JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2004, **81,** 267–288 NUMBER 3 (MAY)

NAMING AND CATEGORIZATION IN YOUNG CHILDREN: II. LISTENER BEHAVIOR TRAINING

PAULINE J. HORNE, C. FERGUS LOWE, AND VALERIE R. L. RANDLE

UNIVERSITY OF WALES, BANGOR

Following pretraining with everyday objects, 1- to 4-year-old children received listener training with three pairs of arbitrary stimuli of differing shapes. For each pair, 9 children were trained to select one stimulus in response to the spoken word /zog/ and the other to the spoken word /vek/. Next, in the look-at-sample category match-to-sample test, none categorized the six stimuli correctly when asked to look at the sample before selecting from five comparisons. Seven of these children failed a subsequent test of corresponding speaker behavior (tact test); following tact training, 5 of them passed either a repeat of the look-at-sample category test (2 subjects) or an alternative category test (3 subjects) in which they were required to tact the sample before selecting comparisons. The remaining 2 failed both category tests. Of the 2 who passed the tact test, 1 passed the tact-sample category test; the other failed to complete category testing. Two children were next given a second stimulus set. One passed the look-at-sample category test and the tact test; the other failed both tests but passed the tact-sample category test after tact training. The results show that 1- to 4-year-old children may learn listener behavior without corresponding speaker behavior. The results also show that common listener behavior is not sufficient to establish arbitrary stimulus classes, and they are consistent with the proposition that naming may be necessary for categorization of such stimuli.

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

How organisms come to form categories, particularly of arbitrary stimuli, is a matter of considerable debate in the psychological literature, not least within the areas of child development and behavior analysis. Although the cognitive developmental literature is replete with studies that suggest a role for object names, perceptual features, and common ''functions'' in the establishment of categories in young children (Gelman & Coley, 1990; Gopnik & Meltzoff, 1992; Gopnik & Sobel, 2000; Kemler Nelson, Frankenfield, Morris, & Blair, 2000; Nazzi & Gopnik, 2000; Smith, Jones, & Landau, 1996), there are few such studies with sufficient experimental controls to determine which of these variables, if any, are necessary for category formation.

Within behavior analysis, the role of stimulus generalization in the establishment of categorization has received considerable attention (Fields, Matneja, et al., 2002; Fields, Reeve, et al., 2002; Herrnstein, Loveland, & Cable, 1976; Sutton & Roberts, 2002). However, in order to determine other ways in which categories can be formed, it is recognized that studies should use stimuli that do not share common physical features, that is, *arbitrary* stimuli. Given that particular physical features of stimuli occasion particular responses in young children (see Bloom, 1996, and Gibson, 1969), the use of arbitrary stimuli should also exclude control by extraexperimentally trained common functions among the stimuli. In addition, it is important to systematically manipulate what has been variously termed *naming* or *labeling*. In cognitive developmental studies of children's categorizing, such manipulation generally consists of simply telling a child what a given stimulus is called (see Gopnik & Sobel, 2000; Liu, Golinkoff, & Sak, 2001; Sloutsky, Lo, & Fisher, 2001; Welder & Graham, 2001). In this literature, the distinction between the child understanding a name (i.e., showing appropriate listener behavior) and producing it is not generally considered relevant. On the one hand, in order to establish the impact, if any, of naming or labeling on categorizing, a behavior analytic account requires at least the demonstration that the participants reliably produce the correct name when, for exam-

We are grateful to Gareth Horne for his help in the preparation of this manuscript and to Pat Lowe for her editorial contribution. The research reported here was supported by a Ph.D. studentship awarded by the Economic and Social Research Council to Valerie Randle.

Requests for reprints should be directed to Dr. Pauline Horne, School of Psychology, University of Wales, Bangor, Brigantia Building, Penrallt Road, Bangor, Gwynedd LL57 2AS, United Kingdom (e-mail: p.j.horne@bangor.ac.uk).

ple, the relevant objects are presented to them and, as each is pointed to, they are asked, "What's this?" On the other hand, if the independent variable under consideration is understanding of the spoken name, then it is necessary to show that the child can point to the correct object when asked, for example, "Where's the [object name]?" To understand the role of verbal behavior in categorization and, indeed, to determine the necessary and sufficient conditions for it to occur, it is vital that these distinctions are made and subjected to controlled investigation.

According to Sidman's (1994, 2000) stimulus equivalence theory, the formation of bidirectional (or symmetry) relations between stimuli and responses that feature in a particular behavioral contingency is a behavioral primitive or "given". However, despite decades of endeavor, evidence of stimulus equivalence in species other than verbal humans has proved exceedingly difficult to find (see Dugdale & Lowe, 2000; Hayes, 1989; Horne & Lowe, 1996, 1997; Lowe & Horne, 1996; K.J. Saunders, Williams, & Spradlin, 1996; Urcuioli, 1996; but see also Kastak, Schusterman, & Kastak, 2001; Schusterman & Kastak, 1993). This remains the case regardless of the variety of procedural adaptations that have been introduced to facilitate the occurrence of symmetric responding in nonhuman organisms. Most recently, for example, Lionello-DeNolf and Urcuioli (2002) found that neither multiple location training during the establishment of baseline relations, nor additional training to ensure that each stimulus was discriminated from each of the others, nor testing for symmetry in the context of reinforcement for other symmetric relations resulted in symmetry performances in pigeons. These authors concluded that differences in stimulus location in training and test trials is not one of the causes of the ubiquitous failure of this species to show equivalence and indeed that the latter "may be beyond the capabilities of the pigeon" (p. 467).

Verbal humans, on the contrary, generally do pass match-to-sample tests of stimulus equivalence (Fields, Reeve, Varelas, Rosen, & Belanich, 1997; Pilgrim, Chambers, & Galizio, 1995; Schenk, 1995; Sidman, 1994), though there are circumstances in which, for reasons yet to be fully determined, some fail (Fields, Reeve, Rosen, et al., 1997; Innis, Lane, Miller, & Critchfield, 1998; Pilgrim et al., 1995; Pilgrim & Galizio, 1995; R. R. Saunders, Drake, & Spradlin, 1999; Spencer & Chase, 1996; but see Horne & Lowe, 1996). Consistent with the developmental literature, naming, either as common naming or as intraverbal naming, is a strong determinant of category formation, as measured by match-to-sample procedures (Beasty, 1987; Bentall, Dickins, & Fox, 1993; Dugdale & Lowe, 1990; Eikeseth & Smith, 1992; Horne & Lowe, 1996; Lowe, Horne, Harris, & Randle, 2002; Mandell & Sheen, 1994; Randell & Remington, 1999; K. J. Saunders, Saunders, Williams, & Spradlin, 1993). What has not so far been demonstrated is that naming is necessary for the establishment of categorizing in humans (Carr, Wilkinson, Blackman, & McIlvane, 2000; Horne & Lowe, 1996; Lowe & Horne, 1996).

Because equivalence theory holds that stimulus equivalence is a basic behavioral given, which may or may not turn out to be restricted to only a subset of species, no special role is accorded to naming in the establishment of category relations (Sidman, 1994, p. 281–307; Sidman, 2000, p. 145). It follows that bidirectional (category) relations should form among arbitrary stimuli regardless of whether those stimuli are established (e.g., via standard reinforcement procedures) as members of a common listener, or a common speaker, relation. So whether a child learns to select particular stimuli on hearing their common name, or alternatively, to produce the common name when he or she sees each of them, should have no critical bearing on whether or not that child will subsequently categorize those stimuli in terms of their common name. These distinctions also are accorded little significance by another theory of categorization behavior, the relational framing account (Hayes, 1996; Hayes & Hayes, 1989, 1992). According to this account, given that a contextual cue specifying the arbitrarily applicable frame of coordination is present, bidirectional relations between the arbitrary stimuli so framed should be formed regardless of whether they are presented as common listener relations or as common name relations. Thus neither the stimulus equivalence nor the relational framing accounts consider that naming is neces-

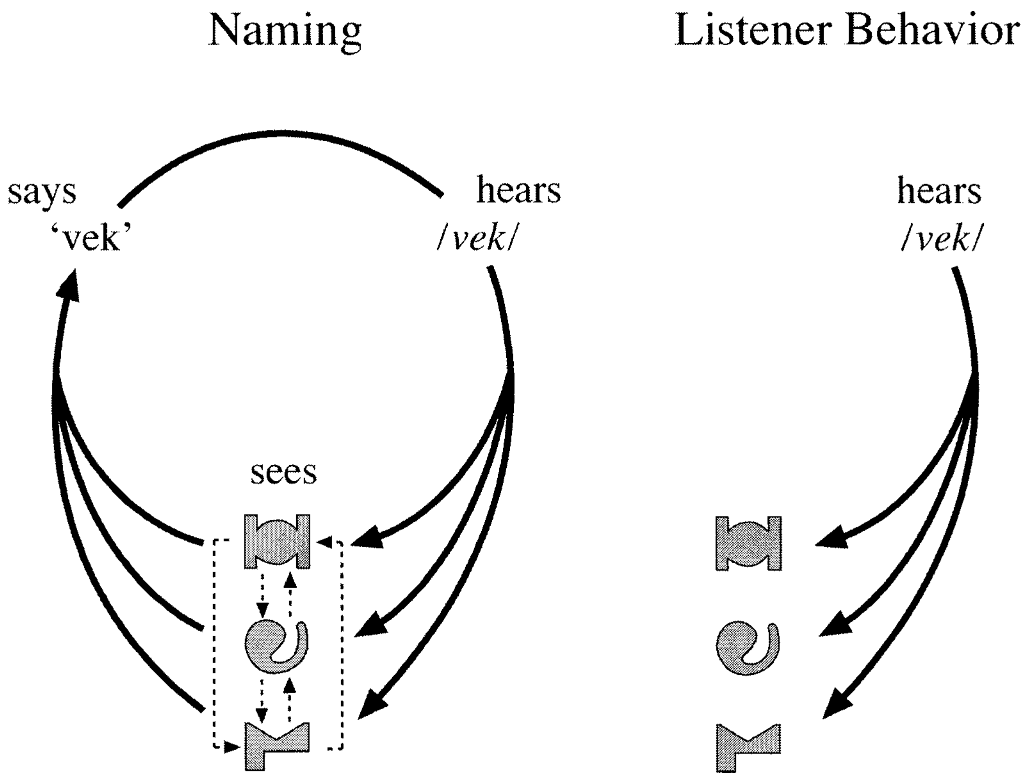

Fig. 1. The left side shows, according to the naming account, that when a subject comes to emit a common name (e.g., "vek") for differently shaped arbitrary stimuli, new relations (shown by broken lines) emerge via the name relation so that, when presented with any one of these stimuli as the sample, the subject may select the others in the class (see text). This is in contrast to what happens when common listener behavior alone is established (right side); no new untrained relations emerge.

sary for the establishment of category relations among arbitrary stimuli.

