THE ROLE OF NAMING IN STIMULUS CATEGORIZATION BY PRESCHOOL CHILDREN

CAIO F. MIGUEL, ANNA I. PETURSDOTTIR, JAMES E. CARR, AND JACK MICHAEL

WESTERN MICHIGAN UNIVERSITY

The purpose of the current study was to assess whether children would categorize pictures when taught the relevant 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/labeling (Test 2) a sample stimulus and selecting the appropriate comparison stimuli. In Experiment 1, 4 children (3–5 years) were taught to tact pictures of six U.S. state maps as either north or south. An assessment was conducted to determine whether they would (1) correctly categorize or sort when presented with a visual sample and (2) select the correct stimuli when hearing their category names (listener behavior). Two of the children categorized the pictures during Posttest 1 after the initial (pairwise) tact training. The other 2 categorized after receiving additional tact training with all pictures presented together. However, one of them categorized only during Posttest 2. In Experiment 2, 4 children (3–5 years) were taught to select pictures when hearing their category names. An assessment was conducted to determine whether they would (1) correctly categorize or sort and (2) tact the stimuli (speaker behavior). One child categorized the pictures during Posttest 1, and two during Posttest 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 Experiments 1 and 2 show that both speaker and listener behavior play an important role in stimulus categorization.
Key words: naming, categorization, stimulus equivalence, verbal behavior, preschool children

---

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 often is assumed that the categorization process is dependent upon the acquisition of specific ''concepts'' (Barsalou, 1992). These concepts are said to be units of mental representation and seem to exist independently of any behavior–environment relation (Zentall, Galizio, & Critchfield, 2002).

In contrast, a concept also could be defined as a group of objects (e.g., stimuli, actions) that control similar responses. When an individual responds similarly to each object in a group of objects, these objects constitute a class, which can then be called a *concept* (Keller & Schoenfeld, 1950). As a result, concepts may be equated to stimulus classes.

Interesting examples of stimulus classes are those whose members do not share physical similarity. Examples of such classes include the relation among pictures, printed words, and spoken words. These stimuli may become substitutable for each other under specific conditions. When asked to point to ''ball,'' someone might point to an actual 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 for such a response (Green & Saunders, 1998). Interestingly, after being taught to respond to some members of a class, humans may behave similarly in the presence of other class members without being directly trained to do so. The search for an understanding of the variables responsible for this emergent repertoire is what drives research in the area of stimulus equivalence (Green & Saunders; Sidman, 1994).

Research on stimulus equivalence has been prevalent in the experimental analysis of behavior for years, but there is still no consensus regarding the mechanism responsible for the emergent behavior observed during equivalence tests (Saunders, Williams, & Spradlin, 1996). According to Horne and Lowe (1996), the formation of equivalence classes is heavily dependent upon the subject's verbal repertoire, or more specifically, upon what is described as the *naming* relation. Horne and Lowe consider naming to be a higher-order class of behavior that involves a bidirectional relation between ''a class of objects and events and the speaker–listener behavior they occasion'' (p. 200). The necessary conditions for the development of naming are present during typical child–caregiver interactions. When naming is acquired, the presence of one member of the class evokes a tact (i.e., a verbal response evoked by a nonverbal S^D), the product of which (e.g., auditory stimulus) in turn evokes 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.

Despite suggestions that the development of equivalence classes may not depend on the ability of humans to name or label stimuli (e.g., Sidman, Cresson, & Wilson-Morris, 1974; Sidman, Kirk, & Wilson-Morris, 1985; Sidman & Tailby, 1982; Sidman, Wilson-Morris, & Kirk, 1986), and the evidence for the development of stimulus equivalence in nonhuman subjects and individuals with minimal verbal repertoires (e.g., Brady & McLean, 2000; Carr, Wilkinson, Blackman, & McIlvane, 2000; Shusterman & Kastak, 1993), it is not unreasonable to suggest that the development of equivalence classes may be influenced by language. Some reports have suggested a positive correlation between verbal competence and successful performance on equivalence tests (e.g., Barnes, McCullagh, & Keenan, 1990; Devany, Hayes, & Nelson, 1986; Eikeseth & Smith, 1992). Additionally, research has demonstrated that the development of equivalence may at least be influenced, if not determined, by language (e.g., Eikeseth & Smith; Goyos, 2000; Horne, Lowe, & Randle, 2004; Lowe, Horne, Harris, & Randle, 2002; Mandell & Sheen, 1994; Randell & Remington, 1999, 2006). It is

possible that equivalence class formation in humans could be either verbally mediated (e.g., rule-governed) or contingency-shaped (de Rose, 1996). Hence, if verbal behavior can mediate the formation of stimulus classes, its direct manipulation should generate or at least improve performance on tests for equivalence or categorization.

Lowe et al. (2002) used a stimulus-sorting procedure to evaluate the development of equivalence classes. They demonstrated that stimulus classes could be established without the use of conventional matching-to-sample procedures by teaching children to tact each member of the target stimulus class. In their first experiment, 9 typically developing 2- to 4-year-old children learned a common vocal response (''zag'') for three arbitrary shapes, and another (''vek'') for three additional shapes. Following tact training, children were exposed to a categorization test. During these category-sort trials, each child was presented with all six shapes; the experimenter selected one of them and asked the child to ''give the others''. A correct category sort was scored if, when presented with a sample stimulus (either a ''zag'' or a ''vek''), the child selected the two positive comparisons. Lowe et al. reported that after training, 4 children passed the categorization test (passing criterion was four out of nine correct sorts per common tact category). In the second experiment, the remaining 5 children were exposed to another categorization test in which they were required to tact the sample before selecting the appropriate comparisons. All 5 participants passed the second categorization test. Moreover, 2 children demonstrated class expansion to two 6-member classes via tact training. The authors also attempted to control for the possibility that having all stimuli presented together during tact training facilitated stimulus class formation. Three children were exposed to tact training in pairwise trials only. Once the training criterion was reached, the children were tested for categorization. Two children passed the first categorization test and one passed the second. Additionally, when probed for listener behavior, all children from the second experiment were able to select the stimuli even though they had never been trained to do so.

In a follow up study, Horne et al. (2004) taught 9 children (1 to 4 years old) to select

specific arbitrary shapes in the presence of the words ''zog'' and ''vek'' in a tabletop two-comparison auditory–visual matching preparation. After this training (i.e., listener training), none of the children categorized the pictures when asked to look at the sample before selecting two of the five comparisons, as in Lowe et al. (2002). When probed for tacts, 7 children failed to tact the stimuli correctly. However, after being directly trained to tact the stimuli, 2 of these children passed categorization tests when required to look at the sample and 3 passed categorization tests when required to tact the sample. Two children were trained on a second stimulus set. After listener training, one child passed the first categorization test, and the other passed the second categorization test only after direct tact training.

A series of recent studies (Horne, Hughes, & Lowe, 2006; Horne, Lowe, & Harris, 2007; Lowe, Horne, & Hughes, 2005) replicated and extended previous findings by showing that common naming was effective in 1) establishing arbitrary stimulus classes in children as young as 1 year and 7 months, and 2) bringing about transfer of function among members of a class. In the Lowe et al. (2005) study, 9 children (1 to 3 years old) learned common tacts for two experimentally defined three-member classes. They also were taught to clap or wave to one of the stimuli from each class. Participants demonstrated listener behavior, categorization, and emitted the specified motor response in the presence of other class members without direct training. When these same motor responses were presented by the experimenter, participants were able to select the corresponding stimuli. Furthermore, when motor responses were trained in the presence of new sets of stimuli, children correctly named, selected, and categorized the corresponding stimuli without training.

In Horne et al. (2006), 14 children (1 to 3 years old) learned to select specific arbitrary shapes in the presence of the words ''zog'' and ''vek''. When tacts were probed, 4 children failed to tact the stimuli correctly. As in the previous study, participants learned to clap or wave in response to one of the stimuli from each class. Participants who demonstrated both listener and speaker (tact) behavior, were able to categorize the shapes accurately, as well as emit the specified motor response in the

presence of other class members. Children who failed to tact, in contrast, were only able to categorize and emit motor responses in the presence of other class members once tacts were directly trained.

In Horne et al. (2007), 8 children (2 to 4 years old) were able to categorize after learning to emit a common manual sign in the presence of each member of the stimulus class. In addition, transfer of function was demonstrated when after learning to tact one exemplar of each class, children correctly tacted the remaining stimuli. The authors concluded that together, these studies add support to the notion that it is only when children acquire both speaker and listener relations that they are able to categorize.

