JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2002, **78,** 527–549 NUMBER 3 (NOVEMBER)

NAMING AND CATEGORIZATION IN YOUNG CHILDREN: VOCAL TACT TRAINING

C. FERGUS LOWE, PAULINE J. HORNE,
FAY D. A. HARRIS, AND VALERIE R. L. RANDLE

UNIVERSITY OF WALES, BANGOR, AND
UNIVERSITY OF NEWCASTLE

In three experiments, 2- to 4-year-old children, following pretraining with everyday objects, were presented with arbitrary stimuli of differing shapes. In Experiment 1A, 9 subjects were trained one common tact response, ''zag,'' to three of these and a second tact, ''vek,'' to another three. In category match-to-sample Test 1, 4 subjects sorted accurately when required only to look at the sample before selecting from five comparisons. The remaining 5 subjects succeeded in Test 2, in which they were required to tact the sample before selecting comparisons. Experiment 1B showed, for 2 of these subjects, that tact training with 12 arbitrary stimuli established two six-member classes that were still intact 6 weeks later. In Experiment 2, 3 new subjects participated in a common tact training procedure that ensured that none of the exemplars from the same class were presented together prior to the test for three-member classes. Two subjects passed category Test 1 and the third passed Test 2. Tests showed subjects' listener behavior in response to hearing /zog/ and /vek/ to be in place. These experiments indicate that common naming is effective in establishing arbitrary stimulus classes and that category match-to-sample testing provides a robust measure of categorization.

Key words: naming, categorization, stimulus classes, stimulus equivalence, tacting, category match to sample, children

How stimulus classes are established has become a central question for behavior analysis. Research has focused particularly on arbitrary classes, that is, classes of stimuli that are not distinguished by any common formal characteristic (e.g., of color, shape, or size). Such classes or categories are usually shown in humans in early childhood and thereafter feature prominently in much of human behavior (Horne & Lowe, 1996; Markman, 1989; Mervis, 1987). For example, it may be evidence of a young child's having learned to categorize ''toys'' that when asked to ''put the toys in the bag,'' she shows the same behavior (i.e., putting in the bag) to all toys present, how ever much they differ physically, while at the same time not responding to other objects. This behavior would not be very remarkable, nor could we describe it as ''categorizing,'' if it had been previously reinforced. If, on the other hand, putting the toys in the bag were novel behavior that had been established with just one or two exemplars that then generalized without reinforcement to the other toys, that behavior could be regarded as evidence of categorization. Even stronger evidence would be if, for the first time and in the absence of reinforcement, one could pick up as a sample stimulus any one of those toys at random, say to the child, ''Give me the other toys,'' or simply, ''Give me the others,'' and find that the child did then select the other items in that class of toys correctly. These two features—stimulus substitutability, in which stimuli can be used interchangeably to occasion selection of other members of the arbitrary class, and transfer of behavior, in which behavior trained to one or more exemplars generalizes to other members of the class—are both of central importance practically as well as theoretically. Because categorization entails so much novel or ''emergent'' behavior without the need for direct and lengthy reinforcement of each response, it plays a central role in human activity in general. But it is how to

We are grateful to Sue James, Gareth Horne, and Sue Peet for their help in the preparation of this manuscript, to Carl Hughes for his very helpful feedback, to Chris Whitaker for advice on statistics, and to Pat Lowe for her invaluable editorial contribution. The research reported here was supported by PhD studentships awarded by the University of Wales, Bangor, and the Economic and Social Research Council to Fay Harris and Valerie Randle, respectively.

Requests for reprints should be addressed to C. Fergus Lowe, School of Psychology, University of Wales Bangor, Brigantia Building, Penrallt Road, Bangor, Gwynedd LL57 2AS, United Kingdom (e-mail: c.f.lowe@bangor.ac.uk).

explain the occurrence of these new and apparently unreinforced responses, which are definitive of categorization, that poses the major theoretical challenge.

The pioneering research of Sidman and colleagues (Sidman, 1994, 2000) has focused attention on the defining features of categorization and, particularly, on bidirectional or symmetrical responding as a key characteristic of such behavior. After almost 20 years of research on stimulus classes, however, there is still no clear consensus about how they are established (K. J. Saunders, Williams, & Spradlin, 1996). It appears that arbitrary stimulus classes may be readily established in humans (e.g., Fields, Reeve, Varelas, Rosen, & Belanich, 1997; Pilgrim, Chambers, & Galizio, 1995; Schenk, 1995; Sidman, 1994) although particular training procedures result in either their failure to form (Fields, Reeve, Rosen, et al., 1997; R. R. Saunders, Drake, & Spradlin, 1999) or in limitations in their untrained within-class substitutability (Fields, Landon-Jimenez, Buffington, & Adams, 1995; R. R. Saunders & Green, 1999; Spencer & Chase, 1996). Furthermore, for reasons yet to be determined, there are some conditions in which class formation unravels and shows conflicting patterns of within-class substitutability, in the match-to-sample performances of both adults (Pilgrim & Galizio, 1995) and children (Pilgrim et al., 1995). In contrast to the innumerable instances, with many variations, of arbitrary stimulus categorization in humans, there is still little evidence that nonhuman animals show such behavior (see Carr, Wilkinson, Blackman, & McIlvane, 2000; Dugdale & Lowe, 2000; Hayes, 1989; Horne & Lowe, 1996, 1997; Lowe & Horne, 1996; K. J. Saunders, 1989; K. J. Saunders et al., 1996; Urcuioli, 1996).

A number of experimental variables have been shown to promote categorization of arbitrary stimuli by humans. These include the use of (a) auditory–visual, as opposed to visual–visual, baseline conditional relations (Green, 1990), (b) nameable as opposed to difficult-to-name stimuli (Mandell & Sheen, 1994), (c) stimuli with names that rhyme along common class lines (Randell & Remington, 1999), and (d) training either common or intraverbal names for the stimuli (Beasty, 1987; Dugdale & Lowe, 1990; Eikeseth & Smith, 1992; Horne & Lowe, 1996; K. J. Saunders, Saunders, Williams, & Spradlin, 1993). None of these studies, however, has shown that any one of these variables is *necessary* for categorization to occur.

Theoretical explanations of arbitrary stimulus class formation include the stimulus equivalence account of Sidman (1994, 2000), the relational frame account of Hayes and Hayes (1989, 1992; Hayes, 1996) and the naming account of Horne and Lowe (1996, 1997; Lowe & Horne, 1996). The experiments reported in the present paper arose from the latter account, which proposes that verbal behavior, particularly naming, may be critical for the establishment of arbitrary stimulus classes. According to Horne and Lowe (1996), naming is a circular relation among an object (or event), a particular speaker behavior, and the corresponding listener behavior. Echoic reproduction of the speaker behavior, which is initially evoked by the object, also occurs within the name relation, particularly when doing so generates reinforcing consequences (Horne & Lowe, 1996, p. 203). A pivotal feature of this account is that a variety of stimuli may occasion the same speaker–listener behavior and thereby become members of a common name relation. Such stimuli may be physically similar or different from one another. For example, following appropriate tact training, the response "ball" may be occasioned by stimuli as different as a ball, the written letters BALL, or a symbol such as ∉. Given that, with respect to each of these stimuli, listener behavior to the auditory stimulus "ball" has been established, the conditions are such that if any one of these stimuli evokes the common speaker behavior "ball," the common listener behavior may also occur; that is, having heard /ball/,[1] the speaker–listener is likely to orient not just to one but to all of the stimuli (the ball, BALL, and ∉). Thus, regardless of which particular stimulus in the name relation first evokes the speaker behavior, the effect of that speaking may be to "carry the listener back" to the objects that feature in the same name relation. It may be said then that, through the operation of a name relation, subjects may

[1] To distinguish a vocal response from the auditory stimulus it produces, we indicate the former as, for example, "ball" and the latter as /ball/ (see Horne & Lowe, 1996, p. 190).

respond to a particular stimulus in terms of the category of stimuli to which that particular stimulus belongs (Mead, 1934, pp. 83–100, 190–191; Vygotsky, 1934/1987, pp. 163, 285). For this reason, we have proposed that naming *is* categorizing (Horne & Lowe, 1996, p. 227).

It follows directly from this account that to establish arbitrary stimulus classes, the child may need only learn to name each member of the designated class using a common name (or other particular forms of naming, see Horne & Lowe, 1996, pp. 218–222). For this reason we have devised a new training procedure that does not rely on the training of conditional discriminations between class exemplars or match-to-sample techniques, and a new test, *category match to sample*, to determine whether arbitrary stimulus classes have been established. In these experiments our initial aim was to train 2- to 4-year-old children a common tact in response to seeing each exemplar of a designated stimulus class. Two three-member sets of arbitrary stimuli were employed, to one of which sets the children were trained the common tact ''zag''; to the other, the children were trained ''vek.'' Having undergone the tact training procedure, the children were then tested for categorization. Though we explicitly trained only tact relations, there is a great deal of theoretical and empirical evidence showing that training tact relations in developmentally normal children of this age also entails the concomitant training of corresponding listener behavior (see Horne & Lowe, 1996, 1997); that is, in training tacting, one is effectively training naming.

Rather than separately testing the relatedness of each of the stimuli to each of the others in a series of standard conditional discrimination tests, our category test simultaneously measures relatedness across all the stimuli in each trial. The child is presented with all of the training stimuli; the experimenter then selects one of these as sample and asks the child to look at it and then to give ''the others.'' In the next trial, the experimenter selects another of the stimuli as sample, and so on, until each of the stimuli has served in turn (for a related procedure, see Pilgrim & Galizio, 1996, p. 188). Throughout these unreinforced category test trials, the child is free to choose from all the stimuli (of whatever number) that have featured in training.

In the experiments reported here, the category test was conducted under two different instructional conditions. In the first, the child was asked only to *look* at the sample and to give the others. In the second condition, which was given to any child who failed to categorize appropriately in the first, the child was shown the sample and asked, ''What is this?'' (i.e., to respond as a speaker to the sample by saying ''zag'' or ''vek'') and was then asked to give the others (i.e., to respond as a listener to his or her own speaker behavior by selecting the other zags or veks). To ensure that the children understood these instructions, they were given several training and test sessions with everyday objects prior to the introduction of the arbitrary stimuli. In Experiment 1A, the two three-member common tact relations were trained and category match-to-sample behavior was tested. In Experiment 1B, to further investigate the potential size and duration of such categories that may be learned, 2 of the children who had previously participated in Experiment 1A were taught two additional three-member tact relations; this time they were taught the same tact responses as had been trained in Experiment 1A but to six new stimuli, each physically different from any of the others, the procedure thus generating two six-member common tact relations. These children's categorizing of the new six-stimulus set and then of all 12 stimuli combined was tested and maintenance of categorization was determined at a 6-week follow-up.

