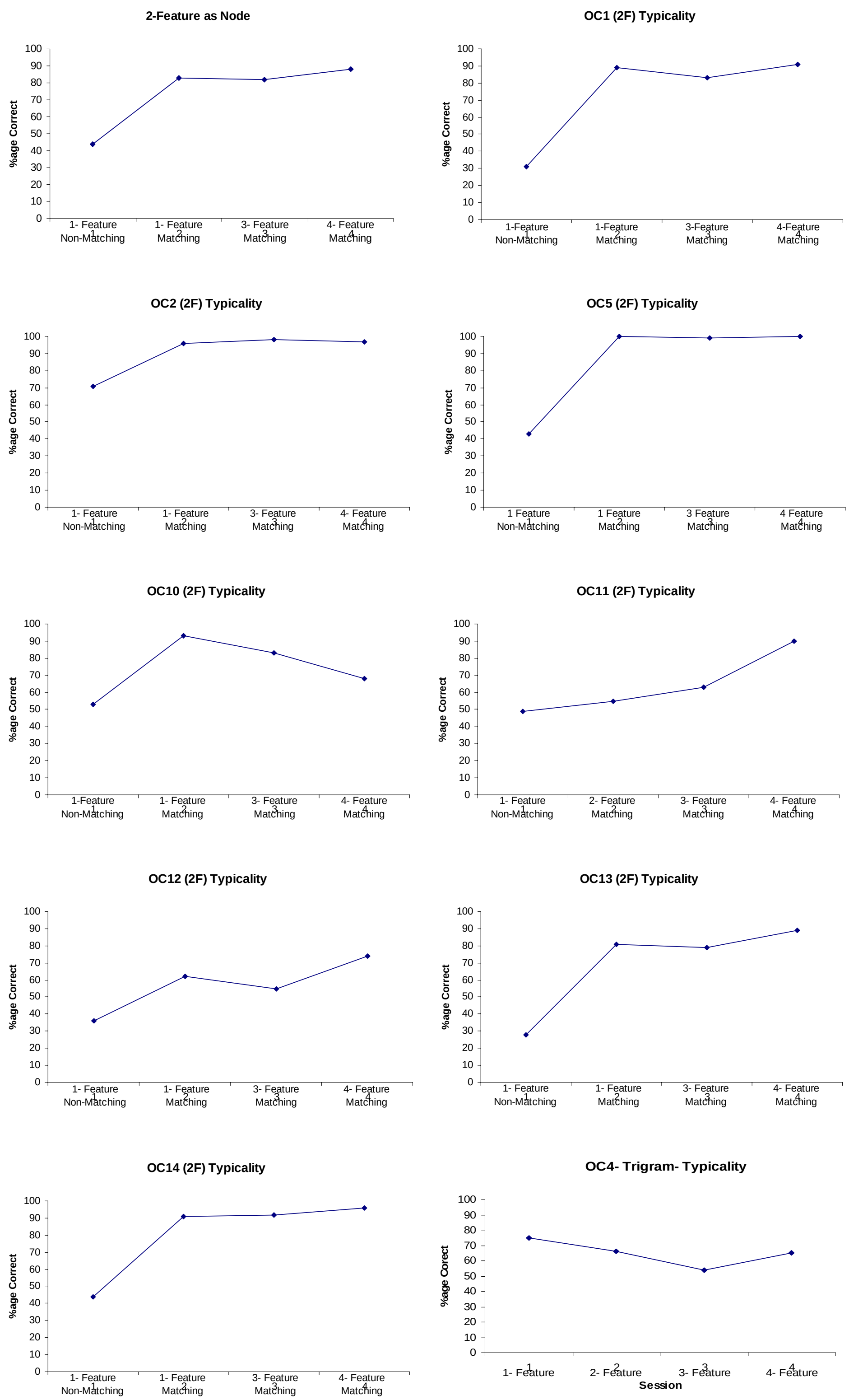


Figure 13. Stimulus class typicality. Each data point represents the percentage of correct responses to each feature level and whether feature matching was possible.