By contrast, the naming account of Horne and Lowe (1996) claims that bidirectional relations are formed between arbitrary stimuli when each evokes the same speaker–listener behavior that constitutes a name relation. How this occurs experimentally is illustrated in the left panel of Figure 1. A young child is first presented, one at a time, with three differently shaped arbitrary stimuli (in the context of one or more other stimuli), upon seeing each of which he or she is taught to say "vek." This training also establishes, at the same time, the corresponding listener relations, that is, upon hearing the sound /vek/ (the auditory product of the child's "vek" response), the child looks at the stimulus that occasioned the response "vek." Once the speaker–listener relation has been learned for each of the arbitrary stimuli, the latter enter into a new functional relationship with each other via the common speaker–listener behavior they each evoke. If the child is next presented with Vek 1 (in the context of veks and other stimuli) and asked, "What's this? Where are the others?" the sight of the stimulus occasions the vocal response "vek." The latter, in turn, automatically produces the sound /vek/, to which the child responds as a listener by orienting to, and selecting, not simply Vek 1 but also Vek 2 and Vek 3. No matter which of the stimuli first occasions saying "vek," the child's attention will be drawn via the common listener behavior to the other stimuli that feature in the common name relation. This relationship between arbitrary (vek) stimuli via the name relation exemplifies how, according to the naming account, categorization is established.

When it is not the name relation but common listener behavior alone that is learned, the naming account predicts a different outcome. For example, if, as shown on the right side of Figure 1, the child had been taught, not to name the stimuli, but to select each vek when asked "give me the vek," then each time the child hears the auditory stimulus /vek/ the child will orient to all of the veks present. Unless common naming has been established, however, when presented with one of the veks as the sample and asked to select the others, no relations exist to enable the child to make the correct selections. According to the naming account, common listener relations alone will not give rise to categorization.

The claim that common naming establishes categorization of arbitrary stimuli, in the manner suggested in Figure 1, was directly tested by Lowe et al. (2002). In this study, 2- to 4-year-old children were trained to produce the common name "zag" for each of three arbitrary shaped wooden stimuli, and to name three other such stimuli "vek." Using a novel category match-to-sample test procedure designed to assay an entire category in each trial, the children were presented with a sample stimulus in each test trial and then were required to select all the remaining members in the stimulus class. This contrasts with standard match-to-sample procedures that test for categorization in an exhaustive pairwise fashion. In the latter case, each trial consists of the presentation of a sample and then a forced choice between one potential category member (i.e., the correct comparison) and one or more noncategory stimuli (the incorrect comparisons). In the category match-to-sample test, however, each trial is a simultaneous free choice between all category and all noncategory members and as such it provides a more rigorous procedure for the functional measurement of categorization (see Innis et al., 1998; Lowe et al., 2002). Following common name training, all 12 children sorted the stimuli correctly; 6 of them did so right away in test trials in which they were asked to look at the sample before selecting comparisons, and the remaining 6, who initially failed the test under the latter instruction, also succeeded in categorizing the stimuli when they were asked to name the sample before selecting. A subset of the children, who were taught the same common names for an additional three zag and three vek stimuli, successfully sorted the resulting 12 stimuli into the two 6-member categories.

Though this study demonstrated that common naming may establish categorization of as many as six arbitrary stimuli, particularly when the child is required to name the sample before sorting, it does not show that the full name relation is necessary for this outcome. If, indeed, naming is neither necessary nor special, it should be possible to demonstrate that categorization can be readily established through the listener behavior route. As proposed by Lowe et al. (2002), this is one means of disproving the naming theory.

The experimental procedure of the present investigation was modeled closely on that of Lowe et al. (2002) except that the pairwise stimulus training procedure was modified in an attempt to establish listener relations without also teaching naming. During training, children were taught to select one of a pair of arbitrary stimuli (Pair 1) when they heard the auditory stimulus /zog/, and the other when they heard /vek/. The same procedure was then conducted for Pair 2 arbitrary stimuli, and then for the Pair 3 arbitrary stimuli. Categorization was then assessed in a category match-to-sample test in which the children were required to look at the sample before selecting comparisons. This was followed by a tact test to determine whether listener behavior alone or both speaker and listener behavior (naming) had been learned during listener training. Children who failed to categorize and to show evidence of naming were next trained to provide a common name for the stimuli in order to determine whether naming would bring about categorization. They were first required to sort the stimuli under the look-at-sample instruction and, if they failed, to name the sample before selecting the comparisons in the category match-to-sample test. Several children who succeeded on the latter tests were retested under the look-at-sample instruction. In addition, to test for generalization of categorization, 2 of the children were trained in common listener relations with a further set of six arbitrary stimuli and the same two auditory stimuli (/zog/ and /vek/). Because it was so important in this study to establish whether the children had learned simply listener relations or also came to name the stimuli, their on-task verbal behavior was recorded throughout.

Table 1

Subjects' sex, ages at start of the study, participation (indicated as √) in everyday objects and arbitrary stimulus training, and ages at start of arbitrary stimuli category testing. This also shows, for the children aged from 16 to 28 months, the listener and speaker scores on the MacArthur Communicative Development Inventory (MCDI) and, for children who completed all testing procedures, the General Quotient scores on the Griffiths Mental Development Scales (GMDS).

Subject	Sex	Age at start (year/ months)	Everyday objects	Arbitrary stimuli training	Age at testing (year/ months)	MCDI score		GMDS general quotient
						Listener	Speaker	
KF	F	1/4	x	—	—	245	90	—
PF	M	1/7	x	—	—	194	70	—
JC	F	1/7	√	√	1/9	254	174	112
HM	F	1/7	x	—	—	195	113	—
JL	M	1/9	x	—	—	*	*	—
MJ	F	1/10	√	√	2/2	340	258	133
SJ	F	1/11	√	x	—	340	340	—
BH	M	1/11	√	√	2/4	234	66	111
HS	F	2/0	√	x	—	379	371	—
BG	M	2/1	x	—	—	246	63	—
CM	M	2/2	√	x	—	336	197	—
CG	M	2/3	x	—	—	314	268	—
TK	M	2/4	√	x	—	*	*	—
CT	M	2/6	√	√	2/11	—	—	124
TP	M	2/8	√	x	—	—	—	—
LN	F	2/10	√	√	2/11	—	—	122
HW	M	2/10	√	√	3/0	—	—	144
NW	F	3/9	√	√	3/11	—	—	129
RE	F	3/10	√	√	4/0	—	—	121
TB	M	4/0	√	√	4/1	—	—	101

* Parents of these children failed to complete and return the MCDI.

METHOD

Subjects

Subjects were 9 female and 11 male children who attended nurseries in Bangor, Wales and were aged from 1 year 4 months to 4 years at the start of the study (see Table 1 for subject descriptions). Receptive and expressive scores on the MacArthur Communicative Development Inventory (Fenson et al., 1993) were obtained for those children who were aged between 1 year 4 months and 2 years 4 months. Of the 20 participants, 6 either withdrew from the study or left the nursery before they completed pretraining with everyday objects, and a further 5 withdrew before they completed arbitrary stimulus listener training. Data are reported only for the 9 children who reached the arbitrary stimulus category-testing phase (see Table 1). These latter children were all tested on the Griffiths Mental Development Scales (Griffiths, 1954). All scores on both developmental tests were in the normal range.

Apparatus and Stimuli

For some children, experimental sessions took place in the University of Wales Daycare Center in an experimental room fitted with two wall-mounted color video cameras with built-in directional microphones. Video and audio inputs were transmitted from the experimental room to video recorders in a central audiovisual console room. For other children, sessions were conducted in a similarly equipped experimental room in a mobile child research laboratory located outside the entrance to the nursery they attended. Recording of the child's and the experimenter's speech during the sessions took place via a radio microphone worn by the experimenter (in the nursery) or a fixed microphone (in the mobile laboratory). The child and experimenter sat on opposite sides of a small red table. In some sessions, a wooden screen (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 transparent 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 used a teddy bear puppet that she wore on her hand to present stimuli to the child through an aperture (height 20 cm, width 50 cm, and partly covered by a red and green crepe paper fringe) in the lower section of the screen. The experimental stimuli were of two kinds: (a) six everyday objects, each of a different color, consisting variously of three different types of toy hat and three different types of toy cup; and (b) six different arbitrary green wooden shapes randomly assigned to each child from a pool of such shapes (for examples, see Lowe et al., 2002, Figure 1). The shapes were approximately 10 cm in length, 0.9 cm in depth, and varied in width from 4 cm to 7.5 cm. The toy hats and cups resembled everyday objects that many young children have learned to name by the time they are 2 years of age (Fenson et al., 1993). They were employed to familiarize the children with the training and testing procedures and to demonstrate that the children were able to respond correctly to the task instructions subsequently employed with the novel, arbitrary stimuli. The scheduled reinforcer was social praise and, occasionally, tokens and stickers. At the end of each session, the child chose one toy from among a selection of small toys to take home and, at the end of the study, the children were presented with a toy of their choosing.

