The Role of Naming in Stimulus Categorization by Preschool Children

Caio Flávio Miguel
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations
Part of the Applied Behavior Analysis Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/3398
THE ROLE OF NAMING IN STIMULUS CATEGORIZATION
BY PRESCHOOL CHILDREN

by

Caio Flávio Miguel

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Psychology

Western Michigan University
Kalamazoo, Michigan
August 2004
ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. James Carr for his outstanding supervision, invaluable support, and most of all, his friendship.

I have been extremely fortunate for having the opportunity to work so closely with Dr. Jack Michael, who served as my intellectual mentor and co-advisor in the current project. Jack has played a vital role in my professional, as well as personal development throughout these years, becoming an inestimable friend.

I would like to thank my committee members, Dr. Linda LeBlanc, Dr. Scott Gaynor, and Dr. Michael Hixson for their helpful comments and suggestions.

I feel extremely in debt to Anna I. Petursdottir for playing a major role at the development and execution of this project. I would also like to thank Amanda Mahoney, Tyler Weeks, Kaneen Geiger, Gina Savo, and Holly Bihler for their help with data collection.

This project would have not been possible without the ongoing support of Deborah Johnston and all the staff at Little Lambs Childcare Center; as well as the financial support of the Western Michigan University Graduate Research Fund.

Finally, I would like to dedicate this work to my parents, Cristina and João Paulo, and to my brother Ciro, for their love and support.

Caio Flávio Miguel
THE ROLE OF NAMING IN STIMULUS CATEGORIZATION
BY PRESCHOOL CHILDREN

Caio Flávio Miguel, Ph.D.
Western Michigan University, 2004

Naming is defined as a higher-order behavior class that involves the establishment of a bi-directional relation between a class of objects/events and the speaker-listener behavior they occasion. The purpose of the current study was to assess whether children would categorize unfamiliar pictures when taught the listener and speaker behaviors separately. A category-sort test was used to assess emergent conditional relations. Category-sort trials consisted of looking at (Test 1) or tacting (Test 2) a sample and selecting the remaining comparisons. In Experiment 1, four children (3-5 yrs.) were taught to tact the category (i.e., north and south) of pictures of six U.S. state maps, and were assessed to see whether they would (1) categorize and (2) select them when hearing their category names (listener behavior). Two of the children categorized the pictures during Test 1 after pairwise tact training. The other two categorized after receiving additional tact training with all pictures grouped together. However, one of them only categorized during Test 2. When children failed to categorize, they also performed poorly on stimulus selection. In Experiment 2, four children (3-5 yrs.) were taught to select pictures when hearing their category names and were assessed to see whether they would (1) categorize and (2) tact them (speaker
behavior). One child categorized the pictures during Test 1, and two during Test 2. The other child required additional training with all pictures grouped together. When participants failed to categorize, they also failed to tact the pictures accurately. Taken together, results from Experiment 1 and 2 show that both speaker and listener behaviors have to be present for categorization to occur. Thus, naming seemed to play an important, if not essential, role in the development of categorizations.
# TABLE OF CONTENTS

**ACKNOWLEDGMENTS** ...................................................................................... ii

**LIST OF TABLES** ........................................................................................ vi

**LIST OF FIGURES** ....................................................................................... vii

**INTRODUCTION** ......................................................................................... 1

- Stimulus Equivalence.................................................................................. 3
- Common Naming and Equivalence Classes ............................................. 9
- Verbal Competence and Equivalence Classes ........................................ 15
- Verbally Mediated Equivalence.................................................................. 20
- Stimulus Categorization.............................................................................. 24
- Purpose of the Current Study....................................................................... 28

**EXPERIMENT 1** ......................................................................................... 29

- Method........................................................................................................ 29
  - Participants, Setting, and Materials.................................................... 29
  - Dependent Variables and Data Collection......................................... 30
  - Interobserver Agreement (IOA)............................................................. 31
  - Experimental Design........................................................................... 31
  - Pretraining............................................................................................ 33
  - Procedure.............................................................................................. 38
  - Independent Variable Integrity (IVI)...................................................... 43
**Table of Contents—continued**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results</td>
<td>44</td>
</tr>
<tr>
<td>Familiar Pictures</td>
<td>44</td>
</tr>
<tr>
<td>Unfamiliar Pictures (Maps)</td>
<td>45</td>
</tr>
<tr>
<td>Vocal-Verbal Behavior</td>
<td>57</td>
</tr>
<tr>
<td>Discussion</td>
<td>58</td>
</tr>
<tr>
<td>EXPERIMENT 2</td>
<td>63</td>
</tr>
<tr>
<td>Method</td>
<td>63</td>
</tr>
<tr>
<td>Participants, Setting, and Materials</td>
<td>63</td>
</tr>
<tr>
<td>Dependent Variables and Data Collection</td>
<td>63</td>
</tr>
<tr>
<td>Interobserver Agreement (IOA)</td>
<td>64</td>
</tr>
<tr>
<td>Experimental Design</td>
<td>65</td>
</tr>
<tr>
<td>Pretraining</td>
<td>65</td>
</tr>
<tr>
<td>Procedure</td>
<td>66</td>
</tr>
<tr>
<td>Independent Variable Integrity (IVI)</td>
<td>70</td>
</tr>
<tr>
<td>Results</td>
<td>71</td>
</tr>
<tr>
<td>Familiar Pictures</td>
<td>71</td>
</tr>
<tr>
<td>Unfamiliar Pictures (Maps)</td>
<td>71</td>
</tr>
<tr>
<td>Vocal-Verbal Behavior</td>
<td>82</td>
</tr>
<tr>
<td>Discussion</td>
<td>83</td>
</tr>
<tr>
<td>GENERAL DISCUSSION</td>
<td>87</td>
</tr>
<tr>
<td>Pairwise vs. All-Stimuli Training</td>
<td>88</td>
</tr>
</tbody>
</table>
# Table of Contents—continued

Interdependence Between Speaker and Listener Repertoires .................. 90
Joint Control ......................................................................................... 92
Acquisition of Listener Behavior/Stimulus Selections ............................ 93
Limitations .......................................................................................... 96
Future Directions ................................................................................. 98
Applied Implications ............................................................................ 100
Conclusion ............................................................................................ 101

## APPENDICES

A. Unfamiliar Pictures ........................................................................ 103
B. Posttest Interviews .......................................................................... 105
C. Sample Datasheets .......................................................................... 115
D. Protocol Clearance from the Human Subjects Institutional Review Board .......................................................... 120

## BIBLIOGRAPHY

.......................................................... 122
LIST OF TABLES

1. Interobserver Agreement Results for Experiment 1 ........................................... 31
2. Summary of Experimental Conditions – Experiment 1 ........................................ 33
3. Average IVI Scores for Experiment 1 ................................................................. 44
4. Number of Trials to Criterion During Pretraining ........................................... 44
5. Number of Trials to Criterion During Tact Training ........................................ 57
6. Interobserver Agreement Results for Experiment 2 ........................................... 64
7. Summary of Experimental Conditions – Experiment 2 ........................................ 66
8. Average IVI Scores for Experiment 2 ................................................................. 71
9. Number of Trials to Criterion During Pretraining ........................................... 72
10. Number of Trials to Criterion During Stimulus Selection Training .................. 82
LIST OF FIGURES

1. Percentage of Correct Categorizations for Tom (upper panel) and Adam (lower panel) ................................................................. 46
2. Percentage of Correct Responses During Stimulus Selection-1 Test for Tom ........................................................................... 49
3. Percentage of Correct Responses During Stimulus Selection-1 Test for Adam ................................................................. 50
4. Percentage of Correct Responses During Stimulus Selection-2 Test for Adam ................................................................. 51
5. Percentage of Correct Categorizations for Rita (upper panel) and John (lower panel) ................................................................. 53
6. Percentage of Correct Responses During Stimulus Selection-1 Test for Rita ........................................................................... 55
7. Percentage of Correct Responses During Stimulus Selection-1 Test for John ................................................................. 55
8. Percentage of Correct Responses During Stimulus Selection-2 Test for John ................................................................. 56
9. Percentage of Correct Categorizations for James (upper panel) and Maria (lower panel) ................................................................. 73
10. Percentage of Correct Responses During Tact-1 Test for James ......................................................................................... 75
11. Percentage of Correct Responses During Tact-1 Test for Maria ......................................................................................... 76
12. Percentage of Correct Categorizations for David (upper panel) and Pam (lower panel) ................................................................. 77
13. Percentage of Correct Responses During Tact-1 Test for David ......................................................................................... 80
14. Percentage of Correct Responses During Tact-2 Test for David ......................................................................................... 80
15. Percentage of Correct Responses During Tact-1 Test for Pam ......................................................................................... 81
INTRODUCTION

The process of determining how to group objects or events together is usually called categorization or classification, while those objects or events that cohere may be regarded as a category or a class. It is often assumed by cognitive psychologists that the categorization process is dependent upon the acquisition of specific "concepts" (Barsalou, 1992). For example, by having the concept of "bird," one would be able not only to categorize previously encountered birds, but also new ones, given that these may be recognizable via comparison to the general information previously acquired. These concepts are said to be units of mental representation and seem to exist independently of any behavior-environment relation (Zentall, Galizio, & Critchfield, 2002).

The practice of attributing the ability of grouping objects or events together to an inferred mental state (i.e., the concept) may be detrimental as it detracts us from identifying causal variables that could be readily available for a scientific analysis (Skinner, 1953, 1974). A better practice is to describe the specific behaviors that would account for terms like concept and categorization. As a result, these behaviors could be analyzed in terms of their particular environmental antecedents and consequences, permitting us to understand, predict, and control their occurrence.

A concept can be defined as a group of objects (stimuli, actions, etc.) that control similar responses. When an individual behaves similarly in response to a group of objects, these objects form a class that can be called a concept. As a result,
concepts may be equated to stimulus classes. Concepts have also been spoken of as
generalization within a class and discrimination between classes of stimuli (Keller &
Schoenfeld, 1950). The concept of “bird,” for example, involves generalization
among all stimuli that we label as birds. The same concept involves the discrimina­
tion between these stimuli and others we do not call birds (Catania, 1992). Most of
these concepts are acquired through formal or informal education, when the same
response (e.g., saying “bird”) is reinforced in the presence of one stimulus (e.g., an
actual bird). After being reinforced to say “bird” in the presence of several different
stimuli/birds, but not in the presence of other animals or objects (i.e., multiple-
exemplar training), the verbal response “bird” would be controlled by the physical
characteristics that are common to all of these stimuli. When a new stimulus that
shares the same or some of these characteristics is presented, one would, through
stimulus generalization, respond similarly by calling that novel stimulus a “bird.”

Early researchers in the field of experimental analysis of behavior were
successful in teaching specific concepts to non-human subjects. Hernstein and
Loveland (1964), for example, taught pigeons to peck a specific key located in an
experimental chamber in the presence of pictures that contained a human form, but
not in the presence of pictures that contained other forms. In a follow-up study,
Hernstein, Loveland, and Cable (1976) reinforced key pecks in the presence of slides
with naturalistic scenes that contained different examples of the critical stimulus (e.g.,
tree for some pigeons, water for others, etc.), but not in the presence of slides which
did not contain these critical stimuli. After training, when presented with a variety of
novel slides, subjects were more likely to peck those that contained some features of the critical stimuli. Additional studies (e.g., Bhatt & Wright, 1992; Malott & Siddal, 1972) have demonstrated the development of complex stimulus classes as a direct result of differential reinforcement.

Far more interesting examples of stimulus classes are those whose members do not share any physical similarity with one another. These stimuli are said to cohere because of their shared functional properties. One example would be the concept of animals. Although physically dissimilar, birds, cats, giraffes, monkeys, and fish all evoke the same response "animals." Of interest is that a child may encounter a never-before-seen exemplar and correctly respond towards it by saying "animal," despite the fact that this new exemplar does not necessarily resemble previously trained ones. Since stimulus generalization cannot account for the development of such stimulus classes, one possibility is that stimuli become members of the same class for serving a similar behavioral function (i.e., they all evoke the same response). Those stimuli that control the same response but have no physical properties in common are usually referred to as functional classes (Sidman, 1994). In some cases, stimuli may become part of the same class without sharing any physical feature or necessarily serving the same behavioral function. These are called equivalence classes (Sidman & Tailby, 1982).

Stimulus Equivalence

In the field of behavior analysis, research on stimulus equivalence has focused on the study of stimulus classes whose members do not share physical features or
serve similar behavioral functions. An example of these classes would be the relation among pictures, printed words, and spoken words. These stimuli may become substitutable under specific conditions. When asked to point to a ball, someone can point to either a ball, the picture of a ball, or the printed word ball. However, if asked to kick the ball, only one stimulus (i.e., the actual ball) would function as an effective discriminative stimulus (Green & Saunders, 1998). Interestingly, after being taught to respond to some of these stimuli, humans may behave similarly in the presence of other stimuli without being directly trained to do so. The understanding of the variables responsible for this emergent repertoire is what drives research in the area of stimulus equivalence (Green & Saunders, 1998; Sidman, 1994).

The matching-to-sample procedure (MTS) is often used in the study of stimulus equivalence. In MTS, a single stimulus (sample) is presented followed by two or more other stimuli (comparisons). Selecting one of the comparisons is always reinforced in the presence of one specific sample, but not in the presence of others. As a result, the discriminative function of the comparison stimuli for the selection response becomes dependent upon the presence of a specific sample. When an organism has learned to conform to contingencies like these, it is said to have learned conditional discriminations (Green & Saunders, 1998).

In stimulus equivalence research, a series of conditional discriminations is established via MTS. The stimuli used during these conditional discriminations are usually arbitrary and physically dissimilar. It is assumed that after a well-established conditional discrimination, the conditional relation between stimuli, and an
equivalence relation are produced (Sidman & Tailby, 1982). If during a series of MTS tasks, subjects learn to “match” comparisons with sample stimuli, it is suggested that stimulus substitutability between sample and comparison has been produced (Sidman, 1994). This substitutability or equivalence among stimuli may be assessed though a series of tests within the MTS procedure. These include the tests for reflexivity, symmetry, and transitivity relations. During reflexivity tests, each stimulus should bear a relation to itself. In this case, subjects should be able to select a comparison stimulus A in the presence of the sample stimulus A, and a comparison B in the presence of a sample stimulus B. During symmetry tests, the reversibility in the relationship between stimuli is assessed. Thus, if during training choosing comparison B was reinforced in the presence of sample A, symmetry would be demonstrated if the comparison A is chosen when the sample B is presented. For transitivity tests, a third stimulus is required (C), so each relation has one stimulus in common. If in the presence of A, the selection of B was reinforced and in the presence of B, the selection of C was reinforced, then transitivity would be demonstrated if, in the presence of A, the comparison C is chosen (Sidman & Tailby, 1982). Additionally, the relations of symmetry and transitivity of AB and BC can be tested simultaneously if, in the presence of C, the comparison A is chosen. This combined test has been referred to as the test for equivalence (Green & Saunders, 1998). Although research on equivalence has been prevalent in the field of experimental analysis of behavior for years, there is still no consensus regarding the mechanism responsible for the
emergent behavior observed during transitivity and equivalence tests (Saunders, Williams, & Spradlin, 1996).

According to Sidman (1994), equivalence is a primitive function that is a direct product of contingencies of survival, as are reinforcement and discrimination. The establishment of equivalent relations is "one of the outcomes of reinforcement contingencies" (p. 387). During reinforcement, reinforcers and responses join the conditional and discriminative stimuli as members of an equivalence class. In other words, equivalence relations consist of ordered pairs of all elements that participate in the reinforcer contingency, namely stimuli and responses (Sidman, 2000). The contingency creates the prerequisites for demonstrating the mathematical properties that define the equivalence relation. Thus, the phenomenon of equivalence should not be derived from more basic behavioral processes, but considered as one of them.

Other behavior analysts (e.g., Hayes, 1991) have suggested that success on equivalence tests is a direct product of one's history of reinforcement specific to the tasks required on these tests. This approach to explain stimulus equivalence and other types of derived stimulus relations has been called Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001). The core assumption of RFT is that equivalence is a form of generalized operant behavior, learned via an appropriate history of multiple exemplar training. According to Hayes et al., some species can learn to respond to relations between stimuli when these relations are not defined by the physical forms of the stimuli, but by contextual cues. Consequently, when these contextual cues are in place (e.g., the word "same"), the organism is capable of
responding relationally to any arbitrary stimuli (e.g., respond as if these stimuli are the same). The emergent behavior observed during equivalence tests is a product of a learned generalized repertoire, much similar to the operant class of generalized imitation (Hayes et al., 2001).

A third explanation for the establishment of equivalence classes has been proposed by Horne and Lowe (1996, 1997). According to these authors, the formation of equivalence classes is heavily dependent upon the subject’s verbal repertoire, more specifically, what they call the naming relation. According to Horne and Lowe (1996), naming is a higher-order behavior class that involves the establishment of a bi-directional relation between “a class of objects and events and the speaker-listener behavior they occasion” (p. 200).

In their seminal article, Horne and Lowe (1996) fully described how the naming repertoire is established during early language acquisition in young children. The necessary conditions for the development of naming are present during typical child-caregiver interactions. Initially, a child may be reinforced to orient to a particular object, for instance, a shoe in the presence of the caregivers’ vocal stimulus “shoe.” The child may learn not only to orient to a particular shoe, but also to all objects that have been called “shoe” by her caregivers. Thus, the listener repertoire regarding the specific object is established. When the child becomes able to echo the vocal production of others, the caregiver may point to the shoe and ask the child to say “shoe.” The echoic response “shoe” or any approximation is either reinforced by the caregiver or automatically reinforced by the auditory product of the child’s verbal
response (Vaughan & Michael, 1982). Because the child has already learned to
behave as a listener (orient towards shoes) in response to the auditory stimulus “shoe”
produced by others, hearing this auditory stimulus as a product of her own echoic
behavior may occasion the child to engage in both listener behaviors and further
echoic utterances. When the caregiver points to a shoe and says, “shoe,” the sight of
the shoe becomes a frequent antecedent for the echoic utterance “shoe.” Consequ­
ently, the sight of the shoe becomes a discriminative stimulus that evokes the
verbal response “shoe” as a tact. Later, when the child is alone, the presence of a
shoe occasions the verbal response “shoe” whose auditory stimulus evokes the
relevant listener behaviors of reorienting to the shoe(s). This bi-directional, and
somehow circular relation between listener and speaker repertoires is what comprises
the object’s name (Horne & Lowe, 1996).

According to Horne and Lowe (1996), when naming is acquired, the presence
of one member of the class would evoke a tact (e.g., “shoe”), whose auditory product
(the sound “shoe”) would, in turn evoke the listener behavior of reorienting and
selecting other members that are part of the same name relation. It follows that to
establish arbitrary stimulus classes, one would need only to learn how to name each
member of the designated class (Horne & Lowe, 1996).