Given the recent literature on the importance of verbal behavior in the development of equivalence classes, and the availability of new measures of stimulus categorization (e.g., Horne et al., 2004, 2006, 2007; Lowe et al. 2002, 2005; Pilgrim & Galizio, 1996), the current study was designed to replicate and extend the work of Horne et al. (2004) and Lowe et al. (2002) by assessing whether slightly older, typically developing children (3 to 5 years old) could categorize unfamiliar pictures when taught listener and speaker behaviors separately. Training and testing procedures were similar to those of previous studies, but there were several important methodological differences that are worth mentioning. First, whereas previous studies used completely abstract stimuli, the current study employed unfamiliar two-dimensional pictures (maps) and their corresponding category names. These stimuli and categories were selected to simulate what might be taught to preschool children within a natural context. Second, a multiple-baseline design was used to control for exposure to experimental stimuli prior to the onset of the study, and history of exposure to the experimental stimuli during probes. Third, listener training and testing was conducted with three, rather than two comparison stimuli to 1) decrease the likelihood that conditional discriminations would be controlled by the incorrect comparison (i.e., reject relation), and 2) decrease the probability that high accuracy was a product of chance responding. Fourth, the categorization–tact test (tact sample match-to-others) was conducted with all participants to assess whether

vocal tacts could have been emitted during categorization tests.

Two experiments were conducted. In the first experiment, children were taught to directly tact (using a common name) individual members of specific categories (categories included maps of northern and southern U.S. states) and were assessed for the ability to (1) sort them into groups without training and (2) select the correct pictures when given the common names (categories). In the second study, children were directly taught to select pictures when given their common names and were assessed to see whether they were able to (1) sort them into groups without direct training and (2) correctly tact them.

EXPERIMENT 1
METHOD
Participants, Setting, and Materials

Four typically developing children, Tom, Adam, Rita, and John participated in Experiment 1. Their ages were 5 years 3 months, 4 years 6 months, 4 years 2 months, and 3 years 1 month, respectively, at the beginning of the study. Participants had no known learning or developmental disabilities and possessed verbal skills that were appropriate to their age. Sessions were conducted behind a divider in a quiet area at the children's preschool. During each session, the child and experimenters sat on the same side of a small table to avoid any inadvertent cueing by the experimenter. Sessions lasted approximately 10 min and were conducted twice each weekday.

Materials included six unfamiliar pictures (state maps) and six familiar pictures. 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. Figure 1 depicts the unfamiliar pictures used during training.

Dependent Variables and Data Collection

The main dependent measure was the percentage of correct category sorts out of six trials. A correct sort was scored when in the presence of the sample (e.g., N1[1]), the child selected the two correct comparisons (e.g., N2 and N3) from a five-stimulus array (e.g., N2, N3, S1, S2, S3). Selection during categorization consisted of removing the picture from the table and giving it to the experimenter. Data on the number of correct category sorts were collected during probe sessions.

Additional dependent measures included (1) selecting a stimulus in response to an instruction given by the experimenter (i.e., listener behavior) and (2) participants' vocal–verbal descriptions of their performance. Listener tests were administered prior to tact training and following categorization tests. Vocal–verbal behavior was assessed during sessions via digital recording, as well as while answering questions following successful completion of categorization tests. The experimenter asked participants to answer openended questions regarding the categorization task. 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 digital recorder.

Interobserver Agreement (IOA)

A second observer was present for at least 37% of all categorization tests, listener tests, and tact training sessions to assess IOA. For each category-sort and tact trial, either an agreement or a disagreement on the dependent measures between the two observers was scored. Point-by-point agreement was calculated by dividing the number of agreements by the sum of agreements and disagreements multiplied by 100%. IOA across conditions averaged 98% (range 83–100%) for Tom, 99% (range 83–100%) for Adam, 99% (range 83–100%) for Rita, and 99% (range 87–100%) for John.

Experimental Design

A nonconcurrent multiple-baseline design across 2 participants (Watson & Workman,

---

[1] N and S refer to the experimenter-defined categories North and South, respectively. Exemplars from each category were numbered (1, 2, and 3).
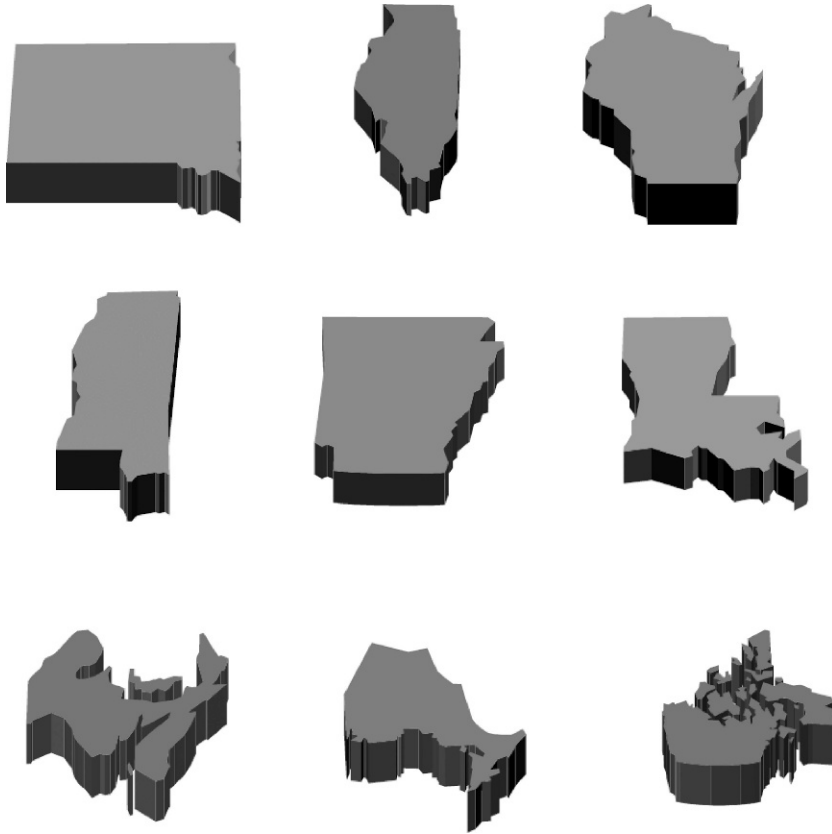
Fig. 1. Unfamiliar pictures. The North, South, and distracter maps are displayed in the top, middle, and bottom panels, respectively.

1981) was used to assess the effects of the independent variable on the percentage of correct categorizations. Much like a concurrent multiple-baseline design, each pair of participants started the experiment at approximately the same time and sessions were conducted with each participant twice a day. The design was chosen so that individual participant absenteeism, as well as individual differences in training (i.e., lengthy training phases with one participant but not the other) would not affect the other participant's progress through the experiment.

Two different categorization tests were administered after training criterion was achieved in experimental conditions. Test 1 was a *look-at-sample categorization test*, and Test 2 consisted of a *tact-sample categorization test*. A description of each of these tests is provided below.

The order of conditions was as follows: Categorization 1 and 2 pretests, listener-1

pretest, tact-1 training, categorization posttests 1 and 2, listener-1 posttest. For those participants who failed to categorize after tact-1 training, the following additional conditions were presented: listener-2 pretest, tact-2 training, categorization 1 and 2 posttests, listener-2 posttest. Interviews were conducted following successful categorization posttests.

Prior to baseline, participants were exposed to pretraining with familiar stimuli to control for the possibility that subsequent failure to categorize with arbitrary stimuli was due to lack of instructional control. Table 1 summarizes the experimental conditions.

Pretraining

Tact-1 pretraining. During tact-1 pretraining, the experimenter randomly separated the familiar pictures into three training pairs. Each pair contained a picture of a bird and a picture of an insect. Pictures were placed

Table 1

Experimental Phases – Experiment 1

Phase	Task	Stimuli per trial	Trials per block	Training Criterion
Pretraining	Tact-1 (pairs)	2	8	2 blocks at 100%
Pretraining	Tact-2 (all stimuli)	6	6	2 blocks at 100%
Pretraining	Listener-1	3	6	2 blocks at 100%
Pretraining	Listener-2	6	6	2 blocks at 100%
Pretraining	Categorization Pretraining	6	6	2 blocks at 100%
1	Categorization Pretest 1	6	6	N/A
	Categorization Pretest 2	6	6	N/A
2	Listener-1 Pretest	3	6	N/A
3	Tact-1 Training (pairs)	2	8	3 blocks at 100%
4	Cat. Pretraining Review	6	4	1 block at 100%
5	Categorization Posttest 1	6	6	N/A
	Categorization Posttest 2	6	6	N/A
6	Listener-1 Posttest	3	6	N/A
7[a]	Listener-2 Pretest	6	6	N/A
8[a]	Tact-2 Training (all stimuli)	6	6	3 blocks at 100%
9[a]	Cat. Pretraining Review	6	4	1 block at 100%
10[a]	Categorization Posttest 1	6	6	N/A
	Categorization Posttest 2	6	6	N/A
11[a]	Listener-2 Posttest	6	6	N/A
12[b]	Posttest Interview	N/A	N/A	N/A

[a] Children were exposed to these phases only if they failed the previous categorization test.
[b] Children were exposed to this phase immediately after successful categorization tests.

horizontally on the table. The experimenter presented the first pair to the child and while pointing to one of the stimuli said, ''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, the experimenter pointed to the stimulus and said, ''This is a bird [insect], can you say it?'' If the child repeated the name of the category, the experimenter said, ''That's right'', or ''OK'' and began 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 met 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 six-picture array. 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, the experimenter pointed to the stimulus and said, ''This is a bird [insect], 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 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 (six-trial blocks) with no corrections.

