JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2006, **85,** 247–273 NUMBER 2 (MARCH)

NAMING AND CATEGORIZATION IN YOUNG CHILDREN: IV: LISTENER BEHAVIOR TRAINING AND TRANSFER OF FUNCTION

PAULINE J. HORNE, J. CARL HUGHES, AND C. FERGUS LOWE

UNIVERSITY OF WALES, BANGOR

Following pretraining with everyday objects, 14 children aged from 1 to 4 years were trained, for each of three pairs of different arbitrary wooden shapes (Set 1), to select one stimulus in response to the spoken word /zog/, and the other to /vek/. When given a test for the corresponding tacts (''zog'' and ''vek''), 10 children passed, showing that they had learned common names for the stimuli, and 4 failed. All children were trained to clap to one stimulus of Pair 1 and wave to the other. All those who named showed either transfer of the novel functions to the remaining two pairs of stimuli in Test 1, or novel function comprehension for all three pairs in Test 2, or both. Three of these children next participated in, and passed, category match-to-sample tests. In contrast, all 4 children who had learned only listener behavior failed both the category transfer and category match-to-sample tests. When 3 of them were next trained to name the stimuli, they passed the category transfer and (for the 2 subjects tested) category match-to-sample tests. Three children were next trained on the common listener relations with another set of arbitrary stimuli (Set 2); all succeeded on the tact and category tests with the Set 2 stimuli. Taken together with the findings from the other studies in the series, the present experiment shows that (a) common listener training also establishes the corresponding names in some but not all children, and (b) only children who learn common names categorize; all those who learn only listener behavior fail. This is good evidence in support of the naming account of categorization.

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

---

This is the fourth in a series of studies designed to investigate how organisms come to categorize stimuli that have no distinguishing physical features in common. This topic has attracted considerable interest; first, because a substantial body of research has shown that when categories of arbitrary stimuli are established, new or ''emergent'' behaviors are observed that have never been directly trained or reinforced (Sidman, 1994). This finding is theoretically challenging and, at the same time, holds considerable promise for the practical application of behavioral principles to real-life problems. Another reason why this work on arbitrary stimulus classes has received so much attention is the view that it may hold the key to understanding much of what is distinctive about human language (Hayes, 1996; Hayes & Hayes, 1989, 1992; Horne & Lowe, 1996, 1997; Lowe & Horne, 1996; Sidman, 1971, 1994, 2000).

Within the behavior-analytic literature, however, there are different theories concerning the origins of arbitrary stimulus classes and their role in verbal behavior. The present study represents one of a series of studies that put to empirical test Horne and Lowe's (1996) naming account, according to which the categorization of arbitrary stimuli is driven by naming and other verbal behavior. Naming is a bi-directional speaker–listener relation that establishes category relations between a set of stimuli, given only that each particular stimulus in the set occasions the same name. Indeed, so defined, naming *is* categorization. The functional properties of name relations are demonstrated by (a) emergent name-based patterns of category sorting among different sets of arbitrary stimuli, and (b) untrained transfer, to all members of the same name relation, of novel behaviors that are trained to just one exemplar. A central theoretical issue concerns whether, as the naming account proposes, naming is necessary for the establishment of arbitrary stimulus classes or wheth-

---

We are grateful to Gareth Horne, Paula Gurteen, and Heather Roberts for their help in the preparation of this manuscript, to Chris Whitaker for his advice on statistics, and to Pat Lowe for her editorial contribution. 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, UK.

Correspondence concerning this article should be addressed to Professor C. Fergus Lowe, School of Psychology, University of Wales, Bangor, Brigantia Building, Penrallt Road, Bangor, Gwynedd LL57 2AS (E-mail: c.f.lowe@bangor.ac.uk).

doi: 10.1901/jeab.2006.125-04

er they can occur in its absence. The present study, which focuses particularly on transfer of stimulus functions, was designed in conjunction with the others in the series to address this question directly. Before considering the naming account in more detail, however, we first outline other behavioral accounts of arbitrary stimulus class formation and our evaluation of their standing in relation to the empirical evidence.