The naming approach for understanding the formation of stimulus classes
draws heavily from Skinner’s analysis of verbal behavior (Skinner, 1957) by empha­
sizing the “individual as speaker and listener within the same skin” (Horne & Lowe,

1 A Tact is a verbal operant whose response form is under control of a non-verbal discriminative
stimulus (Skinner, 1957).
According to some behavior analysts (e.g., Dugdale, 1996; Michael, 1996; Palmer, 1996; Remington, 1996), the naming analysis as proposed by Horne and Lowe (1996) is a plausible and attractive alternative to both Sidman’s (1994) and Hayes’ (1991) approaches to interpret the formation of stimulus classes and explain the phenomenon of derived stimulus relations. It serves as a model for analyzing the effects that the interaction among the various verbal operants (i.e., echoic, tact) has upon the acquisition of complex linguistic skills. It also establishes a much-needed bridge between behavior analysis and developmental psychology. According to Dugdale (1996), even if naming is not a prerequisite for the formation of equivalence classes, it is still likely to be the predominant means by which humans learn how to categorize. Thus, it may be important to understand how naming (or the interaction among verbal operants, as suggested by Michael, 1996) may directly affect stimulus class formation/categorization in humans.

Common Naming and Equivalence Classes

One of the first studies to address the influence of verbal behavior on stimulus class formation was conducted by Sidman, Cresson, and Wilson-Morris (1974). Two institutionalized boys (ages, 14 and 18 yr.) diagnosed with Down syndrome participated in the study. Through a series of conditional discrimination tasks, the participants were exposed to AB and BC matching tasks, in which A was a set of...
auditory stimuli (i.e., dictated words), and B and C were sets of visual stimuli (i.e., pictures and printed words, respectively). Participants were tested on the emergence of the crossmodal-matching task AC: in the presence of the sample dictated word, selecting the corresponding printed word out of several comparisons would be considered a correct response. Participants were also tested on two oral naming/labeling tasks, CD and BD. In other words, they were assessed on whether the pictures or printed words would evoke their respective correct labels - tacts or textual behaviors (e.g., in the presence of the written word “car” [C], participant would say “car” [D], or in the presence of a picture of a car [B], the participant would say “car” [D]). Both participants were able to perform the AC task without being directly trained to do so. In the naming/labeling tasks, one participant was not able to read the words aloud even though he was able to perform all of the other tasks. Although one participant was able to read some of the printed words, Sidman et al. (1974) suggested that this performance (textual) may have been a by-product of the AB training (receptive discrimination) and not a required skill for AC to have emerged.

Another study, conducted by Sidman and Tailby (1982), indirectly addressed the issue of verbal mediation. The authors taught eight typically developing children to match sets of three different visual comparisons (BCD) to dictated words (A). For instance, in the presence of the experimenter-produced sound “Lambda” participants were required to press a center key that would produce the comparison stimuli on

---

2 The term “naming” appears quite often in the behavior analytic literature usually to describe vocal behavior produced in response to a visual stimulus (word or picture). Even though some human performances identified as “naming” may ultimately fit Home and Lowe’s (1996) definition, most authors use the term solely to refer to the participants’ expressive performance - either tact or textual responses.
adjacent keys. Selecting the correct comparison (printed Greek letter) was reinforced. Participants were directly trained on AB, AC relations (auditory-visual matching tasks), and DC relations (visual-visual matching). They were then tested on emergent conditional discrimination performances (CB, BC, CD, AD, DB, BD) and oral naming/labeling. All eight participants responded consistently on tests of emergent performance after training. During the naming/labeling tests, participants were presented with the specific stimuli used during training, and asked, “Tell me what you see” or “What is it?” Oral labeling tests were all conducted under extinction. Consistent labeling was defined as emitting the same vocal topography in the presence of visual stimuli of each class. For example, saying “lambda” in the presence of all stimuli (BCD) of the lambda class would be considered “consistent naming”. Seven out of eight participants consistently labeled the stimuli in each class. Of interest were the participants’ responses to the D stimuli. These stimuli never served as comparisons in the presence of a dictated word (A), but only as samples during the visual-visual DC conditional discrimination training. The authors suggested that during the naming/labeling tests, one participant “hesitated” when asked to label the stimuli from set D, which would suggest that he did not label the stimuli until he was required to do so during the test. According to Sidman and Tailby, this would cast doubts on the hypothesis that participants were naming/labeling stimuli during the conditional discrimination tasks.

In a follow-up study, Sidman, Kirk, and Wilson-Morris (1985) conducted a series of experiments with typically developing children, adults, and mentally
retarded individuals as an attempt to evaluate procedures to expand stimulus classes from three to six-members. Participants were again tested on conditional discrimination performances and oral naming/labeling. All eight participants responded consistently on the emergent performance tests for the six-member classes after training. Six of the participants were able to label the stimuli according to their class membership. Saying “delta” in the presence of all stimuli of class 1, “sigma” for class 2, and “xi” for class 3. The other two participants who did not consistently name/label the stimuli during the oral naming test were still able to pass all the conditional discrimination tests. Consequently, Sidman, Kirk, and Wilson-Morris concluded that equivalence classes could be formed even when individuals do not apply the same name to all members of the class.

Sidman, Wilson-Morris, and Kirk (1986) further demonstrated that common naming/labeling may not be necessary for the formation of equivalence classes. After a series of conditional discrimination tasks, two typically developing children (5 yrs.) and four mentally retarded youths passed equivalence tests for two sets of three member classes, one containing both auditory and visual stimuli and the other containing only visual stimuli. The four participants diagnosed with mental retardation and one of the typically developing children labeled some of the stimuli in the visual classes, but did not give consistent names to them. Also, two of the individuals diagnosed with mental retardation failed to consistently label the stimuli in the auditory-visual class, which would further suggest that the emergence of stimulus equivalence does not require the mediation of common naming/labeling.
Although all participants in the Sidman et al. (1986) study demonstrated emergent relations, their test scores, at least for the participants diagnosed with mental retardation, were much better in the auditory-visual than in the visual-visual conditional discriminations. Also, for the formation of purely visual classes, procedural adjustments needed to be made. Visual classes are readily formed when test trials are interspersed among baseline trials (e.g., Saunders, Saunders, Kirby, & Spradlin, 1988; Saunders, Watcher, & Spradlin, 1988; Sidman et al., 1986). When other procedures are used, for instance, when blocks of test trials are administered after training is finished, failure to form visual classes may be observed (e.g., Devany, Hayes, & Nelson, 1986).

A critical study on the distinction between the development of auditory-visual and visual-visual classes was conducted by Green (1990). The participants of the study were adults diagnosed with mild mental retardation. They were given training on conditional discriminations and tested on four potential equivalence classes. Two of these classes included only visual stimuli, while the other two were auditory-visual. Green’s results indicate that all participants demonstrated the development of equivalence classes, however the purely visual classes developed slower than the ones that included auditory stimuli. It is possible, as suggested by Green, that the use of auditory samples (i.e., dictated names) during training would facilitate the process upon which participants would apply common names to the stimuli. Such a process could be explained based on a possible interdependence between the listener and speaker repertoires. Green (1990) also reported that during oral naming tests, only
two of five participants labeled all stimuli in the class consistently with the same name. The other participants described the stimuli in familiar terms, using a particular word for each stimulus.

Even though the aforementioned studies suggest that the development of equivalence classes may not depend upon the ability to name/label the stimuli, it is possible to argue that oral naming tests may not have been the best way to assess whether participants were engaging in verbal behavior during the experiment. The fact that some participants did not label the stimuli during naming tests does not necessarily mean that verbal behavior did not play an important role in the development of equivalence classes. The conditions present during naming tests were quite different from the ones present during training (Goyos, 1996). In addition, it is possible that participants lacked a history of verbal conditioning to use descriptive terms or follow specific instructions, which would yield false-negative results during tests that required participants to respond to questions such as “What is this?” (Stoddard & McIlvane, 1986). It has been shown that vocalizations made during training are very different from the ones made during naming tests (e.g., Lowe & Beasty, 1987). Unfortunately, none of these studies reported the spontaneous vocalizations that may have occurred during equivalence tests or conditional discrimination tasks.

Of interest is that most participants in the previous studies (especially Green, 1990) used unique words to describe each stimulus. This behavior may indicate that some form of naming/labeling was taking place. Lowe and Beasty (1987) observed
that when children between the ages of 2 and 5 yrs. were required to perform visual-
visual matching to sample tasks, they labeled individual stimuli such as a vertical
line, a horizontal line, a cross, a triangle, a green, and a red stimulus, as “up,”
“down,” “cross,” “triangle,” “green,” and “red,” respectively. The authors report that
the participants who passed the equivalence tests on their study were using the
stimuli’s names to relate them to one another during training (e.g., “up green,” “up
triangle”). Similar results were found with adult participants (e.g., Wulfert, Dougher,
& Greenway, 1991).

When presented with arbitrary visual-visual matching tasks, participants may
search for features in each stimulus that they can name and later relate to one another,
in some sort of intraverbal/autoclitic frame (Skinner, 1957). It is probable that this
form of naming (i.e., intraverbal naming) may be responsible, or at least facilitate, the
success on the development of visual-visual equivalence classes, as suggested by

Verbal Competence and Equivalence Classes

Hitherto, research suggests that verbal competence may be correlated with
performance on equivalence tests (Stromer & Mackay, 1996). Devany, Hayes, and
Nelson (1986), for example, taught a series of four related, arbitrary visual discrimi-
inations (AB, DE, AC, DF) to three groups of children matched on a conventional
measure of mental age: normally developing children, verbal children diagnosed with
mental retardation, and non-verbal children also diagnosed with mental retardation.
Their results indicated that the participants of both verbal groups (i.e., mentally
retarded and normally developing children) required fewer trials to complete conditional discrimination training. Also, verbal children performed significantly better than verbally disabled children on tests of equivalence (i.e., BC, EF, CB, FE). The authors reported no significant difference between the two verbal groups. Additionally, all participants were able to master the conditional discrimination tasks regardless of their verbal ability, however only the verbal children developed stimulus equivalence after training. Devany et al. concluded that in spite of the high correlation between verbal skills and equivalence class formation, it is not possible to affirm that language is a prerequisite for equivalence. It could be that the capacity to form equivalence is a prerequisite for language development, or that both verbal ability and equivalence are produced by a common environmental process. Nonetheless, their results support a close relation between verbal competence and the development of equivalence classes.

In a systematic replication, Barnes, McCullagh, and Keenan (1990) taught similar visual conditional discrimination tasks to six preschool children divided into three groups based on their verbal abilities: normally developing children, normally developing partially hearing children with mental ages above 2 yrs., and normally developing partially hearing children with mental ages below 2 yrs. In addition, they pre-tested all participants on receptive-discrimination and tact performances. For instance, pointing to a table when asked, “Where’s the table?” (receptive-discrimination) or saying, “table” when in the presence of the table and hearing the instruction

---

3 The authors refer to receptive-discrimination and tact trials as word-object and object-word sequences, respectively.
"What is this?" (tact). Regardless of their verbal ability, all children mastered the conditional discrimination tasks within approximately the same number of trials. Only one child from the third group (i.e., partially hearing children with mental ages below 2 yrs) did not develop stimulus equivalence. According to the authors, this participant was rated with the lowest verbal age (1.0-1.5 yrs) and was the only one not to perform during the receptive-discrimination and tact pretests. Barnes et al. (1990) suggested that the history of reinforcement that gives rise to receptive and tacting repertoires may also be responsible for the development of symmetry, transitivity, and thus equivalence. Interestingly, the data presented by Barnes et al. (1990) are consistent with the naming account. As mentioned before, according to Horne and Lowe (1996), both the receptive (i.e., listener) and tact (i.e., speaker) repertoires are necessary for the development of equivalence classes.

Eikeseth and Smith (1992) have also demonstrated a relation between verbal competence and equivalence by testing for the formation of equivalence classes in four children (3-5 yrs.) diagnosed with autism. In the first phase of the experiment, participants were exposed to a series of AB and AC auditory-visual conditional discrimination tasks and then tested on the BC and CB visual-visual emergent relations. After extensive training on conditional discrimination, all participants achieved mastery criterion, however none of them were able to perform the BC and CB emergent relations. In subsequent phases, the experimenters attempted to remedy the failure to form equivalence classes by teaching the participants to name each stimulus (a description of these phases will be presented in detail later).
Although the previous studies support a relation between language and the formation of equivalence classes, recent reports have presented evidence for derived relations in individuals with minimal verbal repertoires (e.g., Brady & McLean, 2000; Carr, Wilkinson, Blackman, & McIlvane, 2000) and nonhuman (thus, nonverbal) subjects (e.g., Barros, Galvão, & McIlvane, 2002; Santos, Barros, & Galvão, 2003; Shusterman & Kastak, 1993).

Brady and McLean (2000) investigated emergent relations in speakers and nonspeakers with developmental disabilities (26-45 yrs.), as well as in typically developing children (3-6 yrs.). Participants were taught to select specific lexigrams (C) in the presence of familiar objects (B), the BC task. All participants were already able to select the objects (B) in the presence of their dictated names (A), as previously assessed. After mastering the baseline conditional discrimination (BC), participants were tested on symmetry (CB), transitivity (AC) and naming (CA). Seven out of eleven participants showed emergence between dictated-name samples and lexigram comparisons (AC). Two of these participants who developed equivalence were nonspeakers and, although able to identify objects and lexigrams given their spoken names (listener behavior), were never capable of naming the stimuli (speaker behavior).

Carr et al. (2000) taught several auditory-visual (AB) and visual-visual (CB, DB) matching performances to three individuals with severe mental retardation (verbal age scores of approximately 2 yrs. each). During conditional discrimination training, the sample stimuli were gradually transformed from one set (e.g., set B) to
another (e.g., set C) over a series of trials. On subsequent tests of equivalence, participants successfully displayed emergent relations. In a second study, two participants with mental retardation and autism were exposed to purely visual MTS tasks. After extensive training on baseline conditional discrimination, one participant developed equivalence.

It has also been reported that under extensive training and fairly specific conditions, capuchin monkeys may be capable of demonstrating the prerequisites for equivalence, namely reflexivity (Barros et al., 2002), and symmetry (Santos et al., 2003). Although promising, positive equivalence tests with these subjects have not yet been obtained.

Shusterman and Kastak (1993) presented the most compelling case for equivalence with non-human subjects. In their study, a sea lion was first directly trained to respond to symmetric and transitive relations with 12 stimulus classes. Secondly, the sea lion was tested for equivalence with 18 novel stimulus classes. The authors reported that its performance on the equivalence test was significantly better than chance. Although impressive, the results presented by Shusterman and Kastak will need to be replicated before any strong claim regarding the capability of nonhumans to form equivalence can be made. To this point, most attempts to generate emergent relations with nonhuman subjects have failed (e.g., D'Amato, Salmon, Loukas, & Tomie, 1985; Dugdale & Lowe, 2000; Hogan & Zentall, 1977; Holmes, 1989).

---

4 *stimulus control shaping method*

5 For other published studies claiming to have shown equivalence with nonhumans see McIntire, Cleary, and Thompson (1987), and Vaughan (1988). Closer analyses of these studies demonstrated their failure to meet the necessary criteria for equivalence class formation (Hayes, 1989).
Given the current evidence on the relation between language and equivalence, attempts have been made to investigate the extent to which language may produce or facilitate the formation of stimulus classes (e.g. Eikeseth & Smith, 1992; Goyos, 2000; Lowe, Horne, Harris, & Randle, 2002; Mandell & Sheen, 1994; Randell & Remington, 1999).

As reported earlier, in the first phase of the study by Eikeseth and Smith (1992), all participants (i.e., children diagnosed with autism) failed the equivalence tests after mastering baseline conditional discrimination (MTS) tasks. In a subsequent phase, participants were taught to produce a common word in the presence of members of the first stimulus class (A1B1C1) and another common word for members of the second stimulus class (A2B2C2). Participants were then required to perform the baseline conditional discrimination tasks again while labeling/tacting the visual stimuli. When reassessed on stimulus equivalence, 3 out of 4 subjects presented positive results.

On a third phase, Eikeseth & Smith (1992) taught children to tact two novel sets of stimuli (D1E1F1 and D2E2F2) using a common word for each set. After learning to tact all stimuli, participants were assessed on the untrained MTS tasks (DE, DF, EF, FE). Two participants showed mastery of these conditional relations. The participants who failed the conditional discrimination tests were directly trained.
to respond to the DE and EF relations and then tested on DF and FE. Both were able to consistently respond to the untrained relations. Next, participants were trained to match novel (G and H) with already known stimuli (E and F). After training, one participant was able to perform the four untrained relations correctly (i.e., GD, HD, GH, and HG), another participant was able to perform three (i.e., HD, GH, HG), another participant performed two (i.e., GD and HD), and the other participant was able to perform only one relation (i.e., HD). When trained again on auditory-visual conditional discriminations with new unnamed stimuli (IJK), 2 out of 4 participants developed equivalence classes.

Results presented by Eikeseth and Smith (1992) strongly suggest that training common names/labels facilitates the development of stimulus equivalence that had previously failed to develop. Additionally, untrained conditional relations may emerge after stimuli are given the same name/label. Besides that, the relation of new unnamed stimuli to already named ones seemed to have produced some untrained conditional relations. These are undoubtedly important data to support the facilitative effects of naming in the development of arbitrary stimulus classes.

Goyos (2000) presented additional evidence for the importance of language in the formation of stimulus classes. Three preschool children (4-5 yrs.) were exposed to visual-visual MTS procedures using two class-specific reinforcers (red and yellow tokens). Participants were trained on identity (AA, BB, CC, DD) and arbitrary matching (AB and BC) as an attempt to produce two stimulus classes. Participants were videotaped throughout the experiment and their vocal-verbal behavior recorded.
After training, participants were tested for the formation of two three-member classes, ABC (BA, CB, AC, and CA relations), and for class expansion, ABCD (AD, DA, BD, DB, CD, and DC relations). After reaching training criteria, all participants showed the formation of the ABC classes. One participant however, failed to show expansion to ABCD classes. According to the authors, this participant was the only one who did not label the stimuli with token color names during the experimental sessions. When directly taught to label/tact the D stimuli with token color names, this participant was able to demonstrate class expansion.

In a second experiment, Goyos (2000) exposed two of the participants to identity matching procedures with the reinforcers reversed for the D stimuli. In other words, the D stimulus that was originally correlated with the yellow token during identity matching trials became, during the second experiment, correlated with the red token, and vice-versa. When tested on the two four member classes, both participants responded to all relations in a manner consistent with the classes established during Experiment 1, suggesting that reversing the reinforcers did not change stimulus class membership. When participants were taught to reverse the names/labels for the D stimuli, their responses reversed during tests. Thus, stimuli evoking the same verbal response ("yellow" or "red") became equivalent. The study by Goyos further suggests that stimulus classes may be established, and failures remedied by teaching common names to each stimulus.

In an interesting study conducted with undergraduate students, Mandell and Sheen (1994) demonstrated that the pronounceability of the stimuli may be an
important factor during the establishment of equivalence classes. Their participants were exposed to conditional discrimination tasks involving three groups of stimuli, namely phonologically correct pronounceable pseudowords (e.g., SNAMB), phonologically incorrect pseudowords (e.g., NSJBM), and punctuation marks (e.g., =]*>!). Participants who were exposed to pronounceable stimuli demonstrated more consistent equivalence performances than those exposed to the other stimuli. The authors also reported that participants exposed to unpronounceable stimuli tended to spontaneously label them with idiosyncratic names. In a second experiment, when participants were trained to label the phonologically incorrect pseudowords, their performances on equivalence tests were better than of other participants.