Listener-1 pretraining. Listener training refers to the behavior of selecting/pointing to visual stimuli in the presence of an instruction provided by the experimenter. This task has also been referred to as receptive discrimination (e.g., Miguel, Petursdottir, & Carr, 2005). During this condition, the experimenter randomly separated the pictures into three pairs. Each pair contained a picture of a bird and a picture of an insect. Three additional pictures from different categories were used as distracters (e.g., tools, kitchen items). One distracter picture was randomly selected and added to each pair as the third comparison during each

trial. The three-choice matching procedure was chosen to decrease the probability of conditional discriminations coming under control of an incorrect comparison (i.e., a reject relation), which could result in failure on subsequent equivalence tests (de Rose, 1996).

The experimenter presented the first three comparisons (set 1) to the child and asked, ''Can you give me the bird [insect]?'' If the child selected the correct picture, and handed it to the experimenter, praise was delivered. If the child responded incorrectly or did not respond, the experimenter pointed to the correct picture and said, ''This is the correct one''. Training for each set was arranged in six-trial blocks in which pictures of birds and insects were presented in a quasirandom order, each picture serving as the correct comparison once to the right, once in the middle, and once to the left of the child. The training criterion was achieved when a child responded correctly to all bird and insect comparisons in two consecutive six-trial blocks. After training set 1, the procedure was repeated for sets 2 and 3. Listener-1 training was completed when responding was at criterion for all trained sets.

Listener-2 pretraining. During this condition, the three pairs of pictures presented during listener-1 pretraining were combined in a six-picture array, and selection responses were trained in this context. Distracters were omitted during this phase. The pictures were randomly ordered in a row and placed in front of the child. The experimenter said to the child, ''Can you give me all the 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 response. If the child selected an incorrect comparison stimulus, 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 praise and then 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 the correct ones''. Bird and insect trials were interspersed in six-trial blocks. Training continued until the child performed at 100% (responded correctly in all trials) in two consecutive blocks.

Categorization pretraining. During this condition, participants were taught to respond to two different tasks that were later used to test categorization with unfamiliar pictures. The two tasks were a *look-at-sample* categorization test and a *tact-sample* categorization test.

During the look-at-sample categorization test, the experimenter placed all six pictures in a predetermined location in front of the child, picked one at random (randomization was predetermined and indicated on the datasheet), and asked the child, ''Look at this, can you give me the others?'' If the child selected the two correct comparisons (two birds or two 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 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 remaining correct comparison after the prompt, the experimenter delivered verbal praise. After each prompted response, the experimenter provided the prompt again. Trials of birds and insects were interspersed in six-trial blocks in which each picture served as a sample once. Training continued until the child responded correctly on all trials for two consecutive blocks.

During the tact-sample categorization test, the experimenter selected a picture and 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?'' Consequences for correct and incorrect selections were the same as in the look-at-sample test. Training continued until the child performed at 100% in two consecutive blocks.

Experimental Procedures

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 training (posttests). Prior to categorization posttests, participants were reexposed to categorization training with familiar pictures until they performed correctly on four consecutive trials. This review was conducted to ensure that performance on categorization tests was under appropriate instructional control.

During test 1, the experimenter used the look-at-sample instruction. All six pictures were placed in a predetermined prerandomized location in front of the child. The experimenter selected one picture and said, ''Look at this. Can you give me the others?'' If the child selected the two correct comparisons, the experimenter waited for five seconds and if no additional pictures were selected, the response was recorded as correct, but no consequences were provided. If during a trial the child selected only one correct comparison, the experimenter prompted another response by saying, ''Are there any more?'' After each prompted response, the experimenter repeated the prompt. If the child selected the correct stimuli, the trial was scored as correct. If the child selected all five comparison stimuli, the experimenter provided feedback to the child by saying, ''I don't want all of the pictures, just some of them'', and scored that trial as incorrect. Testing was conducted in six-trial blocks, with each picture serving as a sample once per block. The location of comparison stimuli was counterbalanced across sessions. In each trial, the experimenter showed a different picture until all pictures had served as a sample. The criterion for ''passing the test'' was set at three consecutive blocks at 66% (four out of six correct categorizations) or higher. The criterion for failure was set at three consecutive blocks at 33% (two out of six correct categorizations) or below.

During test 2, the experimenter used the tact-sample instruction. All procedures were the same as in test 1 except for the instruction. The experimenter asked the child, ''What is this?'' and waited 5 s for the child to tact the sample stimulus. Regardless of the child's response, the experimenter proceeded by asking, ''Can you give me the others?'' Trials were scored as correct when the child selected the two correct comparison stimuli that were consistent with the sample regardless of her tact performance. Categorization passing and failure criteria were the same as in test 1.

Listener-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 pairs. Three additional maps of Canadian provinces that were not part of training were used as distracters (see Figure 1). Each trial included a map of a northern state, a map of a southern state, and a distracter as comparisons. The experimenter presented the first set to the child and asked, ''Give me the North [South]?'' Responses were recorded, but no consequences were provided. Testing for each set was arranged in a six-trial block in which pictures of northern and southern states were presented in a quasirandom order, each picture serving as the correct comparison once on the right, once in the middle, and once on the left of the child. After testing set 1, the procedure was repeated with sets 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 pictures belonged. The condition was similar to tact pretraining. 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 vocal 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 eight-trial blocks in which North and South pictures were presented unsystematically either on the right or left of the child. Training criterion for each pair was correctly tacting all pictures in three consecutive eight-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% (responses on every other correct trial were reinforced). When the 50% criterion was met, the reinforcement probability was reduced to 0% (responses on correct trials were no longer reinforced). Once performance was maintained without reinforcement, training was terminated and categorization tests were administered. Probability of reinforcement was reduced so that performance observed during subsequent tests (conducted under extinction) could not be attributed to a sudden change in the frequency of reinforcement.

Listener-2 pretest/posttest. During this condition, participants were exposed to all pictures at once. The North and South stimuli were combined in a six-picture array, and selection responses to each picture were recorded. 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, but no consequences were delivered. No additional instructions or prompts were provided. Testing was arranged in six-trial blocks in which North and South questions were presented in a quasirandom order three times each. A minimum of two blocks or 12 trials were administered.

Tact-2 training. In this condition, the six North and South 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 vocal 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 responding to each picture had occurred (six trials), the experimenter arranged them in a different order and repeated the training. The training criterion was met when the child correctly tacted all pictures in three consecu-

tive six-trial blocks without prompting. When criterion was met, the probability of reinforcement was reduced from 100% to 50% to 0%. Once the child correctly tacted all pictures in three consecutive arrays without any prompting or reinforcement, training was terminated and categorization tests were 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?'' and ''How did you know that these go together?'' Vocal responses were digitally recorded for later analysis. Although such verbal reports are likely to be under control of variables other than the experimental contingencies, such data may still provide additional information regarding participants' verbal repertoires.