Stimulus Equivalence

Sidman (1971, 1994, 2000) has proposed that arbitrary stimulus categorizing arises from a basic behavioral ''given'' termed *stimulus equivalence*. Category relations are said to emerge when, for example, a subject is trained in a match-to-sample procedure to select a comparison stimulus (e.g., D1) when each of several arbitrary stimuli (e.g., A1, B1, C1) serves as sample and to select another comparison (e.g., D2) when several other arbitrary stimuli (e.g., A2, B2, C2) serve as samples. Given that stimulus equivalence is said to ensure that, for any learned relation, its symmetric counterpart emerges without training, the result should be one behavioral partition or two categories (D1, A1, B1, C1 and D2, A2, B2, C2, respectively) in which, in addition to those that were directly trained, all other possible pairwise combinations of the partition members simply emerge in what are described as *event pairs*. For Sidman, categorizing and novel behavior transfer are one and the same: If a novel behavior is trained to one member of an equivalence class (e.g., R1 is trained to A1), then the novel behavior enters the relevant partition and may subsequently form an event pair with any other members of that partition. Within the bounds of the theory, therefore, the term *transfer* is redundant (Sidman, 1994). Given that partitions may be generated simply by training a minimal number of pairwise conditional relations among a set of arbitrary stimuli, verbal behavior, according to Sidman, is not necessary for the categorizing of arbitrary stimuli in humans. Moreover, unless proven otherwise, it is the case that all animals and not just humans have the potential to learn categories among arbitrary stimuli via equivalence.

However, despite considerable efforts (e.g., Jitsumori, Siemann, Lehr, & Delius, 2002; Lionello-DeNolf & Urcuioli, 2002), there is still little convincing evidence for stimulus equivalence in nonhuman animals (but see Kastak, Schusterman, & Kastak, 2001; Schusterman & Kastak, 1993; for an alternative interpretation of both studies, however, see Horne & Lowe, 1997, pp. 284–288). In contrast, many humans, though by no means all, have succeeded on tests of stimulus equivalence. One key variable in instances of test failure in humans may be the extent to which subjects are instructed, or are indeed able, to respond verbally during the match-to-sample procedures. For example, in their review of human performance in equivalence tests, K. Saunders, Williams, and Spradlin (1996, p. 100) observed that, ''Symmetry appears to be less likely in subjects with severe to moderate mental retardation and profoundly limited language skills''. However, Horne, Lowe, and Randle (2004) have argued that because no study with severely retarded participants has, to date, adequately measured their verbal repertoires, whether vocal or manual, particularly in relation to the stimuli employed in the procedures, the definitive study on how language repertoires affect symmetry and categorization in people with mental retardation has yet to be conducted (see Horne et al., 2004; cf. Carr, Wilkinson, Blackman, & McIlvane, 2000; O'Donnell & Saunders, 2003).

Recently, there has been a renewed focus on rates of attrition in arbitrary match-to-sample studies in other human populations. For example, it appears that unless special additional procedures are employed, normally developing young children often fail completely to learn the baseline relations and so do not go forward to the equivalence tests (Auguston & Dougher, 1991; Pilgrim, Jackson, & Galizio, 2000). Attrition is also prevalent in studies of subjects with mental retardation (O'Donnell & Saunders, 2003). Many potent interventions for facilitating baseline learning and/or equivalence test performance in these populations are explicitly verbal. These include: introducing experimenter-provided names for one or more of the stimuli, and/or specific relational task instructions (Green, 1990; Pilgrim et al., 2000); providing spoken words as samples in auditory-visual match-to-sample procedures (Green, 1990); requiring subjects to name the sample and comparison stimuli in a class-consistent manner (Beasty, 1987; Bentall, Dickins, & Fox, 1993); and training children on common name relations among arbitrary stimuli before embedding the latter in a category sorting task (Lowe, Horne, Harris, & Randle, 2002). There are other less obvious verbal influences: The match-to-sample procedures employed as pretraining in several of the seminal studies by Sidman and colleagues required their subjects to select the appropriate color (e.g., a red square) in response to the spoken word (i.e., /red/) and next to name that color (e.g., see Sidman, Kirk, & Willson-Morris, 1985, pp. 238–239; Sidman, Willson-Morris, & Kirk, 1986, pp. 291–292). This color-naming procedure may in effect have trained the subjects to name the other arbitrary stimuli subsequently presented in the main match-to-sample task (see Pilgrim et al., 2000). Others have found that nameability of the arbitrary stimuli determines success in the tests (Mandell & Sheen, 1994) and that class formation is enhanced when the names of the potential class members rhyme (Randell & Remington, 1999). Lastly, when subjects, prior to conditional discrimination training, undergo extensive listener training with some of the stimuli that subsequently feature in the prospective stimulus class, this listener training may give rise to corresponding naming and, as a result, help to offset attrition in match-to-sample procedures, particularly in populations with mental retardation (Carr et al., 2000; O'Donnell & Saunders, 2003; cf. Horne et al., 2004).