Randell and Remington (1999) conducted another study to show how naming/tacting stimuli may influence the formation of equivalence classes. Thirty verbally able adults were exposed to three different combinations of the same group of pictures. In one of the combinations, the classes to be trained were composed of pictures whose names rhymed with each other (e.g., bear, chair, hair, pear). For this “rhyme group,” the correct comparison always rhymed with the sample during MTS trials. For the other conditions (i.e., control), classes consisted of pictures whose names did not rhyme. In one combination, the names of the comparisons always rhymed with each other, but never with the sample. In the other combination, the name of one incorrect comparison always rhymed with the sample’s name but never with the other pictures presented during a trial. The authors reported that in comparison with the other participants, those in the rhyme group required fewer trials
to achieve mastery criterion during training. In addition, all participants in the rhyme group met criterion for equivalence, whereas only 5 out of 20 participants from the control groups demonstrated equivalence after extensive testing. Participants were also exposed to MTS tests with previously unseen pictures. During testing, almost all participants selected the comparisons that rhymed with the sample. When asked whether they had used names/labels to describe the pictures during the experiment, almost all participants reported having named the pictures.

Results from these studies suggest that verbal participants spontaneously label the experimental stimuli, and that labeling may improve their training and testing performances. It seems reasonable to suggest that the development of equivalence may be determined, or at least influenced by language. It could be possible that equivalence class formation in humans is verbally mediated (i.e., rule-governed) in contrast to being exclusively contingency-shaped (de Rose, 1996). Hence, if language/verbal behavior mediates the formation of stimulus classes, then its direct manipulation should generate or at least improve performance on equivalence tests/categorization.

Stimulus Categorization

It is believed that research on stimulus equivalence serves as a useful model for what cognitive psychologists refer to as “language categories” (e.g., Galizio, Stewart, & Pilgrim, 2001; Horne & Lowe, 1996; Pilgrim & Galizio, 1996; Stromer & Mackay, 1996). When stimuli become members of the same equivalence class, they may be said to be part of the same category. However, as mentioned by Pilgrim and
Galizio (1996), researchers outside of the field of behavior analysis may not necessarily agree that stimulus equivalence is an appropriate model to explain cognitive abilities such as categorization, mostly because research on stimulus equivalence cannot be easily applicable to naturally occurring behaviors. Cognitive psychologists, for instance, have been using measures such as stimulus sorting, and free recall tests to observe the same phenomena studied by behavior analysts via matching-to-sample procedures. Hence, the use of more traditional measures of emergent relations could benefit the field of behavior analysis by promoting a more direct dialogue with other disciplines (Pilgrim & Galizio, 1996). In addition, as suggested by Dymond and Rehfeldt (2001), the use of new, possibly more user-friendly measures of derived stimulus relations could encourage much-needed clinically relevant interventions based on equivalence.

Recently, Lowe, Horne, Harris, and Randle (2002) used a stimulus sorting procedure to evaluate the development of equivalence classes. In addition, they attempted to demonstrate that stimulus classes could be established without the use of conventional conditional discrimination procedures, but by teaching participants how to tact each potential member of the stimulus classes. In their first experiment, 9 typically developing children (2-4 yrs.) were taught a common tact response “zag” in the presence of three arbitrary shapes, and another “vek” in the presence of the other three shapes. Initially, the six shapes were separated into three training pairs, and participants were required to tact members of each pair (“zag” and “vek”), one pair at a time (i.e., pairwise training). Throughout training, the experimenters gradually
presented all stimuli at once, and participants were required to tact each shape in a predetermined random order (i.e., six stimulus trials). After reaching training criterion, participants were exposed to a categorization test. During each test trial, children were presented with all six shapes; the experimenter selected one of them and asked, “Look at this, can you give Teddy the others?” A correct category sort was scored if when presented with a “zag” stimulus, participants selected the remaining two “zag” comparisons or when presented with a “vek,” selected the remaining two “vek” comparisons. After training, 4 children passed the categorization test (passing criterion was 4 out of 9 correct sorts per common tact category). The remaining 5 children were exposed to another category test in which children were required to tact the sample before selecting the comparisons. During this second test, the experimenter asked the children, “What is this? Can you give me the others?” All five participants passed the second categorization test. Moreover, two of the participants who had passed the first test were exposed to similar tact training with 12 shapes (6 were novel). Both children were able to pass categorization tests, demonstrating class expansion via tact training.

In a second experiment, Lowe et al. (2002) attempted to control for the possibility that having all stimuli presented together during tact training had facilitated stimulus class formation. The authors suggested that a sequence of identical tact responses (e.g., “vek”) during the six-stimuli training may have occasioned the

---

6 Prior to the experiment, children were introduced to “Teddy”, which took the form of either a hand puppet or a large teddy bear placed on a chair next to the child. Throughout the experiment, children were instructed to help Teddy by answering the questions.
response "same" which could have played a role in the development of stimulus classes. During this second experiment, three participants (3-4 yrs.) were exposed to tact training in pairwise trials only. Once training criterion was reached, children were tested for categorization. Two children passed the first categorization test and one passed the second, in which a tact response to the sample was required. Additionally, experimenters evaluated the participants' listener repertoire by presenting each stimulus pair to the child and asking, "Where's the vek [zog]?" All three children performed with almost no errors.

Results from the study conducted by Lowe et al. (2002) showed that teaching common tact responses to each exemplar may develop arbitrary stimulus classes. Also, results from Experiment 2 suggest that stimulus classes can be developed without children ever seeing class exemplars grouped together prior to testing. Data presented by Lowe et al., strongly support the hypothesis that naming was the mechanism responsible for the development of stimulus classes. According to the naming explanation, tacting the sample, either overtly or covertly, produces a stimulus, which in turn controls responses of selecting the correct comparisons. Those children who failed the first categorization test, ended up succeeding when required to tact the sample (test 2), which would suggest that tacting the sample may be an important component for the ability to categorize. Plus, when probed for listener behavior, all children from Experiment 2 were able to select the stimuli even though they had never been trained to do so. These results support the required interdependence

\footnote{During Experiment 2 common tacts were "vek" and "zog".}
between speaker (tact) and listener (receptive-discrimination) behaviors, both of which comprise the naming repertoire.

**Purpose of the Current Study**

Given the recent literature on the role of verbal behavior in the development of equivalence classes, and also the availability of new measures of stimulus categorization, the current study was an attempt to extend the work of Lowe et al. (2002) by teaching unfamiliar stimulus classes to preschool children while manipulating their speaker and listener repertoires. It follows from Horne and Lowe's (1996) analysis that children may be able to sort arbitrary stimuli into categories without being directly trained if they can 1) tact all stimuli with the same category name (speaker behavior) and 2) appropriately select these stimuli when hearing their category name (listener behavior).

More specifically, the purpose of the current study was to assess whether children could categorize unfamiliar pictures when taught the listener and speaker behaviors separately. Two studies were conducted. In the first, children were directly taught to tact the category of different pictures (categories included maps of northern and southern U.S. states) and assessed to see whether they were able to (1) sort them into groups without any training and (2) select the correct pictures when asked for their categories. In the second study, children were directly taught to select pictures when asked for their categories and assessed to see whether they were able to (1) sort them into groups without direct training and also (2) correctly label/name them.
EXPERIMENT 1

Method

Participants, Setting, and Materials

Four typically developing children, Tom, Adam, Rita, and John participated in Experiment 1. Their ages were 63, 54, 50, and 37 months, respectively, at the start of the experiment. Sessions were conducted behind a divider in a quiet area at the children’s preschool. During each session, the child and the experimenters sat on the same side of a small table to avoid cueing by the experimenter. Sessions lasted approximately 10 min and were conducted twice a day.

Materials included six unfamiliar pictures (state maps) and six pictures of familiar objects. Familiar pictures were 5 cm by 5 cm color photographs of objects on a white background, obtained from the Picture This© CD-ROM. Each picture was encased in a transparent hard-plastic cover measuring 7.5 cm by 10 cm. Unfamiliar pictures were 6.5 cm by 6.5 cm laminated black-and-white maps on a white background, obtained online from the Microsoft© Clipart database. Each trial was presented by placing the pictures horizontally on the table. As each trial was completed, the experimenter rearranged the pictures on the table and presented the next trial. The unfamiliar pictures used during the experiment can be found in Appendix A.
Dependent Variables and Data Collection

The main dependent measures were the number/percentage of correct category sorts out of six trials. A correct sort was scored when the child, in the presence of the sample (e.g., N1\(^8\)) selected the two correct comparisons (e.g., N2 and N3) from a five-stimulus array (e.g., N2 N3 S1 S2 S3). Data on the number of correct category sorts were collected on site during probe sessions.

Additional dependent variables included (1) the percentage of correct stimulus selections in response to an instruction given by the experimenter (listener's repertoire) and (2) the participants' vocal-verbal descriptions of their performance.

Stimulus selection tests were administered prior to tact training and following categorization training. Participants' vocal-verbal behavior was also assessed following successful completion of category tests. The experimenter asked participants to answer open-ended questions regarding their ability to categorize the pictures. Vocal behavior produced by participants during sessions was also recorded. During all training and probe sessions a digital voice recorder was positioned close to the children in an unobtrusive manner. Data on vocal-verbal behavior were later transcribed from the recorder.

During tact training trials, vocal responses to the stimuli were scored as either correct or incorrect. Data on the number of correct tact responses were also collected on site.

\(^8\) N and S refer to the experimenter-defined categories North and South, respectively. Exemplars from each category were numbered (1, 2 and 3).
**Interobserver Agreement (IOA)**

A second observer was present for at least 37% of all category-sort probes, stimulus-selection probes, and tact training sessions to assess IOA. For each category-sort and tact trial, either an agreement or a disagreement between the two observers was scored. An agreement was calculated when both observers scored a trial in the same way. Point-by-point agreement was calculated by dividing the number of agreements by the sum of agreements and disagreements multiplied by 100%. Table 1 depicts the percentage of sessions in which IOA was calculated, the average IOA and the IOA range for all participants across experimental conditions.

<table>
<thead>
<tr>
<th></th>
<th>Tom</th>
<th>Adam</th>
<th>Rita</th>
<th>John</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pretraining</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessions</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>84%</td>
</tr>
<tr>
<td>Average</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>99%</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td></td>
<td>(87%-100%)</td>
<td></td>
</tr>
<tr>
<td><strong>Category Tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessions</td>
<td>66%</td>
<td>78%</td>
<td>66%</td>
<td>56%</td>
</tr>
<tr>
<td>Average</td>
<td>100%</td>
<td>99%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td>(83-100%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stimulus Selection Tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessions</td>
<td>50%</td>
<td>100%</td>
<td>100%</td>
<td>71%</td>
</tr>
<tr>
<td>Average</td>
<td>95%</td>
<td>100%</td>
<td>98%</td>
<td>100%</td>
</tr>
<tr>
<td>Range</td>
<td>(83%-100%)</td>
<td>(83%-100%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tact Training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessions</td>
<td>55%</td>
<td>72%</td>
<td>83%</td>
<td>37%</td>
</tr>
<tr>
<td>Average</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>100%</td>
</tr>
<tr>
<td>Range</td>
<td>(87%-100%)</td>
<td>(87%-100%)</td>
<td>(87%-100%)</td>
<td></td>
</tr>
</tbody>
</table>

**Experimental Design**

A nonconcurrent multiple-baseline design across two participants was used to assess the effects of the independent variable on the percentage of correct
categorizations. Two different categorization tests were administered after training criterion was achieved in experimental conditions. Categorization tests differed based on the instruction provided to the participant. Instruction/test 1 consisted of *look-at-sample match-to-others* and instruction/test 2 of *tact-sample matching-to-sample* (Lowe et al., 2002). A careful description of each of these instructions is given below. Criterion for failure on categorization tests is discussed below.

The order of conditions was as follows: Baseline categorization pretests, stimulus-selection-I pretest, tact-I training, categorization posttests, stimulus-selection-I posttest. For those participants who failed to categorize after tact-I training, the following additional conditions were presented: Stimulus-selection-2 pretest, tact-2 training, categorization posttests, stimulus-selection-posttest 2, and posttest interview. Interviews were conducted following successful categorization tests. Criteria for failure on training and testing conditions are discussed below.

Prior to training with unfamiliar stimuli/maps, participants were exposed to training conditions with familiar stimuli to control for the possibility that failure to categorize with arbitrary stimuli was due to lack of instructional control. Table 2 summarizes the order of experimental conditions for familiar and unfamiliar stimuli.
Table 2. Summary of Experimental Conditions – Experiment 1

<table>
<thead>
<tr>
<th>Phase</th>
<th>Task</th>
<th>Trials per block</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretraining</td>
<td>Tact-1 (pairs)</td>
<td>8</td>
<td>2 blocks at 100%</td>
</tr>
<tr>
<td>Pretraining</td>
<td>Tact-2 (all stimuli)</td>
<td>6</td>
<td>2 blocks at 100%</td>
</tr>
<tr>
<td>Pretraining</td>
<td>S. Selection-1</td>
<td>6</td>
<td>2 blocks at 100%</td>
</tr>
<tr>
<td>Pretraining</td>
<td>S. Selection-2</td>
<td>6</td>
<td>2 blocks at 100%</td>
</tr>
<tr>
<td>Pretraining</td>
<td>Categorization Pretraining</td>
<td>6</td>
<td>2 blocks/arrays at 100%</td>
</tr>
<tr>
<td>1</td>
<td>Categorization Test 1</td>
<td>6</td>
<td>3 blocks at 33% or below</td>
</tr>
<tr>
<td>2</td>
<td>Categorization Test 2</td>
<td>6</td>
<td>3 blocks at 33% or below</td>
</tr>
<tr>
<td>3</td>
<td>S. Selection-1 Pretest</td>
<td>6</td>
<td>3 blocks</td>
</tr>
<tr>
<td>4</td>
<td>Cat. Pretraining Review</td>
<td>4</td>
<td>1 block at 100%</td>
</tr>
<tr>
<td>5</td>
<td>Categorization Test 1</td>
<td>6</td>
<td>3 blocks at 33% or below</td>
</tr>
<tr>
<td>6</td>
<td>Categorization Test 2</td>
<td>6</td>
<td>3 blocks at 33% or below</td>
</tr>
<tr>
<td>7*</td>
<td>S. Selection-2 Posttest</td>
<td>6</td>
<td>3 blocks</td>
</tr>
<tr>
<td>8*</td>
<td>Tact-2 Training (all stimuli)</td>
<td>6</td>
<td>3 blocks at 100%</td>
</tr>
<tr>
<td>9*</td>
<td>Cat. Pretraining Review</td>
<td>4</td>
<td>1 block at 100%</td>
</tr>
<tr>
<td>10*</td>
<td>Categorization Test 1</td>
<td>6</td>
<td>3 blocks at 33% or below</td>
</tr>
<tr>
<td>11*</td>
<td>Categorization Test 2</td>
<td>6</td>
<td>3 blocks at 33% or below</td>
</tr>
<tr>
<td>**</td>
<td>Posttest Interview</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

*Children were exposed to these phases only if they failed the previous categorization test
**Children were exposed to this phase immediately after successful categorization tests

Pretraining

*Tact-1 pretraining.* The purpose of this condition was to train participants to tact the categories to which picture cards belonged. During tact-1 pretraining, the experimenter randomly separated the pictures into three training pairs. Each pair contained a picture of a bird and a picture of an insect. Pictures were placed horizontally on the table. The experimenter presented the first pair to the child and while pointing to one of the stimuli asked, "What is this?" If the child produced the correct tact response for the category (i.e., saying "bird" or saying "insect"), the experimenter
delivered praise (e.g., good job!). If the child responded incorrectly or did not respond, a correction procedure was implemented. The correction consisted of the experimenter pointing to the stimulus and saying, "This is a bird [insect], can you say it?" If the child responded to the echoic prompt by repeating the name of the category, the experimenter acknowledged the correct response by saying, "That's right", or "ok" and then moved on to the next trial. On any given trial, only one picture in the pair was targeted. Pictures of birds and insects were presented unsystematically either on the right or on the left of the child in 8-trial blocks (each picture was targeted twice on the left and twice on the right). The training criterion for each pair was correctly tacting all pictures in two consecutive 8-trial blocks. After training pair 1, the procedure was repeated for pairs 2 and 3. Tact-1 pretraining was completed when responding was at criterion for all pairs.

Tact-2 pretraining. During this condition, training was conducted with all pictures at once. The three pairs of pictures presented during Tact-1 pretraining were combined in a 6-picture array, and the tact for each picture was trained in this context. The pictures were randomly ordered in a row and placed in front of the child. The experimenter pointed to one of the pictures and asked, "What is this?" If the child produced the correct tact response, the experimenter delivered praise. If the child responded incorrectly or did not respond, a correction procedure was implemented. The correction consisted of the experimenter pointing to the stimulus and saying, "This is a bird [insect], can you say it?" If the child responded to the echoic prompt by repeating the name of the category, the experimenter acknowledged the correct
response and pointed to the next picture. After all pictures were targeted (six trials), the experimenter showed them in a different order, and started the training again. Tact-2 training was completed when the child correctly tacted all pictures in two consecutive arrays (6-trial blocks) with no corrections.

**Stimulus-selection-I pretraining**. Participants were trained to select a specific picture in response to a verbal stimulus specifying the category to which that picture belonged. The experimenter randomly separated the pictures into three groups. Each group contained a picture of a bird and a picture of an insect. Two additional pictures from different categories were used as distractors (e.g., a tool, a kitchen item). One distractor picture was randomly selected and added in a group as the third comparison during each trial. The three-choice matching procedure was chosen to prevent conditional discriminations from coming under the control of an incorrect comparison (i.e., reject relation), which typically results in failure on subsequent equivalence/transitivity tests (de Rose, 1996).

The experimenter presented the first three comparisons (group 1) to the child and asked, “Can you point to the bird [insect]?” If the child selected the correct picture, the experimenter provided verbal praise (e.g., “Good job”). If the child responded incorrectly or did not respond, a correction procedure was implemented. The correction consisted of the experimenter pointing to the correct picture and saying, “This is correct.” Training for each group was arranged in six-trial blocks in

---

9 Stimulus Selection refers to the behavior of selecting/pointing to visual stimuli in the presence of an instruction provided by the experimenter. Some may refer to this task as Receptive Discrimination. This task is being used as a measure of the listener’s repertoire.
which pictures of birds and insects were presented in a quasirandom order, each picture serving as the correct comparison once on the right, once on the middle, and once on the left of the child. The training criterion was achieved when a child correctly responded to all comparisons in two consecutive 6-trial blocks. After training group 1, the procedure was repeated for groups 2 and 3. Stimulus-selection-1 training was completed when responding was at criterion for all trained groups.