EXPERIMENT 1A

METHOD

Subjects

Subjects were 5 girls (GH, FK, JC, EB, and JR) and 6 boys (BJ, CS, JA, WA, RL, and GF) who attended the Daycare Nursery and Center for Child Development at the University of Wales, Bangor. As shown in Table 1, the children were aged from 2 years 3 months to 4 years 3 months old at the start of the study; none had any previous experimental history of conditional discrimination training, and all had normal scores on the Griffiths Mental Development Scales (Griffiths, 1954).

Table 1

Participants' gender and age, and number of sessions. Also shown for each child is the general quotient score on the Griffiths Mental Development Scales (GMDS).

Subject	Gender	Age at start (years/months)	Sessions: Everyday objects	Sessions: Arbitrary stimuli	Age at testing (years/months)	GMDS general quotient
GH	F	2/3	4	29	2/5	147
FK	F	2/6	5	46	2/10	129
JC	F	2/9	4	33	3/1	133
BJ	M	2/9	6	24	3/0	116
CS	M	2/9	5	10	2/11	130
JA	M	2/10	6	30	3/3	127
EB	F	2/11	3	23	3/6	112
RL	M	3/0	2	22		128
GF	M	3/5	3	13		128
WA	M	3/8	4	21	4/1	122
JR	F	4/3	1	3	4/3	131

Apparatus and Stimuli

Experimental sessions were conducted in the daycare center in a purpose-built experimental room fitted with two wall-mounted color videocameras with built-in directional microphones that were controlled remotely from a central audio-visual console room. A radio microphone was also worn by the experimenter. The child and experimenter sat on opposite sides of a small red table. In some sessions (see Procedure), a wooden screen called "Teddy's house" (height 70 cm, width 50 cm, depth 5 cm), supported on two wooden blocks (30 cm by 10 cm by 5 cm), was placed on the table midway between the child and experimenter. The top section of the screen contained a clear plastic window (height 50 cm, width 40 cm), covered with a net curtain on the experimenter's side, through which the experimenter could unobtrusively observe the child's responses. When the screen was in place, the experimenter presented stimuli to the child through an aperture (height 20 cm, width 50 cm), partly covered by a red and green crepe paper fringe, in the lower section of the screen. For 2 of the children it was necessary to use alternative screens between child and experimenter; in the one case this took the form of a large teddy bear, and in the other, a white card (21 cm by 30 cm). A large toy (usually a teddy bear) was seated on a chair to the left of the child or, alternatively, a teddy bear hand puppet was worn by the experimenter. The experimental stimuli were of two kinds: (a) six everyday objects, each of a different color, consisting variously of three different types of child's hat and three different types of child's cup (see Figure 1, top row); and (b) six different arbitrary green wooden shapes randomly assigned to each child from a pool of 12 such shapes (see Figure 1, bottom two rows). The shapes were approximately 10 cm long, 0.9 cm in depth, and varied in width from 4 cm to 7.5 cm. The scheduled reinforcer was social praise and, occasionally, stickers. At the end of each session, the child was given either a small toy of some kind, or stickers, or access to a large "speaking" teddy bear.

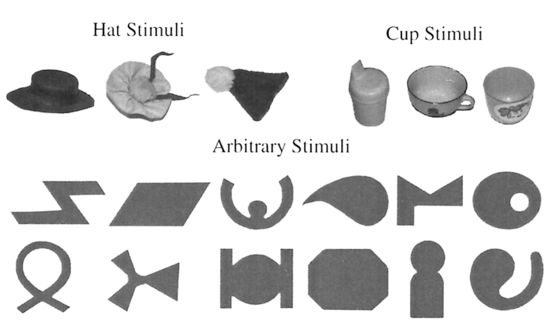

Fig. 1. Top row shows the everyday objects, three differently colored hats and cups. Bottom two rows show the 12 arbitrary stimuli, green wooden shapes.

Procedure

Everyday objects. The experimenter first established good rapport with the children during unstructured daily play sessions in the preschool playroom of the daycare center. Children were then taken one at a time to the experimental room and seated at the table across from the experimenter. The six everyday objects, three hats and three cups, were placed in a predetermined randomized spatial array on the table. The child was introduced to ''Teddy,'' which took the form either of a hand puppet on the hand of the experimenter or of a large teddy bear placed on a chair to the left of the child. The experimenter provided the following instruction: ''Teddy has come to school with you. He wants to learn the names of things. Can you help him? He would like to know what these [indicated hats and cups] are called.'' The experimenter then pointed once, in a predetermined random order, to each of the six stimuli in turn and said, ''This is a hat [cup].'' The experimenter then removed the six stimuli from the table. For each child, the experimenter randomly separated the three hats and three cups into three different training pairs, each consisting of one hat and one cup. Each child's training pairs remained constant throughout the duration of the study. For the purpose of reporting and scheduling stimulus presentations, the first training pair were designated Hat 1 and Cup 1, the second as Hat 2 and Cup 2, and the third as Hat 3 and Cup 3. In this and other conditions throughout the experiment, sessions varied in duration from 15 to 30 min. The children's overt verbal behavior was recorded throughout the study.

Tact overtraining in pairwise trials. The experimenter conducted two test trials to confirm that each child could produce the appropriate echoic responses when asked, ''Can you say hat [cup]?'' The experimenter, having placed Hat 1 and Cup 1 on the table in such a way that one stimulus lay 10 cm to the left of the child's midline and the other 10 cm to the right, then pointed to one of the stimuli and asked, ''What is this? Can you tell Teddy what this is?'' If the child produced the correct vocal tact response (''hat'' for Hat 1 or ''cup'' for Cup 1), the experimenter responded, ''Yes, clever girl [boy]! It is a hat [cup].'' If the child produced an inaccurate tact response or remained silent, the experimenter provided the following corrective instruction: ''This is a hat [cup]. Can you say it?'' On any one trial, only one stimulus in the pair was targeted. Trials were arranged in eight-trial blocks in which Hat 1 and Cup 1, in a predetermined quasirandom order, were each targeted twice on the right of the child's midline and twice on the left. The learning criterion was three of four correct responses to each target stimulus within any sequence of eight consecutive trials. Once criterion was met for Pair 1, the same training procedure was repeated for Pairs 2 and 3. Tact training was completed when responding was at criterion for all three pairs.

Listener overtraining in pairwise trials. These were similar to tact trials except that, when a pair of stimuli was placed in front of the child, the experimenter asked, ''Where is the hat [cup]? Can you give Teddy the hat [cup]?'' If the child pointed to the hat or cup or picked it up, the experimenter responded, ''Clever girl [boy]! Yes, that's the hat [cup].'' If the child made no response or selected the wrong stimulus, the experimenter provided the corrective instruction, ''No, this [pointing to the correct stimulus] is the hat [cup]; this [pointing to the incorrect stimulus] is the cup [hat].'' The learning criterion was the same as for tact training. The same procedure was then repeated with Pair 2 and then with Pair 3. Listener training ceased when responding was at criterion for all three stimulus pairs.

Tact overtraining in six-stimulus trials. The three pairs of stimuli were then combined in a six-stimulus array, and tacting of each hat and cup stimulus was tested and, if necessary, trained, in this context. At this point, to ensure that the child's responses were not cued by the experimenter, the wooden screen was placed on the table, midway between the experimenter and child. On each trial, the experimenter reached through Teddy's door in the lower section of the screen and placed the six everyday objects on the table on the child's side of the screen. The three hats and three cups were randomly ordered in two rows, 7 cm apart, midway between the screen and the child's edge of the table (and equidistant from either side of the table) with approximately 10 cm between each stimulus. The experimenter pointed to each stimulus in turn, asking in each case, ''What is this? Can you tell Teddy what this is?'' Depending on whether the child responded correctly or incorrectly, the experimenter gave either social praise or corrective feedback as in the pairwise tact training trials. After each stimulus in the array had been separately targeted, the experimenter removed all six from the table, replacing them in different locations at the start of each new trial. Trials continued until the child had produced three consecutive correct responses for each of the six stimuli.

Category match-to-sample training. The screen was used throughout the category match-to-sample training phase designed to establish the child's correct responding to each of three different instructions. The first of these was the *tact-sample match-to-sample-tact* instruction. On each trial, the experimenter placed the six everyday objects at predetermined locations on the table in front of the child and then picked up one at random and asked, ''What is this? Can you tell Teddy what this is?'' If the child produced the correct response (i.e., either ''hat'' or ''cup''), the experimenter then asked, ''Can you give Teddy the other hats [cups]?'' If, when given a hat as the sample, the child selected the two other hats as comparisons, this constituted a correct category sort and the experimenter gave verbal praise; likewise, given a cup as sample, the child was required to select the other two cups. If the child responded by selecting a wrong comparison stimulus, the experimenter provided the following corrective verbal feedback, ''No, this [pointing to sample] is a hat [cup]. This [pointing to incorrect comparison stimulus] is a cup [hat].'' If the child selected only one correct comparison stimulus, the experimenter prompted a second response by saying, ''Are there any more?'' If, in response to this question, the child selected the second correct comparison, the experimenter provided verbal praise as before. If the question prompted no second response from the child, the experimenter conducted further six-stimulus tact training trials to criterion before resuming category match-to-sample training and testing. To ensure that the number of comparison selections was not cued by the question, the experimenter asked it again following any prompted second stimulus selection. Trials were repeated until the child performed three consecutive correct trials per category.

The *tact-sample match-to-others* instruction was then introduced. Trials were conducted in the same way as for the previous instruction except that at the start of each trial the child was now asked, ''What is this? Can you give Teddy the others?'' That is, this second instruction, unlike the first, did not specify the category to which the sample belonged. Trials were repeated until the child performed three consecutive correct sorts per category.

In the case of the third instruction, *look-at-sample match-to-others,* the experimenter said to the child, ''Look at this. Can you give Teddy the others?'' Unlike the two previous category match-to-sample instructions, the child was not this time required to tact the sample but only had to look at it and then select ''the others.'' Trials were otherwise conducted in the same manner as for the first and second instructions, and categorizing was tested to the same criterion.