How conditional discrimination training is structured also affects test performance. When subjects, particularly young children and people with mild to moderate retardation, are given many-to-one conditional discrimination training, they generally pass the tests of equivalence more readily than when a one-to-many training structure is employed (see R. Saunders, Drake, & Spradlin, 1999, for a review). Participants in one-to-many baseline procedures are more prone to fail at first testing, often necessitating many sessions of repeated testing before outcomes are successful (e.g., Green, 1990); even so, some never pass the tests (e.g., Subject PM, Sidman et al., 1985). The structure of training and testing also has been shown to affect test outcomes in normally developing adults (Fields et al., 1997; Horne & Lowe, 1996, p. 235). The conventional match-to-sample procedure is a forced-choice task. However, when subjects are given the option of indicating that there is no correct comparison present in a given test trial, the frequency of performances consistent with equivalence diminishes (Innis, Lane, Miller, & Critchfield, 1998). It appears that, given more choice, verbally-able humans often choose not to respond in a manner that can be described as equivalence in match-to-sample test procedures. Even in humans, therefore, success in match-to-sample equivalence tests is far from guaranteed.

The relation between success on match-to-sample tests of equivalence and transfer of function is also less straightforward than Sidman's (1994) version of the theory suggests it should be. In humans, some studies have reported transfer of functions among members of previously established equivalence classes (e.g., Barnes, Browne, Smeets, & Roche, 1995; Barnes & Keenan, 1993; Dougher, Auguston, Markham, Greenway, & Wulfert, 1994; Kohlenberg, Hayes, & Hayes, 1991; Lazar & Kotlarchyk, 1986; Wulfert & Hayes, 1988), whereas others have reported transfer failures (e.g., Bones et al., 2001; Fields, Landon-Jimenez, Buffington, & Adams, 1995; Greenway, Dougher, & Markham, 1995; Sidman, Wynne, Maguire, & Barnes, 1989; Smeets, Barnes, & Roche, 1997).

Relational Frame Theory

Hayes and Hayes (1992) have proposed that when an organism is given an appropriate learning history, relations among stimuli and behavior are governed by means of contextually-specified, arbitrarily-applicable relational framing. These arbitrary relations are initially trained either by operant (e.g., Steele & Hayes, 1991) or respondent-like (e.g., Schenk, 1995) procedures, but thereafter these relations may extend without training to novel arbitrary stimuli. This generalization of relational responding depends on the presence of a discriminative (or, as RFT theorists term it, a ''contextual'') stimulus throughout the training of a particular relation (e.g., sameness) among two or more arbitrary stimuli. After a history of such training a point is reached whereby the discriminative stimulus, when it is presented together with novel arbitrary stimuli, instantiates that relation (i.e., sameness) among them, without training. In the case of a relation of sameness, it is claimed that the relevant contextual stimulus determines that the stimuli it accompanies are related bidirectionally to each other, in what is termed a *frame of coordination*. In addition, any novel function trained to one of the stimuli will transfer, via what is termed *transformation of function*, to any others in the same contextual frame. This is the simplest example of how, according to the theory, a child may categorize a particular collection of novel arbitrary stimuli.

The Naming Account

According to Horne and Lowe (1996), category relations among arbitrary stimuli may be established by training the same name to each potential category member to produce a common name relation. The speaker–listener relation that is naming establishes bidirectional relations between each and every member of the common name relation via the common name that each occasions. That is, if a child is trained to say ''vek'' to each of several arbitrary stimuli, when he or she is later presented one of those stimuli and names it, the resulting listener stimulus, /vek/, should evoke the child's orienting as a listener to all other stimuli that he or she has previously learned to call ''vek''. The common listener behavior the child instigates when he or she produces the name draws together the otherwise disparate stimuli to produce a category relation that may be assayed in terms of emergent, name-based patterns of category sorting and untrained transfer of novel behavior among members of the common name relation. The category sorting prediction was confirmed in a study conducted by Lowe et al. (2002) in which the 2- to 4-year-old children, who first learned two three-member common name relations among arbitrary stimuli, all proceeded to sort the stimuli into common name categories. The naming account would be disconfirmed, however, if the same effects were to occur among children who had learned only the corresponding common listener relations. In a second study, Horne et al. (2004) found that when 7 children aged from 1 to 4 years learned only common listener relations among the same arbitrary stimuli as were employed in Lowe et al., none of them showed name-based sorting in the subsequent category match-to-sample test. However, when those 7 ''listeners'' then were trained to produce the corresponding tact responses, that is, the common names, 5 went on to pass the category tests. Taken together, these two studies of category sorting support the hypothesis that the full bidirectional name relation may be necessary for the categorizing of arbitrary stimuli by young children.