*Stimulus selection-2 pretraining.* During this condition, training was conducted with all pictures at once. The three pairs of pictures presented during stimulus selection-1 pretraining were combined in a six-picture array, and selection responses to each picture serving as a sample were trained in this context. Distractors were omitted during this phase. The pictures were randomly ordered in a row and placed in front of the child. The experimenter asked the child, “Can you give me all birds [insects]?” If the child selected the three correct comparisons (either the three pictures of birds or the three pictures of insects), the experimenter provided praise. Selecting the three correct comparisons was considered a correct trial. If the child selected a wrong comparison stimulus, the following correction procedure was implemented: The experimenter pointed to the correct comparisons and said, “These are the correct ones.” If the child selected some but not all three correct comparisons, the experimenter prompted the next response by saying, “Are there any more?” When the child selected the correct comparison(s) after the prompt, the experimenter delivered verbal praise. After a prompted response, the experimenter provided the prompt again to avoid the possibility that the number of pictures selected was being
controlled by the prompt. When the child did not respond after the prompt, the experimenter pointed to the remaining stimuli and said, “These are correct.” North and South trials were interspersed in a six-trial block. Training continued until the child performed at 100% in two consecutive blocks.

*Categorization training.* During this condition, participants were taught to respond to two different instructions that were later used to test categorization with unfamiliar pictures. The two instructions/tests were *look-at-sample match-to-others* and *tact-sample matching-to-sample*.

During the *look-at-sample-match-to-others* trials, the experimenter placed all six pictures in a predetermined location in front of the child, picked one at random asked the child, “Look at this, can you give me the others?” if the child selected the two correct comparisons (either the two other pictures of birds or the two other pictures of insects), the experimenter provided verbal praise. Selecting the two correct comparisons was considered a correct category-sort trial. If the child selected an incorrect comparison stimulus, the following correction procedure was implemented: The experimenter pointed to the comparison and said, “This is an insect [a bird]” If during category-sort trials, the child only selected one correct comparison, the experimenter prompted the next response by saying, “Are there any more?” If the child selected the correct remaining comparison after the prompt, the experimenter delivered verbal praise. After each prompted response, the experimenter provided the prompt again to avoid the possibility that the number of pictures selected was being controlled by the prompt. Trials of birds and insects were interspersed. Each block
consisted of six trials (each picture served as a sample once). Training continued until the child performed at 100% in two consecutive blocks.

During the tact-sample matching-to-sample trials, the experimenter asked, “What is this?” If the child produced the correct tact response by saying either “bird” or “insect”, the experimenter continued by asking, “Can you give me the others?” When the child selected the two correct comparisons (either the two other pictures of birds or the two other pictures of insects), the experimenter provided verbal praise. Selecting the two correct comparisons was considered a correct category-sort trial. If the child selected an incorrect comparison stimulus, the correction procedure was implemented as described above. Training continued until the child performed at 100% in two consecutive blocks.

Procedure

Categorization tests. This condition was designed to assess conditional relations among stimuli labeled as “north” and “south.” Categorization tests were administered during baseline (pretests) and immediately after tact training (posttests). Prior to categorization posttests, participants were exposed to categorization training with familiar pictures until they performed correctly on four consecutive trials. This review was conducted to guarantee that performance on categorization tests was under appropriate instructional control.

During test 1, the experimenter used the look-at-sample match-to-others instruction. All six pictures were placed in a predetermined location in front of the child, the experimenter picked one at random and asked, “Look at this. Can you give
me the others?” If the child selected the two correct comparisons, the experimenter recorded the trial as correct, although no consequences were provided. If during a trial, the child only selected one correct comparison, the experimenter prompted the next response by saying, “Are there any more?” After each prompted response, the experimenter provided the prompt again. If the child selected all five comparison stimuli, the trial was categorized as null (all other trials were considered valid). When a null trial occurred, the experimenter instructed the child by saying, “I don’t want all the pictures, just some of them.” After six consecutive null trials, this condition was terminated and categorization training with familiar pictures conducted until the child performed four consecutive correct trials. Testing was conducted in a 6-trial block, with each picture serving as a sample once. In each trial, the experimenter showed a different picture, until all pictures had served as a sample. Failing criterion was set at three consecutive blocks at 33% (2 out of 6 correct categorizations) or below. Passing criterion was set at three consecutive blocks at 66% (4 out of 6 correct categorizations) or above.

During test 2, the experimenter used the tact-sample match-to-others instruction. All conditions, including training criteria, were the same as in test 1 except for the instruction. The experimenter asked the child, “What is this? And waited for 5 s to see whether the child would tact the sample stimulus. Regardless of the child’s response, the experimenter proceeded by asking, “Can you give me the others?” Passing and failing criteria were the same as in test 1.
Stimulus selection-1 pretest/posttest. The purpose of this condition was to assess whether participants were able to select a specific picture in response to a verbal stimulus specifying the category to which that picture belonged. The experimenter randomly separated the pictures into three training groups. Two additional maps that were not part of training were used as distractors. One distractor picture was randomly selected and added as the third comparison during each trial. Trials included a map of a northern state, a map of a southern state and a distractor (map of a Canadian province) serving as comparisons. The experimenter presented the first group to the child and asked, “Point to the North [South]?”. Responses were recorded, but no consequences were provided. Testing for each group was arranged in 6-trial blocks in which pictures of northern and southern states were presented in a quasi-random order, each picture serving as the correct comparison once in the right, once in the middle, and once on the left of the child. After testing group 1, the procedure were repeated with groups 2, 3 and with the pictures reassigned to new groups (mixed).

Tact-1 training. The purpose of this condition was to train participants to tact the categories to which picture cards belonged. The condition was similar to tact pretraining described above. Pictures were separated into three training pairs (N1S1, N2S2, N3S3). The experimenter presented the first pair to the child, and while pointing to one of the stimuli asked, “What is this?” When the child produced the correct tact response the experimenter delivered praise. When the child responded incorrectly or did not respond, the experimenter pointed to the stimulus and said,
“This is North [South], can you say it?” If the child repeated the name of the category, the experimenter acknowledged the correct response and presented the next trial. Only one picture in the pair was targeted per trial. Training was arranged in 8-trial blocks in which pictures of north and south were presented unsystematically either on the right or on the left of the child. Training criterion for each pair was correctly tacking all pictures in three consecutive 8-trial blocks (each picture being targeted twice on the left and twice on the right of the child). After training pair 1, the procedure was repeated for pairs 2 and 3. When responding was at criterion for all trained pairs, the six stimuli were randomly reassigned to three new mixed training pairs. Training criterion for the new pairs was the same as for the initial pairs. When criterion was met, the probability of reinforcement was reduced from 100% to 50% (every other correct trial was reinforced). When criterion was met at 50%, reinforcement probability was reduced to 0% (correct trials were no longer reinforced). Once the child correctly tacked all pictures in three consecutive blocks, training was terminated and categorization tests administered. Probability of reinforcement was reduced so performance observed during subsequent tests (conducted under extinction) could not have been attributed to a sudden change in the rate of reinforcement.

*Stimulus-selection-2 pretest/posttest.* During this condition, participants were exposed to all pictures at once. Pictures presented during stimulus selection-1 training (with the exception of the distractors) were combined in a 6-picture array, and selection responses to each picture were assessed. The pictures were randomly ordered in a
row and placed in front of the child. The experimenter asked the child, “Can you give me all of the North [South]?” If the child selected the three correct comparisons, the trial was scored as correct, however no consequences were delivered. Testing was arranged in 6-trial block in which North and South questions were presented in a quasirandom order three times each. A minimum of two blocks (12 trials) were administered.

_Tact-2 training._ This condition was similar to tact-1 training described above. The pictures presented during this condition were combined in one array. The pictures were randomly ordered in a row and placed in front of the child. The experimenter pointed to each picture and asked, “What is this?” When the child produced the correct tact response, the experimenter delivered praise. When the child responded incorrectly or did not respond, the experimenter pointed to the stimulus and said, “This is North [South], can you say it?” If the child repeated the name of the category, the experimenter acknowledged the correct response and pointed to the next picture. After all pictures had been targeted (six trials), the experimenter showed them in a different order, and started the training again. Training criterion was achieved when the child correctly tacted all pictures in three consecutive 6-trial blocks without any prompts. When criterion was met, the probability of reinforcement was reduced from 100% to 50%. When criterion was met at 50%, reinforcement probability was reduced to 0%. Once the child correctly tacted all pictures in three consecutive arrays without any prompts, training was terminated and categorization tests administered.
Posttest interviews. Following the completion of successful categorization tests, the experimenter exposed the child to an additional categorization test and asked questions while the child was selecting the pictures. These questions were: “How did you do that?” “How did you know that these go together?” Vocal responses were digitally recorded for later analysis.

Independent Variable Integrity (IVI)

Independent-variable integrity (IVI) was assessed for at least 34% of tact-1 and tact-2 training sessions by an independent observer. Sessions used in the calculation of IVI were randomly selected. IVI was calculated by dividing the number of correctly implemented trials by the total number of trials conducted by the experimenter. Trials were scored as entirely correct or incorrect based on the following categories: (1) Reinforcement – praise had to be delivered for all correct trials during the 100% praise condition, for every other correct trial during the 50% praise condition, and not delivered during the 0% praise condition; (2) Correction – the correction procedure had to be correctly implemented if the trial was marked as incorrect. Table 3 depicts the percentage of sessions in which IVI was calculated, the average IVI score and the IVI range for all participants across experimental conditions.
Table 3. Average IVI Scores for Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Tom</th>
<th>Adam</th>
<th>Rita</th>
<th>John</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tact-1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessions</td>
<td>37%</td>
<td>67%</td>
<td>78%</td>
<td>34%</td>
</tr>
<tr>
<td>Average</td>
<td>100%</td>
<td>99%</td>
<td>99%</td>
<td>100%</td>
</tr>
<tr>
<td>Range</td>
<td>(87-100%)</td>
<td>(87%-100%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tact-2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessions</td>
<td>-</td>
<td>56%</td>
<td>-</td>
<td>64%</td>
</tr>
<tr>
<td>Average</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
</tbody>
</table>

Results

**Familiar Pictures**

Table 4 shows the number of trials it took each participant to reach criterion during pretraining. The minimum number of trials required during this condition was 148. Data show that the number of trials required for reaching criterion on pretraining was inversely correlated with participants’ age. The older the child, the fewer trials it took her to reach criterion. This speed of acquisition may be related to participants’ verbal abilities, although no direct measures of their verbal abilities were ever taken.

Table 4. Number of Trials to Criterion During Pretraining

<table>
<thead>
<tr>
<th></th>
<th>Age (months)</th>
<th>Trials to Criterion (Minimum =148)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tom</td>
<td>63</td>
<td>182</td>
</tr>
<tr>
<td>Adam</td>
<td>54</td>
<td>196</td>
</tr>
<tr>
<td>Rita</td>
<td>50</td>
<td>222</td>
</tr>
<tr>
<td>John</td>
<td>37</td>
<td>250</td>
</tr>
</tbody>
</table>

All participants made relatively few errors during pretraining, suggesting that they may have already been exposed to similar pictures and/or tasks in the past. The
fast acquisition of tacts, selections, and categorizations during pretraining demonstrated that the instructions and consequences provided were adequate to teach and test these repertoires.

Unfamiliar Pictures (Maps)

*Tom and Adam.* Figure 1 depicts data on correct category sorts for Tom (upper panel) and Adam (lower panel). Filled circles represent the percentage or correct categorizations out of six trials. Bars represent the percentage of correct tacts to the sample. For both categorization tests 1 and 2, all six pictures were placed in a predetermined location in front of the child. The experimenter picked one at random and asked, "Look at this. Can you give me the others?" for test 1, and "What is this? Can you give me the others?" for test 2. Therefore, tacts to the sample were only required during test 2.

During categorization pretest 1, Tom never selected the two correct comparisons in the presence of a sample, either maps of the Northern or Southern hemispheres. During pretest 2, he selected the correct pictures in 1 out of 6 possible trials (16%) during the 6th block/probe, suggesting that he could not categorize these pictures. Immediately after baseline, tact-1 training (pairwise) was implemented. Once criterion on tact-1 training was reached, Tom was exposed to categorization posttest 1 and was able to select the two correct comparisons in the presence of a sample in almost all trials. On probes 8 and 9, Tom selected the correct comparisons 100% of the time, and on probes 10 and 11, he selected the correct comparisons in 5 out of 6 trials (83%). When required to tact the sample (test 2), Tom continued to
select the correct comparisons. Additionally, he was able to correctly tact the samples 100% of the time, suggesting that the pairwise training was successful in producing reliable tacts.

Figure 1. Percentage of Correct Categorizations for Tom (upper panel) and Adam (lower panel).
Adam was exposed to additional categorization pretest probes (Fig. 1, lower panel). This served to control for the possibility that continuous exposure to testing conditions would account for any observed increase in the percentage of category sorts. During pretest 1, Adam selected the correct comparisons in 1 out of 6 trials on the first probe (16%) and in 2 out of 6 trials on the fourth probe (33%), meeting the criterion for failure. During test 2, he selected the correct comparisons only once during probe 8, thus failing the test. After tact-1 training (pairwise), Adam was not able to categorize the pictures into North or South in either of the categorization posttest conditions. Although he was capable of tacting some of the samples when required (posttest 2), his performance was not as accurate as expected, given that he had just met criterion (three consecutive blocks at 100% with no reinforcement) on tact-1 training. During probes 15 and 17, he incorrectly tacted 1 out of 6 samples, and during probe 16, he incorrectly tacted 2 out of 6 samples. Because Adam was unable to categorize following pairwise training (tact-1), a tact training procedure with all pictures presented at once was implemented (tact-2).

After meeting criterion on tact-2 training (3 consecutive blocks at 100% correct with no reinforcement), three categorization probes under testing condition 1 were administered. Adam was able to select the correct comparisons 100% of the time during all three probes. Under categorization posttest 2, Adam was still able to successfully categorize the pictures. He selected the correct comparisons in 5 out of 6 trials on probes 21 and 22, and in all trials on probe 23. His tacts to the sample were also very accurate, more so than when tacts were assessed following tact-1 training
(probes 15, 16, 17), suggesting that the additional tact training with all pictures presented at once increased response accuracy.

Tom and Adam were also exposed to stimulus selection-1 tests. These tests were conducted always after categorization tests, during baseline, and after tact training. During stimulus selection tests, participants were exposed to a series of auditory-visual conditional discrimination tasks in which they were required to point to a correct comparison in the presence of an auditory stimulus presented by the experimenter; the question, “point to north [south].” Comparisons consisted of one north picture, one south picture, and a distractor (picture of a Canadian province). Distractors never served as positive comparisons.

Figure 2 displays data on stimulus selection-1 tests for Tom. During pretest, Tom selected the correct comparison 33%, 16%, 16%, and 33% of the time when presented with the targets N1S1, N2S2, N3S3, and Mixed, respectively (gray bars). These data show that initially, Tom was not able to reliably select the correct comparison in the presence of the question, suggesting the absence of listener behavior in regards to the training stimuli. After tact-1 training, Tom selected the correct comparison 83%, 100%, 100%, and 100% of the time when presented with the targets N1S1, N2S2, N3S3, and Mixed, respectively (black bars). This suggests that tact-1 training generated accurate stimulus selections. Thus, for Tom, a complete transfer between speaker (trained) and listener (untrained) repertoires was seen.
Figure 2. Percentage of Correct Responses During Stimulus Selection-1 Test for Tom.

Figure 3 displays data from Adam’s stimulus selection-1 test. During pretest, Adam selected the correct comparison 50%, 0%, 16%, and 0% of the time when presented with the targets N1S1, N2S2, N3S3, and Mixed, respectively (gray bars). Adam was exposed to an additional block of N1S1, during which he was never able to select the correct comparison (0%), suggesting that his previous performance (50% correct) was due to chance. Pretest data show that, similarly to Tom, Adam was not able to select a correct comparison in the presence of a question. After tact-1 training, Adam selected the correct comparison 50%, 50%, 16%, and 66% of the time when presented with the targets N1S1, N2S2, N3S3, and Mixed, respectively (black bars). In contrast with the results obtained with Tom, tact-1 training did not generate accurate stimulus selections for Adam. Although an increase from pretest to posttest
was observed, especially when target stimuli were mixed (from 0% to 66%), a complete transfer between speaker (trained) and listener (untrained) repertoires was not seen. The implications of these results will be discussed later.

Figure 3. Percentage of Correct Responses During Stimulus Selection-1 Test for Adam.

Since Adam had to be exposed to tact-2 training, stimulus-selection-2 pretests and posttests were conducted. During stimulus-selection-2, Adam was presented with all pictures at once (with the exception of the distractors) in a 6-picture array. The behavior of selecting the correct pictures as a result of the question, “Can you give me all of the North [South]?” was assessed.

Figure 4 displays the data from stimulus-selection 2 tests for Adam. During pretest, Adam was only able to select the correct comparisons in the presence of the question in 4 out of 12 trials (33%), supporting the assumption that tact-1 training did
not produce reliable selection responses. After tact-2 training, Adam was capable of selecting the correct comparisons during stimulus-selection-2 posttests in 11 out of 12 trials (91%). This suggests that the additional tact training with all pictures presented at once facilitated the development of listener behavior with regards to the target stimuli.

![Graph showing percentage of correct responses for Adam](image)

Figure 4. Percentage of Correct Responses During Stimulus Selection-2 Test for Adam.

*Rita and John.* Figure 5 depicts data on correct category sorts for Rita (upper panel) and John (lower panel). Filled circles represent the percentage of correct categorizations out of six trials and bars represent the percentage of correct tacts to the sample.

During categorization pretests 1 and 2, Rita never selected the two correct comparisons in the presence of a sample (Probes 1-6). After reaching criterion on
tact-1 training, Rita was exposed to three probes of categorization posttest 1 (probes 7-9). She was able to select the two correct comparisons in the presence of a sample 100% of the time, even though she was never required to tact the sample. Rita’s accurate performance was maintained when categorization posttest 2 was implemented. In addition, Rita was able to correctly tact the samples 100% of the time during probes 10 and 12, and in 5 out of 6 trials (83%) during probe 11. Interestingly, Rita still selected the correct comparisons during probe 11, even though she incorrectly tacted the sample.

Categorization data for John are also displayed on Figure 5 (lower panel). During categorization pretest 1, John selected the correct comparisons in only 1 out of 6 trials (16%) during probe 5. In the subsequent probes (pretest 2), John never selected the correct comparisons in the presence of a sample, suggesting that he was not able to categorize the pictures into North and South prior to the intervention.

As seen on probes 11-13 (Figure 5, lower panel), pairwise tact training (tact-1) did not generate correct category sorts for John. His performance was at 0% through posttest 1 conducted immediately after tact-1 training. Even though John was not required to tact the samples during posttest 1, he spontaneously tacted one sample on probe 11. When required to tact (posttest 2) the samples, John did it 100% of the time during all three probes (probes 14-16). Although able to tact the samples, his categorization performance was still inaccurate. John was able to correctly categorize the pictures only in 1 out of 6 trials (16%) during all three posttest-2 probes. Because John was unable to categorize after tact-1 training, tact-2 training was implemented.
Figure 5. Percentage of Correct Categorizations for Rita (upper panel) and John (lower panel).