Arbitrary stimuli: Tact training in pairwise trials. For each child, a set of six stimuli was selected at random from the pool of 12 wooden shapes (see Figure 1). The nonsense names ''zag'' and ''vek'' were selected as suitable target tact responses because developmental studies (e.g., Vihman, 1996) of children's minimal echoic repertoires suggest that English-speaking 2-year-olds can produce a reasonably accurate echoic response (and so, potentially, the corresponding tact) to each of these verbal stimuli. The six wooden shapes allocated to each child were separated into three training pairs; the children were then trained to tact one member of each pair ''zag'' and the other ''vek.'' In reporting stimulus presentations, the first training pair are designated as Zag 1 and Vek 1, the second pair as Zag 2 and Vek 2, and the third pair as Zag 3 and Vek 3.

To introduce the child to the target tact responses ''zag'' and ''vek,'' a block of eight echoic pretraining trials was presented for Pair 1 (Zag 1 and Vek 1). These trials were conducted in the same way as for the everyday objects pairwise tact training trials except that in this case the experimenter pointed to the appropriate stimulus, and said, ''This is a zag [vek]. Can you tell Teddy what this is?'' If the child produced the correct vocal response, the experimenter responded, ''Yes, clever girl [boy]. It is a zag [vek].'' If the child produced an inaccurate echoic response or remained silent, the experimenter pointed to the appropriate stimulus and said, ''This is a zag [vek]. Can you say it?'' If the child produced a good echoic approximation to ''zag'' and ''vek'' in all eight trials, tact training was then introduced. Children who failed to echo reliably were presented with additional pretraining trials and did not proceed to tact training until their echoing of the target tact responses met the criterion. Each child's tact responses of ''zag'' to Zag 1 and ''vek'' to Vek 1 were trained in blocks of eight in the same way as for the everyday objects except that, in the case of these trials, the learning criterion for each stimulus was three consecutive correct responses out of four in each block of eight trials. The same procedure (i.e., echoic pretraining trials followed by tact training) was repeated with Pair 2 (Zag 2 and Vek 2) with the addition that, during Pair 2 tact training, performance on Pair 1 was reviewed and, if necessary, retrained to ensure that tacting was continually maintained at criterion for this pair.

Tacting in four-stimulus test trials. In these trials the screen was introduced between child and experimenter. In each trial the four stimuli making up Pairs 1 and 2 were placed on the table in predetermined positions that varied randomly from trial to trial. The experimenter then pointed to each stimulus in turn and asked, ''What is this? Can you tell Teddy what this is?'' Correct responses were not reinforced and no corrective feedback was given. When the child produced the correct tact responses in three consecutive trials for each of the four stimuli, tact training for the Pair 3 stimuli began. (One child, CS, was omitted from this stage of the experiment and proceeded directly to Pair 3 tact training.) Children whose performance in the four-stimulus test was below criterion received further pairwise testing and, if necessary, retraining on Pair 1 or Pair 2 stimuli and were then retested on all four stimuli. This procedure was repeated until criterion performance was attained. Any child who failed to meet criterion after 24 four-stimulus test trials had been presented did not proceed to the next stage of the study.

Pair 3 tact training trials. Once criterion performance in the four-stimulus test was established, tact training trials were conducted without the screen in place for Pair 3 (Zag 3 and Vek 3) in the same way as for Pair 1 and Pair 2.

Tact training in six-stimulus trials. When the child's tact responses to Pair 3 stimuli were at criterion, the six arbitrary stimuli that together constituted Pairs 1, 2, and 3 were presented as a group, in a predetermined random spatial array, with the screen in place. The experimenter pointed to each stimulus in turn, asking, ''What is this? Can you tell Teddy what this is?'' If the child responded correctly, verbal praise was given; if the child responded incorrectly the experimenter provided corrective feedback, as in the pairwise tact training trials. When each stimulus had been so targeted, all six were removed from the table and then replaced at different locations before the experimenter once again proceeded to target each in turn. The learning criterion was three consecutive correct tact responses to each of the six stimuli. When the criterion had been met for all six stimuli, the probability of reinforcement was reduced to 50%, then 33%, and, finally, 0%. At each reinforcement level, testing was continued until the criterion was met. If performance fell below criterion on the six-stimulus test during trials with a reduced level of reinforcement, 100% reinforcement was reintroduced; if criterion performance was not reinstated, tacting for Pairs 1 and 2, singly and combined, and for Pair 3, was retested and, if necessary, retrained. Tact testing for all six stimuli was then implemented once again until criterion performance was reached for the six-stimulus tests in the absence of reinforcement, or until the child had failed to meet the criterion after being presented with 48 six-stimulus training and test trials. In the case of 1 child, EB, who showed evidence of failing to discriminate between Zag 2 and Vek 3 in the six-stimulus test trials, the Pair 2 stimuli were replaced with new stimuli for the remainder of the experiment. Pairwise training to criterion was first conducted with replacement Pair 2, followed by four-stimulus and then six-stimulus test trials as described above.

Category match-to-sample tests. These tests were carried out in four steps, each entirely in the absence of reinforcement and differential feedback. In all cases the possibility of inadvertent experimenter cueing was eliminated by employing either (a) a second experimenter who was familiar to the child but had no knowledge of the particular relations that had been established or of the purpose of the experiment, or (b) a screen to obscure the experimenter's face and body, or both.

Step 1 consisted of an *arbitrary six-stimulus tact review,* in which the second ("blind") experimenter was introduced to the child by the first experimenter at the beginning of the first session: "[Second experimenter's name] has heard how clever you are and would like to see what you do in here. Would you teach her, as well as Teddy?" The second experimenter then sat in the first experimenter's usual position behind the screen. The first experimenter, sitting to the right of the child, conducted one six-stimulus tact test; the criterion required was correct tacting of all six stimuli. If the child performed correctly, the first experimenter then left the room and observed all subsequent trials from the console room or, if the child became apprehensive at this point in the procedure, remained in the experimental room sitting behind the child.

Step 2 was an *everyday objects category match-to-sample review* in which the second experimenter, using the look-at-sample match-to-others instruction, conducted four category sorting test trials with the six everyday objects, two trials for hats and two for cups. The criterion was correct category sorting of the hats and cups, respectively, in all four trials. Progress to Step 3 was dependent on the child's responding at criterion level in both Steps 1 and 2; if performance fell below criterion in either of these steps, the session was terminated and the first experimenter conducted retraining until criterion performance was regained, at which point testing resumed.

Step 3 was the *arbitrary stimuli category match-to-sample Test 1.* These trials were conducted by the second experimenter using the look-at-sample match-to-others instruction. In each test trial the child was presented with all six stimuli; the experimenter selected a randomly predetermined one of these and asked the child, "Look at this. Can you give Teddy the others?" If the child, when presented with a zag (or vek) stimulus as sample selected the remaining two zag (or vek) comparisons from the five-stimulus display, the selection qualified as a correct category sort.

Test trials were classified as either valid or null. A valid trial consisted of the child selecting from one to four of the stimuli available in the five-stimulus array. When a child stopped selecting comparisons the experimenter asked, "Are there any more?" To ensure that the number of responses was not cued by the experimenter, this question was repeated following every subsequent stimulus selection until the child indicated that he or she did not wish to select any others. Null trials were those in which the child selected all five comparison stimuli. If a null trial occurred the child was instructed as follows: "Teddy doesn't want all of them, only some." The trial was then repeated. If the child continued to select all of the stimuli, the session was terminated and the first experimenter then repeated Step 1 and Step 2 procedures, after which Step 3 was reintroduced.

The outcome of the first six valid test trials determined whether the child was presented with a further 12 test trials under the Step 3 instructional conditions or proceeded to category match-to-sample testing under the different instructional conditions of Step 4. If, in the first six valid trials, the child produced at least one correct category sort for each common tact (zag or vek) category, the Step 3 procedure continued for a total of 18 valid test trials, during which each of the six stimuli served as sample three times. The criterion for success on this test was set at four of nine correct sorts per common tact category (the probability of obtaining four or more correct sorts by chance is .008). Any child who failed to produce at least one correct category sort for each common tact (zag or vek) relation in the first six valid trials proceeded to Step 4. For some children (particularly the older ones) it was possible to carry out a session in which all 18 trials were presented; for others a number of shorter sessions were required. In the latter case, each new session began with Step 1 and Step 2 before Step 3 or Step 4 (see below) testing resumed.

Step 4 was the *arbitrary stimuli category match-to-sample Test 2.* Step 1 was repeated and then followed by the Step 2 everyday objects category match-to-sample review, but on this occasion using the tact-sample match-to-others instruction "What is this? Can you give Teddy the others?" This was followed by category match-to-sample tests with the arbitrary stimuli, once again using the tact-sample

Table 2

Use of screen and "blind" experimenter (E) during category MTS test trials. Y indicates presence and N indicates absence of each device.

	Type of sorting instruction			
	Look at sample		Tact sample	
Subject	Screen	Blind E	Screen	Blind E
GH	Y	Y		
FK	Y	N	Y	N
JC	Y	Y	Y	Y
BJ	Y	N	Y	N
CS	Y	Y		
JA	Y	Y	Y	N
EB	Y	N	N	Y
WA	Y	N		
JR	Y	Y		

match-to-others instruction, but otherwise recapitulating the Step 3 procedure.

Procedural exceptions. In some of the test conditions, some of the children were apprehensive or distracted from the experimental task when the cueing control procedures (i.e., the blind experimenter or the wooden screen) were introduced. In one such case (EB), to reestablish reliable test responding, the first experimenter used a large white card instead of the wooden screen during Step 3, match-to-sample Test 1; the blind experimenter conducted match-to-sample Test 2 in the absence of a screen. For another child, BJ, a large teddy bear was used as the screen for Steps 3 and 4. In the case of those children who did not have a blind experimenter, the final test trials were presented by the first experimenter with a screen in place; Table 2 shows for each child whether a screen or blind experimenter (or both) was employed in each of the two types of category test.

Interobserver reliability. An independent observer scored a randomly selected 40% of training and test trials. Interobserver agreement was 99%.

Results

Everyday Objects

Table 1 shows for each child the total number of training and test sessions conducted with everyday objects and arbitrary stimuli, respectively.

Tact and listener behavior overtraining. In the pairwise tact trials, all the children met the criterion within the first four to five presentations of each stimulus. Of the 11 children, 7 met the listener behavior criterion within the first eight-trial block for each pair; the remaining 4 (GH, FK, BJ, and CS) required one to three additional trials for one or more of the pairs. Nine of the children met the six-stimulus test criterion in the first three trials; JC and CS required four trials.