Procedure

Everyday objects. The experimenter first established a good rapport with the children during unstructured daily play sessions. Children were then taken one at a time to the experimental room and seated at the table across from the experimenter. The child was introduced to "Teddy," the teddy bear glove puppet that the experimenter wore on her right hand. 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; these remained constant for the duration of the study. For the purpose of reporting and scheduling stimulus presentations, the first training pair was 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 the other experimental conditions, sessions varied in duration from 15 to 30 min. The children's verbal behavior was recorded throughout the study.

Listener training in pairwise trials. The experimenter, having placed Hat 1 and Cup 1 on the table so that one stimulus lay 10 cm to the left of the child's midline and the other 10 cm to the right, then said to the child, "Look at these. Can you give Teddy the hat/cup?" If the child gave the hat/cup, the experimenter responded, "Yes, clever girl [boy]!" If the child gave the wrong object or did not respond at all, the experimenter provided the following corrective instruction: "No, that's not a hat/cup. Can you give Teddy the hat/cup?" and, if necessary, modeled the required selection response. In 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, quasi-random order, were each targeted twice on the right of the child's midline and twice on the left. The learning criterion was seven out of eight correct responses within any sequence of eight consecutive trials. Once criterion was met for Pair 1, the same training procedure was repeated first for Pair 2 and then for Pair 3.

Category match-to-sample training. The screen was used throughout this phase, which was designed to establish the child's correct responding to different types of instructions. The first of these was the *look-at-sample match-to-others* instruction. On each of six trials, 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, "Look at this. Can you give Teddy the others?" If, when given a hat as the sample, the child selected the two other hats as comparisons, this constituted a correct category sort; likewise, given a cup as sample, the child was required to select the other two cups. Each of the everyday objects served once as the sample stimulus. No reinforcers or corrective feedback were delivered during these trials. If responding in the latter was not 100% correct, the experimenter presented a further six trials in the first of which a modified instruction: ''Can you give Teddy the other *hats/cups*?'' was employed. If the child responded incorrectly, the experimenter provided the following corrective verbal feedback: ''No, this [pointing to the sample] is a hat [cup]. This [pointing to the incorrect comparison stimulus] is a cup [hat],'' and also modeled the correct comparison selection response(s); in the next trial the instruction remained the same as in the previous trial. Following a correct trial, however, the instruction in the subsequent trial reverted to: ''Look at this. Can you give Teddy the others?'' In this manner, further six-trial blocks were conducted until the child responded 100% correctly to the first instruction: ''Look at this. Can you give Teddy the others?'' in six consecutive trials, and in the absence of reinforcement.

Arbitrary stimuli: Set 1. For each child, a set of six stimuli was selected at random from the pool of wooden shapes.

Listener training—Initial pairs. The nonsense sounds /zog/ and /vek/ were selected to be the listener stimuli. The six wooden shapes allocated to each child were separated into three training pairs; the children were then trained to select one pair member when they heard the auditory stimulus /zog/ and the other pair member when they heard /vek/. In reporting presentations of the wooden shapes, the members of the first training pair are designated as Zog 1 and Vek 1, the second pair as Zog 2 and Vek 2, and the third as Zog 3 and Vek 3.

Each child's listener responses to the /zog/ and /vek/ stimuli were trained in blocks of eight in the same way as for the everyday objects. When responding was at criterion for the Pair 1 stimuli, the same listener training procedure was then implemented first for Pair 2 and finally for Pair 3. Correct responding was reinforced to a criterion of seven out of eight consecutive correct responses. To help ensure that the listener relations were well established, the 3 oldest subjects, NW, RE, and TB, aged approximately 4 years, received overtraining on initial pairs under zero reinforcement to a criterion of seven out of eight correct responses in two successive eight-trial blocks; these subjects then proceeded directly to the category match-to-sample Test 1. Likewise, in order to overtrain the younger subjects' listener relations, while at the same time preventing them from becoming bored with the listener trials, these children proceeded to mixed pairs listener training.

Listener training—Mixed pairs. The six stimuli were reassigned so as to constitute three new *mixed* pairs (e.g., Zog 1/Vek 3, Zog 2/Vek 1, and Zog 3/Vek 2). Listener training for each mixed pair was then conducted in the same way as for the initial pairs to a criterion of seven out of eight consecutive correct responses. Those children whose performances met the criterion were presented with the category match-to-sample Test 1. The exceptions were Subjects JC, BH, and HW who, due to a period of absence from the nursery, were once again tested on the initial pairs under zero reinforcement until their performance met the zero reinforcement learning criterion before proceeding to the category test.

Category match-to-sample tests. These tests were conducted entirely in the absence of reinforcement and differential feedback. In all cases, the possibility of inadvertent experimenter cueing was eliminated by employing: (a) a second experimenter who was familiar to the child but had no knowledge of the particular relations that had been trained, and (b) the screen to obscure the experimenter's face and body.

Step 1 was an *everyday objects category match-to-sample review* in which the second experimenter, using the look-at-sample match-to-others instruction, conducted two category sorting test trials with the six everyday objects, one trial for hats and one for cups. The criterion was correct category sorting of the hats and cups, respectively, in both trials. If performance fell below criterion, the session was terminated and the first experimenter conducted retraining until criterion performance was regained, at which point testing resumed.

Step 2 was the first *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 as in Step 1. In each test trial, the child was presented with all six arbitrary 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 zog [vek] stimulus as sample, selected the remaining two zog [vek] comparisons from the five-stimulus display, this qualified as a correct category sort. Test trials were classified as either valid or null. A valid trial consisted of the child selecting two and only two of the stimuli available in the five-stimulus array. Null trials were those in which the child selected either fewer or more than two comparison stimuli. When too many stimuli were selected, the child was instructed, ''Teddy doesn't want all of them, only some'' and, on the few occasions when only one comparison was given, the child was asked, ''Are there any more?'' The trial was then repeated. The Step 2 procedure continued for a total of 18 valid test trials, during which each of the six stimuli served as sample three times. Six or more correct sorts out of 18 is significantly different from chance; the probability of achieving six correct sorts by chance is .0052. A more stringent criterion, ensuring that both categories (i.e., zog and vek, respectively) have been fully established, is four out of nine (i.e., 44.44%) correct sorts for both categories; the probability of obtaining four correct sorts out of nine is .008. If participants met the latter criterion they were considered to have succeeded on both categories and proceeded to the next phase. If, however, they failed this criterion but met the overall categorization criterion (i.e., over all 18 stimuli) then a further test session was conducted. In the case of children who failed to complete all 18 test trials, only those who completed at least two trials with each sample stimulus (i.e., 12 trials in all) were included in this and subsequent analyses, the criterion for success being set at four or more correct (i.e., 66.66%) sorts over six trials for both categories (the probability of obtaining four or more correct sorts by chance is .001). For some children (particularly the older ones) it was possible to carry out a session in which all the test trials were presented; for others, a number of shorter sessions were required. In the latter case, each new session began with Step 1 before Step 2 testing resumed.

Tact test. Immediately following the category match-to-sample test, the experimenter presented all six arbitrary stimuli, pointed to each one in turn, and asked the child, ''What's this?'', waited 5 s, and if no response occurred, repeated the question once more. If the child still failed to respond, the trial was scored as incorrect and the next scheduled trial instituted. Over three 6-trial blocks, three tact test trials were conducted for each stimulus, making 18 test trials in all. The child was deemed to have tacted the stimuli correctly if he or she produced at least eight out of a possible nine correct tact responses for zog and vek stimuli, respectively. Because Subject LN produced 11 correct tact responses in the first 12 trials ($p < .01$) but then became inattentive and responded at chance level in the last six trials, additional tact test trials were conducted. In the case of another child (MJ), who showed distress at failing to tact the stimuli, only two test blocks were conducted.

Given that the child had passed the previous categorization test then, irrespective of the outcome of the tact test, his or her participation in the study ceased at this point. For children who failed the categorization test, the next phase of the procedure depended on their performance on the tact test trials. Children who failed the latter proceeded to *tact training*. The only exception was Subject RE for whom additional listener training trials were conducted to criterion under zero reinforcement; the category match-to-sample Test 1 and the tact test were then repeated in that order.

The 2 subjects who passed the tact test went on to either category Test 2 (LN) or category Test 1 (HW).

Tact training in pairwise trials. In order to introduce the child to the target tact responses ''zog'' and ''vek,'' a block of eight echoic pretraining trials were presented for Pair 1 (Zog 1 and Vek 1). In each trial, the experimenter placed the Zog 1 and Vek 1 stimuli on the table, pointed to one of these and said, ''This is a zog [vek]. What's this?'' If the child produced the correct vocal response, the experimenter responded, ''Yes, clever girl [boy]. It is a zog [vek].'' If the child produced an inaccurate echoic response or remained silent, the experimenter pointed to the appropriate stimulus once more and provided the additional echoic prompt, ''This is a zog [vek]. Can you say it?''