After meeting criterion on tact-2 training, three categorization probes under testing condition 1 were administered. John was able to select the correct comparisons 0%, 33%, and 16% of the time on probes 17, 18 and 19, respectively. Although his categorization performance on probe 18 was slightly better than before, John’s performance still met failing criterion. On categorization posttest 2, John was able to correctly categorize 100% of the time on probe 22 and 66% (4 out of 6 trials) on probes 20, 21 and 23. His tact performance during categorization posttest-2 after tact-
2 training was slightly worse than during posttest 2 after tact-1 training. John incorrectly tacted 1 out of 6 samples on probes 21, 22, and 23, which may have contributed to the incorrect categorizations observed. Data obtained with John suggest that he was only able to categorize after additional tact training (tact-2) was conducted, and when he was required to tact the sample (test 2).

Figure 6 depicts the results for the stimulus selection-1 test for Rita. During pretest, Rita selected the correct comparison 50%, 16%, 33%, and 16% of the times when presented with the targets N1S1, N2S2, N3S3, and Mixed, respectively (gray bars). Rita was exposed to an additional block of N1S1, in which she selected the correct comparison 33% of the time, suggesting that her previous performance (50% correct) was due to chance. These data suggest the absence of listener behavior in regards to the training stimuli. After tact-1 training, Rita selected the correct comparisons 100%, 83%, 66%, and 100% of the time when presented with the targets N1S1, N2S2, N3S3, and Mixed, respectively (black bars). Thus, a high degree of transfer between speaker (trained) and listener (untrained) repertoires was seen.

Figure 7 displays data from John’s stimulus selection-1 test. During pretest, John selected the correct comparison 50%, 16%, 0%, and 0% of the times when presented with the targets N1S1, N2S2, N3S3, and Mixed targets, respectively (gray bars). After tact-1 training, John selected the correct comparisons 83%, 66%, 50%, and 50% of the time when presented with the targets N1S1, N2S2, N3S3, and Mixed, respectively (black bars). Although some transfer from speaker to listener repertoires was seen, John’s performance on the stimulus selection posttest was far from accurate.
Figure 6. Percentage of Correct Responses During Stimulus Selection-1 Test for Rita.

Figure 7. Percentage of Correct Responses During Stimulus Selection-1 Test for John.
Because John had to be exposed to tact-2 training, the stimulus-selection-2 test was also conducted. Figure 8 displays the data for stimulus-selection 2 tests for John. Prior to tact-2 training, John was able to select the correct comparisons 75% of the time (9 out of 12 trials). This suggests that that tact-1 training alone did not produce accurate listener/selection responses. After tact-2 training, John was capable of selecting the same number of correct comparisons (9 out of 12), suggesting that tact-2 training had little impact on his already acquired listener repertoire.

Figure 8. Percentage of Correct Responses During Stimulus Selection-2 Test for John.

*Trials to criterion.* Table 5 displays the number of tact trials each participant was exposed to until training criterion was reached. Criterion consisted of three consecutive blocks (8 trials for tact 1, and 6 trials for tact 2) with no errors.
As with the familiar pictures, the number of trials required for reaching criterion on tact-1 was inversely correlated with the participants’ age. The older the child, the fewer trials it took her to reach criterion on tact-1. The same cannot be said about tact-2. John, the youngest participant (37 months) required fewer trials to master tact-2 training than his older peer, Adam (54 months). As discussed earlier, only the participants who failed to categorize the maps after tact-1 training were exposed to tact-2 training.

**Vocal-Verbal Behavior**

Participants’ vocal behaviors regarding the task were digitally recorded during all categorization test sessions and later transcribed. Observers also documented all vocalizations related to the task that occurred during sessions. A posttest interview was also conducted after successful categorization tests. A transcription of these interviews is attached (Appendix B).

During sessions, Tom tacted the stimuli as the experimenter was placing them on the table on several occasions before the session began, especially during tact training. Adam tacted one of the samples (N1) as “book” and another (S3) as “tree” during baseline categorization tests. During stimulus selection-I pretest, Adam tacted the sample S3 as “chicken bone,” and during tact-1 training, tacted the sample N3 as
“cat.” Although not required to emit any tact responses, Adam correctly tacted the samples on two trials and the comparisons on seven trials during categorization-1 posttest. Comparisons were also correctly tacted during categorization-2 posttest, even though, Adam was only required to tact the samples.

Rita and John did not emit many vocalizations during the sessions. Rita correctly tacted a sample once during categorization posttest 1 and John correctly tacted different samples three times during categorization posttest 1. John also referred to some pictures as “north” and “south” during stimulus-selection-1 posttest, mostly in response to the experimenter’s instruction, “point to north [south],” suggesting that John’s behavior consisted of an echoic response.

During posttest interviews, only Adam was able to justify his selection by naming the category to which the pictures belonged. When asked by the experimenter, “Why did you give me this one?” after successfully categorizing the “north” pictures, Adam said, “Because it is north.” Both Tom and John justified their selections by affirming that the stimuli selected “matched” with the sample. Tom also said, “They go together” when asked, “Why did you pick these two pictures?” Of note, the experimenters never used the word “match” during sessions.

Discussion

Data obtained on Experiment 1 replicate those obtained by Lowe et al. (2002) in that teaching a common tact response to each unfamiliar picture established those pictures as a category or a class. All participants were able to categorize after tact training, although only Tom and Rita passed categorization tests immediately after
pairwise tact training (tact-1). These two participants made almost no errors during stimulus selection-1 posttest, suggesting that for Tom and Rita, tact training generated accurate selection responses.

Both Adam and John required additional tact training with all pictures present (tact-2 training) before demonstrating stimulus class formation. According to the naming hypothesis (Horne & Lowe, 1996), the ability to categorize depends on the combination of the listener and speaker behaviors. In Adam’s case, his selection/listener performance was not improved after pairwise tact-1 training, suggesting that the absence of the listener repertoire may have contributed to his failure to categorize. After being exposed to tact-2 training, Adam passed categorization tests, as well as selected the pictures accurately, suggesting that his categorization skills may have depended upon listener and speaker behaviors with regards to the target stimuli.

John, who also failed categorization tests, showed inaccurate listener behavior after tact-1 training. Similarly to Adam, John passed categorization tests after being trained to tact when all pictures were on the table (tact 2), however this additional training did not seem to have improved his selection/listener performance. John passed categorization tests even though he was still making errors during stimulus selection tests.

Results obtained with John seem to deviate from predictions based on the naming hypothesis, since he was able to categorize without accurately selecting the pictures during stimulus selection-2 posttest. A possible interpretation of John’s results is that tact-2 training facilitated spontaneous tacts to the comparisons in
addition to the already required tacts to the sample. During tact-2 training John learned how to tact each picture when they were all present at the same time. When exposed to categorization tests, John may have not only tacted the sample held by the experimenter, but before selecting the pictures, he may have also tacted the comparisons accurately, as north or south.

According to Lowenkron (1996, 1997), categorization occurs if participants not only tact the sample, but also the comparisons. He suggests that the categorization response is jointly controlled by the response product of tacting and self-echoing the sample plus the response product of tacting the comparison. In other words, John would have had to tact the sample by saying (covertly or overtly) "north," and covertly rehearsed this topography (a self-echoic). He would then have had to look down and tact each of the comparisons. When the auditory stimulus produced by tacting a comparison matched the auditory stimulus produced by the self-echoic rehearsal, the correct selection/categorization occurred. Although there were no data collected on John's tacts to the comparison stimuli, it is possible that he may have covertly tacted the comparisons during categorization posttests, especially after extensive tact training when all stimuli were presented together (tact 2). Given that another participant (Adam) overtly tacted comparisons on several occasions during categorization-1 posttest, it would not be unreasonable to assume that John was engaging in similar behavior at the subvocal level.

The data collected on participants' vocal-verbal behavior adds support to the idea that participants were verbally relating the stimuli with each other during the
tasks. Adam, for example, tacted some of the pictures as familiar objects (e.g., a tree), suggesting the development of intraverbal naming (Horne & Lowe, 1996). This form of intraverbal naming could have played an important role had the study used purely visual conditional discrimination tasks.

During posttest interviews, only Adam was able to justify his selection by affirming that the pictures belonged to a certain category because they had the same name either “north” or “south.” The fact that other participants were not able to justify their selection may not necessarily mean that they did not name the stimuli or that verbal behavior did not play an important role in the development of categories. As suggested earlier, the conditions present during posttest interviews were quite different from the ones during training. This difference could have suppressed verbal responding that would have otherwise occurred during training. Additionally, it is possible that participants lacked a history of verbal conditioning to answer questions such as, “Why did you select this picture?” or, “Why do these go together?”

In summary, data obtained on Experiment 1 support Home and Lowe’s (1996) assumption that naming may have led to, or at least facilitated, stimulus class formation. Participants who failed to demonstrate name-object bi-directionality (e.g., learned to tact but not to select the pictures) did not succeed on categorization tests. Although a causal relation between speaker-listener repertoires and the ability to categorize cannot be directly drawn from the current data, there seems to be some evidence that the absence of either listener or speaker repertoires yields failures in categorization tests. Experiment 2 was designed to further investigate this hypothesis.
by training participants to respond as listeners (stimulus selection) and observe whether increases in speaker behaviors (tacts) would accompany increases in categorization performances.
EXPERIMENT 2

Method

Participants, Setting, and Materials

Four typically developing children, James, Maria, David, and Pam participated in Experiment 2. Their ages were 48, 53, 41, and 56 months, respectively, at the start of the experiment. Sessions were conducted behind a divider in a quiet area at the children’s preschool. During each session, the child and the experimenters sat on the same side of a small table to avoid cueing by the experimenter. Sessions lasted for approximately 10 min and were conducted twice a day. Materials were the same as describe in Experiment 1.

Dependent Variables and Data Collection

The main dependent measure was the number/percentage of correct category sorts. A correct sort was scored when the child, in the presence of the sample (e.g., N1) selected the two correct comparisons (e.g., N2 and N3) from a 5-stimulus array (e.g., N2 S1 N3 S3 S2). Data on the number of correct category sorts was collected on site during probe sessions.

Additional dependent variables included (1) the percentage of correct tacts produced by participants in the presence of unfamiliar stimuli used during training (speaker repertoire), and (2) the participants’ vocal-verbal descriptions of their performance. Tact probes were administrated prior to stimulus-selection training and following categorization tests. Participants’ vocal-verbal behavior was also assessed.
following successful completion of category tests. The experimenter asked participants to answer open-ended questions regarding their ability to categorize the pictures. Vocal-verbal behavior produced by participants during sessions was also recorded. During all training and probe sessions a digital voice recorder was positioned close to the children in an unobtrusive manner. Data on vocal-verbal behavior was later transcribed from the recorder.

**Interobserver Agreement (IOA)**

A second observer was present for at least 29% of all category-sort probes, tact probes, and stimulus-selection training sessions to calculate point-by-point IOA. For each trial, either an agreement or a disagreement between the two observers was scored. An agreement was calculated when both observers scored a trial in the same way. Table 6 depicts the percentage of sessions in which IOA was calculated, the average IOA, and the IOA range for all participants across experimental conditions.

<table>
<thead>
<tr>
<th>Table 6. Interobserver Agreement Results for Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pretraining</strong></td>
</tr>
<tr>
<td>Sessions</td>
</tr>
<tr>
<td>Average Range</td>
</tr>
<tr>
<td><strong>Category Tests</strong></td>
</tr>
<tr>
<td>Sessions</td>
</tr>
<tr>
<td>Average Range</td>
</tr>
<tr>
<td><strong>Tact Tests</strong></td>
</tr>
<tr>
<td>Sessions</td>
</tr>
<tr>
<td>Average Range</td>
</tr>
<tr>
<td><strong>Stimulus Selection Training</strong></td>
</tr>
<tr>
<td>Sessions</td>
</tr>
<tr>
<td>Average Range</td>
</tr>
</tbody>
</table>
Experimental Design

A nonconcurrent multiple-baseline design across two participants was used to assess the effects of the independent variable on the percentage of correct categorizations. Two different categorization tests were administered after training criterion was achieved on experimental conditions. Categorization tests differed based on the instruction provided to the participant. Test 1 consisted of look-at-sample match-to-others and test 2 of tact-sample matching-to-sample. The criterion for failure is discussed later.

The order of conditions was as follows: Baseline categorization pretests, tact-1 pretest, stimulus selection-1 training, categorization posttests, and tact-1 posttest. For the participant who failed to categorize after stimulus selection-1 training, the following additional conditions were presented: Tact-2 pretest, stimulus selection-2 training, categorization posttests, and tact-2 posttest. Interviews were conducted after successful categorization tests. Criteria for failure on training and testing conditions are discussed below. Prior to training with unfamiliar stimuli, participants were exposed to pretraining conditions with familiar stimuli, as described in Experiment 1 to control for the possibility that failure to categorize with arbitrary stimuli was due to lack of instructional control. Table 7 summarizes the order of experimental conditions for familiar and unfamiliar stimuli.

Pretraining

Pretraining was conducted exactly as described in Experiment 1.
Table 7. Summary of Experimental Conditions – Experiment 2

<table>
<thead>
<tr>
<th>Phase</th>
<th>Task</th>
<th>Trials per block</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretraining</td>
<td>Tact-1 (pairs)</td>
<td>8</td>
<td>2 blocks at 100%</td>
</tr>
<tr>
<td>Pretraining</td>
<td>Tact-2 (all stimuli)</td>
<td>6</td>
<td>2 blocks at 100%</td>
</tr>
<tr>
<td>Pretraining</td>
<td>S. Selection-1</td>
<td>6</td>
<td>2 blocks at 100%</td>
</tr>
<tr>
<td>Pretraining</td>
<td>S. Selection-2</td>
<td>6</td>
<td>2 blocks at 100%</td>
</tr>
<tr>
<td>Pretraining</td>
<td>Categorization Pretraining</td>
<td>6</td>
<td>2 blocks at 100%</td>
</tr>
<tr>
<td>1</td>
<td>Categorization Test 1</td>
<td>6</td>
<td>3 blocks at 33% or below</td>
</tr>
<tr>
<td></td>
<td>Categorization Test 2</td>
<td>6</td>
<td>3 blocks at 33% or below</td>
</tr>
<tr>
<td>2</td>
<td>Tact-1 Pretest</td>
<td>8</td>
<td>4 blocks</td>
</tr>
<tr>
<td>3</td>
<td>S. Selection-1 Training</td>
<td>6</td>
<td>3 blocks at 100%</td>
</tr>
<tr>
<td>4</td>
<td>Cat. Pretraining Review</td>
<td>4</td>
<td>1 block at 100%</td>
</tr>
<tr>
<td>5</td>
<td>Categorization Test 1</td>
<td>6</td>
<td>3 blocks at 33% or below</td>
</tr>
<tr>
<td></td>
<td>Categorization Test 2</td>
<td>6</td>
<td>3 blocks at 33% or below</td>
</tr>
<tr>
<td>6</td>
<td>Tact-1 Posttest</td>
<td>8</td>
<td>4 blocks</td>
</tr>
<tr>
<td>7*</td>
<td>Tact-2 Pretest</td>
<td>6</td>
<td>2 blocks</td>
</tr>
<tr>
<td>8*</td>
<td>S. Selection-2 Training</td>
<td>6</td>
<td>3 blocks at 100%</td>
</tr>
<tr>
<td>9*</td>
<td>Cat. Pretraining Review</td>
<td>4</td>
<td>1 block at 100%</td>
</tr>
<tr>
<td>10*</td>
<td>Categorization Test 1</td>
<td>6</td>
<td>3 blocks at 33% or below</td>
</tr>
<tr>
<td></td>
<td>Categorization Test 2</td>
<td>6</td>
<td>3 blocks at 33% or below</td>
</tr>
<tr>
<td>11*</td>
<td>Tact-2 Posttest</td>
<td>6</td>
<td>2 blocks</td>
</tr>
<tr>
<td>**</td>
<td>Posttest Interview</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

*Children were exposed to these phases only if they failed the previous categorization test
**Children were exposed to this phase immediately after successful categorization tests

Procedure

* **Categorization tests.** Categorization tests were conducted exactly as described in Experiment 1.

* **Tact-1 pretest/posttest.** The purpose of this condition was to assess whether participants were able to tact the categories to which the picture cards belonged. Pictures were separated into three training pairs (N1S1, N2S2, N3S3). The experimenter presented the first pair to the child and while pointing to one of the stimuli asked, “What is this?” If the child produced the correct tact response, the
experimenter scored the trial as correct. If the child produced an incorrect tact or did not respond for 10 s, the next trial was presented. Only one picture in the pair was targeted per trial. Sessions were arranged in 8-trial blocks in which the locations of pictures of North and South were presented unsystematically either on the right or on the left of the child.

**Stimulus-selection-1 training.** The purpose of this condition was to teach participants to select a specific picture in response to a verbal stimulus specifying the category to which that picture belonged. The experimenter randomly separated the pictures into three training groups. Three additional maps of Canadian provinces that were not part of the training were used as distractors. One distractor picture was randomly selected and added as the third comparison during each trial. Trials included a map of a northern state, a map of a southern state, and a distractor (map of a Canadian province) serving as comparisons. The experimenter presented the first group to the child and asked, “Can you point to the North [South]?” If the child pointed to the correct picture, the experimenter delivered praise (e.g., “good job”). If the child selected the incorrect picture or did not respond within 10 s, a correction procedure was implemented. The correction consisted of the experimenter pointing to the correct stimulus and saying, “This is correct!” If the child pointed to the correct picture after the prompt, the experimenter acknowledged the correct response by saying, “That’s right” and then presented the next trial. Sessions were arranged in 6-trial blocks in which pictures of northern and southern states were presented in a quasirandom order, with each picture serving as the correct comparison once in the
right, once in the middle, and once on the left of the child. After training group 1, the procedure was repeated with groups 2 and 3. The training criterion was three consecutive blocks at 100%. When responding was at criterion for all trained groups, the stimuli were randomly reassigned to three new mixed groups. The training criterion for the new groups was the same as for the initial groups. Once criterion was met, the probability of reinforcement was reduced from 100% to 50%. When criterion was met at 50% reinforcement probability was reduced to 0%. This condition was terminated when criterion was met with no reinforcement.

*Tact-2 pretest/posttest.* This condition was similar to tact-1 pretest/posttest described above. The pictures presented during this condition were combined in one array. The pictures were randomly ordered in a row and placed in front of the child. The experimenter pointed to each picture in a pre-specified order and asked, “What is this?” If the child produced the correct tact response, the experimenter scored the trial as correct. No consequences were provided during this condition. After all pictures have been targeted (6 trials), the experimenter showed them in a different order. At least two 6-trial blocks were presented.