Category match-to-sample training. When the tact-sample match-to-sample-tact instruction was employed, 6 children responded correctly on all trials in the first three-trial block; the remaining 5 required a further such block. Seven of the children immediately met the criterion with the tact-sample match-to-others instruction, and the remaining 4 (FK, CS, JA, and GF) required one or two further trials. The criterion for the look-at-sample match-to-others instruction was met immediately by 4 children (CS, EB, JR, and RL), and the remaining 7 required one to three additional trials.

Arbitrary Stimuli

Number of trials to criterion for each child's pairwise, four-stimulus, and six-stimulus tact training are shown in Table 3.

Tact training in pairwise trials. Following one block of eight echoic pretraining trials in the context of the Pair 1 stimuli, 5 children (CS, WA, EB, GF, and JR) produced good echoic approximations to the experimenter's utterances of "zag" and "vek." The remaining children required 8 to 72 additional trials before the echoics were established. All the children echoed "zag" and "vek" appropriately within one block of Pair 2 and one block of Pair 3 echoic pretraining trials. Table 3 shows that the children varied considerably in the number of trials taken to achieve criterion tact performance for each stimulus pair; for Pair 1, this ranged from 6 to 160 trials. Most children required fewer training trials to achieve criterion performance on Pair 2, ranging from 6 to 23.

Tacting in four-stimulus test trials. Subjects BJ, JA, WA, JR, and EB achieved criterion performance in the minimum of three trials, and GH, FK, JC, RL, and GF required retraining on Pair 1 or Pair 2 (or both). (Subject CS was omitted from this particular stage.) Despite receiving 120 retraining trials on Pair 1 and 168 retraining trials on Pair 2 stimuli, GF failed to achieve criterion responding on the

Table 3

For each child, the number of trials to meet the tact training criterion in each training phase with the arbitrary stimuli. Superscript R indicates trials for EB with a replacement stimulus pair. F indicates the point at which training was terminated and the subject failed to achieve criterion (numbers in parentheses indicate number of trials completed in that condition). W indicates 1 child's withdrawal from the experiment.

	Pairwise			Pairwise review		Pair-wise		Pairwise and 4 stimulus review trials			
Subject	Pair 1	Pair 2	4 stimulus	Pair 1	Pair 2	Pair 3	6 stimulus	Pair 1	Pair 2	4 stimulus	Pair 3
GH	120	7	6	16	8	10	23	16	16	2	16
FK	160	7	6	48	64	7	48	32	16	2	40
JC	12	6	12	32	24	7	38	8	8	1	24
BJ	45	6	3			6	41	8	24	3	16
CS	12	21				6	12				
JA	8	6	3			6	48	24	32	4	40
EB	142	23	3			6	F(48)	16	120	15	136
EB^R		6	3				12				
RL	49	6	9	112	112	14	W(18)	24	24	18	24
GF	13	8	F(25)	120	168						
WA	32	6	3			7	15	24	24	3	32
JR	6	7	3			6	12				
M	54.5	9.9	5.3			7.5	29.6				

four-stimulus test; his participation in the study ended at this point. That is, for 5 of the 11 children, the discrimination learning achieved in pairwise training broke down when the pairs were combined in a four-stimulus array. For all but GF, however, pairwise review training was successful in establishing criterion responding.

Pair 3 tact training. All children succeeded in meeting the criterion within 6 to 14 trials. Evidence of "learning to learn" (Harlow, 1949) comes from mean scores across subjects; it took an average of 54.5 training trials to reach criterion with Pair 1 stimuli, 9.9 trials with Pair 2, and 7.5 with Pair 3.

Tact training in six-stimulus trials. Except for RL (who was withdrawn before completing this phase for reasons unconnected with the study) and EB (see below), all the children succeeded in meeting the criterion. The number of six-stimulus trials conducted with each subject ranged from 12 to 48. Only 2 (CS and JR) of the 10 children showed no deterioration in discrimination learning when all six stimuli were combined for the first time in the six-stimulus tact test. With the exception of EB, pairwise and four-stimulus review trials together with six-stimulus training trials were sufficient to establish criterion six-stimulus tacting in the 9 children who continued to participate in this training phase. Subject EB failed to reach criterion on the 100% reinforced six-stimulus trials; extensive review trials for Pair 2 (120) and Pair 3 (136) failed to establish effective discrimination between Zag 2 and Vek 3. Accordingly, this child was allocated replacement Pair 2 stimuli, and she met the pairwise tact criterion with the new Pair 2 in the minimum of six trials (Table 3). Pairwise review trials showed that, throughout replacement Pair 2 training, EB maintained criterion responding for Pair 1 and also met criterion in the minimum three trials in the subsequent four-stimulus tests. She maintained criterion responding in pairwise review trials for Pair 3 and met the tact criterion in the six-stimulus tests within the minimum number of 12 trials.

Category match-to-sample tests. Figure 2 (gray bars) shows that under the look-at-sample match-to-others instruction, 4 of the children (GH, CS, WA, and JR) succeeded in producing at least one correct category sort for each common tact category (zag or vek) and so proceeded to the full 18-trial test. All 4 of these subjects met the test criterion of four of nine correct sorts for each common tact category. Indeed, 3 of the 4 performed all 18 trials errorlessly; WA performed 15 correct of 18. It can therefore be concluded that these 4 children had learned the zag and vek categories. Of the remaining 5 children (shown

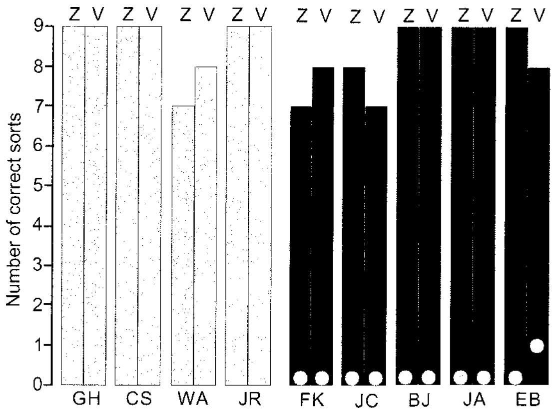

Fig. 2. For each subject, number of correct category sorts to each zag (Z) and vek (V) sample. Sorting under the look-at-sample instruction is shown by gray bars for subjects who completed all 18 test trials and by gray filled circles for subjects who failed to meet criterion in the first six trials. The performances of these latter subjects on the 18 trials under the tact-sample instruction are shown by the black bars.

by gray circles), 4 failed to make any correct sorts, and 1 (EB) made just one correct vek sort. They proceeded to the second categorization test. Under the tact-sample match-to-others instruction (black bars), all 5 subjects succeeded, with 15 to 18 correct category sorts of 18 (see Figure 2). Taking both categorization tests together, all 9 children who were taught common tacts for the two sets of three arbitrary stimuli succeeded in categorizing these stimuli correctly.

Verbal Behavior

In general, the children did not produce much task-related verbal behavior apart from that requested by the experimenter. There were, however, a number of instances of unprompted tacting of the sample or comparison stimuli. For example, in several trials of the category match-to-sample Test 1, GH and WA, although only asked to look at the sample, also produced the appropriate tact before going on to correctly sort the stimuli. Similarly, in category match-to-sample Test 2, FK and JC tacted each comparison stimulus as they selected it and, following a correct sort to a zag sample, JC also tacted the category to which the remaining three vek stimuli belonged saying, "There's only veks now."

Discussion

All 9 children, ranging in age from 2 years 3 months to 4 years 3 months, who completed tact training went on to pass one or the other of the two arbitrary stimulus category tests. These tests were conducted entirely in the absence of reinforcement; to exclude inadvertent cueing (see McIlvane & Dube, 1996), test sessions for all children had a screen (8 subjects) or a second blind experimenter (5 subjects). On those tests in which categorization was successful, it occurred immediately and with relatively few errors. These data show that common tact training may be an effective means of establishing arbitrary stimulus classes.

One of the most interesting features of these results, however, was that 5 of the 9 children failed the look-at-sample version of the test but passed when they were required to tact the sample prior to selecting the other stimuli. This shows that tact training alone may not always be sufficient for successful categorization and that it may be necessary to ensure that the subject first emits the appropriate speaker behavior (e.g., "zag") when he or she sees the sample (a zag) before he or she can successfully select the others.

Of course, although the 4 children who succeeded in categorizing successfully under look-at-sample instructions were not explicitly required by those instructions to emit speaker behavior, that does not mean that they did not do so, either overtly or covertly. For example, on several of these test trials, GH and WA were heard to tact the sample stimulus before sorting the comparison stimuli. On the second category test, some of the children tacted the comparison stimuli individually (FK, JC) and as a class (JC).

An important issue in experimental studies of young children's behavioral repertoires is that of controlling for their understanding of the task instructions. Accordingly, in this study, the children were given preliminary training using familiar everyday objects (hats and cups) in tasks that were later employed with novel arbitrary stimuli. None of the children could proceed to any of the category test trials without first passing the everyday objects categorizing review trials in which they were required to categorize the everyday objects in response to exactly the same instructions as those employed in the two arbitrary stimulus category tests that followed. Because all of the children responded correctly to these instructions in pretraining and pretesting, this effectively rules out an explanation that attributes failure on arbitrary stimulus tests under the look-at-sample condition to a failure to understand the test instruction.

Neither can it be claimed that whether the children succeeded or failed under the look-at-sample condition might be age related. There was, in fact, considerable overlap between the group that failed and the group that succeeded, both in age and in scores on the Griffiths Mental Development Scales. At the time of testing, the ages of the 5 who failed ranged from 2 years 10 months to 3 years 6 months (with Griffiths scores ranging from 112 to 133), whereas the ages of those who succeeded ranged from 2 years 5 months to 4 years 3 months (with Griffiths scores ranging from 122 to 147). These data suggest that age-related factors do not account for the differences observed.

The next experiment was designed to determine whether verbally driven categorization shown by young children could be extended from the three-member classes of arbitrary stimuli established in this experiment to six-member classes. The children were interviewed at the end of testing and, to determine whether categorizing behavior would remain intact, they were tested again 6 weeks after the initial 12-stimulus category tests.

EXPERIMENT 1B

Method

Subjects

Subjects were 1 girl (JR) aged 4 years 3 months and 1 boy (CS) aged 3 years 1 month, both of whom had participated in Experiment 1A and had passed category match-to-sample Test 1.