Independent-Variable Integrity

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 was 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; and (2) Correction—the correction procedure had to be correctly implemented if the trial was marked as incorrect. IVI averaged 100% for Tom, 99% (range 87–100%) for Adam, 99% (range 87–100%) for Rita, and 100% for John. Additionally, experimenters were trained to avoid cueing during test sessions by not engaging in any specific motor movements. They also were taught to avoid providing any additional instructions or feedback other than what was previously described. Experimenters were always seated in a chair positioned next to the child on the same side of the table (as opposed to in front of the child) to prevent eye contact or any other subtle cue (i.e., specific hand movements towards the correct comparisons). Finally,
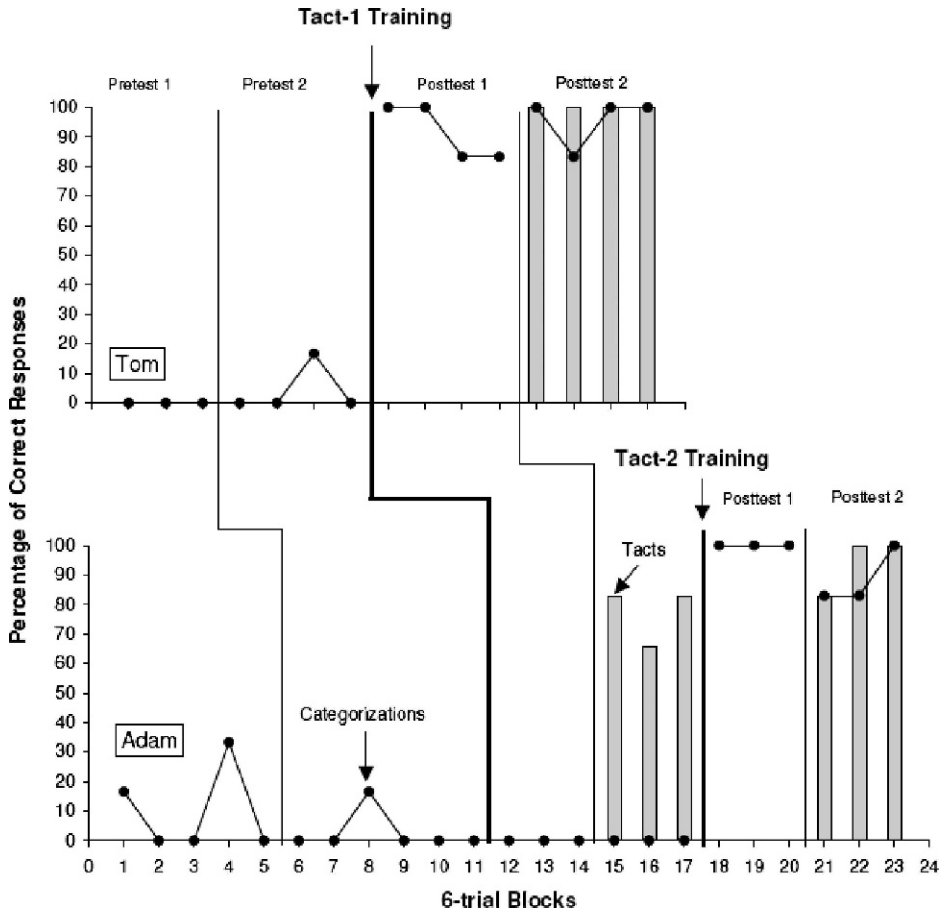
Fig. 2.   Percentage of correct responses (circles) during categorization pre- and posttest conditions and tacts (bars) to sample for Tom (upper panel) and Adam (lower panel).

experimenters were trained not to expect any specific outcomes, and were observed for compliance to the protocol by one of the first two authors for at least 60% of the sessions.

RESULTS

Pretraining with Familiar Pictures

Tom, Adam, Rita, and John required 182, 196, 222, and 250 trials, respectively, to reach the training criterion. All participants made relatively few errors during pretraining. This rapid acquisition of tacts, selection responses, and categorizations indicated that the instructions and consequences provided were adequate to teach and test these repertoires.

Unfamiliar Pictures (maps)

Tom and Adam. Figure 2 depicts the percentage of correct category sorts (circles) and tacts (bars) of the sample for Tom (upper panel) and Adam (lower panel). Tom failed categorization pretests 1 and 2. Once criterion on tact-1 training was reached, Tom selected the two correct comparisons in the presence of a sample in almost all trials on categorization posttest 1 (look at sample). When required to tact the sample (posttest 2), Tom continued to select the correct comparisons. Additionally, he tacted the samples correctly 100% of the time, suggesting that the pairwise training produced reliable tacts.

Adam was exposed to additional categorization pretest probes (Figure 2, lower panel). This served to assess the possibility that extended exposure to testing conditions would result in any observed increase in the percentage of correct category sorts. Adam failed both pretests. After tact-1 training (pairwise), Adam did

Table 3

Percentage of correct responses during listener tests for Experiment 1.

	Listener-1 Test (6-trial blocks)								Listener-2 Test (12-trial blocks)	
	N1S1		N2S2		N3S3		Mixed			
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Tom	33%	83%	16%	100%	16%	100%	33%	100%	-	-
Adam	50%–0%	50%	0%	50%	16%	16%	0%	66%	33%	92%
Rita	50%–30%	100%	16%	83%	33%	66%	16%	100%	-	-
John	50%	83%	16%	66%	0%	50%	0%	50%	75%	75%

not categorize the pictures into North or South in either of the categorization posttest conditions. Although he tacted some of the samples correctly during posttest 2 (tact sample), his tacting performance was inconsistent. Because Adam did not categorize following tact-1 training, the tact-2 training procedure with all pictures presented at once was implemented. After meeting criterion during tact-2 training, Adam selected the correct comparisons on 100% of trials during posttest-1 and 80% or better of posttest-2 probes.

Table 3 displays data on listener tests for all participants. During pretests, both Tom and Adam responded at or below chance levels. 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, suggesting that tact-1 training generated accurate listener behavior. For Adam, by contrast, tact-1 training did not generate accurate listener behavior. He selected the correct comparison 50%, 50%, 16%, and 66% of the time when presented with the targets N1S1, N2S2, N3S3, and Mixed, respectively, on the posttest.

Since Adam had to be exposed to tact-2 training, listener-2 pretests and posttests were conducted. Table 3 also displays the data from listener-2 tests for Adam. During pretest, Adam selected the correct comparisons in 4 out of 12 trials. After tact-2 training, Adam selected the correct comparisons during listener-2 posttests on 92% of trials.

Rita and John. Figure 3 depicts data on correct category sorts (circles) and tacts (bars) to the sample for Rita (upper panel) and John (lower panel). During categorization pretests, Rita never selected the two correct comparisons in the presence of a sample. After reaching criterion on tact-1 training, Rita selected the

two correct comparisons in the presence of a sample 100% of the time. Her accurate performance was maintained when categorization posttest 2 (tact sample) was implemented.

John also failed both pretests, and for him, pairwise tact training (tact-1) did not generate correct category sorts. His performance was at 0% through posttest 1 (look at sample). Even though John was not required to tact the samples during posttest 1, he spontaneously tacted a sample once. During posttest 2 (tact sample), John correctly categorized the pictures in only one out of six trials (16%); however, he tacted all samples correctly. Because John did not categorize after tact-1 training, tact-2 training was implemented. After meeting criterion on tact-2 training, John's performance still met the failing criterion during the posttest-1 condition. On categorization posttest 2, however, John correctly categorized the pictures most of the time. This suggests that John categorized correctly only after additional tact training (tact 2), and when required to tact the sample (test 2).

Table 3 shows that 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. In contrast, John's performance on the listener posttest after tact-1 training was far from accurate. He selected the correct comparisons 83%, 66%, 50%, and 50% of the time when presented with the targets N1S1, N2S2, N3S3, and Mixed, respectively. Prior to tact-2 training, John selected the correct comparisons 75% of the time (9 out of 12 trials). Correct listener behavior remained the same following tact-2 training.

Trials to criterion. The number of trials required to reach the training criterion on tact-1 training was inversely correlated with the
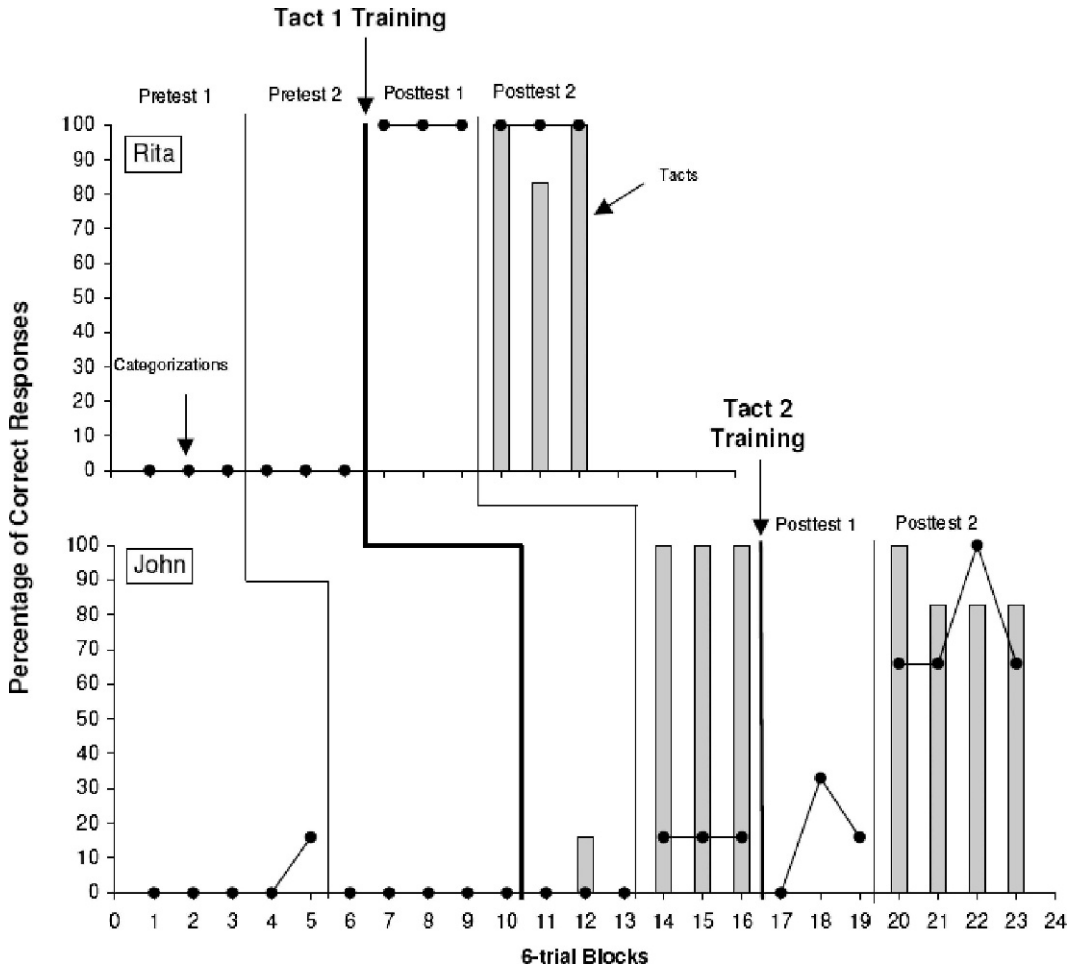
Fig. 3.   Percentage of correct responses (circles) during categorization pre- and posttest conditions and tacts (bars) to sample for Rita (upper panel) and John (lower panel).