Next, each child's tact responses of ''zog'' to Zog 1 and ''vek'' to Vek 1 were trained. Stimulus presentations in blocks of eight trials were scheduled in the same way as for pairwise listener training. In each trial, the experimenter placed Zog 1 and Vek 1 on the table, pointed to one of the stimuli and said, ''What's this?'' The order of targeting for each stimulus was quasi-random as for everyday objects training. As in the echoic pretraining block, correct responses were reinforced and incorrect responses were followed by corrective feedback. The criterion was seven out of eight unprompted correct responses in an eight-trial block. When the child responded at criterion for the Pair 1 stimuli, tact training was then introduced for Pair 2 and then for Pair 3.

Tact training in six-stimulus trials. The six 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's this?'' 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 targeted in this way, 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 0%. If performance fell below criterion on the six-stimulus test during trials with a reduced level of reinforcement, 100% reinforcement was reintroduced until the child demonstrated criterion tacting at this reinforcement level. Probability of reinforcement was then once more reduced to zero. Alternation between 100% and 0% reinforcement continued until the child tacted all six stimuli correctly in 18 consecutive trials (i.e., three trials per stimulus) in the absence of reinforcement. Following tact training to criterion, the child was once again presented with the category match-to-sample Test 1.

Repeat of arbitrary stimuli category match-to-sample Test 1. This was conducted in three steps, each in the absence of reinforcement. Step 1 consisted of an arbitrary six-stimulus tact review in which tacting of each of the six arbitrary stimuli was tested; the criterion was consecutive correct tacting of each of the six stimuli under zero reinforcement. If criterion was not achieved, the child recommenced six-stimulus tact training before returning to Step 1. Step 2 was the everyday objects category match-to-sample review conducted in the same way as for the first category match-to-sample Test 1. Step 3 was a repeat of the arbitrary stimulus category match-to-sample Test 1 that employed the look-at-sample instruction. The children who passed the categorization test ceased their participation in the study at this point. The children who failed were next presented with category match-to-sample Test 2.

Arbitrary stimuli category match-to-sample Test 2. The procedure was the same as for the repeat of category match-to-sample Test 1 except that in Step 2, the everyday objects category match-to-sample review, the tact-sample match-to-others instruction: ''What's this? Can you give Teddy the others?'' was employed. This was followed by Step 3, the category match-to-sample test with the arbitrary stimuli, also using the tact-sample instruction. If performance met the overall criterion (i.e., 6 or more correct sorts out of 18) but not the mastery criterion (i.e., four out of nine sorts) for both categories (MJ and NW), then a further block of test trials was conducted. Though Subject BH met the criterion on both categories, a further block of 12 trials was conducted for this child (see Results). Children who failed this category test ceased their participation in the study at this point. For children who passed the test, a further category match-to-sample Test 1 was given, the one exception being LN who proceeded directly to the *Arbitrary stimuli: Set 2* condition.

Second repeat of arbitrary stimuli category match-to-sample Test 1. The children were presented with further trials of the look-at-sample version of the categorization test, completion of which marked the end of all scheduled procedures with the Set 1 arbitrary stimuli.

Arbitrary stimuli: Set 2. Two subjects (JC and LN) participated in Set 2 listener training; the *everyday objects* phase was omitted for these children. Each child was randomly assigned a second set of six wooden shapes, which had not been included in their respective Set 1 training and testing, from the pool of arbitrary stimuli. For each child, the six stimuli were separated into pairs and training with the auditory stimuli /zog/ and /vek/ then proceeded in the same manner as for Set 1.

For Subject JC, correct listener responding was reinforced first for each of the initial pairs, and then for each mixed pair, to a criterion of seven out of eight consecutive correct responses. Following a period of absence due to illness, she was then trained again on initial pairs until her performance in unreinforced test trials met the zero reinforcement criterion. She then progressed to the category match-to-sample Test 1 followed by the tact test, after which her participation ended.

Subject LN also completed initial pairs and mixed pairs listener training to criterion under zero reinforcement before proceeding to the category match-to-sample Test 1. She was given the tact test (48 trials) and additional listener testing before being given category Test 1 and the tact test again. This was followed by tact training and repeats of category Test 1 and Test 2.

Interobserver reliability. The experimenter scored her presentation of the scheduled stimuli and the subjects' responses and, for the entire study, an independent observer scored both these variables in a randomly selected 25% of training trials and 100% of test trials. The independent observer reported no discrepancies between the scheduled and implemented procedures. Interobserver agreement for the subjects' responses was 97.2% across training trials and 100% across test trials.

RESULTS

Everyday Objects

All 9 children completed pairwise listener training in the minimum 24 trials and category sorting to criterion in a mean of 20.7 trials.

Arbitrary Stimuli Set 1

Figures 2 and 3 show numbers of listener and tact training trials, and performance in category tests (filled gray bars) and tact tests (unfilled bars). Table 2 presents subjects' tact behavior during listener training and (unless indicated otherwise) the first tact test.

Listener training—Initial and mixed pairs. The 3 oldest children (NW, RE, and TB) met the initial pairs criterion in a mean of 26.7 trials (range, 24 to 32) and the initial pairs overtraining criterion in a mean of 336 trials (range, 72 to 792). The 6 younger subjects met the initial pairs criterion in a mean of 129.3 trials (range, 32 to 224) and the mixed pairs overtraining criterion in a mean of 62.7 trials (range, 24 to 160). As Table 2 shows, several of the children (BH, LN, HW, NW, RE, and TB) produced extraexperimental names for one or more of the stimuli.

First category match-to-sample Test 1. Figures 2 and 3 show that all 9 children failed the look-at-sample category test. During test trials, 5 children (JC, MJ, BH, CT, and TB) frequently chose as comparisons the two stimuli closest to their right hands; the remaining 4 children showed no systematic selection patterns. None of the children produced any overt task-relevant verbal behavior except for Subject CT, who in one Zog 1 sample trial said, "That's a zog isn't it?" and in another in which Zog 2 was the sample, "That looks like a television."

Tact test. Seven of the 9 children failed the tact test. Of these, MJ produced no tact responses and JC, as shown in Table 2, only produced correct tact responses to Vek 3. Subjects NW and RE alternated their production of the "zog" and "vek" responses across test trials regardless of the stimulus. The other 3 children (BH, CT, and TB) produced predominantly extraexperimental names either for individual stimuli or as common names for several stimuli.

The children who failed the tact test next proceeded to tact training, with the exception of RE. In the course of testing for maintenance of listener behavior at the end of the tact test, her listener performance was found to have deteriorated and so she received additional listener training trials to the zero reinforcement criterion before being tested again on category Test 1, which she failed. She also failed the subsequent repeat tact test in which she again produced "zog" and "vek" responses on alternate trials; she then proceeded to tact training.

Two of the 9 children (LN and HW) met criterion on the tact test. Subject LN, on the second block of 18 trials (shown in Figure 2),