OC5, OC10, OC13, and OC14) had higher accuracy on the matching stimuli for all three feature levels, than on the 1-Feature non-matching stimuli. An analysis compared the accuracy of responses on 1- feature matching trials to the accuracy of responses on 3- and 4- feature stimuli. The results of the analysis showed that there was no significant difference between the 1- feature matching trials and the 3- and 4- feature stimuli, (t (12) = 2.18, p<.05). This indicates that the children were not demonstrating typicality because accuracy was lower on the non-matched trials than on matched trials regardless of the number of relevant features. The group means for the children in the 2-Feature-as-node condition demonstrate this effect clearly (see Figure 13). OC4, the only child in the trigram-as-node condition, did not show typicality. She actually had higher accuracy on trials that involved 1-feature stimuli than trials that involved 3-feature stimuli.

Typicality effects during novel probe performances were also assessed. The two children who received novel probes were OC2 and OC4 (see Table 11). These children were both exposed to the same stimulus set (Stimulus Set 1), with different training arrangements. OC2 was trained with the 2-Feature-as-node and OC4 was trained with the Trigram-as-node. Another analysis was completed to determine if there was a difference in correct responding on 1- and 2- feature novel stimuli versus 3- and 4- feature novel stimuli. Both children showed some evidence of typicality (see Figure 14). OC2 had a progression of increasing accuracy from 1-Feature through 4-Feature matching trials responding with greater accuracy on the 3-and 4- feature trials. OC4 responded more accurately to the 3- and 4-Feature novel stimuli, than to the 1- and 2-Feature novel stimuli.

Table 11

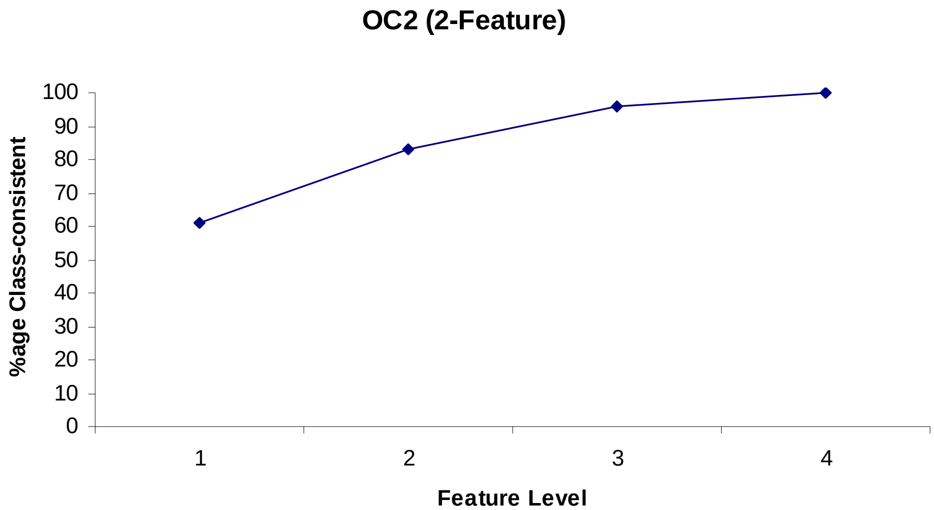
Number of novel probe sessions included in each child's typicality analysis.

Name	# of Sessions
OC2	18
OC4	6

Discussion

One of the questions addressed by Experiment 2 was whether the children would acquire the baseline conditional discriminations regardless of the stimuli presented as node. Within the time constraints provided here, the 2-Feature-as-node training led to acquisition more readily than the trigram-as-node training. The majority of the children who showed some acquisition were from the 2-Feature-as-node condition (OC1, OC2, OC5, OC10, OC11, OC12, OC13, and OC14). However, OC4 who was in the trigram-as-node condition cannot be ignored. OC4 not only showed some acquisition, she finished equivalence and novel probes. Given this example, it is possible that all of the children who were presented with the trigram-as-node condition would have shown acquisition, given sufficient training.

Training with the 2-Feature stimulus as node was more likely to result in the acquisition of the baseline conditional discriminations than training with the trigram functioning as the node. A likely explanation of this outcome involves feature matching. Feature matching involves identity matching based on individual features of the stimuli. Children exposed to the 2-Feature-as-node training seemed to become dependent on feature matching. Most of these children performed at chance levels on all of the trials for which feature matching was not possible, until forced to do otherwise by specialized training. For example, in Figure 13, all eight of the children in the 2-Feature training group had greater accuracy when responding to relations in which they could feature match. The lowest accuracy scores were seen for the relations in which feature matching was not possible. Feature matching might be supported because, by feature matching alone, the child could produce reinforcers on 63% of the trials in each session.