participants' ages. Tom (5 years 3 months), Adam (4 years 6 months.), Rita (4 years 2 months), and John (3 years 1 month) required 448, 536, 640, and 880 trials to reach criterion for tact-1 training, respectively. However, John (the youngest participant) required fewer trials to master tact-2 training than his older peer, Adam.

Vocal–Verbal Behavior

During sessions, Tom tacted the stimuli correctly as the experimenter was placing them on the table on several occasions before the session began, particularly during tact training. Adam tacted one of the samples (N1) as ''book'' and another (S3) as ''tree'' during baseline categorization tests. During the listener-1 pre-

test, Adam tacted the sample S3 as ''chicken bone'', and during tact-1 training, tacted the sample N3 as ''cat''. Although not required, Adam correctly tacted the samples and the comparisons during categorization posttests.

Rita and John emitted few 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 listener-1 posttest, mostly in response to the experimenter's instruction, ''point to north [south]''.

During posttest interviews, only Adam provided a verbal description of his selections. 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

Experiment 1 replicated the results of Lowe et al. (2002) in that teaching a common vocal response (i.e., tact) 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 2 participants made few errors during the listener-1 posttest, suggesting that for Tom and Rita, tact training generated accurate listener behavior.

Both Adam and John required additional tact training with all pictures (tact-2 training) before demonstrating stimulus class formation. John, who also failed categorization tests, demonstrated inaccurate listener behavior after tact-1 training. As was the case for Adam, John passed categorization tests only after being trained to tact in the presence of all pictures simultaneously (tact 2); however, this additional training did not seem to improve his listener performance, and John passed categorization tests even though he was still making errors during listener tests. The reasons for the facilitative effects of additional tact training over categorization performance will be discussed later.

In summary, data obtained in Experiment 1 support the assumption that the acquisition of both speaker and listener repertoires may have facilitated stimulus class formation. Experiment 2 was designed to train participants to respond as listeners. The goal of Experiment 2 was to observe whether increases in speaker behavior (tacts) would accompany increases in categorization performances after participants had been directly trained to receptively discriminate pictures (listener behavior).

EXPERIMENT 2
METHOD
Participants, Setting, and Materials

Four typically developing children—James, Maria, David, and Pam—participated in Experiment 2. Their ages were 4 years 0 months, 4 years 5 months, 3 years 5 months, and 4 years 8 months, respectively, at the beginning of the experiment. Sessions were conducted as in Experiment 1.

Dependent Variables and Data Collection

The main dependent measure was correct category sorts as described in Experiment 1. Additional dependent variables included (1) tacts of the stimuli, and (2) the participants' vocal–verbal descriptions of their performance. Tact tests were administered prior to listener training and following categorization tests.

Interobserver Agreement

A second observer recorded data on at least 29% of all categorization tests, tact tests, and listener training sessions. IOA was calculated as described in Experiment 1. IOA across all conditions averaged 99% (range 87–100%) for James, 99% (range 83–100%) for Maria, 99% (range 83–100%) for David, and 100% for Pam.

Experimental Design

A nonconcurrent multiple-baseline design across 2 participants was used as described in Experiment 1. The order of conditions was as follows: baseline categorization pretests, tact-1 pretest, listener-1 training, categorization posttests, and tact-1 posttest. For the participant who failed to categorize after listener-1 training, the following additional conditions were presented: tact-2 pretest, listener-2 training, categorization posttests, and tact-2 posttest. Interviews were conducted after successful categorization tests. Prior to training with unfamiliar stimuli, participants were exposed to pretraining conditions with familiar stimuli, as described in Experiment 1. Table 2 summarizes the experimental conditions.

Pretraining

Pretraining was conducted as described in Experiment 1.

Experimental Procedures

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

Tact-1 pretest/posttest. Pictures were separated into three training pairs (N1S1, N2S2,

Table 2

Experimental Phases – Experiment 2

Phase	Task	Number of stimuli	Trials per block	Training Criterion
Pretraining	Tact-1 (pairs)	2	8	2 blocks at 100%
Pretraining	Tact-2 (all stimuli)	6	6	2 blocks at 100%
Pretraining	Listener-1	3	6	2 blocks at 100%
Pretraining	Listener-2	6	6	2 blocks at 100%
Pretraining	Categorization Pretraining	6	6	2 blocks at 100%
1	Categorization Pretest 1	6	6	N/A
	Categorization Pretest 2	6	6	N/A
2	Tact-1 Pretest	2	8	N/A
3	Listener-1 Training	3	6	3 blocks at 100%
4	Cat. Pretraining Review	6	4	1 block at 100%
5	Categorization Posttest 1	6	6	N/A
	Categorization Posttest 2	6	6	N/A
6	Tact-1 Posttest	2	8	N/A
7[a]	Tact-2 Pretest	6	6	N/A
8[a]	Listener-2 Training	6	6	3 blocks at 100%
9[a]	Cat. Pretraining Review	6	4	1 block at 100%
10[a]	Categorization Posttest 1	6	6	N/A
	Categorization Posttest 2	6	6	N/A
11[a]	Tact-2 Posttest	6	6	N/A
12[b]	Posttest Interview	N/A	N/A	N/A

[a] Children were exposed to these phases only if they failed the previous categorization test.
[b] Children were exposed to this phase immediately after successful categorization tests.

N3S3) and reassigned to three new mixed pairs (Mixed). 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. No additional instructions or prompts were provided. Only one picture in the pair was targeted per trial. Sessions were arranged in eight-trial blocks in which North and South pictures were presented unsystematically either on the right or left of the child. Participants were exposed to one block of each pair and one block of mixed pairs.

Listener-1 training. During this condition, the experimenter randomly separated the pictures into three training pairs. Three additional maps of Canadian provinces were used as distracters. Each trial included a map of a northern state, a map of a southern state, and a map of a Canadian province as comparisons. The experimenter presented the first set to the child and said, ''Can you give me the North [South]?'' If the child selected 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, the experimenter pointed to the correct stimulus and said, ''This is correct!'' If the child pointed to the correct picture after the prompt, the experimenter acknowledged the correct response and said, ''That's right'' but scored that trial as incorrect. 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 on the right, once in the middle, and once on the left of the child. After training set 1, the procedure was repeated with sets 2 and 3. The training criterion was three consecutive blocks at 100%. When responding was at criterion for all trained sets, the stimuli were randomly reassigned to three new mixed sets. The training criterion for the new sets was the same as for the initial sets. Once criterion performance was attained, the probability of reinforcement was reduced from 100% to 50%. If performance was maintained at 50% reinforcement, probability was reduced to 0%. This condition was terminated when performance was maintained with no reinforcement.

Tact-2 pretest/posttest. This condition was similar to the tact-1 pretest/posttest described for Experiment 1. 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 prespecified order and asked, ''What is this?'' If the child produced the correct tact response, the experimenter scored the trial as correct. No additional instructions or prompts were provided. No consequences were provided. After all pictures were presented (six trials), the experimenter showed them in a different order. Two 6-trial blocks were presented.