The prediction that naming is also a powerful means of transferring novel functions among members of a common name relation was next investigated by Lowe, Horne, and Hughes (2005). Nine children, aged from 1 to 4 years, were trained, in a pairwise procedure, to tact three arbitrary stimuli as ''zog'' and another three as ''vek''. Following common tact training, they were next trained, for example, to clap to one zog stimulus and to wave to one of the veks. All 9 children showed name-based transfer of clapping and waving when presented with the remaining zogs and veks in transfer Test 1. Seven of the children also participated in, and passed, transfer Test 2 in which, for each stimulus pair, they selected the correct stimulus when the experimenter clapped (or waved) and then asked, ''Can you show me which one goes like this?'' Four of the children were next given the category match-to-sample test, which they all passed. When 3 of the children were trained on two additional stimulus sets, both the transfer- and category-sorting measures showed that nine-member arbitrary stimulus classes were established.

Though the naming account predicts that common naming brings about category transfer of novel function, such as was observed in the Lowe et al. (2005) name-training study, it also claims that these effects should not be observed if the children were to learn only common listener relations among the arbitrary stimuli. Accordingly, the main aim of the present study was to test whether common listener relations alone could establish category transfer of function. The procedure was the same as that employed in Lowe et al., except that, at the outset, the children were given common listener, but not common tact, training with six arbitrary stimuli (Set 1). A subsequent tact test determined which of the children had learned to name the stimuli with the common names ''zog'' and ''vek'' by the time they had completed listener training, and which had only listener repertoires. As in Lowe et al., all the children were trained, for

Table 1

Participants' sex, age at start of procedure, age at first category test, and score on the Griffiths Mental Development Scale (GMDS).

Subject	Sex	Age at start (years/ months*)	Age at testing (years/ months*)	GMDS General Quotient
RH	M	1/10	2/06	113
BB	F	1/10	2/00	
CM	F	2/00	2/07	121
FJ	F	2/01	2/03	
MH	F	2/02	2/03	
SH	F	2/03	2/05	138
SO	F	2/07	2/09	119
MW	F	2/08	2/11	
AJ	M	2/08	3/00	109
PW	F	3/00	3/02	116
MD	M	3/00	3/03	
CD	M	3/06	3/07	
FLJ	F	3/10	4/00	108
CG	M	4/00	4/04	120

* age in days rounded to nearest whole month (i.e., 16 days and more rounded up).

example, to clap to one zog and wave to one vek exemplar and then proceeded first to transfer Test 1 and then to transfer Test 2; a subset of the children also were given the category match-to-sample test. Any child who failed the transfer tests and had learned only the listener relations was next trained to produce the corresponding names; these children then were retested on the transfer-of-function tests and a subset were given a repeat category-sorting test. The Set 1 procedures provided the opportunity to establish whether a novel function trained to members of Pair 1 would transfer to those of Pair 2 and Pair 3. In order to determine whether the function trained to Pair 1 would extend to additional arbitrary stimuli, 3 children were given listener training (as for the Set 1 stimuli) with three further arbitrary stimulus pairs (i.e., Set 2). Following the tact test, and without novel function training to any of the Set 2 stimuli, transfer of the Set 1 novel functions, both as production (transfer Test 1) and comprehension (transfer Test 2), was tested among the six Set 2 stimuli.

METHOD

Subjects

Subjects were 9 girls and 5 boys who attended the Daycare Nursery and Center for Child Development at the University of Wales, Bangor. The children were aged from 1 year 10 months to 4 years at the start of the study (see Table 1). A Griffiths Mental Development Scales (Griffiths, 1954) assessment was conducted with 8 of the children, all of whom scored within the normal range for children of their age group (see Table 1). The 6 remaining children, who left the nursery before they completed the developmental tests, showed no sign of developmental delay.