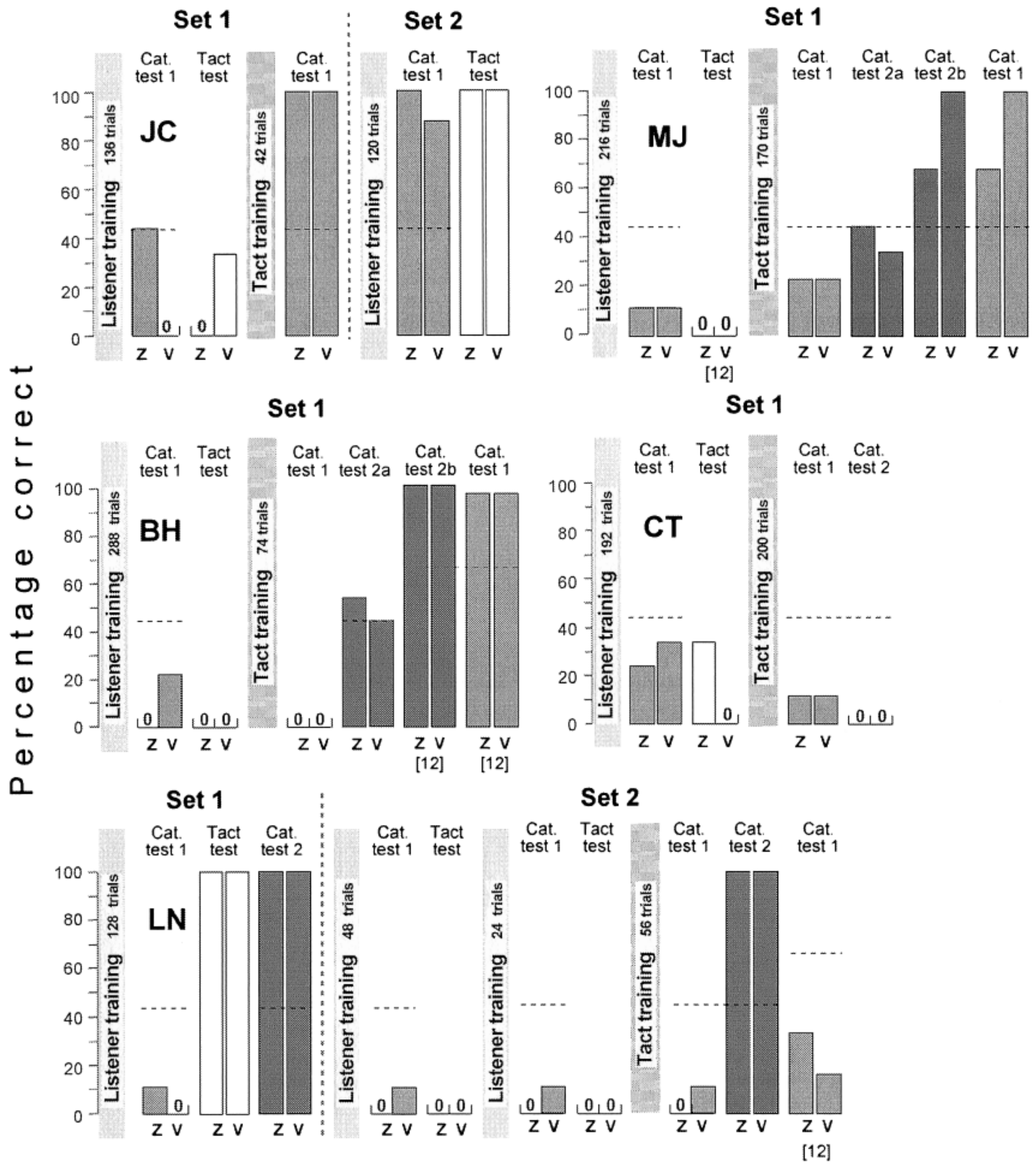

Fig. 2. For 5 of the 9 subjects who reached the category testing stage, performances in the main phases of the study with the Set 1 and Set 2 stimuli, in chronological order for each child. Numbers of listener training trials are shown by light gray bars. Percentage of correct category sorts (dark gray bars) to each zog (Z) and vek (V) sample are presented for the look-at-sample (Cat. test 1) and tact-sample (Cat. test 2) category tests; the test criterion is indicated by a horizontal broken line. If additional testing was given on Test 2, the first set of 18 test trials is labeled Cat. test 2a, and the second set, Cat. test 2b. Percentage correct trials on the tact test are shown by unfilled bars. Where 12, rather than 18, test trials were conducted is indicated by [12]. Numbers of tact training trials are given in the hatched bars.

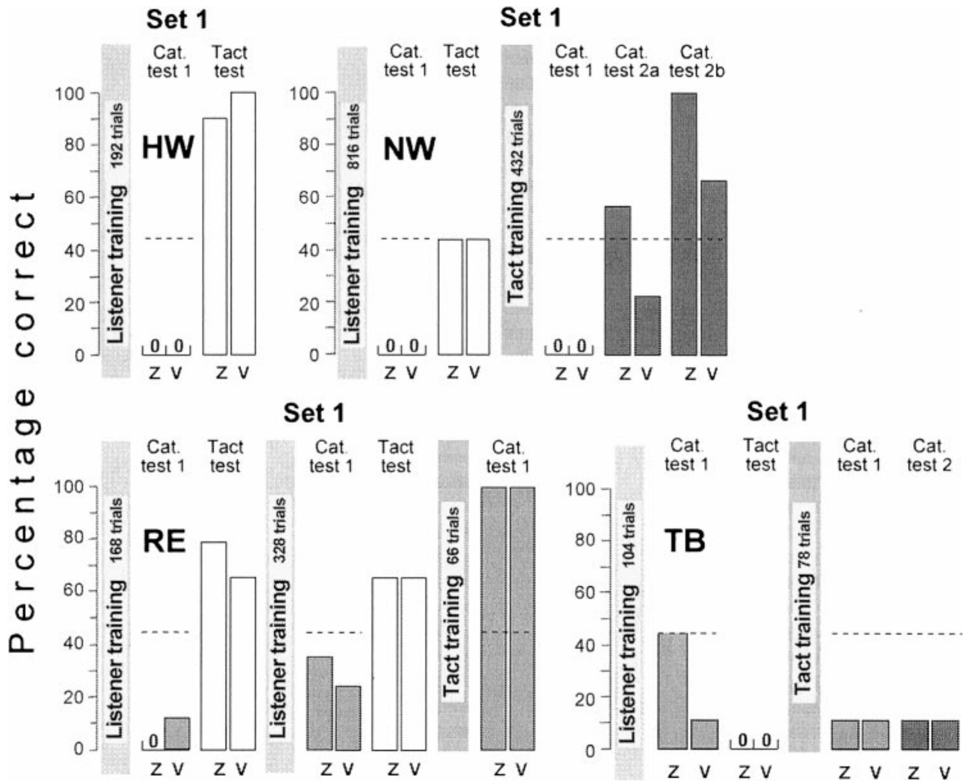

Fig. 3. Performances of the remaining 4 subjects who reached the category testing stage (see Figure 2).

said "u" to each zog stimulus and "i" to each of the veks.

Tact training. All 7 children who failed the tact test successfully completed tact training, the mean number of trials per child being 152 (range, 42 to 432). During training, TB at first continued to call Zog 1 "moon"; when the experimenter corrected him saying, "This is a zog," he responded, "No, this a moon!" Subject CT also initially continued to call Zog 2 "television." In four of the six-stimulus tact training trials, Subject RE ordered the stimuli into two rows, one of zogs and one of veks, saying aloud, "These are zogs and these are veks" as she arranged each row, respectively.

Second category match-to-sample Test 1. Following tact training, 2 of the 7 children (JC and RE) passed the look-at-sample category test. During testing, 2 of the remaining children (BH and CT) continued to show a right hand selection pattern, and 3 (MJ, NW, and TB) selected comparisons unsystematically. None of the children produced experimental or other names for the stimuli.

Subject HW, who passed the tact test, thereafter became increasingly inattentive in test trials, claiming to be bored and refusing to complete the minimum number of 12 trials on category Test 1. No further data are reported for this child.

Category match-to-sample Test 2. Six subjects completed this phase and, unless stated otherwise, made at most only one error when asked to tact the sample. Having already passed the tact test, Subject LN sorted perfectly and she proceeded to Set 2 arbitrary stimulus training. Subject BH produced sample-tact errors on 3 and sorted incorrectly on 9 of the first 12 test trials. In the six-stimulus tact review conducted prior to presentation of the last six test trials, he rearranged the stimuli into two rows, one of veks and one of zogs, naming aloud each vek in the top row and each zog in the bottom row. From this point onward, he tacted the stimuli correctly

Table 2

The tact responses of each subject in response to each arbitray stimulus during listener training and the first tact test. The subject from whom no tacts were recorded (i.e. MJ), and stimulus presentations that were not followed by a tact, are excluded. When a tact was uttered more than once, it is followed by the appropriate number (e.g., Zog x 3). Data are for the Set 1 stimulus set except where indicated as Set 2 (in parenthesis).

Subject	Stimulus	Listener training	First tact test	Subject	Stimulus	Listener training	First tact test
JC	Zog 1		finger		Zog 3		zog; vek x 2
	Zog 3		finger		Vek 1		vek x 2; zog
	Vek 3		vek x 3		Vek 2		vek; zog x 2
BH	Zog 1	big one	square x 2		Vek 3		vek; zog x 2
	Zog 2		circle x 2	RE	Zog 1		zog x 3
	Zog 3		circle x 3		Zog 2		zog x 3
	Vek 1		circle		Zog 3	rain; raindrop; wheel	zog; vek x 2
	Vek 2		circle x 2		Vek 1	like a vek	vek x 3
	Vek 3		zog x 3		Vek 2	ring (x 2); round; twirly	vek x 2; zog
CT	Zog 1		zog x 3		Vek 3	that's a big one	vek; zog x 2
	Zog 2		television x 2	TB	Zog 1	moon	moon x 3
	Zog 3		circle		Zog 2		ball x 3
	Vek 1		buckle x 3		Zog 3		ball x 3
	Vek 2		it's a name		Vek 1		gate x 2; square
	Vek 3		food		Vek 2	people	people x 3
LN	Zog 1		u x 12; en; i		Vek 3		gate x 3
	Zog 2		u x 14	JC (Set 2)	Zog 4	zog x 6	zog x 3
	Zog 3		u x 14		Zog 5	zog	zog x 3
	Vek 1		i x 13; u		Zog 6	zog x 10	zog x 3
	Vek 2		1 x 13; u		Vek 4	vek x 2	vek x 3
	Vek 3	hole	i x 14		Vek 5		vek x 3
HW	Zog 1		zog x 3		Vek 6	vek x 6	vek x 3
	Zog 2		zog x 2	LN (Set 2)	Zog 4		du x 2; a; ai
	Zog 3	bird; triangle whistle	zog x 3		Zog 5		du; n x 5
	Vek 1		vek x 3		Zog 6		b x 2; w x 2; a
	Vek 2		vek x 3		Vek 4		a; w; octopus; dolphin
	Vek 3		vek x 3		Vek 5		du x 2; a; octopus
NW	Zog 1	alphabet; a name? (x 2)	zog; vek x 2		Vek 6		du x 2; a; w
	Zog 2		zog x 2; vek				

and sorted with 100% accuracy in the six remaining trials of Test 2a. Though at this point he had met the mastery criterion for both categories, a further 12 test trials were conducted to determine whether the accuracy displayed on the previous six trials would be sustained, and it was (see Figure 2, Test 2b).