Apparatus and Stimuli

The experimental setting was the same as in Experiment 1A. From the pool of 12 arbitrary stimuli (see Figure 1), each child was randomly assigned six new stimuli (i.e., Set 2). The Set 2 stimuli were divided randomly into three training pairs. For purposes of reporting and scheduling stimulus presentations, members of the first training pair were designated as Zag 4 and Vek 4, the second as Zag 5 and Vek 5, and the third as Zag 6 and Vek 6.

Procedure

Arbitrary stimuli: Tact training with Set 2 stimuli. The procedure was the same as for the Set 1 arbitrary stimuli in Experiment 1A. The child was taught to produce the tact response "vek" to each of the Set 2 stimuli designated as veks and the tact "zag" to each of the Set 2 zag stimuli. As in Experiment 1A, tact training was conducted first for Pair 4, then for Pair 5, for all four stimuli (Pairs 4 and 5 combined), for Pair 6, and lastly for Pairs 4, 5, and 6 combined.

Category tests: Set 2. Categorization test trials were conducted in exactly the same way as for the Set 1 arbitrary stimuli in Experiment 1A.

Pairwise tact review trials: Set 1 and Set 2. Each of the six pairs of stimuli (Pair 1 through Pair 6) was presented separately, and tact responding to each stimulus was reviewed in one block of eight trials for each pair. If the criterion of three consecutive correct trials with both the zag and vek stimuli in each stimulus pair was not met, then that pair was re-presented, and retraining and retesting were conducted until criterion was reached.

Six-stimulus tact review trials: Set 1 and Set 2. Performance in six-stimulus tact test trials for Set 1 was reviewed without reinforcement or corrective feedback and using the criterion of three consecutive tact responses to each of the six stimuli. Any child whose responses in the six-stimulus tests were below criterion received further pairwise testing and, if necessary, retraining for the three separate pairs of stimuli. Each child was then retested on the array of six stimuli; if the criterion was not met, the six-stimulus training procedure with reinforced responding that was employed in Experiment 1A was introduced, and the six-stimulus test trials were presented once again, and so on, until the six-stimulus criterion was met. The same procedure was then conducted for the Set 2 stimuli.

Twelve-stimulus tact test trials. The second (blind) experimenter placed all 12 arbitrary stimuli (Set 1 and Set 2) randomly in two rows, approximately 7 cm apart, midway between the screen and the child's edge of the table and equidistant from either side of the table. The experimenter pointed to each of the stimuli in turn and asked, ''What is this? Can you tell Teddy what this is?'' No reinforcement or corrective feedback was given; the criterion was two consecutive correct tact responses to each of the 12 stimuli.

Set 1 and Set 2 category match-to-sample review trials. The second experimenter conducted six category match-to-sample trials (each stimulus serving as sample once) with the Set 1 stimuli. In this and all subsequent categorization tests, each child sorted to the look-at-sample match-to-others instruction. Two correct sorts (one from each category) were required to show that the learning criterion was maintained. Six category sorting test trials were then presented for Set 2 in the same way as for the Set 1 stimuli.

Six-stimulus category match-to-sample test: Random selection from Sets 1 and 2. From the 12 stimuli (Set 1 and Set 2 combined) three zag and three vek stimuli were randomly selected without replacement, and with the constraint that no two stimuli from the same tact training pair could serve (e.g., if Vek 1 was selected, Zag 1 was not). Six category sorting trials, in which each stimulus served once as sample, were carried out by the second experimenter in the same manner as for the separate Set 1 and Set 2 category tests. The criterion for success was the same as for stimulus Sets 1 and 2.

Twelve-stimulus category match-to-sample test. All 12 stimuli were presented together in the same array as for the 12-stimulus tact test trials, and the second experimenter conducted 12 category sorting test trials. Selection of the five comparison stimuli that evoked the same tact response as the sample constituted a correct category sort. In each trial, the child was presented with a different sample stimulus until each of the 12 stimuli had been targeted once in this manner. The criterion for test success was set at one correct sort per category (i.e., to a vek and a zag sample, respectively). The probability of a correct sort (i.e., of all five correct comparisons being selected) to either a vek or a zag sample by chance is .002. The blind experimenter conducted all test trials with the screen in place.

Posttest interview. The first experimenter conducted two additional 12-stimulus category sorting test trials (one to a zag sample and one to a vek) without the screen in place. The experimenter asked a number of open-ended questions such as, ''How did you do that?'' or ''Why did you give Teddy these?'' or ''What should Teddy look for now?'' The child's verbal responses were recorded and the session then terminated.

Delayed posttest. This test was designed to measure maintenance of the category match-to-sample behavior and was conducted 6 weeks after the 12-stimulus category test phase. The first experimenter, with the screen in place, conducted a further session consisting of (a) three 12-stimulus tact test trials and (b) three 12-stimulus category sorting tests. In the latter, the sample stimuli were chosen randomly with the constraint that both trial types (zag and vek) had to be included.

Interobserver reliability. A randomly selected 31% of the experimental trials were included in the reliability assessment; interobserver agreement was 100%.

RESULTS

Training and testing were completed in eight sessions for CS and five for JR.

Tact training with Set 2 stimuli. The 2 children produced good approximations to the two vocalizations ''zag'' and ''vek'' in all eight tact exposure and echoic prompt trials for each of Pairs 4, 5, and 6. For Pairs 4 and 5 both children attained criterion in the minimum of six trials. Subject CS reached the four-stimulus (unreinforced) test trial criterion in the minimum of three trials, whereas JR required six four-stimulus and six pairwise review trials for both Pairs 4 and 5. Both subjects met the criterion for Pair 6 in the minimum of six trials and that for the six stimuli in the minimum of 12 trials.

Category test: Set 2. Under the look-at-sample match-to-others instruction, both children sorted the six Set 2 stimuli correctly on all nine test trials for each category (i.e., zag and vek) and thus met the categorization criterion.

Pairwise review tact trials: Set 1 and Set 2. Both subjects attained criterion in the minimum six trials for each of the Set 1 and Set 2 pairs.

Six-stimulus tact review trials: Set 1 and Set 2. Both subjects reached criterion for Set 1 and

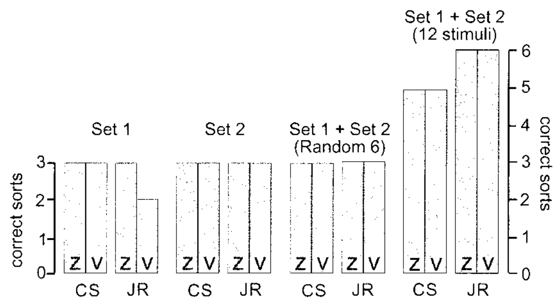

Fig. 3. For each of the 2 subjects, CS and JR, number of correct category sorts under the look-at-sample instruction in the Set 1 and Set 2 category match-to-sample review trials; in a random selection of six from the Set 1 and Set 2 stimuli (Random 6); and in the full 12-stimulus test combining all Set 1 and Set 2 stimuli.

Set 2, respectively, in the minimum three trials for each set.

Twelve-stimulus tact trials. Both subjects immediately met the criterion of two consecutive correct tact responses to each of the 12 stimuli.

Set 1 and Set 2 category match-to-sample review trials. Figure 3 indicates that, with the six Set 1 and then the six Set 2 stimuli, CS and JR maintained criterion responding. CS sorted all the stimuli correctly on the three test trials with each set; JR produced only one error, in Set 1.

Six-stimulus category match-to-sample test: Random selection from Set 1 and Set 2. Both subjects sorted the six stimuli correctly into common tact categories on all three test trials (see Figure 3).

Twelve-stimulus category match-to-sample test. Figure 3 shows that in the 12-stimulus tests, both subjects exceeded the criterion requirement of one correct sort for each category. Subject CS selected all five correct comparisons on 10 of the 12 trials, and JR's performance was errorless on all 12 test trials.

Posttest interview. On the zag trial, CS tacted first the sample stimulus and then, as he selected them, each of the five correct comparison stimuli. When given the vek sample and asked what should be looked for, he replied, "vek." JR performed on the category sorting tests without speaking. When asked why she sorted as she did she replied, "Because they are zag [vek]," as appropriate to the particular trial.

Delayed posttest. When retested after an interval of 6 weeks, CS's performance on the 12-stimulus tact test was 100% correct on the 18 trials, whereas JR made four errors in tacting the zag stimuli and 3 tacting the veks. When the children then were given the 12-stimulus category sorting trials, the performance of both was errorless on the three trials.

Discussion

This study shows that the 2 children learned, in relatively few sessions, to tact the second set of arbitrary stimuli (Set 2) and then to categorize them, without errors, under the look-at-sample instruction. When the Set 1 and Set 2 stimuli were combined in a 12-stimulus array, the children tacted them errorlessly and then went on, under the look-at-sample instruction, to categorize them with a very high level of accuracy. These data show that very young children, given an arbitrary wooden shape as sample, can select the five others, from an array of 11, for which they have been trained a tact that is common to those five and the sample alike; any of the 12 stimuli functioned as an effective sample for the selection of the remaining five in its class. That such categorization is a robust phenomenon is further indicated by the fact that 6 weeks after the completion of initial testing it was still in place, with 100% accuracy.

EXPERIMENT 2

Although Experiments 1A and 1B appear to show that common tact training may establish arbitrary stimulus classes, it is important to rule out other possible artifacts that may have contributed to test success. One such possibility concerns the manner in which tact training was conducted in these experiments. In both of them, following the pairwise tact training the stimulus pairs were combined and the subjects were required to tact each of the stimuli in the resulting six-stimulus (or 12-stimulus) array. It is possible that the spatial proximity of the stimuli, albeit random with respect to the common tact categories, together with the tact responding to each in turn, may have somehow facilitated the learning of the target stimulus classes. For example, during the six-stimulus tact test the arrangement of the stimuli from trial to trial would, on occasion, result in the child producing two or even three consecutive "zag" (or ''vek'') tact responses. A sequence of identical tact responses may have evoked the autoclitic response ''same''; the latter response, in turn, may have occasioned looking once again at the subset of stimuli that had evoked this autoclitic response (see Horne & Lowe, 1997, pp. 280–281; cf. Lowenkron, 1996). Given the failure of more than half the subjects to pass the first category test, it seems unlikely that this mechanism was in operation across all subjects during six-stimulus tact training. Nevertheless, this aspect of the procedure merits direct empirical investigation. Accordingly, in this experiment, which used two three-member stimulus classes, 3 new subjects were given only pairwise tact training; they never saw class exemplars grouped together until testing. To ensure that the children learned to discriminate each of the six experimental stimuli reliably, tact training was conducted with six different stimulus pairings. For example, following criterion tact performance for each of three initial pairs (Zog 1 Vek 1, Zog 2 Vek 2, and Zog 3 Vek 3), the six stimuli were rearranged into three new mixed pairs (which varied across subjects) such as Zog 1 Vek 3, Zog 2 Vek 1, and Zog 3 Vek 2. When criterion tacting was achieved with these new pairings, the 3 subjects proceeded directly to the category sorting test without the opportunity to see, and, therefore, to sequentially tact, the six training stimuli combined in a spatial array. (Because of the introduction of a character called ''Zag'' on children's television at the time of this study, in Experiment 2 the tact ''zog'' replaced the ''zag'' employed in Experiment 1.)