*Stimulus-selection-2 training (all stimuli).* During this condition, pictures presented during stimulus-selection-1 training (with the exception of the distractors) were combined in a 6-picture array, and selecting (not pointing) responses were assessed. The pictures were randomly ordered in a row and placed in front of the child. The experimenter asked the child, “Can you give me all of the North [South]?” Selecting the three correct comparisons was considered a correct trial for which the
experimenter provided enthusiastic praise. Correct selections of each individual picture were acknowledged by the experimenter saying, “That’s right”, until all three correct pictures were selected. Enthusiastic praise was only provided when all three correct pictures had been selected independently. If the child selected an incorrect comparison stimulus, the experimenter said, “No” and waited approximately 10 s to see if the child selected a correct comparison instead. If so, the experimenter acknowledged the response by saying, “That’s right”, and waited for other selections. If the child selected another incorrect comparison, the experimenter said, “No,” pointed to one of the correct comparisons and said, “This is correct.” In case there was still one correct comparison to be selected, the experimenter waited approximately 10 s to see if the child would select it. Selecting a remaining correct comparison was acknowledged. If the child once again selected an incorrect comparison, the correction procedure described above was repeated.

When the child selected only one or two correct comparisons, the experimenter waited 10 s to prompt the next response by saying, “Are there any more?” If the child selected the remaining correct comparison(s) after the prompt, the experimenter acknowledged the response, and provided the prompt again to avoid the possibility that the number of pictures selected was being controlled by the prompt. If only correct comparisons were selected, the experimenter provided enthusiastic praise and scored the trial as correct. If the child did not respond after the prompt, the experimenter implemented the correction procedure as described above.
Training was arranged in 6-trial blocks in which North and South questions were presented in a quasirandom order during three trials each (a 6-trial block). Training continued until the child selected all pictures in both categories three times in a row (three 6-trial blocks at 100%) When criterion was met, the probability of reinforcement was reduced from 100% to 50%. When criterion was met at 50%, reinforcement probability was reduced to 0%. This condition was terminated once criterion was met with no reinforcement.

Posttest interviews. Following the completion of successful categorization tests, the experimenter exposed the child to an additional categorization test and asked questions while the child was selecting the pictures. The questions were: “How did you do that?” and “How did you know that these go together?” Vocal responses were digitally recorded for later analysis.

Independent Variable Integrity (IVI)

Independent-variable integrity (IVI) was assessed by an independent observer for at least 29% of the stimulus-selection-1 and stimulus-selection-2 training conditions for all participants. Sessions used in the calculation of IVI were randomly selected. IVI was calculated by dividing the number of correctly implemented trials by the total number of trials conducted by the experimenter. Trials were scored as entirely correct or incorrect based on the following categories: (1) Reinforcement – praise had to be delivered for all correct trials during the 100% praise condition, for every other correct trial during the 50% praise condition, and not delivered during the 0% praise condition; (2) Correction – the correction procedure had to be correctly
implemented if the trial was marked as incorrect. Table 8 depicts the percentage of sessions in which IVI was calculated, the average IVI scores and the IVI range for all participants across experimental conditions.

Table 8. Average IVI Scores for Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>James</th>
<th>Maria</th>
<th>David</th>
<th>Pam</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessions</td>
<td>29%</td>
<td>44%</td>
<td>52%</td>
<td>58%</td>
</tr>
<tr>
<td>Average</td>
<td><strong>100%</strong></td>
<td>98%</td>
<td><strong>99%</strong></td>
<td><strong>100%</strong></td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td>(66%-100%)</td>
<td>(83%-100%)</td>
<td></td>
</tr>
<tr>
<td>SS-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessions</td>
<td>-</td>
<td>-</td>
<td>33%</td>
<td>-</td>
</tr>
<tr>
<td>Average</td>
<td>-</td>
<td>-</td>
<td><strong>100%</strong></td>
<td>-</td>
</tr>
<tr>
<td>Range</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

Results

Familiar Pictures

In contrast with Experiment 1, data on the number of trials to criterion on pretraining (Table 9) were not consistent with participants’ age. Participants made very few errors and mastered all tasks in a comparable number of trials (range, 154-200). This suggests that they may have been exposed to similar pictures and tasks in the past. Similarly to Experiment 1, results obtained during pretraining indicate that instructions and consequences provided during stimulus selection training, tact and categorization tests were adequate to teach and test these skills.

Unfamiliar Pictures (Maps)

James and Maria. Figure 9 depicts data on correct category sorts for James (upper panel) and Maria (lower panel). Filled circles represent the percentage of

---

10 Experimenter said the name of the category during the correction procedure in 2 out of 6 trials.
correct categorizations out of six trials. Bars represent the percentage of correct tacts to the sample. For both tests 1 and 2, all six pictures were placed in a predetermined location in front of the child. The experimenter picked one at random and asked, “Look at this. Can you give me the others?” for test 1, and “What is this? Can you give me the others?” for test 2. Therefore, tacts to the sample were only required during test 2.

Table 9. Number of Trials to Criterion During Pretraining

<table>
<thead>
<tr>
<th></th>
<th>Age (months)</th>
<th>Trials to Criterion (Minimum =148)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pam</td>
<td>56</td>
<td>156</td>
</tr>
<tr>
<td>Maria</td>
<td>53</td>
<td>200</td>
</tr>
<tr>
<td>James</td>
<td>48</td>
<td>154</td>
</tr>
<tr>
<td>David</td>
<td>41</td>
<td>190</td>
</tr>
</tbody>
</table>

During categorization pretests 1 and 2, James never selected the two correct comparisons in the presence of a sample nor tacted the samples. Immediately after baseline, stimulus-selection-1(SS-1) training was implemented. Once criterion on SS-1 training was reached, James was exposed to categorization posttest 1. On the second and third probes (probes 8 and 9) James was able to select the two correct comparisons in the presence of a sample in 3 out of 6 trials (50%) and in 2 out of 6 trials, respectively. On probes 10 and 11 James never selected the correct comparisons, and on probe 12 he did so once (16%). Although James’ performance met failure criterion on posttest 1 (three consecutive sessions at 33% or below), he certainly performed above baseline level. When required to tact the sample (posttest 2), James selected the correct comparisons in 5 out of 6 trials (83%) on the first probe.
(probe 13), and selected all the correct comparisons on the subsequent probes (probes 14-17). During categorization posttest 2, James only failed to correctly tact the sample once, on the same trial in which he failed to select the correct comparisons (probe 13).

Figure 9. Percentage of Correct Categorizations for James (upper panel) and Maria (lower panel).
Figure 9, lower panel, depicts categorization data for Maria. During categorization pretest 1, Maria selected the correct comparisons in 1 out of 6 trials on the second probe (16%). During pretest 2, she selected the correct comparisons only once during probe 8, demonstrating that she had not learned to categorize these pictures prior to the experiment. After SS-1 training, Maria did not categorize the pictures into north or south during the first probe on posttest 1 (probe 12), but did so twice (33%) in the subsequent probe (probe 13). She once again failed to select any correct combination of comparison stimuli on probe 14, meeting criterion for failure. When required to tact the samples (posttest 2), Maria’s performance on categorizations improved considerably, she selected the correct comparisons once on probes 15 and 18 (16%), and three times (50%) on probes 16 and 17. Interestingly, Maria’s tact responses during these probes were not unequivocal; she tacted the samples incorrectly twice on probes 15 and 16, and once on probes 17 and 18. Every trial in which the sample was tacted incorrectly, categorization performance was also incorrect. On probes 19, 20, and 21, the number of correct tacts to the sample increased together with the number of correct categorizations emitted by Maria. On probes 19 and 21 she categorized all pictures correctly, as well as correctly tacted all samples. On probe 20, she selected the correct comparisons in 4 out of 6 trials, and tacted 5 out of 6 samples correctly, thus succeeding on categorization posttest 2.

James and Maria were also exposed to tact-1 pre/posttests. These tests were always conducted immediately after categorization tests. During tact tests the
experimenter presented a pair of pictures to the child (e.g., N1S1) and assessed whether she could correctly tact one of the pictures.

Figure 10 displays data on tact-1 tests for James. During pretest, James did not tact any of the pictures when presented with the targets N1S1, N2S2, N3S3, and Mixed, suggesting the absence of speaker behavior in regards to the training stimuli. After SS-1 training, James tacted the pictures 100% of the times when presented with the combinations N1S1, N2S2, N3S3, and Mixed (black bars). This suggests that after learning how to select the pictures in an auditory-visual task (SS-1), James was also able to label them. Thus, for James there was a complete transfer between listener (trained) and speaker (untrained) repertoires.

Figure 10. Percentage of Correct Responses During Tact-1 Test for James.
Figure 11 displays data from tact-1 test for Maria. During pretest, Maria was never able to tact the pictures. After SS-1 training, Maria tacted the pictures correctly 100%, 100%, 75%, and 87% of the times when presented with the targets N1S1, N2S2, N3S3, and Mixed, respectively (black bars). Data show that most incorrect tacts occurred in the presence of the N3 and S3 pictures. The lack of stimulus control exerted by these two stimuli contributed for the inaccuracies observed when she was required to tact the pictures in the reassigned (mixed) pairs.

![Graph showing percentage of correct responses for Maria](image)

Figure 11. Percentage of Correct Responses During Tact-1 Test for Maria.

*David and Pam.* Figure 12 depicts data on correct category sorts for David (upper panel) and Pam (lower panel). Filled circles represent the percentage of correct
categorizations out of six trials and bars represent the percentage of correct tacts to the sample.

Figure 12. Percentage of Correct Categorizations for David (upper panel) and Pam (lower panel).

During baseline categorization pretest 1 and 2, David never selected the two correct comparisons in the presence of a sample (Probes 1-6). After reaching criterion on SS-1 training, David was still not able to select the comparisons in the
presence of the sample (posttest 1). When exposed to categorization posttest 2, David continued to select the incorrect comparisons, even though he was able to correctly tact some of the samples when required. He correctly tacted 50%, 66%, and 66% of the samples on probes 10, 11, and 12, respectively (categorization posttest 2). Because David failed to categorize after SS-1 training during both categorization test conditions, stimulus selection-2 (SS-2) training was implemented.

Immediately after meeting criterion in SS-2 training, David responded correctly on the first probe (probe 13) and in 5 out of 6 trials on probes 14 and 15, thus successfully passing categorization posttest 1. Under categorization posttest 2, David correctly selected the comparisons four times (66%) on probe 16, six times (100%) on probe 17, and five times on probes 18 and 19. David correctly tacted four samples on probes 16 (66%), and all of the six samples presented on the subsequent probes (probes 17, 18 and 19). Of note, on probe 16, when David incorrectly tacted the sample, he also categorized the pictures incorrectly. His tact performance during categorization posttest 2 following SS-2 training was much more accurate than when he was required to tact the samples during categorization posttest 2 following SS-1 training. Data obtained with David suggest that he was only able to categorize after additional stimulus selection training (SS-2) was conducted and his tact performance was accurate.

Categorization data for Pam are displayed on the lower panel of Figure 12. During baseline category pretests 1 and 2, Pam never selected the correct comparisons in the presence of a sample, suggesting that she was not able to categorize the
pictures prior to the intervention. As seen on probes 11-13, Pam’s categorization performance after SS-1 training seemed to have increased as a function of test exposure. Pam correctly selected the comparisons twice on probe 11 (33%), five times on probe 12 (83%), three times on probe 13 (50%), six times on probe 15 (100%), five times on probe 16 (83%), and six times on probe 17 (100%). After being exposed to six probes, Pam successfully passed the categorization posttest 1. When required to tact (posttest 2), Pam continued to categorize, as well as correctly tact all samples (probes 17, 18 and 19).

Figure 13 depicts the results for the tact-1 test for David. During pretest, David did not correctly tact any of the pictures. After SS-1 training, David correctly tacted 75%, 0%, 100%, and 62% of the times when presented with the targets N1S1, N2S2, N3S3 and Mixed, respectively (black bars), suggesting limited transfer between listener and speaker repertoires. Most of David’s incorrect tacts occurred in the presence of the N2S2 pair. He tended to respond by saying, “North” in the presence of S2 and “South” in the presence of N2, when these two pictures were presented together, as well as when they were mixed with other pictures.

Given that David had to be exposed to SS-2 training, pretests and posttests on tact-2 were also conducted (Figure 14). Prior to SS-2 training, David tacted 7 out of 12 pictures correctly (58%). Given that this test was administered after SS-1 training, this percentage suggests that that SS-1 alone did not produce accurate speaker responses. After SS-2 training, David tacted all the pictures correctly suggesting that this additional selection/listener training produced accurate tacts.
Figure 13. Percentage of Correct Responses During Tact-1 Test for David.

Figure 14. Percentage of Correct Responses During Tact-2 Test for David.
Figure 15 displays data from tact-1 test for Pam. During pretest, Pam never tacted the pictures correctly. After SS-1 training, Pam’s correct performance was at 100%, 87%, 87%, and 100% when presented with the targets N1S1, N2S2, N3S3, and Mixed, respectively (black bars). Data show that after learning how to select the pictures (SS-1), Pam was also able to label them with almost no errors, suggesting a complete transfer between listener (trained) and speaker (untrained) repertoires.

![Graph showing percentage of correct responses during tact-1 test for Pam.](image)

**Figure 15.** Percentage of Correct Responses During Tact-1 Test for Pam.

**Trials to criterion.** Table 10 displays the number of stimulus selection trials each participant was exposed to until training criterion was reached. Criterion consisted of three consecutive 6-trial blocks with no errors.
Table 10. Number of Trials to Criterion During Stimulus Selection Training

<table>
<thead>
<tr>
<th>Age (months)</th>
<th>SS-1</th>
<th>SS-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pam</td>
<td>56</td>
<td>198</td>
</tr>
<tr>
<td>Maria</td>
<td>53</td>
<td>1,116</td>
</tr>
<tr>
<td>James</td>
<td>48</td>
<td>162</td>
</tr>
<tr>
<td>David</td>
<td>41</td>
<td>1,062</td>
</tr>
</tbody>
</table>

There seemed to be no relation between the number of trials to reach criterion and the participants’ age. Pam and James, who were ages 56 and 48 months, respectively, took approximately the same number of trials to reach criterion on SS-1 training. Similarly, Maria and David also took approximately the same number of trials to reach criterion on SS-1 training, however, these two participants had to be exposed to over 1,000 trials before they could select the pictures in the presence of the auditory stimuli, “North” and “South.” David was also exposed to SS-2 training, and reached criterion on 126 trials.

Vocal-Verbal Behavior

Participants’ vocal behaviors regarding the task were digitally recorded during all categorization test sessions and later transcribed. Observers also documented all vocalizations related to the task that occurred during sessions. A posttest interview was also conducted after successful categorization tests. A transcription of these interviews is attached (Attachment B).

Neither James nor Pam emitted any vocalizations during the sessions other than the required tacts. Maria tacted the north and south stimuli once after the experimenter implemented the correction procedure during SS-1 training. On another trial, Maria said, “South is the highway” after hearing the experimenter say, “Point to
south.” David correctly tacted the comparisons five times during SS-1 training trials, and on a trial in which his selection response was scored as incorrect he spontaneously tacted N2 as “South,” and S1 as “North.” Additionally, in the presence of S3, David said, “it looks like a L,” and in the presence of N1, “It looks like a D.” Although not required to emit any tact responses, David correctly tacted one sample and one comparison on two different trials during categorization-1 posttest.

During posttest interviews, Maria and David were able to justify their selections by naming the category to which the pictures belonged. When asked by the experimenter, “Why did you give me this one?” after successfully categorizing, Maria said, “Because it is north [south].” Similarly, David pointed to the sample and comparisons while labeling them (e.g., “north, north, north”). Both Maria and David said that the stimuli selected “matched” with the sample, even though participants or experimenters never mentioned the word “match” during sessions.

Discussion

During Experiment 2, 3 out of 4 participants (James, Maria, and Pam) were able to correctly categorize the pictures into north and south after being exposed to stimulus-selection training (SS-1). Two of these participants (James and Maria) were only able to do so when required to tact the sample (categorization posttest 2). On the other hand, Pam successfully categorized during categorization posttest-1, but not until repeated exposure to the probes.

In addition to passing categorization tests after SS-1 training, James, Maria, and Pam showed near perfect scores on tact-1 posttests, suggesting an increase in
both categorization and speaker behaviors (tacts) as a function of training the listener repertoire. The participant who failed to categorize after SS-1 training (David), also failed to produce accurate tacts. After being trained on SS-2, David passed both categorization tests, and correctly tacted all of the stimuli during the tact-2 posttest, showing that the additional listener training with all pictures present at the same time (1) improved his tact performance, and (2) facilitated stimulus class formation.

It seemed that Pam’s categorization performance increased as a result of being exposed to categorization posttest 1. It is possible to speculate that, initially, Pam was not tacting the samples, given that she was not required to do so, and thus was not able to correctly categorize the pictures. After repeated exposure to category-sort trials, Pam may have started to covertly tact the samples, which could have contributed to the increase in her categorization performance.

Maria’s categorization performance also increased as a result of repeated exposure to category-sort trials, but not until categorization posttest 2. Not only Maria’s categorization, but also her tact performance increased during posttest 2. It is possible to assume that the fact that Maria was not tacting the samples during categorization posttest 1, accounted for her poor categorization performance during this condition. When asked to tact (posttest-2), she was not able to categorize right away, but as the number of correct tacts emitted by Maria increased, so did the number of correct categorizations.

Results obtained with James were similar to the ones obtained with Maria. His categorization performance also increased as soon as he started to produce accurate
tacts. As mentioned before, in most cases, when participants failed to correctly tact the sample, they also failed to correctly select the correct pictures (categorize).

Nevertheless, there were trials in which participants correctly tacted the sample, but still failed to select the correct pictures. Such performance may be related to a failure to correctly tact the comparisons as previously discussed (Lowenkron, 1996; 1997).

As mentioned earlier, there seemed to be no relation between the number of trials to reach criterion and the participants' age. Pam and James took approximately the same number of trials to reach criterion on SS-1. Similarly, Maria and David also took approximately the same number of trials to reach criterion on SS-1, however, these two participants had to be exposed many more trials before they could select the pictures in the presence of the auditory stimuli.

It is possible to take the number of trials to criterion on the auditory-visual matching/stimulus selection tasks as a measure of participants' receptive skills. Pam and James may have been previously exposed to higher number of auditory-visual tasks, similar to the ones used in the current experiment. Therefore, even if participants' performance was not a function of their age, performance during training could still be related to their language abilities. Unfortunately, participants' language skills were not directly assessed via standardized measures.

The slow acquisition of stimulus selections by Maria and David could have also been a function of lack of accurate observing responding, insensitivity to the consequences provided by the experimenter, and reinforcement of error patterns, as discussed later.
Results from Experiment 2 corroborate the hypothesis that the listener and speaker repertoires played an important role in producing stimulus categorization. It was only when participants behaved accurately as listeners (stimulus selection) and speakers (tact) that they passed categorization tests. These results support the assumption that the absence of one of these repertoires was correlated with failure in the development of stimulus classes.
GENERAL DISCUSSION

The purpose of the current study was to assess whether children could categorize unfamiliar pictures when taught the listener and speaker behaviors separately. Results suggest that children between 3 and 5 years of age were only able to categorize when they behaved both as a speaker and a listener in regards to the visual stimuli. Moreover, some children successfully categorized only when tacts to the sample were required. Taken together, results from Experiment 1 and 2 seemed to support Horne and Lowe's (1996) assumption that naming plays an important role in the development of stimulus classes/categorization by typically developing children. Based on the naming analysis, during categorization tests children had to tact the sample (speaker behavior), either overtly or covertly, producing an auditory or proprioceptive stimulus, which in turn controlled responses of selecting the correct comparisons (listener behavior). Participants in the current study who successfully passed categorization also demonstrated the development of the full name relation (speaker + listener).