Two children, NW and MJ, met the overall criterion but not the mastery criterion for both categories. When Subject NW was presented with a further 18 test trials (Test 2b), her performance met the sorting criterion for both categories. For MJ, though her tact responses up to this point had been accurate, in the six-stimulus tact review conducted prior to the repeat of the category Test 2, she tacted the veks accurately but produced no tacts to the zogs. In all nine trials in which a vek was the sample, she said ''vek'' and immediately selected the other two veks. In six of the nine zog trials, however, she produced no tact response when the experimenter pointed to the zog sample and asked, ''What's this?'' but first touched all the vek stimuli before selecting the two zog comparisons. In the three remaining zog sample trials, she again said nothing when asked to tact the sample and selected incorrectly. Her performance in Test 2b met the criterion for both categories.

The performance of the 2 remaining children, CT—who continued his right-hand comparison selection pattern throughout testing—and TB, failed to meet the overall criterion and their participation in the study ended at this point.

Third category match-to-sample Test 1. Of the 3 children who reached this stage, MJ and BH passed the test. In the six-stimulus tact review, MJ first named the vek stimuli and then the zogs. Subject BH correctly named the sample in the first 2 trials. NW failed to complete the minimum number of test trials.

Summary. Following listener training, all 9 children failed the look-at-sample Test 1. When tacting of the arbitrary stimuli was established, 6 out of 8 children passed either Test 1 or Test 2. Two children (CT and TB) failed both tests, and following the tact test, 1 child (HW) failed to complete category Test 1.

Arbitrary Stimuli: Set 2

Subject JC. JC met the zero reinforcement listener criterion for the new set of six stimuli in 120 trials (see Figure 2). During listener training trials for Set 2, JC frequently echoed the listener stimulus, and in 25 of the trials she tacted the stimuli as she handed them to the experimenter (see Table 2). In the subsequent category Test 1, JC made only one incorrect sort, and she performed errorlessly in the subsequent tact test.

Subject LN. This child learned the listener relations for the Set 2 stimuli in 48 trials. She failed category match-to-sample Test 1 and produced no correct tact responses in the subsequent tact test. Because LN's performance on the tact test differed so markedly across Set 1 and Set 2, she was given an additional 30 tact test trials, all of which she failed. Her tact responses, which included extraexperimental names, in the first two 18-trial blocks of the tact test are shown in Table 2. Her listener performance was then retested to the zero reinforcement criterion. She was then retested on category Test 1 and the tact test, both of which she failed. Her tact responses in the second tact test were similar to those produced in the first. She was then given tact training. Following this she failed category Test 1 but performed at 100% accuracy on category Test 2.

DISCUSSION

Listener and speaker behavior. A key aim of the study was to use a pairwise listener training procedure with arbitrary stimuli to establish common listener behavior, but not common tacting, in young children. This was achieved for 7 participants, ranging from as young as 1 year 7 months to 4 years 1 month, all of whom completed listener training but failed the tact test. The finding that for most of the children it was possible to train listener behavior without also establishing the corresponding speaker behavior is consistent with Skinner's account of the early stages of children's learning of verbal behavior, according to which listener and speaker behavior are functionally independent (Skinner, 1957, p. 195; and see Lee, 1978; 1981). This claim for functional independence of listener and tact repertoires is supported by research conducted with normally developing young infants (Bell, 1999; Lipkens, Hayes, & Hayes, 1993), but less is known about 2- to 4-year-olds. In the Lowe et al. (2002) study, however, when tact training was given to 2- to 4-year-old children, they learned both the speaker and the corresponding listener behavior, despite the fact that the latter was not directly trained. Consistent with this finding, there are many studies in the developmental literature showing that when a particular tact is present in the verbal repertoires of normally developing young children, so also is the corresponding listener behavior (Harris, Yeeles, Chasin, & Oakley, 1995; Huttenlocher & Smiley, 1987); this is also the case for some autistic children (Keller & Bucher, 1979). Similarly, there are many developmental studies documenting learning of the listener repertoire prior to that of the corresponding speaker behavior (Anisfeld, 1984; Baldwin, 1991; Baldwin, et al., 1996; Foster, 1990; Greenfield & Smith, 1976; Hollich, Hirsh-Pasek, & Golinkoff, 2000; Schafer & Plunkett, 1998; Werker, Cohen, Lloyd, Casasola, & Stager, 1998; Woodward, Markman, & Fitzimmons, 1994).

This asymmetry between the acquisition of listener behavior on the one hand, and of speaker–listener behavior (i.e. naming) on the other, is an important feature of the naming account. Horne and Lowe (1996, p. 202) proposed that whether the corresponding speaker behavior emerges depends on whether the child echoes the listener stimulus while looking at the referent object. This is supported by reports in the developmental literature of a positive correlation between 13-

and 14-month-old infants' rates of echoing and the size of their concurrent naming repertoires (e.g., Bates, Bretherton, & Snyder, 1988), and that 13-month-old infants' echoing of words outside their current naming repertoires is a strong predictor of their naming repertoires at 17 months (Masur, 1995). More recently, in a behavior analytic study, Bell (1999) has demonstrated that when 20- to 23-month-old infants learn listener relations, the corresponding speaker behavior does not emerge until the infants are also trained to echo the listener stimulus (see also Smith, Michael, & Sundberg, 1996) and then only when the infant does so while looking at the referent object. An example of this comes from Subject JC in the present study who, during listener training with the Set 2 stimuli, repeatedly echoed the listener stimulus while looking at the object being named by the experimenter, and then proceeded to name the stimuli herself. Such echoing, however, need not be overt. If a child (or an adult) covertly echoes the listener stimulus in the course of listener training and looks at the named stimulus while doing so, this may be sufficient to establish naming (Horne & Lowe, 1996; Skinner, 1957, pp. 141–146). This may account for the fact that 2 of the children (i.e., LN and HW) produced the common names for the Set 1 stimuli following listener training. As children grow older and, particularly, as their listener and corresponding naming repertoires grow, one might expect that listener training would increasingly also give rise to naming (Horne & Lowe, 1996).

Categorization. According to the naming account, common listener behavior alone is not sufficient to establish categorization of arbitrary stimuli and, given that the children in the present study did not have the corresponding speaker behavior, they should not have categorized. By the end of listener training, all of the children were highly proficient in selecting the correct stimulus from zog/vek pairs when presented with the listener stimuli /*zog*/ or /*vek*/. But when presented with one of the zog or vek stimuli in the look-at-sample categorization test, they failed to select the others in the experimenter-defined stimulus class. None of the 7 children passed this test. Because these children had failed to tact these stimuli, there was no point in conducting the tact-sample version of the test, requiring as it did participants to tact each sample before comparison selections could be made.

It is revealing to compare the failure to categorize of these children, who had learned only listener behavior as opposed to the complete name relation, with the performance of children in whom common naming has been established. In the experiments reported by Lowe et al. (2002), all of the 12 children who were taught to common name the zogs and veks, subsequently went on to pass category match-to-sample tests, 6 passing Test 1 and 6 Test 2. Although the 12 participants in the Lowe et al. experiments underwent common name training from the outset, it also should follow from the naming account that teaching the tact component of the name relation to children, such as the 7 in the present study, who already have common listener behavior but not naming, should produce better success in categorization tests. And, indeed, when the 7 children were given common tact training, 5 of them proceeded to pass either Test 1 or 2. The failure of the remaining 2 (CT and TB) to pass either version of the test, along with some of the failures of other participants who had learned common naming, points to another aspect of the role of naming.

According to the naming account (Horne & Lowe, 1996; Lowe et al., 2002), naming may be necessary for the categorization of arbitrary stimuli, but it is not sufficient. Whether or not the establishment of naming repertoires results in categorization is not a mathematical given but depends on whether, in any given trial, the child overtly or covertly produces the relevant common name and, in turn, whether the latter exerts greater stimulus control than that from competing sources such as an alternative name or a positional selection preference; which source of stimulus control dominates will depend upon the individual's learning history and current stimulus conditions. For example, Lowe et al. found that 6 of their 12 participants who had been taught common naming failed the look-at-sample test and only succeeded when they were prompted to name the stimuli on the tact-sample version of the test. This was also true of 4 children (NW, MJ, BH, and LN-Set 2) in the present study.

The participants in the Lowe et al. (2002)

study, who had tact training from the outset, had little difficulty succeeding on Test 2. One possible explanation for the poorer Test 2 performance in the present study is that over the course of the listener training, category testing, and tact testing that took place, other sources of control may have been established to compete with the common naming that was later trained. For example, in the absence of training to produce experimental names for the wooden shapes during listener training, the children may nevertheless have given their own names either to individual stimuli or to groups of stimuli (see Dugdale & Lowe, 1990; Keller & Bucher, 1979) that in later tests competed with the more recently established experimental names for the control of category sorting. Certainly, at various stages all the children, with the exception of MJ, overtly produced their own individual names for at least some of the stimuli. In the case of BH and TB, they also frequently produced their own common names (e.g., ''circle,'' ''gate,'' and ''ball'') for two or more of the stimuli, as well as other well-established individual names for the remainder.

How stimulus class naming and categorization may occur prior to the formal categorization test was evident in the performance of Subject BH who, in the tact review preceding Trial 13 of category Test 2, rearranged the six stimuli into a row of zogs and a row of veks, naming the stimuli in each of the rows. From this point on he sorted the stimuli without error. Similarly, Subject RE, toward the end of tact training, rearranged the six stimuli into two rows, one of zogs and one of veks, and named them accordingly; she then went on to succeed on category Test 1.

Position preferences are another source of control of stimulus selection that may compete with experimental names. Although initially evident in the performances of just over half of the children in the present study, it was a factor that was most marked in the performance of CT, who from the outset selected the stimuli closest to his right hand regardless of the sample stimulus and who continued to show this bias to the end of the study. This may account for his failure to pass either of the categorization tests.