Another feature of Experiment 2, absent from Experiments 1A and 1B, was the addition of listener behavior probes immediately following category testing. The children were presented with various stimulus pairs (e.g., Zog 1 Vek 1) and asked, ''Where is the vek [zog]?'' The aim of the probes was to provide a separate and direct test of the children's listener behavior to the auditory stimuli /zog/ and /vek/.

Method

Subjects

One girl (AH) and 2 boys (ES and CW) aged 3 years 8 months, 3 years 9 months, and 4 years, respectively, at the start of the experiment, were recruited from local daycare nurseries. Their Griffiths Mental Developmental scores, which were all in the normal range, were 124, 118, and 114, respectively.

Apparatus and Stimuli

The experimental setting was similar to that of Experiment 1. Each child was randomly assigned a set of six arbitrary wooden stimuli from the pool of 12 (see Figure 1). The six stimuli were divided randomly into three initial training pairs. For the purpose of reporting and scheduling stimulus presentations, the experimenter designated members of the three initial training pairs as Zog 1 and Vek 1, Zog 2 and Vek 2, and Zog 3 and Vek 3.

Procedure

Everyday objects. The procedure was a shortened version of the everyday objects procedure used in Experiment 1A, and consisted of the tact training in pairwise trials phase followed by category match-to-sample training and testing with the look-at-sample match-to-others instruction; for all 3 children, a second (blind) experimenter and a screen were introduced for test sessions. The tact learning criterion was more stringent than in Experiment 1A; here, the requirement was seven of eight consecutive correct tact responses for each stimulus pair.

Arbitrary stimuli. The procedure followed that of Experiment 1A; any differences will be noted.

Tact training in pairwise trials. Pairwise training was conducted first for Initial Pair 1 and Initial Pair 2. Then, omitting the four-stimulus stage of Experiment 1A, training was conducted with Initial Pair 3. The tact criterion was as for everyday object training, described above.

When this criterion had been reached, the three zog and three vek stimuli were, without replacement, randomly reassigned as three new mixed training pairs. For the purposes of reporting and scheduling stimulus presentations these pairs were termed Z/V_a, Z/V_b, and Z/V_c, respectively, where Z refers to a zog stimulus and V refers to a vek stimulus. Tact training to criterion was then conducted with the mixed pairs as for the initial pairs. At this point, the probability of reinforcement was reduced from 100% to 0% and the learning criterion was made yet more stringent, requiring seven of eight consecutive correct trials in two consecutive eight-trial blocks, spaced over two separate experimental sessions. If, when reinforcement for correct responding was withdrawn, tact performance on any of the mixed pairs fell below the criterion, reinforcement was reintroduced until criterion performance was reestablished, then was once again withdrawn in the next trial block, and so on until criterion tacting was achieved in its absence for all three mixed training pairs.

Category match-to-sample tests. The procedure employed was the same as in Experiment 1A except that, in place of the six-stimulus tact review for the arbitrary stimuli, this experiment included a pairwise tact review in which a four-trial block of tact test trials was presented for each of the mixed training pairs; the criterion required was 100% correct tacting of all the stimuli. In addition, all subjects in this experiment completed the full 18-trial test with the look-at-sample match-to-others instruction, irrespective of their performance in the first six trials (cf. Experiment 1A). For any child who passed this test, procedures were terminated at this point. If the child failed the look-at-sample test, 18 category sorting trials were initiated with the tact-sample match-to-others instruction. Success on the latter test trials resulted in 18 further look-at-sample match-to-others test trials, and failure resulted in termination of all category match-to-sample procedures. A second (blind) experimenter and a screen were in place for all test sessions with all 3 children.

Listener behavior probe trials. When category sorting trials were completed, the children were presented with a test block in which each of the three mixed training pairs was presented separately to each child. Conducting four trials for each pair, the experimenter presented the two stimuli and in each trial asked either, ''Where is the zog?'' or ''Where is the vek?'' The left–right position of the stimuli and the order of presentation of the two questions were randomized and counterbalanced across trials so that each question was asked twice for each stimulus pair. Thus a total of 12 listener probe trials were conducted across the three stimulus pairs.

Then, in a second, main block of listener behavior test trials, the six arbitrary stimuli were presented in all possible pairings (i.e., Z1/V1, Z1/V2, Z1/V3, Z2/V1, Z2/V2, Z2/V3, Z3/V1, Z3/V2, and Z3/V3). For both stimuli in each of the nine pairs, the listener behavior of selecting the appropriate stimulus in response to either the auditory stimulus ''zog'' or ''vek'' was probed for twice in a randomized but counterbalanced order across the pairs (36 trials total). The criterion for success on this test was set at 25 or more correct trials of 36; the probability of this occurring by chance is .014. There was no scheduled reinforcement for any listener responses in either of the two listener behavior testing blocks.

RESULTS

Everyday objects. The 3 children achieved the tact criterion for each of the three hat–cup pairs in the minimum number of test trials (i.e., eight) per stimulus pair. Two children, AH and CW, passed the category test using the look-at-sample match-to-others instruction in the minimum of six trials, and the 3rd child, ES, required a further six trials to achieve the sorting criterion.

Arbitrary stimuli: Pairwise tact trials. Except for ES, who required one extra eight-trial block for the first of the three initial pairs, the children achieved the 100% reinforcement criterion for both initial and mixed pairs in one eight-trial block per pair. Although 1 child, CW, reached the 0% criterion in the minimum of 16 trials, ES and AH required additional 100% reinforced, interspersed with 0% reinforced, test trials before the two-block tact criterion was attained in the absence of reinforcement. The number of retraining and testing trials per pair ranged from 64 to 88 for ES and from 160 to 200 for AH.

Category match-to-sample Tests 1 and 2. Figure 4 shows the number of trials in which each child sorted along common tact lines in the category tests. In the look-at-sample match-to-others version of the test (gray bars), both ES and CW exceeded the criterion of four of nine correct sorts in each category; CW sorted correctly on all 18 trials and ES was correct on 15 of the 18. The remaining child, AH (gray circles), failed to sort correctly on any of the trials; consequently, she proceeded to the tact-sample match-to-others test (shown by black bars) in which she met criterion,

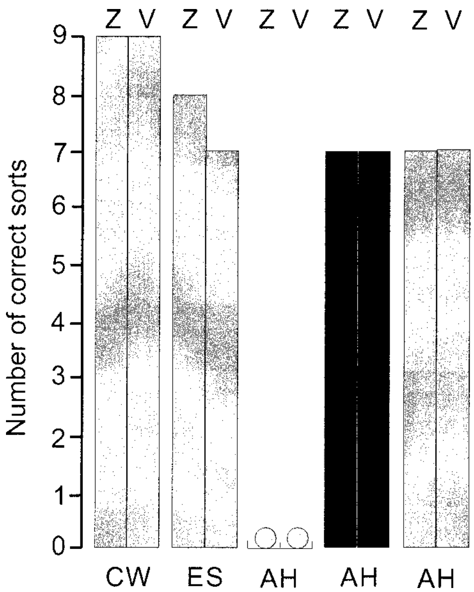

Fig. 4. For each subject, number of correct category sorts under the look-at-sample instruction (gray bars; zero responding is indicated by gray filled circles). Performance under the tact-sample instruction (AH) is shown by the black bar.

with seven of nine correct in each category. To determine whether she could now also categorize under the look-at-sample match-to-others instruction, she was tested once again in that condition (gray bars) and sorted with the same accuracy as in the tact-sample category test.

Listener behavior probe trials. In the first mixed pairs listener behavior test trials, 2 of the children, ES and CW, produced errorless performances on the 12-trial test block; AH made two incorrect responses. In the main pairwise listener behavior test that included all possible zog and vek combinations, all 3 children exceeded the criterion of 25 correct responses over the 36 test trials and performed without any errors. This shows that in all 3 subjects, appropriate listener behavior to /zog/ and /vek/ was at strength.

Discussion

The ages and Griffiths Mental Development scores of the 3 children in the present experiment overlapped considerably with those of the children in Experiment 1A and, in general, the pattern of results was similar across the two studies. Following only pairwise tact training to establish the two three-member common tact relations with the arbitrary stimuli, all 3 children sorted the six stimuli along common tact lines in the absence of reinforcement. These findings suggest that the four-stimulus and six-stimulus tact training trials employed in Experiments 1A and 1B, in which several members of the same class were presented together, are not necessary for the establishment of this kind of sorting behavior. These results also provide no evidence for the necessity of either autoclitic naming (cf. Horne & Lowe, 1997) or a joint control process such as that proposed by Lowenkron (1996), at least insofar as the latter might be envisaged to occur during the six-stimulus tact training stage of the procedure prior to the category tests.

As in Experiment 1A, a history of common tact training did not produce successful categorization with the look-at-sample instruction in all subjects. Like 5 of the children in that experiment, AH required the instruction to tact the sample before she produced correct category sorting behavior. In her case, however, following 18 trials of the tact-sample match-to-others instruction, explicit instruction to tact the sample was no longer required; she passed a second category test with only the look-at-sample match-to-others instruction.

The fact that during pairwise training, tacting was learned to criterion under 100% reinforcement but deteriorated when the probability of reinforcement was reduced to 0% suggests a "win-stay lose-shift" behavior pattern in which the presence or absence of reinforcement for the tact response made in the first trial of an eight-trial block may to some extent have determined subsequent tacting in the remaining seven trials. This interpretation seems possible, particularly in the cases of AH and ES, who both required a considerable number of reinforced pairwise tact review trials before criterion under 0% reinforcement was finally achieved. It also suggests that, in discrimination studies with young children, testing under 0% reinforcement conditions may be an important measure of the kind of learning that has taken place.

In the listener behavior probe trials, all 3 children responded virtually errorlessly to the auditory stimuli /zog/ and /vek/, showing almost perfect listener behavior to these stimuli.