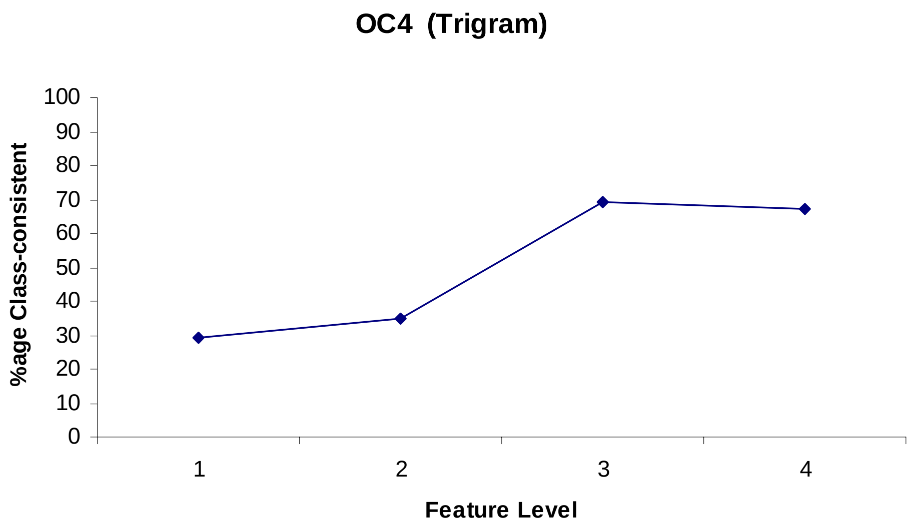


Figure 14. Stimulus class typicality for the novel stimuli presented. Each data point represents the percentage of correct responses for each feature level.

Performing at chance levels on trials that did not allow for feature matching would add reinforcers for an additional 12% of the trials, giving an overall reinforcement density of approximately 75% for the session. This level of reinforcement could be sufficient to maintain non-criterion patterns of responding. In such a scenario, the child has not learned the arbitrary stimulus relations; instead they have learned the relation of identity.

Based on these results, it appears as though having the 2-Feature stimulus serve as node can be an efficient way to approach simultaneous conditional discrimination training, if there are overlapping features as in the present case. Depending on the experimental or applied goals, this could be a good procedure to use when time cannot be devoted to taking small steps that build upon each other, as suggested earlier. However, one should be cautious about drawing this conclusion. It is reasonable to argue that the children here did not learn the conditional discriminations needed to form equivalence (at least not without specialized training). What was learned here were not arbitrary relations, but rather identity matching with a large number of stimuli in each class.

Stimulus Control Topography (SCT) coherence (McIlvane and Dube, 2003), describes the degree of concordance between the stimulus properties that control the behavior of the behavior analyst and those that control the behavior of the organism, and provides another way to consider the issue of feature matching in this experiment. Here the training was designed to establish arbitrary conditional discriminations between relevant features. However, performance came under the control of only those features held in common with the comparisons (i.e., identity relations were established). It is the arbitrary relations that are critical to class formation and in this case, the arbitrary relations were not learned.

When these arbitrary relations were not learned, specialized training procedures were employed for the children in Experiment 2. These included the correction procedure and training phases in which a smaller number of conditional discriminations were presented simultaneously. Results with the correction procedure were not promising. In fact, accuracy scores were often lower with this procedure than before the specialized training was implemented. Many of the children responded to the same incorrect stimulus on multiple successive trials, demonstrating that repeating the trial did not appear to lead to “correct” responses. None of the children exposed to the correction procedure mastered the conditional discriminations until additional specialized training procedures were added. Another specialized training procedure utilized with children in Experiment 2 involved alternating intermixed sessions of two conditional discriminations with and without delay. The alternating delay training did appear to aid in acquisition. This held true for OC1, OC2, and OC3.

A second question addressed by this experiment was whether each training structure would produce strong evidence of equivalence-class formation as demonstrated by a high percentage of class-consistent responding on equivalence probes. Only three subjects were presented with equivalence probes. With only three subjects, very little can be said regarding this question, although some interesting patterns were observed. OC4, who received equivalence probes in the trigram-as-node condition, chose the class-consistent comparison on almost 100% of the trials for all reflexivity, symmetry and transitivity probes. The two children in the 2-Feature-as-node condition performed very differently on their equivalence probes. While OC2 and OC5 chose the class-consistent comparison on almost 100% of the reflexivity probes, class-consistent responses were made on 50-100% of the symmetry probes and on only 0-60% of the transitivity probes. Based on these results, the 2-Feature training used here did not appear to facilitate class formation.