Generalization of naming and category sorting performances from Test 2 to Test 1 was investigated in 3 subjects (BH, MJ, and LN-Set 2) who, following tact training, had passed the tact-sample category test. Two of these (BH and MJ) then passed the repeat of the look-at-sample test, but LN failed. Though not required to do so, Subject BH named the sample in the first two trials of the final look-at-sample test. This indicates that, once having learned to name the sample in the tact-sample test and to select so-named comparisons, this behavior may, for at least some children.

.

The introduction of common listener training with a second set of stimuli (Set 2) for 2 of the participants (JC and LN) provided further opportunities to consider generalization effects. Having undergone tact training and having succeeded on the subsequent look-at-sample category test, Subject JC frequently echoed the listener stimulus (/zog/ or /vek/) while looking at the correct stimulus in the training pair during the Set 2 listener training trials. These are precisely the conditions under which the naming account predicts that the object stimulus will come to evoke the appropriate tact response (see Horne & Lowe, 1996, p. 202). Though not required to do so, JC tacted the stimuli in 25 of the trials before she completed her listener training. This is clear evidence that she had learned not simply the common listener behaviors but also the corresponding tacts, that is, common naming. Consistent with this observation, Subject JC went on to pass with a high level of accuracy both the look-at-sample category test and the subsequent tact test. Aged only 1 year and 9 months at the start of category testing, JC's performance throughout the training and testing phases of both Sets 1 and 2 is consistent with the naming account, with respect to how naming develops and to how it gives rise to categorization of arbitrary stimuli.

Subject LN was 1 of only 2 children who passed the tact test for the Set 1 stimuli without direct training, and she went on to pass the tact-sample category test. However, when given common listener training on the second set of stimuli, she did not learn to tact them or to sort the stimuli correctly, despite repeated listener training, look-at-sample category testing, and tact testing. It is not clear why, having produced common tacts to the Set 1 stimuli without being directly trained to do so, she did not do likewise with the Set 2 stimuli. One possibility is that she failed to echo the listener stimulus during listener training for Set 2. Having been trained the common tacts for the latter, however, she proceeded to pass the tact-sample version of the test.

Other studies of listener behavior and categorization. The present study appears to be the first within the behavior analytic literature to examine the effects of common listener training on naming and categorization with a group of very young children. Other studies of categorization (e.g., Green, 1990; Sidman, Kirk & Willson-Morris, 1985; Sidman & Tailby, 1982; Sidman, Willson-Morris & Kirk, 1986) have trained common listener relations but in normally developing older children and adults, and in adults with mental retardation. Comparison of their findings with those of the present experiment is further complicated by the fact that their subjects were also given visual-visual conditional discrimination training before, after, or in some cases intermixed with, the common listener trials. We will focus here upon the findings from the subjects in these studies whose auditory-visual training trials were not interspersed with visual-visual trials.

In the three studies by Sidman and colleagues, a total of 16 children aged 5 to 10 years passed the equivalence tests, and all produced a common name for members of each common listener stimulus class. This suggests that children of 5 years and above are more likely than younger subjects, such as those in the present study, to produce a common name for the visual stimuli in trained common listener relations. Certainly, it is expected that as children's naming skills develop, they will increasingly learn to name stimuli to which they might previously have only responded as listeners (Horne & Lowe, 1996). However, it should also be noted that, prior to the main arbitrary stimulus training procedures, the subjects in the studies conducted by Sidman and colleagues were pretrained to select hues in response to the corresponding color names and to produce these color names. Occurring as it did within the experimental context, this pretraining may have increased the likelihood that these children would go on to name the visual stimuli presented during subsequent arbitrary common listener training. Because the major aim of the present study was to test for categorization in the absence of naming, any pretraining that might require the children to name the stimuli was strictly avoided.

In Green's (1990) study of mildly retarded female adults, 2 were given common listener (AB, AC) training, with dictated names for one stimulus set, followed by visual-visual (DE, DF) training for a second set, whereas the converse order of training was given for the remaining 3 subjects. When equivalence testing for both sets was completed, all 5 were asked to name the visual stimuli and, finally, to sort those "that go together in stacks." For the auditory-visual set, all 5 passed the equivalence tests virtually immediately and sorted the visual stimuli correctly. Subjects A1 and A2, who were trained first on the auditory-visual stimulus set, both produced common names for the B and C stimuli. In the case of Subject A1, however, common naming occurred not on the naming test, where she provided individual names for each stimulus, but in the course of the sorting test. Subject V2 individually named the stimuli in the first trial of the naming test but produced the common names in the second and third. The remaining 2 subjects gave individual names. Given the complex verbal instructions used in this study, it is clear that these participants had well-developed verbal repertoires that certainly were sufficient to produce common names, intraverbal sequences of individual names, or other forms of verbally governed behavior known to facilitate success on equivalence tests (see Horne & Lowe, 1996, pp. 215–222; Sidman, 1994, p. 308).

Because auditory-visual training involves dictated common names for the stimuli, common naming is more likely to occur when such training is given before, rather than after, visual-visual training. The study shows again, however, that for a particular set of stimuli, subjects may produce a common name (or even two—see Sidman et al., 1985, 1986), as well as an individual name, for each stimulus; which will predominate in a particular experimental context is not always easy to determine. As the case of Subject A1 demonstrates, though subjects may not answer with a common name in a formal naming test, they may use a common name during the sorting tests themselves (see also Horne

& Lowe, 1996, pp. 216–218). Sidman and colleagues also conducted conditional discrimination training with moderately retarded subjects. In Sidman et al. (1986), there were 2 subjects (AA and PA) who only had auditory-visual training, unmixed with visual-visual trials and who had mental ages of 4 years 5 months and 3 years 1 month, respectively. Subject AA passed the common listener equivalence test and common named the stimuli. PA's data are more difficult to interpret. With retesting, he also eventually passed the common listener equivalence tests and, though he produced some common names as well as individual names for the visual stimuli, it appears that his articulation problems made it difficult to reliably score his responses in the naming test.

Is naming necessary for categorization? Carr et al. (2000) investigated the formation of stimulus classes in severely retarded individuals. Their aim was to assess the naming account by testing for stimulus equivalence in humans ''who have never developed the capacity for speech. Such a deficit effectively eliminates the route by which the naming process operates, and equivalence test outcomes should be negative'' (p. 102). They conducted two experiments with participants who had poorly developed language skills and who were aged from 13 to 21 years old. Their findings that 4 out of the 5 subjects passed tests of stimulus equivalence, they conclude, thus poses major problems for the naming account. Their fundamental premise, however, is based on a misreading: The naming account has never, in fact, equated verbal behavior with speech. Rather, it has, in line with Skinner's (1957, 1969) analysis, explicitly indicated that modalities other than speech (e.g., manual signing) can serve for naming just as well as oral repertoires and, whatever the modality, naming may occur overtly or covertly (Horne & Lowe, 1996, pp. 208–209; 1997, p. 294; Lowe & Horne, 1996, pp. 324–325; see also Catania, 1996). Neither does the naming account, as Carr et al. maintain, require that subjects have ''well developed language'' (p. 110) in order to pass tests of equivalence. It is enough simply to have either a common name for each of the designated classes, or individual names for class members that are linked intraverbally (Horne & Lowe, 1996). This is borne out by the results of the present study and those of Lowe et al. (2002), which show that children as young as 1 year 7 months, who have limited verbal repertoires, readily learn to produce common names for sets of arbitrary stimuli, and to pass tests of equivalence and categorization.

As Carr et al. (2000) acknowledge, the use of standard language tests to compare the naming skills of older mentally retarded people with those of very young children is not straightforward and may underestimate the much greater language experience of the older subjects. Nevertheless, regardless of which of the language tests were used in the Carr et al. study, none of their 4 subjects who passed equivalence tests had lower age-equivalent scores than 2 years. This was true of the 3 participants in Experiment 1, who not only had some oral naming skills, but also had been taught manual signing (DJB and JRV) or picture-based augmentative communication (DJB and IVB). Unfortunately, 2 of the 3 subjects were not given the language tests in the medium of their non-oral verbal repertoires. The third, DJB, who had achieved the age-equivalent score of 2 years 1 month on an oral vocabulary test, was scored at 2 years 5 months when retested on a version of the test that only measured signing. Although it may not be clear, then, exactly how these subjects might compare with young normally developing children, particularly when both oral and manual communication skills are considered together, the evidence does not support the contention that they lacked the basic naming skills necessary for categorization.

Indeed, Experiment 1 appears to have been designed expressly to capitalize upon such skills. If the aim were to show that equivalence could occur in the absence of naming, then it would have been advisable to use unfamiliar arbitrary stimuli for which these subjects had no existing names. Instead, as the nodal stimuli, Carr et al. (2000) used, variously, pictures of a shirt, dog, and cake and, as auditory stimuli, the corresponding spoken names (i.e., ''shirt,'' ''dog,'' and ''cake''). These were all stimuli familiar to the subjects prior to the study and they responded appropriately as listeners to the spoken words though Carr et al. did not assess whether they could also produce these particular names. If these names were in place, then because the three familiar visual stimuli always served as the comparisons (B stimuli) in the many-to-one (AB, CB, DB) procedure, common naming of the stimuli conditionally related to them is likely to have developed. Moreover, in baseline training the ''shaping'' procedure involved presenting in the sample location each B stimulus (e.g., the shirt) merged first with its corresponding C stimulus (e.g., the letter ''s'' [for shirt]), and then later its corresponding D stimulus (a shape). Presenting each C and D stimulus in turn in a compound with the familiar nameable B stimulus creates ideal conditions for the establishment of common naming of these stimuli (see Skinner, 1957, pp. 99–102, on metonymical extension).