Four aspects of the current study warrant further discussion. First, children who failed categorization tests after tact-1 training (Exp. 1) or stimulus-selection 1 training (Exp. 2) later learned to categorize after training with all pictures grouped together. Second, interdependence between listener and speaker repertoires was observed across all participants. Third, there was some evidence that children were not only tacting the samples, but also the comparisons. Fourth, for two of the
participants in Experiment 2, acquisition of stimulus selections was relatively slow, despite the fact that receptive skills are easily acquired during early language development (Horne & Lowe, 1996; 2000). These four findings, along with limitations of the study, applied implications, and directions for future research are discussed below.

**Pairwise vs. All-Stimuli Training**

For two participants in Experiment 1, stimulus classes were formed without them ever seeing class exemplars grouped together prior to testing. The two participants who failed to categorize after pairwise training (tact 1), were able to pass categorization tests after learning to tact the stimuli when grouped together (tact-2 training). In Experiment 2, only one participant required additional stimulus selection training with pictures grouped together (SS-2 training) before passing categorization tests.

There are two possible ways in which tact-2 and SS-2 training may have facilitated categorization. First, when children emitted a sequence of identical tact responses or heard a sequence of auditory samples emitted by the experimenter (e.g., “north,” “north”...), the topography “same” may have emerged. This response could have covertly controlled children’s correct selections during categorization tests (Lowe et al., 2002). Although there was no direct evidence that the response “same” controlled children’s selections, at least one participant (Maria) said that pictures looked the “same” during the posttest interview. Some of the other children said that pictures belonging to the same category, “matched.”
Second, tact-2 and SS-2 training may have facilitated categorization by promoting stimulus control topography coherence (McIlvane & Dube, 2003), and generalization between training and testing conditions. During tact-1 and SS-1 training, children learned to tact or select pictures when they were presented in pairs or groups of three, respectively, but never when they were grouped together. During tact-1 and SS-1 training, a correct response could have come under control of the presence or absence of a specific stimulus feature. As pointed out by Ray and Sidman (1970), "all stimuli are complex in the sense that they have more than one element, or aspect, to which a subject may attend. To ask that the experimenter be aware of all the possibilities is already, perhaps, an impossible demand" (p. 199).

During tact-1 training, for instance, a map containing a round corner could have evoked the response "north." This would have allowed positive results on pairwise (tact-1) training, given that stimuli in the pair could have become discriminable based on this specific feature. Similarly, children’s behavior may have become under control of the presence/absence of stimulus features during SS-1 training.

When stimuli were grouped together during categorization posttests, features that earlier served as the basis for discrimination (during pairwise or training with three pictures) may have been shared by exemplars belonging to different stimulus classes. For example, if the "round corner" controlled the tact "north," pictures containing round corners would have evoked the same tact, regardless of the category to which they belonged. If this was the case, additional tact (or stimulus selection) training with all stimuli grouped together, may have encouraged control by the
experimenter-specified (more complex) stimulus differences. Given that pictures were similarly grouped together during tact-2 and SS-2 training and categorization tests, stimulus generalization between training and testing conditions may have occurred.

Interdependence Between Speaker and Listener Repertoires

When describing the development of naming in early language training, Horne and Lowe (1996) suggested that after the receptive, echoic and tact behaviors are acquired separately, caregiver’s tacting of a new object comes to evoke the whole sequence of behaviors comprising the name relation. A qualitatively new bi-directional relation between listener and speaker is then established in the child’s repertoire. As a result, appropriate listener behavior would be invariably shown when a verbal response is made in the presence of a particular object (i.e., tact). Naming becomes a higher order behavioral relation, dispensing with reinforcement of both speaker and listener behaviors for each new name to be established.

In the current study, the establishment of one behavior (i.e., listener or speaker) consistently led to the emergence of the other, suggesting that participants had already the higher-order naming relation established in their repertoire. In Experiment 1, all participants showed emergence of the listener repertoire after tact training. Lowe et al. (2002) suggested that when teaching tacts to typically developing children, listener behaviors are usually concurrently learned. This is because children’s own utterances may continually precede (listener) behaviors of
(re)orienting to the tacted object, establishing the auditory stimulus produced by vocal responses as an S^D for these behaviors.

In Experiment 2, all participants showed the emergence of the speaker repertoire after stimulus-selection training. It is possible that when trained to point to/select pictures in the presence of an auditory stimulus (e.g., “point to north [south]”), participants oriented towards the correct comparison while engaging in self-echoic behavior (repeating the name of the category spoken by the experimenter). When participants looked and pointed to the correct comparison, the experimenter who was delivering reinforcement contingent upon the correct selection may have also reinforced children’s’ subvocal echoic behavior in the presence of the comparison, accidentally training a tact.

It seems reasonable to assume that the emergence of one repertoire (after the other had been trained) is dependent upon the acquisition of echoics, tacts and listener behaviors (the full name relation), however, specific contingencies of reinforcement for the direct training of both listener and speaker repertoires can be identified in the procedures used in both experiments. Interestingly, in a previous study conducted with the same population, using similar tabletop teaching procedures, a clear functional interdependence between speaker and listener repertoires was not observed (Miguel, Petursdottir, & Carr, 2004).

Hitherto, specific conditions responsible for the transfer between speaker and listener repertoires are unknown (Wynn & Smith, 2003). Researchers may face
procedural limitations when studying this phenomenon, given that teaching one repertoire gives rise to contingencies that may shape the other.

Joint Control

There were several instances in which participants from both experiments performed incorrectly on a specific category-sort trial despite the fact that they had been seen to correctly (1) tact the sample, and (2) select the pictures given their names. This outcome is inconsistent with Horne and Lowe's (1996) analysis, which would predict that categorizations would occur as a function of (1) tacting the sample, and (2) responding as a listener to the response product generated by this tact. Results obtained with one participant in Experiment 1 (John) also seemed to somehow deviate from predictions based on the naming hypothesis given that he correctly categorized the pictures even though he was observed to, sometimes, perform inaccurately as a listener.

As previously discussed, the notion of joint control may circumvent inconsistencies with the naming hypothesis. According to Lowenkron (1998) joint control can be defined as "a discrete event, a change in stimulus control that occurs when a response topography, evoked by one stimulus (e.g., the sample) and preserved by rehearsal, is emitted under the additional (and thus joint) control of a second stimulus, (e.g., the comparison)" (p. 332). In the current study, the tact evoked by the presence of a sample (e.g., the vocal response "north") would be preserved by rehearsal (e.g., child would covertly self-echo the word "north"), and emitted under the additional (and thus joint) control of a correct comparison (e.g., a north picture). The selection
response (i.e., categorization) would be evoked by the occurrence of joint control over the topography rehearsed as an echoic (e.g., “north”) (Lowenkron, 1998).

Based on this analysis, when children failed to tact the comparisons, they would have not been able to categorize. A failure to correctly tact the comparisons would have prevented the additional stimulus (the auditory stimulus produced by tacting the comparison) to jointly control the specific topography, evoking the categorization response. On the other hand, the participant who categorized despite his lack of accurate listener behavior (John) may have done so by tacting the sample, covertly rehearsing this topography, and tacting each of the comparisons, as required to produce joint control. In the joint control analysis, the listener behavior seems to play a minor role in the formation of arbitrary stimulus classes.

Unfortunately, Horne and Lowe’s (1996, 1997), and Lowenkron’s (1996, 1997, 1998) analyses describe processes that are not directly observable, but inferred from what is currently known about stimulus control. Data from the current study clearly support the idea that categorizations were verbally mediated. The specific verbal processes involved in this mediation remain unclear, and should be the focus of future investigations as described later.

Acquisition of Listener Behavior/Stimulus Selections

In Experiment 2, two of the participants made relatively few errors during listener training, mastering SS-1 training in less than 200 trials. On the other hand, participants from Experiment 1 required over 400 trials to master tact-1 training. These results suggest that listener behavior was acquired faster than speaker behavior.
Given that both procedures (tact and stimulus selection training) have ultimately yielded stimulus categorization, it would not be unreasonable to recommend stimulus selection training as the most efficient means to develop stimulus classes in typically developing children. Nevertheless, two of the participants from Experiment 2 required over 1,000 trials to reach mastery criterion on SS-1. Even though the literature on child development suggests that the listener behavior may be mastered before the speaker behavior (e.g., Fraser, Bellugi, & Brown, 1963), acquisition of the listener behavior may, sometimes, be difficult given that it involves a series of conditional, as opposed to simple, discriminations (Michael, 1985).

During tact training, a distinguishable topography (either "north" or "south") was reinforced in the presence of a specific map, after which, each map served to evoke one of the topographies. During stimulus selection training, an undistinguishable response (pointing) was reinforced in the presence of two stimuli; one visual (the picture pointed at), and one auditory (the word spoken by the experimenter). As an outcome, the auditory stimulus altered the controlling strength of the visual stimulus over the selection response. This additional level of conditionality, present in stimulus selection, may require additional training (Michael, 1985). Moreover, stimulus selection training misses an important feedback component present in tact training. During tact training, when participants emitted a vocal response in the presence of a picture, there was a correspondence between the details of the vocal muscle action and the auditory stimulus that results (Michael, 1985). This correspondence may have functioned as an additional source of (automatic) reinforcement, hastening acquisition.
Other variables, inherent to the current procedure, could have played a role in delaying acquisition of stimulus selections. These were: (1) lack of a distinguishable observing response, (2) ineffective consequences, and (3) accidental reinforcement of error patterns.

Research suggests that effective observing behavior is a prerequisite for accurate visual discriminations (Dinsmoor, 1985). Ordinarily, observing behavior to visual stimuli includes orienting the head and moving the eyes so that light reflected by the stimulus falls on the fovea (Dube et al., 2003). In auditory visual discriminations, observing behavior may not include eye movements, but the vocal repetition of the sample (Saunders & Williams, 1998). Typically, observing behavior is maintained by the production of discriminative stimuli (i.e., comparisons) as the consequence. During stimulus selection training, no attempt was made to ensure that participants were “observing” the stimulus. It is possible that, during some of the trials, participants were not attending to the auditory stimulus. This problem would have been remedied had participants been required to repeat the auditory stimulus before comparisons were placed on the table.

It is also possible, that for these participants, the consequence provided during training (i.e., praise) was ineffective as a form of reinforcement. If this were the case, participants would have performed poorly on tact training, as well as on stimulus selection training. Nonetheless, the use of more powerful consequences such as tokens exchangeable for tangibles, or edibles contingent upon correct selections could have accelerated acquisition of auditory-visual discriminations/stimulus selection.
Finally, incorrect selections during training may have been accidentally maintained by the experimenter's praise. During stimulus selection training, responses (i.e., pointing) to the positive comparison were reinforced on a FR1 basis; in other words, every correct response was followed by praise. Therefore, if participants emitted a sequence of incorrect selections followed by a correct selection, they would still receive praise, presumably contingent upon the correct selection. Thus, it is possible that the delivery of praise after a sequence of incorrect + correct selections may have reinforced errors. A possible solution would have been the implementation of a time-out procedure contingent upon incorrect selections (i.e., a longer intertrial interval, or changeover delay, in which the experimenter would turn away from the child). This may have prevented accidental reinforcement of error patterns, possibly by punishing incorrect selections.

Limitations

Procedures in Experiment 1 were similar to those used by Lowe et al. (2002), with the exception of the stimuli used and responses trained. Lowe et al. used 3-dimensional arbitrary objects and trained arbitrary vocal responses, while the current study used unfamiliar two-dimensional pictures (maps) and their corresponding familiar category names ("north" and "south").

Procedures in Experiment 2 were similar to what has been previously suggested by Horne, Lowe and colleagues as a way to evaluate the role of naming on stimulus categorization while training the listener behavior alone (Horne & Lowe, 1996; 2000; Lowe et al., 2002). Although, data have never been published, Horne and
Lowe (2000) reported having conducted a similar experiment with nine children between ages 1 year 7 months and 4 years 1 month, finding very little transfer between listener and speaker behaviors. Regardless of its procedural consistency with previous research, at least three limitations of the current study are noteworthy.

First, it is possible that categorization performance was influenced by the specific instructions provided. On categorization tests, the experimenter showed a picture to the child and asked, “Look, can you give me the others?” (Test 1), or “What is this? Can you give me the others?” (Test 2). During this condition children were expected to tact the sample, by saying (overtly or covertly) “north” or “south.” The auditory covert product of this tact would in turn evoke the listener behavior of reorienting and selecting other members that were part of the same class. However, immediately before selections had to occur, children heard the experimenter saying, “Can you give me the others? The auditory stimulus “others” instead of the name of the category (“north” or “south”) may have controlled children’s behavior of selecting all pictures from the array. Thus, the instructions provided may have contributed for some of the errors observed in category-sort trials. Future replications should try to eliminate this form of instruction, guaranteeing that the last sound heard by the child is the auditory product of her tact.

As discussed earlier, a second possible limitation was that the procedure used to teach stimulus selections could have been made more efficient had a distinguishable observing response been required, and a time out correction used. Furthermore, edible or tangible consequences could have been used during both stimulus selections
and tact training. The choice of praise as a form of reinforcement was made because it replicates conditions present during language learning of typically developing children (Horne & Lowe, 1996).

A third limitation was that correlations between participants’ performance and their verbal abilities could not be made given that participants’ language skills were never directly assessed via standardized measures. In a recent review of the equivalence literature with individuals with language limitations, O’Donnell and Saunders (2003) suggested that better documentation of participants’ characteristics, including pre-experimental verbal skills, would enhance the contributions to the literature. Future research on naming should attempt to test participants’ verbal skills prior to the onset of the study.

Future Directions

There is much to be explored in the area of naming and stimulus categorization. Results obtained in the current study suggest different avenues for future research. Below are some questions that future researchers should attempt to answer.

As suggested by Lowe et al. (2002), it is possible that participants could have passed categorization tests with only the trained repertoire (either the listener or the speaker), and that the emergence of the untrained one is a mere correlate of the training procedure, playing no causal role in the formation of stimulus classes. Naming or the word-object bi-directionality may not be necessary for developing categorization. Future researchers should try to prevent the development of the untrained repertoire and assess whether stimulus classes would still be formed. If the
naming relation does not emerge, participants should not be able to demonstrate stimulus class formation.

The current study succeeded in assessing stimulus class formation by using an alternative measure to the standard matching-to-sample (MTS) procedure; the categorization test. During categorization tests all possible relations among stimuli were assessed (e.g., symmetry and transitivity) given that each picture served, at some point, as a sample and as one of the positive comparisons. As discussed before, more traditional measures of emergent relations could benefit the field of behavior analysis by promoting a more direct dialogue with other disciplines. Future researchers should better evaluate the appropriateness of the categorization test used as a measure of stimulus equivalence, as well as explore other tests involving conditions similar to those encountered in a child’s natural environment.

As previously discussed, results from the current study conform to the naming (Horne & Lowe, 1996), as well as joint control accounts (Lowenkron, 1998). Future researchers should attempt to evaluate the role that the verbal processes described by each of these accounts have on stimulus class formation. Researchers should evaluate the effects that tacting the comparisons (vocal as well as non-vocal tact) has on accurate category-sort trials. Given the similarities between the naming and joint control hypotheses, the experimental analysis of verbal behavior has much to gain by a closer examination of data independently gathered by these two distinct lines of research.
Applied Implications

Procedures derived from the stimulus equivalence literature have increasingly been used to teach a number of different academic skills such as, reading (e.g., de Rose, de Souza, & Hanna, 1996), mathematics (e.g., Lynch & Cuvo, 1995), and geography (e.g., LeBlanc, Miguel, Cummings, Goldsmith, & Carr, 2003) to both typically developing and developmentally delayed individuals. Results from the current study provide additional information about the role of verbal behavior in stimulus class formation. This information may contribute to applied research on derived stimulus relations in, at least three ways.

First, difficulties in establishing the prerequisite/baseline conditional discriminations in typically and developmentally delayed individuals have been reported (e.g., Eikeseth & Smith, 1992; Pilgrim, Jackson, & Galizio, 2000). Such difficulties pose an important limitation on interventions aiming at establishing equivalence classes and generating emergent performance. Based on the current and previous data (e.g., Eikeseth & Smith), teaching children how to tact/label stimuli may facilitate acquisition of baseline conditional discriminations, leading to the formation of equivalence classes. Moreover, if naming were found to be sufficient for generating emergent relations, such as the ones observed in stimulus equivalence research, procedures involving tact training similar to the one used in Experiment 1, may dispense with conditional discrimination training, especially when teaching typically developing children.
Second, the *ability to form categories* has been viewed as an essential language skill to be acquired by typically developing individuals (Barsalou, 1992). While interacting with their natural environment, as well as in structured educational activities, children are constantly learning to categorize new objects such as toys, animals, fruits, and musical instruments. Results from the current study suggest that categorization may be indirectly learned via tact or stimulus selection training. Therefore, procedures that are similar to the ones used here could be devised to formally teach categorization to preschoolers.

Third, it became clear that the procedure used to teach stimulus selection/receptive discrimination could have been improved. Data suggest that practitioners designing interventions to teach receptive language should (1) require an observing response (repetition of the sample), (2) place comparisons on the table after child has repeated the sample, and (3) use powerful reinforcers.

**Conclusion**

It is important to note that the data presented in the current study are correlational in nature, thus the conclusion that *naming* played an essential role in the development of categorization should be taken with caution. Additionally, given that the behavioral mechanisms described as important for the development of categorization were not directly observed, but inferred, it is possible that something else, other than naming, was responsible for the performances observed. Further clarifications will depend on additional demonstrations that equivalence classes only emerge when the repertoires described by Horne and Lowe (1996) are present.
Unfortunately, research on naming suffers from a major methodological limitation, namely, the impossibility of direct measurement of the putative controlling variable (Pilgrim, 1996). Although it may be impossible to directly assess covert verbal behavior, it is hoped that future researchers would concentrate on procedural refinements for the study of naming.