There are several reasons why the 2-Feature-as-node training condition may have failed to facilitate class formation. As stated above, feature matching may have worked against the formation of equivalence classes and therefore, equivalence probe performances. In order for the children to progress to the probe sessions, mastery criterion had to be met. This included the trials in which feature matching was not possible. According to this criterion, it appeared that the required classes had formed. Despite this, the feature matching SCT may have been reinforced often enough to increase its probability over that of equivalence. Therefore on all probe trials for which feature matching was possible, class-consistent responses were made. However, on probe trials for which feature matching was not possible, performances were at chance levels and there was no evidence that equivalence classes had formed. Control by identity relations proved to be stronger than control by equivalence.

The naming theory (Dugdale & Lowe, 1996) could also be used to explain the failure of the 2-Feature-as-node training condition to facilitate class formation. Training with the trigrams-as-node required the subject to match each trigram (a potential name) to every stimulus in its intended class during every session. In this way, a "name" might be established for each stimulus in a class and, according to naming theory, every stimulus that has a common name is considered equivalent. With 2-Feature-as-node training, the trigrams appeared as comparisons on only three trials total during each session and were presented with only the E stimuli as samples (one trial with each of the E stimuli as a sample). The “name” was not matched with any of the other stimuli in the intended class. Thus, a common name for each stimulus within the class was not directly established. According to Dugdale and Lowe (1990), the naming process is necessary for equivalence class formation. The naming position does seem consistent with OC4’s strong equivalence performances and with the lack of class-consistent performances for OC2 and OC5. While the 2-Feature training condition did not prevent naming from occurring, it did not require it as explicitly, as did the trigram-as-node training condition, and therefore may have been less likely to generate strong equivalence performances.

The third question explored in Experiment 2 was whether generalized equivalence would result from each of the training conditions. To show generalized equivalence, a novel stimulus that had one to four features relevant for a class would be categorized as part of that class. This would demonstrate that the novel stimulus had become a member of the equivalence class though the common features. Natural language classes work in the same manner (Stewart, 1999; Madden, 2001).

The child in the many-to-one, trigram-as-node condition (OC4) performed less class-consistently on the novel stimulus inclusion than the child in the many-to-one, 2-feature-as-node condition. These patterns are exactly opposite those shown on the equivalence probes. Feature matching may again provide an explanation. The same novel stimuli were used for the probe sessions that followed each of the training condition; only the way they were presented differed. In both conditions the novel stimuli were always presented as the sample. In the trigram-as-node condition, the comparisons were trigrams and in the 2-Feature-as-node condition, the comparisons were the 2-Feature stimuli. This enabled the children to match many, but not all, novel stimuli with the 2-Feature stimulus based on common features. The feature matching would allow for the higher class-consistent responding for the 2-Feature-as-node condition. At this point observations are limited to the two children who received novel-probe testing. These data will be expanded as the children whose training is still in progress complete their equivalence and novel probe phases.

With regard to typicality, Experiment 2 showed some interesting results. Typicality was calculated for all nine of the subjects (OC1, OC2, OC4, OC5, OC10, OC11, OC12, OC13, and OC14) who either progressed to probe testing or achieved at least 80% accuracy on full baseline sessions (i.e., those presenting all eight conditional discriminations). However, typicality effects were not consistently demonstrated across baseline training (see Figure 13). In order for typicality to be demonstrated, the percentage of correct responses would need to be higher on the trials which involved three and four feature stimuli and lower on the trials which involved stimuli with one feature, regardless of feature matching. OC11 and OC12 did appear to show some evidence of typicality. The other children’s responses on these trials did not demonstrate this phenomenon (see Figure 13). They instead tended to show more support for feature matching. Correct responses on the 1 feature-matching trials were as high or higher than for the three and four feature stimuli.

With regard to typicality effects on the novel probes, no direct comparison can be made given data from only two children. However, interesting patterns were observed with the novel probes and typicality. OC2, who was trained in the 2-Feature-as-node training group showed an increasing progression in accuracy from 1-Feature to 4-Feature stimuli (see Figure 14). OC4, who was trained in the Trigram-as-node training group, showed more accurate responding to the 3 and 4-Feature stimuli than to the 1 and 2-Feature stimuli (see Figure 14).