GENERAL DISCUSSION

Taken together, all three experiments show that, in children 2 to 4 years old, simply training a common tact response to each of a number of arbitrary stimuli establishes those stimuli as a class or category. Children in these studies learned both three-member and six-member categories. Through training a common tact for each of six stimuli, for example, many new relations between these stimuli emerged untrained and unreinforced. Given any of the zags (e.g., Zag 1) as sample, the children selected the remaining five (Zag 2, Zag 3, Zag 4, Zag 5, Zag 6) as the related stimuli. This was the case for all possible such combinations, symmetric and otherwise, of the zags, and it was likewise for the veks. Experiment 1B showed that categorization was 100% correct at follow-up, 6 weeks after initial testing. Experiment 2 also showed that accurate categorization occurred without the children ever having seen class exemplars grouped together prior to testing. Thus, common tact training is an effective means of establishing stimulus classes in children of this age.

The present studies arose from Horne and Lowe's (1996) account of verbal behavior and categorization, in which it was proposed that common naming of arbitrary stimuli is sufficient to establish stimulus classes. The present procedure ostensibly trained only tacting. It is almost invariably the case, however, that when a normally developing 2- to 4-year-old child is taught a tact relation, he or she also learns the corresponding listener behavior, that is, the full name relation. By the time the child learns the earliest tact relations, he or she generally has an extensive listener repertoire and learns listener behavior rapidly with minimum social support (Baldwin, 1991); when an attempt is made to train a new tact by reinforcing the speaker behavior of, for example, ''zag'' in the presence of a stimulus, the child not only says ''zag'' but, in so doing, produces the auditory stimulus /zag/ to which she responds as a listener by attending to or looking again at the stimulus. So, as well as the tact, the corresponding listener behavior of looking at or otherwise attending to the tacted stimulus also is reinforced. There is a substantial experimental literature that supports this view (see Horne & Lowe, 1996, 1997), and further evidence is provided by the present studies. This was shown (a) in Experiment 2, in the near-perfect performance of all subjects on the direct tests of listener behavior to the auditory stimuli /zog/ and /vek/, a finding replicated with an additional 11 children aged 2 to 4 years old by Hughes (2000); and (b) in the successful performances on the category match-to-sample tests, particularly the tact-sample test in which the child was required to respond as a listener to her own sample tact response to select the ''others'' (i.e., the correct comparisons).

According to the naming account, naming is a principal if not, indeed, the sole means of establishing arbitrary stimulus classes. For example, in the category match-to-sample tests used here, if the child names the sample, that name should also control the listener behavior of looking at or reaching for the other members of the class also included in that name relation. Particularly when this entails selecting as many as five other stimuli from the array, the child may not only emit, overtly or covertly, the sample name but may also echo and reecho it, thereby sustaining the appropriate listener behavior of selecting each of the other five (e.g., CS, Experiment 1B). Apart from common naming, verbal control may, of course, be supplemented by autoclitic naming (Horne & Lowe, 1997, pp. 280–281), intraverbal naming, and other more complex ''rule-governing'' types of behavior that include other forms of naming (e.g., JC, Experiment 1A).

That the results of the three experiments are consistent with the naming account is as true of the test failures as of the successes. The fact that several of the 12 subjects who learned the prerequisite arbitrary tact relations in Experiments 1a and 2 did not succeed on the category tests until they were required to name the stimuli is what the account predicts. We have proposed previously (Lowe & Horne, 1996) that whether or not subjects will categorize in such circumstances is not a matter of mathematical or logical necessity but of behavior, that is, of whether they name the stimuli appropriately at the time of testing. In the look-at-sample test, the experimenter simply saying to the child, ''Look at this. Where are the others?'' may not be sufficient to occasion the sample name in all 2- to 4-year-old children. According to the naming account, if the child does not name the sample, either overtly or covertly, he or she should not be able to select the other stimuli in that class. On the other hand, if we ensure that the child names the sample then, because that name is also the name of all the correct comparisons, it should occasion the listener behavior of attending to and selecting those comparisons. If, as with the look-at-sample instruction, we do not explicitly require such naming, then whether children will ''spontaneously'' name the sample stimulus will most likely depend on their differing individual histories of naming. For example, AH in Experiment 2 failed to categorize any of the stimuli in Test 1 (look-at-sample) but was successful when required to name the sample in Test 2. When she was then given Test 1 again, she passed it with equal success, presumably because she had learned during Test 2 to name the sample stimulus prior to sorting.

Of course, although some of the children who succeeded on the look-at-sample version of the category test (i.e., GH and WA) were heard to name the sample stimuli overtly, others were not. Although we have no direct evidence that these latter children named the sample covertly, the naming account would assume that they did.

A striking feature of the category test results presented in Figures 2, 3, and 4 is the ''all-or-none'' nature of the outcomes: that is, there are no intermediate scores. The children either categorized almost perfectly or they failed almost completely to do so. This dichotomous outcome is also consistent with the naming account's prediction of chance responding if sample naming does not occur and accurate categorizing (allowing for occasional attentional lapses in children of this age) if it does.

The present findings are also consistent with other experimental work on naming. For example, in an early study, Birge (1941, Experiments 1 and 2) presented 152 children aged 8 to 9 years old with four boxes each with a physically different stimulus on the upper surface. The children were first taught to name each of two of the boxes ''meef'' and each of the other two ''towk''; that is, they were taught two-member common name relations. For each child, two of the boxes, one ''meef'' and one ''towk,'' were then employed in a simple discrimination task in which candy was reliably hidden under either the ''meef'' or the ''towk'' (counterbalanced across children). A transfer test was then conducted in which the child was required to say which of the remaining boxes, one ''meef'' and one ''towk,'' would contain candy. When the child was not required to name the boxes during the simple discrimination task and the transfer trials, correct responding in the latter was at chance levels (around 50% correct). When overt naming was required in both experimental phases, however, 84% of the children predicted the location of the candy correctly (see also Kendler, 1972). Results such as these suggest that in the case of children, to maximize the probability that common naming will exert its categorizing effects, they should, at least initially, be required to name the relevant stimuli overtly during task performance (see also K. J. Saunders & Spradlin, 1990, 1993).

The present experiments also should be compared with that of Eikeseth and Smith (1992). Their subjects were 4 autistic children, aged 3 years 6 months to 5 years 6 months who, compared to normal children, had language deficits. These children were first taught baseline conditional discriminations in a match-to-sample procedure and then were tested for the emergence of untrained conditional discriminations. All of them failed the test. In the next phase, however, when they were trained a common tact for each of the two three-member stimulus sets, 2 of the subjects passed the test and 1 performed at above-chance levels. The third experimental phase was the one that, in terms of the training procedures, most closely resembled that of the present experiments. The children were first taught two three-member tact relations with six new stimuli (D1, E1, F1 and D2, E2, F2), and then were tested for the emergence of stimulus classes using the two-choice match-to-sample procedure as before. Two of the 4 children immediately passed these tests following just tact training, a result that is similar to that reported for the 12 children in the present experiments. The 2 other children were given training on D-E and D-F conditional relations, after which one passed and the other showed above chance performance in a combined test of symmetry and transitivity (E-F, F-E). That these 2 children needed some match-to-sample training on two of the relations before the remaining relations emerged may have been due to the sensitivity of autistic children to small changes in procedure such as occurred in the transition from Phases 1 and 2 to Phase 3 (see Kelly, Green, & Sidman, 1998). We would add to this that simply training tact relations to autistic children with low-level language skills does not guarantee that corresponding listener behavior is also established, that is, that full naming relations will be in place (Charlop, 1983; Fay & Butler, 1968; Lewis, 1987, pp. 143–150). The fact that the 1 subject, Danny, who passed the tests did not overtly name the stimuli during testing does not, of course, mean that he did not do so covertly. Indeed, of all the subjects, Danny was the most verbally able and learned the tact relations in the fewest trials. This study not only provides evidence that common naming is effective in establishing stimulus classes but also shows that there are particular problems in doing research of this kind with subjects who are autistic or otherwise intellectually disabled (Sidman, 1994, pp. 305–307; Stromer & Mackay, 1996, pp. 234–240).

The main aim of the present study was not to resolve controversial theoretical issues in the area of arbitrary stimulus classes but rather to investigate in young children how naming might serve to establish such classes. Thus, the present experiments have little to say about the question of whether naming is *necessary* for the establishment of arbitrary stimulus classes. The findings do, however, have theoretical implications. In the present experiments, some of the children failed to categorize with the look-at-sample instruction but they all succeeded when required to name the sample prior to sorting. But although these findings are predicted by the naming account (Horne & Lowe, 1996), they are not predicted by Sidman's (1994, 2000) stimulus equivalence theory or by relational frame theory (Hayes & Hayes, 1989, 1992). Indeed, it is not clear how either of these latter theories could account straightforwardly for the differing outcomes of the two category sorting tests.

As regards the nature of future research conducted on stimulus classes, there is much to recommend the training and test procedures developed in the present experiments. The sorting tests used here, which we have termed *category match to sample,* mark a substantial departure from standard techniques of paired stimulus selections (and see Pilgrim & Galizio, 1996, p. 188). The limitations of standard match-to-sample procedures, in which subjects are required, in response to presentation of a single sample, to select one from among two or three comparisons, have been noted previously and their robustness as measures of categorization questioned (Innis, Lane, Miller, & Critchfield, 1998). To overcome some of these limitations and ensure a more reliable measure of stimulus classes, some investigators have included default-response options, in which subjects can indicate a "don't know" or "none of the above" response (Fields, Reeve, Adams, Brown, & Verhave, 1997; Innis et al., 1998; Roche & Barnes, 1996). As Innis et al. have shown, however, adding a default option in itself appears to decrease, often substantially, the number of subjects who pass the tests for emergent conditional discriminations between stimulus pairs. This raises the more general question of why it is that so much behavior-analytic research on stimulus classes and categorization chooses to concentrate so exclusively on stimulus or "event" pairings. It is a methodological focus that certainly is not unrelated to the theories that emanate from it. Sidman (2000), for example, speaks of an equivalence relation as being "a bag that contains ordered pairs of all events that the contingency specifies" (p. 144). But, if we are really concerned with whole categories and classes of stimuli, why "pairs"? Innis et al. make the point as follows: "Everyday acts of communication and conceptualization emerge under relatively unconstrained conditions in which the range of possible responses is far greater than the two or three provided in conditional discrimination trials" (p. 98). In the category tests used in the present experiments, when the sample stimulus was presented in, say, the six-member tests, the child was free to select any number of stimuli from 11 remaining comparisons. The relation here was clearly not an event pair but, at best, a one-to-five relation. And, if the naming account is correct, the sample merely set the occasion for a common name which in turn occasioned the child's selection of the five other stimuli. When it comes to categorizing, in other words, stimulus pairings may be more in the mind (or verbal behavior) of the experimenter than in the behavior of the subject.