As pointed out by Dugdale (1996), “even if naming as they [Horne & Lowe] define it is not necessary for the emergent behavior we study via the equivalence paradigm, it is likely to be, by far, the most predominant means by which most humans accumulate such repertoires” (p. 274). If nothing else, research on the naming relation will contribute to a better understanding of language development, in particular, the interaction among the various verbal operants described by Skinner (1957).
APPENDIX A
Unfamiliar Pictures
South Dakota (N1)

Mississippi (S1)

Nova Scotia (D1)

North Stimuli

Illinois (N2)

Arkansas (S2)

Ontario (D2)

Wisconsin (N3)

Louisiana (S3)

Nunavut (D3)

Distractors
APPENDIX B
Posttest Interviews
Posttest Interview_Tom

E: Okay, we’re just going to do a couple more with them on the table like that, okay?
E: [name], look at this, can you give me the others?
E: Why did you pick those two pictures?
C: Because... they go together
E: Oh, they go together? Why do they go together?
C: (no response)
E: [name], why do these go together? Why did you pick them?
C: Because they both match
E: Oh, they both match. How do they match?
C: They go together.
E: They go together? Okay.
E: Let’s try one more of these, okay?
C: Can I get another sticker when I’m done?
E: Yes, you can. I am just going to show you one more of these.
E: Hey... look at this, can you give me the others?
E: Okay... very good... how did you do it?
C: (no response)
E: How did you know to pick these?
C: [unreadable]
E: They both match? How do they match?
C: (no response)
E: How did you know how to pick them?
C: (no response)
E: How did you decide to pick them? What did you do?
C: They match.
E: Why do you think they match?
C: (no response)
E: Do you think they look the same?
C: (no response)
E: Do they... do they look the same?
C: (no response)
E: Or... do they have the same name?
C: Yes.
E: They do? Is that why you picked them?
C: (no audible response)
E: Yeah? What is their name?
C: South
E: Aaah, south. Okay, so when you were picking them out, when I said “look at this, can you give me the others?”, what did you do?
C: I picked them because they have the same name.
E: I see. So did you look for the other south pictures?
C: I don’t know.
E: Okay... all right, [name], we are all done.
Posttest Interview _Adam

E: "Look, can you give me the others?"
C: [Child selects]
E: "Why did you give me these?"
C: "...because..."
E: "Why did you put these two together?"
C: "Because you said..."
E: "Why did I say?"
C: "...you said..."[child looks very surprised]
E: "Why didn’t you pick this one?" [Experimenter points to a non-selected picture]
C: "Because that’s not south"
E: "Ok. Look, can you give me the others?"
C: Child selects
E: "Why did you give this one?"
C: "Because that’s north"
E: "Look Can you give me the others?"
C: Child selects one correct picture.
E: "Are there any more?"
C: Child selects the other one
E: "Why did you put these two together?"
C: "Because that’s south"
E: "Ok, we are all done"
Posttest Interview_Rita

E: “Look, can you give me the others?”
C: Child selects the correct ones
E: “How did you do that?”
C: No response
E: Why did you give me these two? [points to selected stimuli]
C: No response
E: “Why didn’t you give me these other two?”
C: No response
E: “Look, can you give me the others?”
E: “Which one would you pick?”
C: Child selects correct ones
E: “Why did you pick these two?”
C: No response
E: “Why do they go together?”
C: No response
E: “Look, can you give me the others?”
C: Child selects correct pictures
E: “Why did you pick these two and not these?” (point to incorrect)
C: No response
E: “Ok, we are all done”
Posttest Interview John

E: “What is this?”
C: “South”
E: “Can you give me the others?”
C: Child selects the correct ones
E: “Why did you give me these two?”
C: “because...” Long pause
E: “Why do they go together?”
C: No response
E: “Why do these three go together?”
C: No response
E: “What is this?”
C: “North”
E: “Can you give me the others?”
C: Child selects the correct ones
E: “Why did you give me those two?”
C: “They match”
E: “What is this?”
C: “North”
E: “Can you give me the others?”
C: Child selects the correct ones.
E: “Why did you give me these?”
C: “Because...” long pause
E: “What is this?”
C: “South”
E: “Can you give me the others?”
C: Child selects the correct ones
E: “Why did you give me these two?” Showing the selected comparisons
C: No response
E: “Why do they go together?”
C: No response
E: “Do you know why they go together?”
C: No response
E: “What is this?”
C: “South”
E: “Can you give me the others?”
C: Child selects the correct ones
E: “Why did you give me these two?”
C: “They match”
E: “Why do they match?”
R: No response
E: “Do you know why they match?”
R: No response
E: “What is this?”
C: “North”
E: “Can you give me the others?”
C: Child selects the correct pictures
E: “Why did you give me this?”
C: “They match”
E: “How do they match?”
C: No response
Posttest Interview_James

E: “Look, can you give me the others?”
C: Child selects incorrect pictures
E: “Look, can you give me the others?”
C: Child selects incorrect pictures
E: Why did you give me these two? [points to selected stimuli]
C: No response
E: Do you know why?
C: No response
E: Let’s try this again. Look, can you give me the others?”
C: Child selects incorrect pictures
E: “Why do these go together?”
C: No response
E: “Look, what is this? Can you give me the others?”
C: Says “North”, and selects correct comparisons
E: So, why did you give me these [points to selected comparisons]
C: No response
E: “Why do they go together? These three pictures [points to North pictures]
C: No response
E: What is the difference? [sorts pictures into North and South]?
C: No response
E: “Do you know the difference?”
C: No response.
Posttest Interview_Maria

E: “Look, can you give me the others?”
C: Child selects correct comparisons
E: “Why did you give me these?”
C: “Because that matches with it”
E: “Why do they match?”
C: “Because its north and north”
E: “Can you look? Can you give me the others?”
C: Child selects one correct comparison
E: “Why did you give me this one?”
C: “Cause it matches, south and south”
E: “Are there any more?”
C: Child selects the remaining correct comparison
E: “Why did you give me this one?”
C: “Because it matches”
E: “How does it match?”
C: “Because they are the same”
E: “Why are they the same?”
C: “Because its south and south”
E: “Look can you give me the others one at a time?”
C: Selects the correct comparisons
E: “Why do they go together?”
C: “Because it matches”
E: Why do they match?
C: “Because, I gave the right ones?”
E: “Why are they right?”
C: “Because they match”
E: “Why?”
C: “Because they are north and north”
Posttest Interview  David

E: “Look this can you give me the others?”
C: Child selects correct comparisons
E: “How did you do that?”
C: “I don’t know”
E: “How do you know these go together?”
C: “Because… maybe… because…maybe…because they match”
E: “How do they match?”
C: Child points to N pictures and says, “North, north, north”
E: “Look, can you give me the others?”
C: Child selects the correct comparisons
E: “How did you do that?”
C: “South, south, south”
E: “Look, can you give me the others?”
C: Child selects correct comparisons
E: “Why did you give me these?”
C: “I don’t know”
E: “Look, can you give me the others?”
C: Child selects the correct comparisons
E: “Why do these go together?”
C: “I don’t know”
E: “Look can you give me the others one at a time?”
C: Selects one correct comparison
E: “Why did you give me this one?”
C: “I don’t know”
E: “Why do these go together?” (hold North pics together)
C: “I don’t know”
E: “What is this?”
C: “North”
E: “Can you give me the others?”
C: Selects correct comparisons
E: “Why did you give me these ones instead of these” (points to South pics)
C: “I don’t know”
E: “Why this one (points to one North) instead of the others?”
C: “I don’t know”
APPENDIX C
Sample Datasheets
Categorization Test 1 (6 pictures)

Participant: ______________ Date: ______________ Session (circle): 1 2 3 4 AM PM

(Circle) Experimenter/Observer: __________________ Condition (circle): Baseline / Posttest

Instruction:

Correct trial: If the child selects the two remaining correct pictures
If a child only selects one correct picture, WAIT FOR 5 SECONDS and then say, “are there any more” after each response. If child selects the 2 correct pics, trial is correct

Incorrect trial: Child selects at least 1 incorrect picture. If first picture selected is incorrect, wait for the child to select a second picture before ending the trial.

Null trial: If the child selects all pictures – Experimenter should say, “I don’t want all the pictures, just some of them”

After 3 consecutive null trials or 3 consecutive blocks with 2 correct trials or below test is finished.

<table>
<thead>
<tr>
<th>Block 1</th>
<th>N1 S2 N3 N2 S1 S3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trial</strong></td>
<td><strong>Sample</strong></td>
</tr>
<tr>
<td>1</td>
<td>N1</td>
</tr>
<tr>
<td>2</td>
<td>N3</td>
</tr>
<tr>
<td>3</td>
<td>S3</td>
</tr>
<tr>
<td>4</td>
<td>S2</td>
</tr>
<tr>
<td>5</td>
<td>N2</td>
</tr>
<tr>
<td>6</td>
<td>S1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Block 2</th>
<th>S2 N1 S3 N1 N2 N3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trial</strong></td>
<td><strong>Sample</strong></td>
</tr>
<tr>
<td>1</td>
<td>N3</td>
</tr>
<tr>
<td>2</td>
<td>S1</td>
</tr>
<tr>
<td>3</td>
<td>N1</td>
</tr>
<tr>
<td>4</td>
<td>N2</td>
</tr>
<tr>
<td>5</td>
<td>S2</td>
</tr>
<tr>
<td>6</td>
<td>S3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Block 3</th>
<th>S3 N1 S1 N2 S2 N3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trial</strong></td>
<td><strong>Sample</strong></td>
</tr>
<tr>
<td>1</td>
<td>S1</td>
</tr>
<tr>
<td>2</td>
<td>S2</td>
</tr>
<tr>
<td>3</td>
<td>N2</td>
</tr>
<tr>
<td>4</td>
<td>S3</td>
</tr>
<tr>
<td>5</td>
<td>N3</td>
</tr>
<tr>
<td>6</td>
<td>N1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Block 4</th>
<th>N3 N1 S3 N2 S1 S2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trial</strong></td>
<td><strong>Sample</strong></td>
</tr>
<tr>
<td>1</td>
<td>N2</td>
</tr>
<tr>
<td>2</td>
<td>S3</td>
</tr>
<tr>
<td>3</td>
<td>S1</td>
</tr>
<tr>
<td>4</td>
<td>N3</td>
</tr>
<tr>
<td>5</td>
<td>S2</td>
</tr>
<tr>
<td>6</td>
<td>N1</td>
</tr>
</tbody>
</table>
**Categorization Test 2**

**Participant:** ______ _ Date: _________ Session (circle): 1 2 3 4 AM PM

(Circle) Experimenter/Observer: ___________________________________ Condition (circle): Baseline / Posttest

**Instruction:**

"What is this? [Wait 3 sec for opportunity to respond] Can you give me the others?" ("give me one at a time")

**Correct trial:**

If the child selects the two remaining correct pictures
If a child only selects one correct picture, WAIT FOR 5 SECONDS and say, "are there any more" after each response. If child selects the 2 correct pics, trial is correct

**Incorrect trial:**

Child selects at least 1 incorrect picture. If first picture selected is incorrect, wait for the child to select a second picture before ending the trial.

**Null trial:**

If the child selects all pictures – Experimenter should say, "I don't want all the pictures, just some of them"

After 3 consecutive null trials or 3 consecutive blocks with 2 correct trials or below test is finished

**Block 1 - N1 S2 N3 N2 S1 S3**

<table>
<thead>
<tr>
<th>Trial</th>
<th>Sample</th>
<th>Named</th>
<th>Comparisons</th>
<th>C</th>
<th>I</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N1</td>
<td>Y</td>
<td>N</td>
<td>S2 N3 N2 S1 S3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>N3</td>
<td>Y</td>
<td>N</td>
<td>N1 S2 N3 N2 S1 S3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>S3</td>
<td>Y</td>
<td>N</td>
<td>N1 S2 N3 N2 S1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>S2</td>
<td>Y</td>
<td>N</td>
<td>N1 S2 N3 N2 S1 S3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>N2</td>
<td>Y</td>
<td>N</td>
<td>N1 S2 N3 S1 S3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>S1</td>
<td>Y</td>
<td>N</td>
<td>N1 S2 N3 N2 S3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Block 2 - S2 N1 S3 N2 N1 S3**

<table>
<thead>
<tr>
<th>Trial</th>
<th>Sample</th>
<th>Named</th>
<th>Comparisons</th>
<th>C</th>
<th>I</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N3</td>
<td>Y</td>
<td>N</td>
<td>S2 N1 S3 S1 N2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>S1</td>
<td>Y</td>
<td>N</td>
<td>S2 N1 S3 N2 N3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>N1</td>
<td>Y</td>
<td>N</td>
<td>S2 S3 S1 N2 N3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>N2</td>
<td>Y</td>
<td>N</td>
<td>S2 N1 S3 S1 N3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>S2</td>
<td>Y</td>
<td>N</td>
<td>N1 S3 S1 N2 N3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>S3</td>
<td>Y</td>
<td>N</td>
<td>S2 N1 S3 N2 S3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Block 3 - S3 N1 S1 N2 S2 N3**

<table>
<thead>
<tr>
<th>Trial</th>
<th>Sample</th>
<th>Named</th>
<th>Comparisons</th>
<th>C</th>
<th>I</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S1</td>
<td>Y</td>
<td>N</td>
<td>S3 N1 N2 S2 N3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>S2</td>
<td>Y</td>
<td>N</td>
<td>S3 N1 S1 N2 N3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>N2</td>
<td>Y</td>
<td>N</td>
<td>S3 N1 S1 S2 S3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>S3</td>
<td>Y</td>
<td>N</td>
<td>N1 S2 N3 N2 S1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>N3</td>
<td>Y</td>
<td>N</td>
<td>S3 N1 S1 N2 S2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>N1</td>
<td>Y</td>
<td>N</td>
<td>S3 S1 N2 S2 N3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Block 4 - N3 N1 S3 N2 S1 N2**

<table>
<thead>
<tr>
<th>Trial</th>
<th>Sample</th>
<th>Named</th>
<th>Comparisons</th>
<th>C</th>
<th>I</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N2</td>
<td>Y</td>
<td>N</td>
<td>N3 N1 S3 S1 S2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>S3</td>
<td>Y</td>
<td>N</td>
<td>N3 N1 N2 S1 S2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>S1</td>
<td>Y</td>
<td>N</td>
<td>N3 N1 S3 N2 S3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>N3</td>
<td>Y</td>
<td>N</td>
<td>N1 S3 S1 N2 S1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>S2</td>
<td>Y</td>
<td>N</td>
<td>N3 N1 S3 N2 S1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>N1</td>
<td>Y</td>
<td>N</td>
<td>N3 S1 S2 N1 S2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Stimulus-Selection Training 1
N1 S1

Participant: _______ Date: _________ Session (circle): 1 2 3 4 AM PM

(Circle) Experimenter/Observer: ____________________________ Condition: 100% praise

Instruction: "Point to the North/South"

Correct trial: If the child selects the correct picture (north or south)

Incorrect trial: If the child selects the incorrect picture or if she does not select a picture within 10 sec.
Apply correction procedure

Correction: Point to the correct picture and say, "This is the correct" (DO NOT say "North" or "South")

After 2 blocks at 100% correct, training is finished

### Block 1

<table>
<thead>
<tr>
<th>Trial</th>
<th>Point to</th>
<th>Order</th>
<th>C</th>
<th>I</th>
<th>Sr</th>
<th>cor</th>
<th>C/I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>North</td>
<td>S1 D1 N1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>North</td>
<td>S1 N1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>South</td>
<td>N1 D1 S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>South</td>
<td>S1 N1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>South</td>
<td>D1 S1 N1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>North</td>
<td>N1 S1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Block 2

<table>
<thead>
<tr>
<th>Trial</th>
<th>Point to</th>
<th>Order</th>
<th>C</th>
<th>I</th>
<th>Sr</th>
<th>cor</th>
<th>C/I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>South</td>
<td>S1 N1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>North</td>
<td>N1 S1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>South</td>
<td>N1 S1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>North</td>
<td>D1 S1 N1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>North</td>
<td>S1 N1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>South</td>
<td>D1 N1 S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Block 3

<table>
<thead>
<tr>
<th>Trial</th>
<th>Point to</th>
<th>Order</th>
<th>C</th>
<th>I</th>
<th>Sr</th>
<th>cor</th>
<th>C/I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>North</td>
<td>S1 N1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>South</td>
<td>N1 S1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>North</td>
<td>N1 S1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>South</td>
<td>D1 N1 S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>North</td>
<td>S1 D1 N1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>South</td>
<td>N1 N1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Block 4

<table>
<thead>
<tr>
<th>Trial</th>
<th>Point to</th>
<th>Order</th>
<th>C</th>
<th>I</th>
<th>Sr</th>
<th>cor</th>
<th>C/I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>North</td>
<td>N1 S1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>South</td>
<td>N1 S1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>South</td>
<td>S1 N1 D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>North</td>
<td>D1 N1 S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>North</td>
<td>D1 S1 N1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>South</td>
<td>N1 D1 S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Tact-1 Pre/Posttest

Participant: ___________________ Date: ___________________ Session (circle): 1 2 3 4 AM PM

(Circle) Experimenter/Observer: ___________________ Condition: no reinforcement

Instruction: "What is this?" (wait 5 seconds)

Correct trial: If the child names the picture correctly (north or south). No praise

Incorrect trial: If the child does not respond within 5 sec. or if she makes an incorrect response

Run all blocks

<table>
<thead>
<tr>
<th>Trial</th>
<th>Order</th>
<th>Point to</th>
<th>C</th>
<th>I</th>
<th>Sr</th>
<th>cor</th>
<th>C/I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>S1N1</td>
<td>S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>S1N1</td>
<td>N1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>N1S1</td>
<td>S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>S1N1</td>
<td>N1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>S1N1</td>
<td>S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>N1S1</td>
<td>N1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>N1S1</td>
<td>S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>N1S1</td>
<td>N1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial</th>
<th>Order</th>
<th>Point to</th>
<th>C</th>
<th>I</th>
<th>Sr</th>
<th>cor</th>
<th>C/I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>S2N2</td>
<td>S2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>N2S2</td>
<td>N2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>N2S2</td>
<td>S2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>S2N2</td>
<td>S2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>S2N2</td>
<td>N2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>N2S2</td>
<td>S2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>S2N2</td>
<td>N2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>N2S2</td>
<td>N2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial</th>
<th>Order</th>
<th>Point to</th>
<th>C</th>
<th>I</th>
<th>Sr</th>
<th>cor</th>
<th>C/I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>S3N3</td>
<td>N3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>N3S3</td>
<td>S3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>N3S3</td>
<td>N3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>S3N3</td>
<td>S3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>N3S3</td>
<td>S3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>S3N3</td>
<td>S3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>S3N3</td>
<td>N3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>N3S3</td>
<td>N3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial</th>
<th>Order</th>
<th>Point to</th>
<th>C</th>
<th>I</th>
<th>Sr</th>
<th>cor</th>
<th>C/I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>N1S3</td>
<td>N1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>N3S2</td>
<td>S2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>S3N1</td>
<td>S3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>N2S1</td>
<td>N2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>S2N3</td>
<td>N3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>S1N2</td>
<td>S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>N3S2</td>
<td>S2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>S1N2</td>
<td>N2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX D
Protocol Clearance from the Human Subjects Institutional Review Board
Date: May 1, 2003

To: James Carr, Principal Investigator
    Caio Miguel, Student Investigator for dissertation

From: Mary Lagerwey, Chair

Re: HSIRB Project Number 03-04-03

This letter will serve as confirmation that your research project entitled “Teaching Picture Categories to Preschool Children: The Role of Naming” has been approved under the full category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may only conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: April 16, 2004
BIBLIOGRAPHY


29th annual convention of the Association for Behavior Analysis, San Francisco, CA.