DISCUSSION & CONCLUSIONS

These experiments investigated attempts to train eight conditional discriminations simultaneously. This was done with different training structures, different stimulus sets, and with different stimuli acting as the node. Across all training arrangements, the one for which there was the greatest initial acquisition was the many-to-one 2-feature stimulus-as-node, regardless of the stimulus set presented or the order of training. The children trained in this condition acquired more of the baseline relations, without requiring specialized training, than did children in any other group. While this was probably due to feature matching, as discussed earlier, it is an interesting fact to note. The children in the one-to-many and many-to-one, trigram-as-node conditions all required extensive specialized training in order to master any of the baseline conditional discriminations. The one exception to this was OC4 who mastered the full baseline without any specialized training in the many-to-one, trigram-as-node condition.

In looking at the two stimulus sets used in these experiments the main difference was in how the relevant and irrelevant features were displayed. The second set displayed the relevant features in a color-coded fashion. The rationale for adding color-coded features in Stimulus Set 2 can be related to Sidman and Stoddard's (1966, 1967) circle-ellipse discrimination studies. In these studies the subjects were taught through a stimulus-control shaping procedure (see McIlvane & Dube, 1992) to discriminate the relevant differences in the stimuli by beginning with stimuli that were physically extremely different. Through this stimulus-control shaping procedure subjects were increasingly more accurate in discriminating between stimuli. The present experiment attempted to utilize this idea by making the stimuli physically different in some clear, easily seen method, color. To facilitate discrimination between the relevant and irrelevant features, the relevant features were color coded. However, in observing the children's performance in Experiment 1, this stimulus set did not appear to facilitate acquisition.

The subject of naming was addressed in the present study by manipulating the nature of the stimulus employed as the training node. In the many-to-one, trigram-as-node training condition children were directly taught to match the trigram (name) with the other stimulus for every trial of each session. In the many-to-one, 2-feature-as-node training condition, training sessions included the trigram in only three of the 24 trials. Children in both conditions learned the conditional discriminations. Since children in the condition that did not require trigram matching were still able to learn the conditional discriminations, these experiments could demonstrate that naming is not necessary for conditional discrimination (e.g., Carr, Wilkinson., Blackman, & McIlvane, 2000, & Sidman & Cresson, 1973).

An idea proposed by Saunders and Green (1999) deserves attention here as well. They propose that the many-to-one training structure results in quicker acquisition of the baseline conditional discriminations than does the one-to-many training structure. The reason for this, according to the authors, is that this structure provides all of the simple discriminations within the training that are needed for consistently positive outcomes on all of the tests for equivalence. This idea is neither proved nor disproved by the current experiment. The only children in the one-to-many training structure in these experiments were the younger children, of whom none managed to acquire the baseline conditional discriminations without the aid of the specialized training procedures. In order to have a better chance of directly observing the effects of the training structures, children of all ages should be included within the two different training-structure conditions, one-to-many and many-to-one.

Further research should look at training with one conditional discrimination and building upon that, using the block formation. This may greatly aid in the facilitation of the conditional discriminations, especially in younger children, and allow for observing probe performances of various types in less time. However, if the goal of the conditional discrimination training is to observe typicality, this would not be a preferred method. The reason for this being that the subjects would have more exposure to the conditional discriminations that were trained initially and less exposure to the conditional discriminations that were trained lastly. Therefore, responding would be based upon exposure rather than typicality.

With regard to the problem of subject attrition, it might also be useful to simplify the stimulus set to one-to-three relevant stimuli instead of the four levels used in this experiment. Using only one of the 1-Feature stimuli, instead of all four as in this experiment, would also aid in the speed of acquisition.

Future research should not only address this issue, but should also address the main confound that was contained within this experiment, feature matching. This could be a difficult problem to address if one wanted to stay clear of anything involving a name for the node. The easiest solution seems to be to use a stimulus that has nothing to do with the set at all. Thus, there would be no relevant features that would lead to feature matching.

Simultaneous training with multiple conditional discriminations can be a difficult to achieve with children. Many different aspects need to be carefully observed and regulated. These include, but are not limited to: training structure, stimulus set, color of stimulus, node, the number of conditional discriminations, the number of relevant and irrelevant features, and age of the child. Though the obstacles are great, the rewards and interesting questions that can be answered are bountiful.