Listener-2 training (all stimuli). In this condition, pictures presented during listener-1 training (with the exception of the distracters) were combined in a six-picture array, and selecting 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 response. When a correct picture was selected the experimenter acknowledged the response by saying, ''That's right''. 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 for the child to select a correct comparison. If a correct selection was made, the experimenter acknowledged the response by saying, ''That's right'', and waited for other selections. If the child selected a second incorrect comparison, the experimenter said, ''No'', pointed to one of the correct comparisons and said, ''This is correct''. If there was still a correct comparison to be selected, the experimenter waited approximately 5 s for the child to select it. Selecting the remaining correct comparison was acknowledged. If the child 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 5 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. Training was arranged in six-trial blocks in which North and South questions were presented in a quasirandom order during three

trials each (a six-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 as described for listener-1 training. This condition was terminated once performance was maintained with no reinforcement.

Posttest interviews. Interviews were conducted as described in Experiment 1.

Independent Variable Integrity

IVI was assessed by an independent observer for at least 29% of the listener-1 and listener-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 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; and (2) Correction—the correction procedure had to be correctly implemented if the trial was marked as incorrect. IVI averaged 100% for James, 98% (range 66–100%) for Maria, 99% (range 83–100%) for David, and 100% for Pam. As described in Experiment 1, experimenters were extensively trained to avoid inadvertent cueing during test sessions.

RESULTS

Pretraining with Familiar Pictures

Pam, Maria, James, and David required 156, 200, 154, and 190 trials to reach the training criterion, respectively. As in Experiment 1, the rapid acquisition during pretraining indicated that instructions and consequences provided during listener, tact, and categorization training were adequate to teach these skills.

Unfamiliar Pictures (maps)

James and Maria. Figure 4 depicts data on correct category sorts (circles) and tacts (bars) to the sample for James (upper panel) and Maria (lower panel). During categorization pretests 1 and 2, James never selected the two correct comparisons in the presence of a sample nor tacted the samples. Although
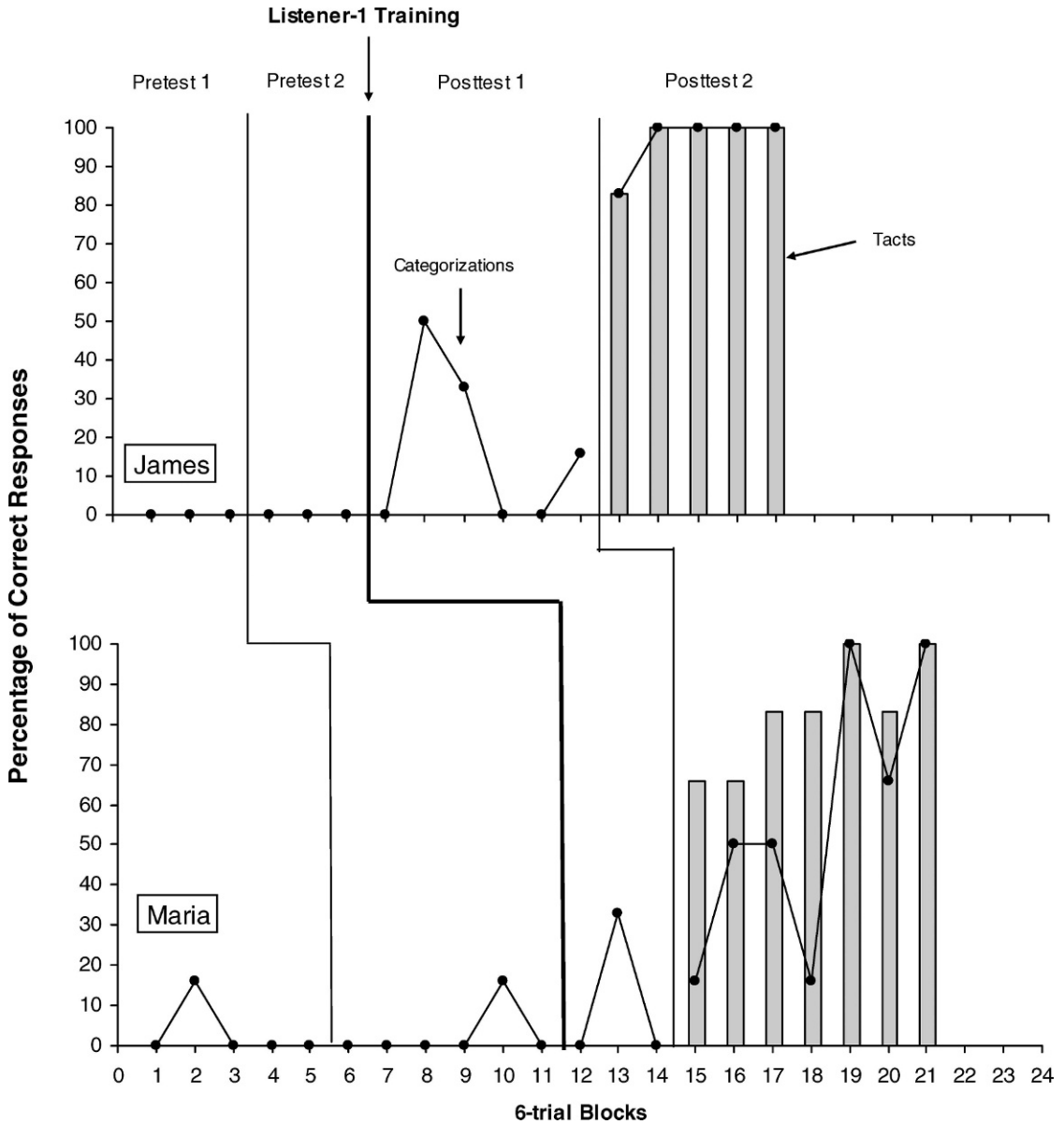
Fig. 4. Percentage of correct responses (circles) during categorization pre- and posttest conditions and tacts (bars) to sample for James (upper panel) and Adam (lower panel).

James's performance met the failure criterion on posttest 1 (three consecutive sessions at 33% or below), he performed above baseline level. When required to tact the sample (posttest 2), James passed the categorization test. During this test, James responded incorrectly once and incorrectly tacted the sample.

Maria did not meet performance criterion on posttest 1. When required to tact

the samples (posttest 2), her performance improved considerably. Interestingly, Maria's tact responses during these probes were not reliably correct; in every trial in which the sample was tacted incorrectly, categorization performance was also incorrect. As the number of correct tacts to the sample increased, so did the number of correct categorizations.

Table 4

Percentage of correct responses during tact tests for Experiment 2.

| | Tact-1 Test (8-trial blocks) | | | | | | | | Tact-2 Test (12-trial blocks) | |
| | N1S1 | | N2S2 | | N3S3 | | Mixed | | | |
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
James	0%	100%	0%	100%	0%	100%	0%	100%	-	-
Maria	0%	100%	0%	100%	0%	75%	0%	87%	-	-
David	0%	75%	0%	0%	0%	100%	0%	62%	58%	100%
Pam	0%	100%	0%	87%	0%	87%	0%	100%	-	-

Table 4 displays data on tact-1 tests for James and Maria. During the pretest, James did not tact any of the pictures when presented with the targets, suggesting the absence of speaker behavior with regard to the training stimuli. After listener-1 training, James tacted the pictures 100% of the time. During pretest, Maria never tacted the pictures correctly, but after listener-1 training, she tacted the pictures correctly most of the time. Most of the incorrect tacts occurred in the presence of the N3 and S3 pictures. The lack of stimulus control exerted by these two stimuli contributed to the inaccuracies observed when she was required to tact the pictures in the reassigned (mixed) pairs.

David and Pam. Figure 5 depicts data on correct category sorts (circles) and tacts (bars) to the sample for David (upper panel) and Pam (lower panel). During pretests 1 and 2, David never categorized correctly. After reaching criterion for listener-1 training, David failed posttest 1. When exposed to categorization posttest 2, David continued to select incorrect comparisons, even though he was able to correctly tact some of the samples. Because David did not categorize after listener-1 training during either categorization test condition, listener-2 training was implemented. After meeting criterion in listener-2 training, David successfully passed categorization posttest 1. Under categorization posttest 2, when David incorrectly tacted the sample, he also categorized the pictures incorrectly.

Categorization data for Pam are displayed on the lower panel of Figure 5. During pretests 1 and 2, Pam never categorized correctly. Her categorization performance seemed to have been acquired during posttest. After being exposed to six probes, Pam successfully passed categorization posttest 1. When required to tact (posttest 2), Pam continued to categorize, as well as to correctly tact all samples.

As seen in Table 4, during tact-1 pretest, David did not correctly tact the pictures. After listener-1 training, David correctly tacted 75%, 0%, 100%, and 62% of the time when presented with the targets N1S1, N2S2, N3S3, and Mixed, respectively. Because David had to be exposed to listener-2 training, tact-2 pretests and posttests also were conducted. During tact-2 pretest, David tacted correctly 58% of the time (7 out of 12 trials). After listener-2 training, he tacted the pictures correctly 100% of the time. Pam never tacted the pictures correctly during tact-1 pretest, but made almost no errors on tacts following listener-1 training.