Apparatus and Stimuli

The sessions were conducted in the daycare center, in a research room equipped with two wall-mounted color videocameras that were controlled remotely from an audiovisual console room; one camera recorded the behavior of the child, the other of the experimenter, and both outputs were recorded using split-screen format. All audio output during sessions was recorded via a radio microphone worn by the experimenter. The child and the experimenter sat at opposite sides of a small table. To conceal the experimenter from the child's view during test sessions, a wooden screen was placed on the table between the child and the experimenter (see Lowe et al., 2002 for dimensions and procedural details). The experimental stimuli were of two types: (a) six everyday objects—three different hats and three different cups; and (b) 18 arbitrary green wooden shapes (for examples, see Lowe et al., 2002, Figure 1). The main scheduled reinforcer was social praise, supplemented occasionally with stickers. At the end of each session, the children chose either to receive several stickers for their personal sticker books, or to play for several minutes with a teddy bear.

Procedure

Everyday objects. Experimenter 1 first established a good rapport with the children during unstructured daily play sessions, following which they were taken into the experimental room one at a time. During the first session the experimenter introduced ''Teddy'', a hand puppet, to the child. For each child, the everyday objects, three hats and three cups, were randomly divided into three pairs, each of which consisted of one stimulus from each category. Each child's training pairs remained