REFERENCES

Arntzen, E. & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record*, **47**, 309-320.

Carr, D., Wilkinson, M., Blackman, D., & McIlvane, W. J. (2000). Equivalence classes in individuals with minimal verbal repertoires. *Journal of the Experimental Analysis of Behavior*, **74**, 101-114.

Dube, W.V. (1991). Computer software for stimulus control research with Macintosh computers. *Experimental Analysis of Human Behavior Bulletin*, **9**, 28-30.

Dugdale, N. A., & Lowe, C. F. (1990). Naming and stimulus equivalence. In D. E. Blackman and & H. Lejeune (Eds.), Behavior analysis in theory and practice: Contributions and controversies (pp. 115-138). Hove, England: Erlbaum.

Hayes, S., Fox, E., Gifford, E., Wilson, K., Barnes-Holmes, D. & Healy, O. (2001) Derived relational responding as learned behavior. In Hayes, S., Barnes-Holmes, D., & Roche, B. (Eds.), Relational frame theory: A Post-Skinnerian account of human language and cognition. (pp. 21-49). Plenum Press, USA.

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

Kastak, C. R., Schusterman, R. J., & Kastak, D. (2001). Equivalence classification by California sea lions using class-specific reinforcers. *Journal of the Experimental Analysis of Behavior*, **76**, 131-158.

Lipkens, R., Hayes, S. C., & Hayes, L. J. (1993) Longitudinal Study of the Development of Derived Relations in an Infant. *Journal of Experimental Child Psychology*, **56**, 201-239.

Madden, Ashley (2001). Developmental changes and typicality effects in the formation of generalized equivalence classes. Unpublished master's thesis, University of North Carolina at Wilmington.

Margolis, E. (1994). A reassessment of the shift from the classical theory of concepts to prototype theory. *Cognition*, **51**, 73-89.

McIlvane, W.J. & Dube, W.V. (2003). Stimulus Control Topography Coherence Theory: Foundations and Extensions. *The Behavior Analyst,* **26**, 195-213.

Mervis, C., & Pani, J. (1980). Acquisition of basic object categories. *Cognitive Psychology,* **12**, 496-522.

Mervis, C., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of Psychonomic Society*, **7(3)**, 283-284.

Pilgrim, C., Chambers, L. & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: II. Children. *Journal of the Experimental Analysis of Behavior,* **63**, 239-254.

Pilgrim, C. & Galizio, M. (1990). Relations between contingencies and equivalence probe performances. *Journal of the Experimental Analysis of Behavior,* **54**, 213-224.

Pilgrim, C., Jackson, J., & Galizio, M. (2000). Acquisition of arbitrary conditional discriminations by young normally developing children. *Journal of the experimental Analysis of Behavior*, **73**, 177-193.

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

Saunders, K. J., Saunders, R. R., Williams, D. C., & Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record*, **43**, 725-744.

Saunders, R. R. & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior,***72**, 117-137.

Sidman, M. (1971b). Reading and auditory visual equivalences. *Journal of Speech and Hearing Research*, **14**, 5-13.

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

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, **74**, 127-146.

Sidman, M. & Cresson, O. (1973). Reading and crossmodal transfer of stimulus equivalences in severe retardation. *American Journal of Mental Deficiency*, **77**, 515-523.

Sidman, M & Stoddard, L.T. (1967). The effectiveness of fading and learning for retarded children. *Journal of the Experimental Analysis of Behavior*, **10**, 3-15.

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**, 47-63.

Stewart, K. (1999). Typicality effects in contingency-shaped stimulus classes. Unpublished master's thesis, University of North Carolina at Wilmington.

Urcuioli, P. J. Avian Visual Cognition. Categorization & Acquired Equivalence. Retrieved February 18, 2003, from http://www.pigeon.psy.tufts.edu/avc/urcuioli/default.htm.

Vaughan, W. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology, Animal Behavior Processes*, **14(1)**, 36-42.

Wittgenstein, L. *Philosophical investigations*. New York: Wiley, 1962.

Zentall, T. (1996). An analysis of stimulus class formation in animals. In Zentall, T.R. & Smeets, P. M. (Eds.) Stimulus Class formation in Humans and Animals (pp 15-34). Elsevier Science B. V.