Trials to criterion. There was no apparent relation between the number of trials to reach criterion and the participants' age. Pam and James, who were approximately the same age, required 198 and 162 trials, respectively, to reach criterion on listener-1 training. Maria and David were separated in age by over a year but required approximately the same number of trials to reach criterion on listener-1 training: 1,116 and 1,062, respectively.

Vocal–Verbal Behavior

Neither James nor Pam emitted any unprompted vocalizations during the sessions. Maria tacted the north and south stimuli once after the experimenter implemented the correction procedure during listener-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 listener-1 training trials, and on a trial in which his selection response was scored as incorrect he tacted N2 as ''South'', and S1 as ''North''. Additionally,
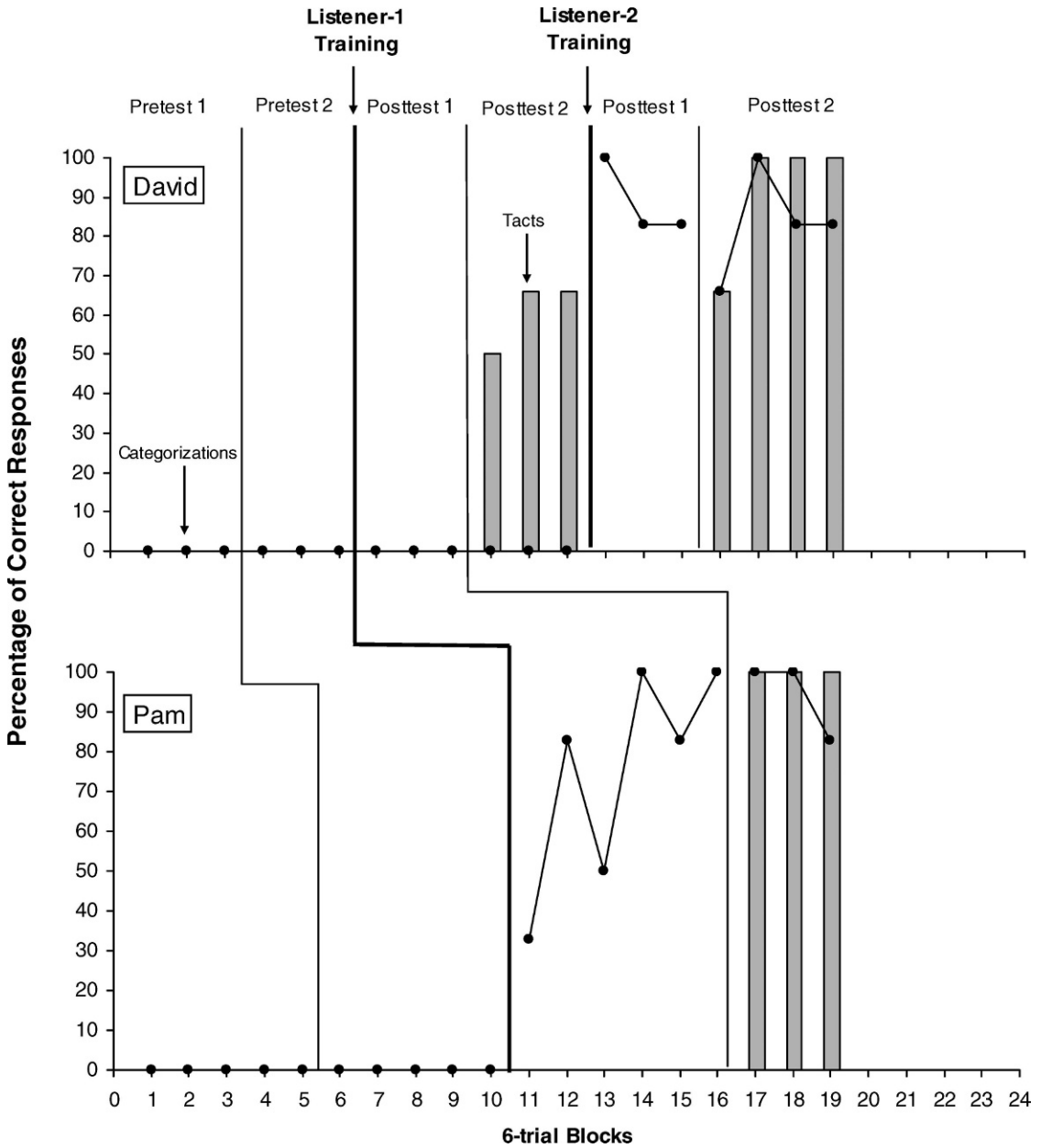
Fig. 5. Percentage of correct responses (circles) during categorization pre- and posttest conditions and tacts (bars) to sample for David (upper panel) and Pam (lower panel).

in the presence of S3, David said, ''It looks like an L'', and in the presence of N1, ''It looks like a D''. David correctly tacted one sample and one comparison on different trials during the categorization-1 posttest.

During posttest interviews, Maria and David provided verbal descriptions of their selections. 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'').

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 listener-1 training. Two of these 3 participants (James and Maria) were only able to do so during categorization posttest 2, although Maria seemed to have learned the task during the categorization test. Although Pam successfully categorized during categorization posttest-1, like Maria, this occurred only after repeated exposure to the test.

In addition to passing categorization tests after listener-1 training, James, Maria, and Pam showed near perfect scores on tact-1 posttests, suggesting an increase in both categorization and tacts as a function of training the listener repertoire. The participant who did not categorize after listener-1 training, David, did not accurately tact. After being trained on listener-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.

Pam's categorization performance appeared to increase as a result of being exposed to the categorization posttest. It is possible that, initially, Pam was not tacting the samples, given that she was not required to do so, and this may have resulted in her not being able to correctly categorize the pictures.

Maria's categorization performance also increased as a result of repeated exposure to category-sort trials, but not until categorization posttest 2. Her tact performance also increased during posttest 2. It is possible that Maria was not tacting the samples during categorization posttest 1, and this contributed to her poor categorization performance. When asked to tact the sample (posttest 2), she initially did not categorize, but as the number of correct tacts emitted by Maria increased, so did the number of correct categorizations.

Results obtained with James suggest that his categorization performance increased as soon as he was required to tact. As mentioned before, in most cases, when participants failed to correctly tact the sample, they also failed to correctly select the pictures. Nevertheless, there were trials in which participants correctly tacted the sample, but still failed to select the correct pictures.

It is important to note that two of the participants from Experiment 2 required over 1,000 trials to reach mastery criterion for listener-1 training. It is possible that the lack of a specifically trained observing response could have contributed to the delayed acquisition of listener behavior.

Results from Experiment 2 replicate those obtained by Horne et al. (2004) in that the participants who responded accurately as both listeners and speakers (tact) also passed categorization tests.

GENERAL DISCUSSION

The purpose of the current study was to assess whether children could categorize unfamiliar pictures when taught 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 speakers and listeners with regard to the visual stimuli. Moreover, some children categorized only when tacts to the sample were required. These results were consistent with the results from Horne et al. (2004, 2006, 2007) and Lowe et al. (2002, 2005) who found that younger children categorized dissimilar objects when able to behave as speakers and listeners with regard to those objects.

Taken together, results from Experiments 1 and 2 appear to support the assumption that *naming* can play an important role in the development of stimulus classes and 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 a stimulus (e.g., the sound heard), 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). However, as discussed below, the results are not unequivocal and alternative interpretations are possible.

Three aspects of the current study warrant further discussion. First, children who failed categorization tests after tact-1 training (Experiment 1) or listener-1 training (Experiment 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. These three findings are discussed in the following sections.

Pairwise versus Group Stimuli Training

Tact-2 and listener-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 listener-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 listener-1 training, a correct response could have come under control of the presence or absence of a specific stimulus feature. 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, the children's behavior may have come under control of the presence or absence of stimulus features during listener-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 were the case, additional tact (or listener) training with all stimuli grouped together may have encouraged control by the experimenter-specified (more complex) stimulus differences. Given that pictures were similarly grouped during tact-2 and listener-2 training and categorization tests, stimulus generalization between training and testing conditions may have occurred.

Interdependence between Speaker and Listener Repertoires

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 learned concurrently. 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 a discriminative stimulus for these behaviors.

In contrast to the results obtained by Horne et al. (2004), all participants in Experiment 2 showed the emergence of the speaker repertoire after listener training. Because the children in the current study were older than the ones in the Horne et al. study, it seems that this emergence of tacts after listener training was dependent upon their more sophisticated verbal repertoires including echoic, tacts, and listener behaviors. However, specific contingencies of reinforcement for the direct training of tacts can be identified in the procedure used in Experiment 2. It is possible that during listener training, when required to select pictures in the presence of an auditory stimulus (e.g., ''Give me north [south]''), participants oriented toward the correct comparison while engaging in covert self-echoic behavior (repeating the name of the category spoken by the experimenter). When participants looked and selected the correct comparison, the experimenter who was delivering reinforcement contingent upon the correct selection also may have reinforced children's subvocal echoic behavior in the presence of the comparison, resulting in the emergence of a tact.