There are also practical considerations in favor of the present procedure. Difficulties in establishing conditional discriminations in young children are well documented in both normally developing (e.g., Augustson & Dougher, 1991; Pilgrim, Jackson, & Galizio, 2000) and developmentally disabled (e.g., Eikeseth & Smith, 1992) populations (but see Devany, Hayes, & Nelson, 1986). According to Pilgrim et al. (2000), the fact that ''arbitrary matching is not readily acquired by normally developing children without special training procedures'' is an issue that ''has received little explicit attention in the empirical literature, and thus may be underappreciated by scientists working outside of this immediate field'' (p. 190). The present category match-to-sample procedure, however, circumvents the need for special interventions to overcome young children's difficulties in learning the baseline discriminations prerequisite for tests of ''emergent'' relations. In this procedure, name relations constitute the baseline behavior, and these are simple discriminations that are readily learned by young children. Even in the lengthiest version of the training procedure (Experiment 1A), the youngest child (GH, aged 2 years 3 months) required only 240 training trials to demonstrate learning of six novel name relations to a stringent criterion (testing under 0% reinforcement conditions), whereas the oldest child (JR, aged 4 years 3 months) required only 34. Consistent with this outcome, Pilgrim et al. found that the young children who participated in their study readily acquired consistent use of sample-specific names (which facilitated learning of arbitrary matching). Even the autistic children who participated in the Eikeseth and Smith (1992) study learned the six baseline tact relations in approximately 80 trials, averaged across children. This compares favorably with the average of approximately 900 trials they took to learn four baseline match-to-sample conditional discriminations. Training common tact relations for physically different stimuli is thus a relatively quick and effective means of investigating emergent behavior in young children even as young as 2 years of age.

In addition, the implementation of this procedure by behavior analysts is likely to contribute to our being able to communicate more effectively with researchers outside that tradition who, for their part, at present make much more extensive use of stimulus sorting tests (albeit often less well controlled than here) to measure categorization and concept formation (and see Pilgrim & Galizio, 1996). And, what is perhaps more important is that, as the present findings show, it is a robust measure of categorization that can be conducted (e.g., in six-member classes) in as little as a single trial.

The use of these procedures in future research could also help to resolve some important theoretical issues not addressed by the present experiments. Although we have proposed that these findings are consistent with the naming account, it is possible that subjects could pass the look-at-sample test with just a tact repertoire and that the emergent corresponding listener behavior relations are mere correlates of tact training and play no causal role in establishing stimulus classes. Naming, in other words, may not be necessary to establish these stimulus classes. The practical difficulty lies in how experimenters might manage to establish only tacts and not full name relations when, as we know, tact training procedures with normally developing young children also concurrently train listener behavior. Given that it is far easier to establish listener behavior in young children without also training the corresponding speaker behavior (Horne & Lowe, 1996), one of the ways in which future research might test the naming account would be to train, instead of common tact relations, common listener relations in response to the same stimuli as were used in the procedures here. If the corresponding speaker behavior (and thus the full name relation) did not emerge, then the naming account would predict that arbitrary stimulus classes would not be established in category sorting tests. If naming did emerge, then so should categorization.

As Horne and Lowe (1996) and Carr et al. (2000) have noted, there are other critical tests of the naming account that need to be conducted, particularly with preverbal infants and nonverbal animals (see, e.g., Kastak, Schusterman, & Kastak, 2001; Schusterman & Kastak, 1993; and, for a critique of these experiments, Horne & Lowe, 1996, 1997; Lowe & Horne, 1996). If arbitrary stimulus classes could be established in such subjects, then it would show that naming is not necessary in all cases for such emergent behavior to occur. This would, of course, still leave unresolved the extent to which arbitrary stimulus classes are contingency shaped as opposed to verbally governed (see Horne & Lowe, 1996, pp. 228–233). As the present experiments and others show, naming is a powerful determinant of categorization, but its role in such behavior has yet to be systematically explored.

REFERENCES

Augustson, K. G., & Dougher, M. J. (1991). Teaching conditional discrimination to young children: Some methodological successes and failures. *Experimental Analysis of Human Behavior Bulletin, 9,* 21–24.

Baldwin, D. A. (1991). Infant contributions to the achievement of joint reference. *Child Development, 62,* 875–890.

Beasty, A. (1987). *The role of language in the emergence of equivalence relations: A developmental study.* Unpublished doctoral dissertation, University of Wales, Bangor.

Birge, J. S. (1941). *The role of verbal response in transfer*. Unpublished doctoral dissertation, Yale University.

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

Charlop, M. H. (1983). The effects of echolalia on acquisition and generalization of receptive labeling in autistic children. *Journal of Applied Behavior Analysis, 16,* 111–126.

Devany, J., 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,* 243–258.

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

Dugdale, N., & Lowe, C. F. (2000). Testing for symmetry in the conditional discriminations of language-trained chimpanzees. *Journal of the Experimental Analysis of Behavior, 73,* 5–22.

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.

Fay, W. H., & Butler, B. V. (1968). Echolalia, IQ, and the developmental dichotomy of speech and language systems. *Journal of Speech and Hearing Research, 11,* 365–370.

Fields, L., Landon-Jimenez, D. V., Buffington, D. M., & Adams, B. J. (1995). Maintained nodal distance effects in equivalence classes. *Journal of the Experimental Analysis of Behavior, 64,* 129–145.

Fields, L., Reeve, K. F., Adams, B. J., Brown, J. L., & Verhave, T. (1997). Predicting the extension of equivalence classes from primary generalization gradients: The merger of equivalence classes and perceptual classes. *Journal of the Experimental Analysis of Behavior, 68,* 67–91.

Fields, L., Reeve, K. F., Rosen, D., Varelas, A., Adams, B. J., Belanich, J., et al. (1997). Using the simultaneous protocol to study equivalence class formation: The facilitating effects of nodal number and size of previously established equivalence classes. *Journal of the Experimental Analysis of Behavior, 67,* 367–389.

Fields, L., Reeve, K. F., Varelas, A., Rosen, D., & Belanich, J. (1997). Equivalence class formation using stimulus-pairing and yes-no responding. *The Psychological Record, 47,* 661–686.

Green, G. (1990). Differences in development of visual and auditory-visual equivalence relations. *American Journal of Mental Retardation, 95,* 260–270.

Griffiths, R. (1954). *The abilities of babies.* London: The Test Agency Limited.

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

Hayes, S. C. (1989). Nonhumans have not yet shown stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 51,* 385–392.

Hayes, S. C. (1996). Developing a theory of derived stimulus relations. *Journal of the Experimental Analysis of Behavior, 65,* 309–311.

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

Hayes, S. C., & Hayes, L. J. (1992). Verbal relations and the evolution of behavior analysis. *American Psychologist, 47,* 1383–1395.

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.

Hughes, J. C. (2000). *Naming and generalisation of behaviour in infants.* Unpublished doctoral dissertation, University of Wales, Bangor.

Innis, A., Lane, S. D., Miller, E. R., & Critchfield, T. S. (1998). Stimulus equivalence: Effects of a default-response option on emergence of untrained stimulus relations. *Journal of the Experimental Analysis of Behavior, 70,* 87–102.

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

Kelly, S., Green, G., & Sidman, M. (1998). Visual identity matching and auditory-visual matching: A procedural note. *Journal of Applied Behavior Analysis, 31,* 237–243.

Kendler, T. S. (1972). An ontogeny of mediational deficiency. *Child Development, 43,* 1–17.

Lewis, V. (1987). *Development and handicap.* Oxford: Basil Blackwell.

Lowe, C. F., & Horne, P. J. (1996). Reflections on naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65,* 315–340.

Lowenkron, B. (1996). Joint control and word-object bidirectionality. *Journal of the Experimental Analysis of Behavior, 65,* 252–254.

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

Markman, E. M. (1989). *Categorization in children: Problems of induction.* Cambridge, MA: Cambridge University Press.

McIlvane, W. J., & Dube, W. V. (1996). Naming as a facilitator of discrimination. *Journal of the Experimental Analysis of Behavior, 65,* 267–272.

Mead, G. H. (1934). *Mind, self and society.* Chicago: University of Chicago Press.

Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization. Emory symposia in cognition, 1* (pp. 201–233). New York: Cambridge University Press.

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

Pilgrim, C., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I. Adults. *Journal of the Experimental Analysis of Behavior, 63,* 225–238.

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: Elsevier.

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

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.

Roche, B., & Barnes, D. (1996). Arbitrarily applicable relational responding and sexual categorization: A critical test of the derived difference relation. *The Psychological Record, 46,* 451–475.

Saunders, K. J. (1989). Naming in conditional discrimination and stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 51,* 379–384.

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

Saunders, K. J., & Spradlin, J. E. (1990). Conditional discrimination in mentally retarded adults: The development of generalized skills. *Journal of the Experimental Analysis of Behavior, 54,* 239–250.

Saunders, K. J., & Spradlin, J. E. (1993). Conditional discrimination in mentally retarded subjects: Programming acquisition and learning set. *Journal of the Experimental Analysis of Behavior, 60,* 571–585.

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: Elsevier.

Saunders, R. R., Drake, K. M., & Spradlin, J. E. (1999). Equivalence class establishment, expansion, and modification in preschool children. *Journal of the Experimental Analysis of Behavior, 71,* 195–214.

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

Schenk, J. J. (1995). Complex stimuli in nonreinforced simple discrimination tasks: Emergent simple and conditional discriminations. *The Psychological Record, 45,* 477–494.

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

Spencer, T., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65,* 643–659.

Stromer, R., & Mackay, H. (1996). Naming and the formation of stimulus classes. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 221–252). Amsterdam: Elsevier.

Urcuioli, P. J. (1996). Acquired equivalences and mediated generalisation in pigeon's matching-to-sample. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 55–70). Amsterdam: Elsevier.

Vihman, M. M. (1996). *Phonological development: The origins of language in the child.* Cambridge, MA: Blackwell.

Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carlton (Eds.), *The collected works of L. S. Vygotsky* (Vol. 1, pp. 39–285). New York: Plenum. (Original work published 1934)

Received November 5, 2001
Final acceptance May 23, 2002