Interpretation of the results from Experiment 2 is even more problematic. For example, although the oral verbal assessment of the 13-year-old subject (BN) reported to have passed the equivalence test yielded an age-equivalent score of 2 years, this assessment was conducted 3 years prior to the study taking place (see Carr, 1997). Similarly, the assessment of the second subject (HF) was conducted up to a year before the study. (Carr et al.'s [2000] paper does not specify when the Experiment 1 verbal assessments were conducted.) Furthermore, BN also had extensive training in the Makaton signing system and used ''a large number of signs'' (p. 105). Unfortunately, however, his signing repertoire, particularly as it related to the stimuli employed in Experiment 2, was not assessed. It also should be noted that BN's performance on the equivalence test is likewise difficult to interpret given that (a) on both tests he failed to show BA symmetry and only passed the combined equivalence and transitivity tests on the fourth 8-trial block, and (b) the stimuli were not randomly assigned, and no pretests were conducted for existing color or form-based relations among the stimuli (e.g., Green, 1990; see also Zentall, 1996, on natural categories). In summary, then, Carr et al.'s study, although providing interesting data on categorization in severely retarded people, does little to clarify the role of naming in this phenomenon (cf. O'Donnell & Saunders, 2003).

Though the present findings and those of Lowe et al. (2002) are consistent with, and were predicted by, the naming account, they pose problems for theories centered on the notions of stimulus equivalence (Sidman, 1994; 2000) and relational framing (Hayes, 1996; Hayes & Hayes, 1989, 1992). Neither of these latter two theories predict that common naming, but not common listener training, will give rise to categorization—nor is it clear how they could account for such a fundamental difference in behavioral outcome. The equivalence account specifically predicts that common listener training should as readily establish equivalence relations among the constituent arbitrary stimuli as would common speaker training and, indeed, much of the early work on equivalence employed common listener training to establish the baseline relations among arbitrary stimuli (see Sidman, 1994, pp. 19–80). It is also not clear how these theories could explain why the tact-sample form of the category test, in which the subject is required to respond verbally to the sample before selecting the comparisons, should be more effective in establishing stimulus classes than its look-at-sample counterpart. According to the naming account, naming is indeed special—it may be the only way that organisms can categorize arbitrary stimuli. Learning common listener responses simply will not do. This is in marked contrast to stimulus equivalence theory that assigns no special role to verbal behavior or to naming in particular. Sidman (2000) is clear on this point:

> Any name we apply to stimuli is a defined discriminative response. Our theory states explicitly that any defined response components of the contingencies have a status that is equal in every way to the stimulus and reinforcer members of the classes. Although just as important, responses require no separate treatment. (p. 145)

The data presented here and by Lowe et al. (2002) do not accord with this theory (and see Horne & Lowe, 1996, 1997; Lowe & Horne, 1996).

Whatever the merits or demerits of the naming account, it does have the virtue of being disconfirmable. For a key feature of the account to be disproved it is only necessary, for example, to show that categorization of arbitrary stimuli occurs reliably in organisms, such as preverbal infants and nonverbal animals, that cannot name these arbitrary stimuli, or in verbal children who have learned listener behavior but not naming. A central purpose of the present study, in combination with that of Lowe et al. (2002), was to subject the account to this latter test. If it had been shown that children who did not common name the stimuli were as successful as those who did, then the account would have to be abandoned or, at least, radically modified. But far from there being no difference between these conditions, the gap was absolute—only those children who showed evidence of common naming categorized. These findings provide strong support for the proposition that naming is indeed necessary for the establishment of arbitrary stimulus classes. At the very least, the detailed results presented from the 21 participants across both sets of studies show how naming arbitrary stimuli, as opposed to merely looking and pointing at them, is a powerful determinant of categorization (see Dugdale & Lowe, 1990; Eikeseth & Smith, 1992; Lowe & Beasty, 1987; Mandell & Sheen, 1994; Randell & Remington, 1999; K. J. Saunders et al., 1993).

REFERENCES

Anisfeld, M. (1984). *Language development from birth to three.* London: Erlbaum.

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

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development, 67,* 3135–3153.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms.* Cambridge, England: Cambridge University Press.

Beasty, A. (1987). The role of language in the emergence of equivalence relations: A developmental study. *Index to Theses with Abstracts, 39,* 1149B.

Bell, M. H. (1999). *The determinants of naming in human infants.* Unpublished doctoral dissertation, University of Wales, Bangor.

Bentall, R. P., Dickins, D. W., & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 46B,* 187–214.

Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition, 60,* 1–29.

Carr, D. (1997). *Stimulus equivalence, naming and contextual control: Studies with language disabled, autistic adolescents.* Unpublished doctoral dissertation, University of Wales, Cardiff.

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.

Catania, A. C. (1996). Natural contingencies in the creation of naming as a higher order behavior class. *Journal of the Experimental Analysis of Behavior, 65,* 276–279.

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.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993). *MacArthur communicative development inventories: User's guide and technical manual.* San Diego: Singular Publishing.

Fields, L., Matneja, P., Varelas, A., Belanich, J., Fitzer, A., & Shamoun, K. (2002). The formation of linked perceptual classes. *Journal of the Experimental Analysis of Behavior, 78,* 271–290.

Fields, L., Reeve, K. F., Matneja, P., Varelas, A., Belanich, J., Fitzer, A., et al. (2002). The formation of a generalised categorization repertoire: Effect of training with multiple domains, samples, and comparisons. *Journal of the Experimental Analysis of Behavior, 78,* 291–314.

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.

Foster, S. H. (1990). *The communicative competence of young children.* London: Longman.

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26,* 796–804.

Gibson, E. (1969). *Principles of perceptual learning and development.* New York: Appleton.

Gopnik, A., & Meltzoff, A. N. (1992). Categorization and naming: Basic-level sorting in eighteen-month-olds and its relation to language. *Child Development, 63,* 1091–1103.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development, 71,* 1205–1222.

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

Greenfield, P. M., & Smith, J. H. (1976). *The structure of communication in early language development.* New York: Academic Press.

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

Harris, M., Yeeles, C., Chasin, J., & Oakley, Y. (1995). Symmetries and asymmetries in early lexical comprehension and production. *Journal of Child Language, 22,* 1–18.

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 Press.

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

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 4,* 285–301.

Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development, 63* (3, Serial No. 262).

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.

Huttenlocher, J., & Smiley, P. (1987). Early word meanings: The case of object names. *Cognitive Psychology, 19,* 63–89.

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.

Keller, M. F., & Bucher, B. D. (1979). Transfer between receptive and productive language in developmentally disabled children. *Journal of Applied Behavior Analysis, 12,* 311.

Kemler Nelson, D. G., Frankenfield, A., Morris, C., & Blair, E. (2000). Young children's use of functional information to categorize artifacts: Three factors that matter. *Cognition, 77,* 133–168.

Lee, V. L. (1978). Teaching generalised receptive and productive behindfront discriminations to two retarded children. *Journal of Applied Behavior Analysis, 11,* 529.

Lee, V. L. (1981). Prepositional phrases spoken and heard. *Journal of the Experimental Analysis of Behavior, 35,* 227–242.

Lionello-DeNolf, K. M., & Urcuioli, P. J. (2002). Stimulus control topographies and tests of symmetry in pigeons. *Journal of the Experimental Analysis of Behavior, 78,* 467–496.

Lipkens, R., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology, 56,* 201–239.

Liu, J., Golinkoff, R. M., & Sak, K. (2001). One cow does not an animal make: Young children can extend novel words at the superordinate level. *Child Development, 72,* 1674–1694.

Lowe, C. F., & Beasty, A. (1987). Language and the emergence of equivalence relations: A developmental study. *Bulletin of the British Psychological Society, 40,* A42.

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

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

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

Masur, E. (1995). Infants' early verbal imitation and their later lexical development. *Merrill-Palmer Quarterly, 41,* 286–306.

Nazzi, T., & Gopnik, A. (2000). A shift in children's use of perceptual and causal cues to categorization. *Developmental Science, 3,* 389–396.

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

Pilgrim, C., 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.

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.

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., 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). North-Holland: 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.

Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development, 69,* 309–320.

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

Schusterman, R. K., & 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.

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

Sidman, M., & Tailby, W. (1982) Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 5–22.

Sidman, M., Willson-Morris, M., & Kirk, B. (1986). Matching-to-sample procedures and the development of equivalence relations: The role of naming. *Analysis and Intervention in Developmental Disabilities, 6,* 1–9.

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

Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis.* New York: Appleton-Century-Crofts.

Sloutsky, V. M., Lo, Y.-F., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. *Child Development, 72,* 1695–1709.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition, 60,* 143–171.

Smith, R., Michael, J., & Sundberg, M. L. (1996). Automatic reinforcement and automatic punishment in infant vocal behavior. *The Analysis of Verbal Behavior, 13,* 39–48.

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

Sutton, J. E., & Roberts, W. A. (2002). Failure to find evidence of stimulus generalization within pictorial categories in pigeons. *Journal of the Experimental Analysis of Behavior, 78,* 333–344.

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

Welder, A. N., & Graham, S. A. (2001). The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development, 72,* 1653–1673.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology, 34,* 1289–1309.

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month olds. *Developmental Psychology, 30,* 553–566.

Zentall, T. R. (1996). An analysis of stimulus class formation in animals. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 15–34). Amsterdam: Elsevier.

Received June 4, 2003
Final acceptance May 16, 2004