So far, specific conditions responsible for the transfer between speaker and listener repertoires are unknown (e.g., 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 observed to correctly (1) tact the sample, and (2) select the pictures given their names. On other occasions, participants performed correctly on a specific category-sort trial despite the fact that they had been observed to

incorrectly tact the sample. These outcomes seem 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 deviate from predictions based on the naming hypothesis given that he correctly categorized the pictures even though he was observed to occasionally perform inaccurately as a listener.

The notion of *joint control* may serve to explain some of the results that were inconsistent 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., the child would covertly 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 behavior (e.g., ''north'') (Lowenkron, 1998).

Based on this analysis, the children may have been unable to categorize when they failed to tact the comparisons. A failure to correctly tact the comparisons would have prevented the additional stimulus (the auditory stimulus produced by tacting the comparison) from jointly controlling the specific topography, evoking the categorization response. In contrast, 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, 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.

Alternatively, weak tact performance during categorization tests could implicate poor stimulus discrimination. If this were the case, then additional discrimination training would be sufficient to produce consistent correct performance, despite participants' speaker and listener repertoires. This far more parsimonious interpretation dispenses with the use of unobservable processes to interpret categorization and suggests that the listener and speaker repertoires are corollary to discrimination training. The specific verbal processes involved, if any, remain unclear and should be the focus of future investigations.

Limitations

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, the children were expected to tact the sample by saying (overtly or covertly) ''north'' or ''south''. The stimulus produced by this tact would in turn evoke the listener behavior of reorienting and selecting other members of the same class. However, immediately before selections had to occur, children heard the experimenter say, ''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 to 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.

A second limitation was that correlations between participants' performance and their verbal abilities could not be observed given that participants' language skills were not directly assessed via standardized measures. O'Donnell and Saunders (2003) suggested that better documentation of participants' characteristics, including preexperimental verbal skills, would enhance contributions to the stimulus equivalence literature. Future research on naming should attempt to test

participants' verbal skills prior to the onset of the study.

A third limitation consists of the fact that the maps used as distracters during listener training and testing never served as correct comparisons. Although the task could have been procedurally defined as three-choice matching-to-sample, it was functionally a two-choice matching-to-sample. Therefore, it is possible that discriminations made by participants were under control of reject relations. Future replications should attempt to establish three as opposed to two equivalence classes to reduce the likelihood of spurious stimulus control.

Future Directions

The current study succeeded in assessing stimulus class formation by using an alternative measure to the standard matching-to-sample procedure; the categorization test. During categorization tests each picture served, at some point, as a sample and as one of the positive comparisons. 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 could be used to support both the naming (Horne & Lowe, 1996) and 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 nonvocal tacts) 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.

Conclusion

The results from the current study seem to add to the current literature on stimulus categorization by replicating previous findings suggesting that the speaker and listener repertoires may contribute to categorizing by capably speaking children. Results also suggest that the specific procedures, as well as the single-subject experimental design (i.e., multiple baseline) used, may have been adequate to study stimulus categorization.

It is important to note that the repertoire described as important for the development of categorization (i.e., naming) was not directly observed, but inferred. Thus, it is always possible that something else was responsible for the observed performances. 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 will concentrate on procedural refinements for the study of naming. 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).

REFERENCES

Barnes, D., McCullagh, P. D., & Keenan, M. (1990). Equivalence class formation in non-hearing impaired and hearing impaired children. *The Analysis of Verbal Behavior, 8*, 19–30.

Barsalou, L. W. (1992). *Cognitive psychology: An overview for cognitive scientists*. Hillsdale, NJ: Erlbaum.

Brady, N. C., & McLean, L. K. S. (2000). Emergent symbolic relations in speakers and nonspeakers. *Research in Developmental Disabilities, 21*, 197–214.

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

de Rose, J. C. (1996). Controlling factors in conditional discriminations and tests of equivalence. In T. R. Zentall, & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 253–277). Amsterdam, Holland: Elsevier.

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

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

Goyos, C. (2000). Equivalence class formation via common reinforcers among preschool children. *The Psychological Record, 50*, 629–654.

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal, & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229–262). New York: Plenum Press.

Horne, P. J., Hughes, J. C., & Lowe, C. F. (2006). Naming and categorization in young children: IV: Listener behavior training and transfer of function. *Journal of the Experimental Analysis of Behavior, 85,* 247–273.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65,* 185–241.

Horne, P. J., & Lowe, C. F. (1997). Toward a theory of verbal behavior. *Journal of the Experimental Analysis of Behavior, 68,* 271–296.

Horne, P. J., Lowe, C. F., & Harris, F. D. A. (2007). Naming and categorization in young children: V. Manual sign training. *Journal of the Experimental Analysis of Behavior, 87,* 367–381.

Horne, P. J., Lowe, C. F., & Randle, V. R. L. (2004). Naming and categorization in young children: II. Listener behavior training. *Journal of the Experimental Analysis of Behavior, 81,* 267–288.

Keller, F. S., & Schoenfeld, W. N. (1950). *Principles of psychology: A systematic text in the science of behavior.* New York: Appleton-Century-Crofts.

Lowe, C. F., Horne, P. J., Harris, F. D. A., & Randle, V. R. L. (2002). Naming and categorization in young children: Vocal tact training. *Journal of the Experimental Analysis of Behavior, 78,* 527–549.

Lowe, C. F., Horne, P. J., & Hughes, J. C. (2005). Naming and categorization in young children: III. Vocal tact training and transfer of function. *Journal of the Experimental Analysis of Behavior, 83,* 47–65.

Lowenkron, B. (1996). Joint control and word-object bidirectionality. *Journal of the Experimental Analysis of Behavior, 65,* 252–255.

Lowenkron, B. (1997). The role of joint control in the development of naming. *Journal of the Experimental Analysis of Behavior, 68,* 244–247.

Lowenkron, B. (1998). Some logical functions of joint control. *Journal of the Experimental Analysis of Behavior, 69,* 327–354.

Mandell, C., & Sheen, V. (1994). Equivalence class formation as a function of the pronounceability of the sample stimulus. *Behavioural Processes, 32,* 29–46.

McIlvane, W. J., & Dube, W. V. (2003). Stimulus control topography coherence theory: Foundations and extensions. *The Behavior Analyst, 26,* 195–213.

Miguel, C. F., Petursdottir, A. I., & Carr, J. E. (2005). The effects of multiple-tact and receptive-discrimination training on the acquisition of intraverbal behavior. *The Analysis of Verbal Behavior, 21,* 27–41.

O'Donnell, J., & Saunders, K. J. (2003). Equivalence relations in individuals with language limitations and mental retardation. *Journal of the Experimental Analysis of Behavior, 80,* 131–157.

Pilgrim, C. (1996). Can the naming hypothesis be falsified? *Journal of the Experimental Analysis of Behavior, 65,* 284–286.

Pilgrim, C., & Galizio, M. (1996). Stimulus equivalence: A class of correlations or a correlation of classes. In T. R. Zentall, & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 173–195). Amsterdam, Holland: Elsevier.

Randell, T., & Remington, B. (1999). Equivalence relations between visual stimuli: The functional role of naming. *Journal of the Experimental Analysis of Behavior, 71,* 395–415.

Randell, T., & Remington, B. (2006). Equivalence relations, contextual control, and naming. *Journal of the Experimental Analysis of Behavior, 86,* 337–354.

Saunders, K. J., Williams, D. C., & Spradlin, J. E. (1996). Derived stimulus control: Are there any differences among procedures and processes? In T. R. Zentall, & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 93–109). Amsterdam, Holland: Elsevier.

Shusterman, R. J., & Kastak, D. (1993). A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *The Psychological Record, 43,* 823–839.

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

Sidman, M., Cresson, O., & Wilson-Morris, M. (1974). Acquisition of matching to sample via mediated transfer. *Journal of the Experimental Analysis of Behavior, 22,* 261–273.

Sidman, M., Kirk, B., & Wilson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior, 43,* 21–42.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 23–44.

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

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Watson, P. J., & Workman, E. A. (1981). The nonconcurrent multiple baseline across individuals design: An extension of the traditional multiple baseline design. *Journal of Behavior Therapy and Experimental Psychiatry, 12,* 257–259.

Wynn, J. W., & Smith, T. (2003). Generalization between receptive and expressive language in young children with autism. *Behavioral Interventions, 18,* 245–266.

Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorization, concept learning and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior, 78,* 237–248.