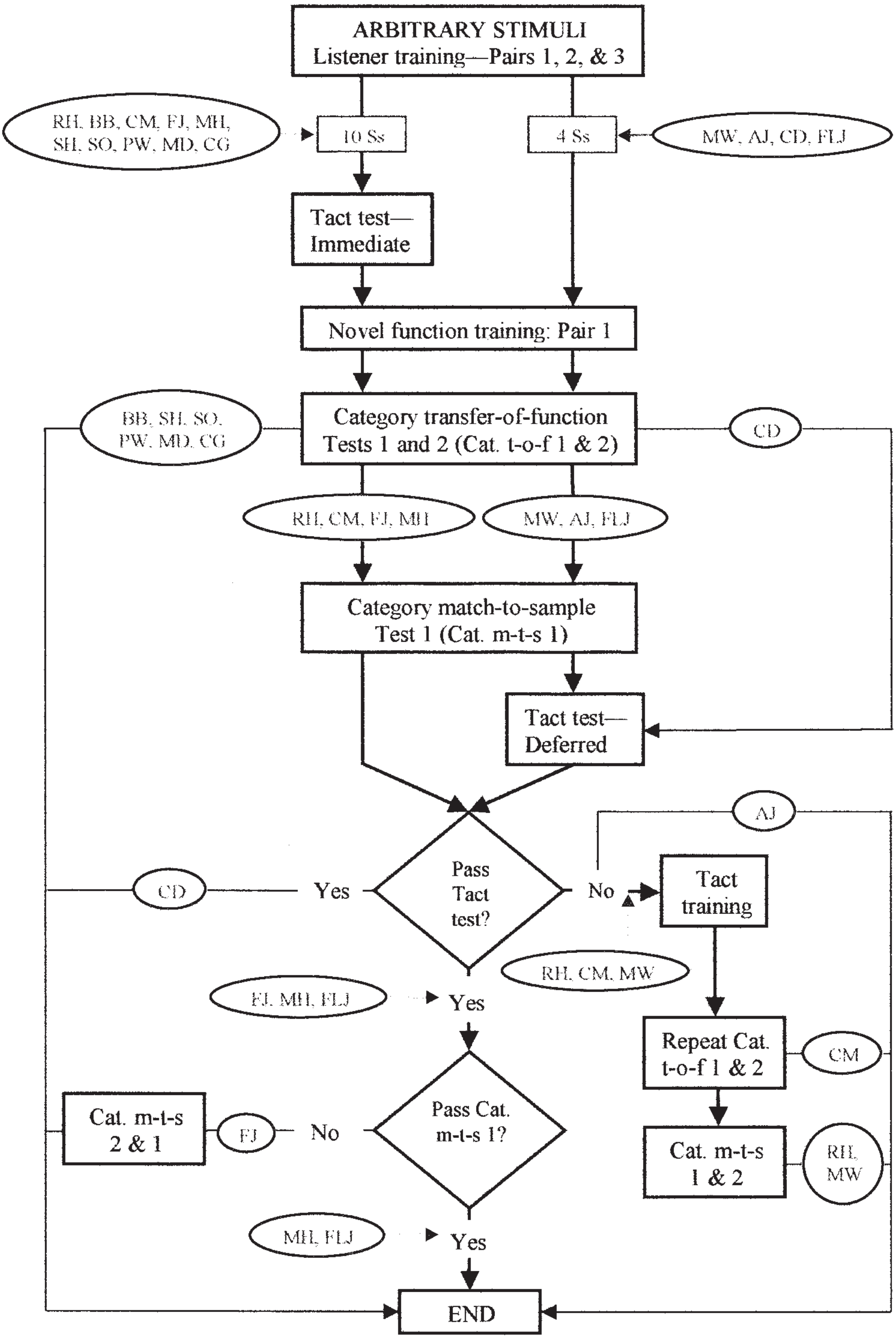

Fig. 1. Outline sequence of arbitrary stimulus training and testing phases for each participant. Tact testing was conducted for 10 subjects (see left-hand track) immediately after tact training (immediate) and for the remaining 4 subjects (see right-hand track) after their final category test (deferred).

constant throughout the experiment. Experimental sessions varied in duration from 5 to 20 min, dependent on the individual child. The children's verbal behavior was recorded during all phases.

Listener overtraining in pairwise trials. The first pair of stimuli (Hat 1 and Cup 1) was presented on the table in front of the child. Only one stimulus was targeted on each trial, following which both stimuli were removed from the table and repositioned for the next trial. The stimuli were placed equidistant from the child's midline about 20 cm apart and approximately 20 cm away from the table edge closest to the child. In each trial the experimenter said, ''Look at these; can you give me the hat [cup]?'' If the child selected the correct object, the experimenter delivered social praise (e.g., ''Clever girl [boy]''). If the child selected the incorrect object, or produced no response, the experimenter provided corrective feedback, saying, ''No, that's not the hat [cup]. Can you give me the hat [cup]?'' and, if necessary, performed the correct response. The stimuli then were removed from the table. Trials were scheduled in a quasi-random order in blocks of eight in which each target stimulus was presented twice on the left of the child's midline and twice on the right, and the same trial type did not appear twice in succession. The learning criterion was seven out of eight correct responses within one eight-trial block. The same training then was conducted for Pair 2 and Pair 3.

Conventional function overtraining: Pair 1. Only the Pair 1 stimuli were employed in this phase. The two stimuli were placed in front of the child, the experimenter pointed to either Hat 1 or Cup 1 and said, ''Look at this; it goes like this.'' The experimenter then performed the conventional function appropriate for that particular stimulus. For the hat, this was putting it on his head; for the cup, he mimed drinking from it. The experimenter then pointed to the hat or cup and asked the child, ''Can you show me how this goes?'' If the child responded correctly, the experimenter delivered social praise; if the child did not respond or responded incorrectly, the experimenter gave corrective feedback saying, ''It goes like this [the experimenter performed the correct action]; can you show me how it goes?'' Once the child had responded to this instruction correctly over one block of trials for both the Hat 1 and Cup 1 stimuli, subsequent instructions were abbreviated to, ''Can you show me how this one goes?'' The scheduling of trials within a block and the learning criterion were as in pairwise listener overtraining.

Category transfer-of-function test: Pairs 2 and 3. This stage was designed to establish correct responding to the instructions that would be employed later during testing with the arbitrary objects. The screen was placed on the table between the child and experimenter who reached through the aperture to place the Pair 2 stimuli on the table in their predetermined positions in front of the child, then pointed to either the hat or the cup and said, ''Can you show me how this one goes?'' Each stimulus was targeted four times in one randomized and counterbalanced block of eight test trials. The same procedure then was carried out with the Pair 3 stimuli. No feedback was given during this test. If any child failed the test, the conventional hat and cup behaviors were trained to Pair 2 and/or Pair 3, as required.

Arbitrary stimuli set 1: Listener training in pairwise trials—Initial pairs. For each child, a set of six stimuli was selected at random from the pool of wooden shapes. The listener stimuli employed were the spoken words /zog/ and /vek/ (see Lowe et al., 2002). The experimenter randomly divided the six stimuli into three initial pairs; in each pair, one stimulus was designated to serve as a zog and the other as a vek. The general pairwise listener training procedure was as described for the everyday objects except that the experimenter presented the first stimulus pair and said, ''Look at these; can you give me the zog [vek]?'' The mastery criterion was seven out of eight correct responses over two consecutive eight-trial blocks. When criterion had been reached with the Pair 1 stimuli, the procedure was repeated with Pair 2 and then Pair 3. In order to consolidate the children's listener relations, while at the same time preventing their becoming bored with the listener trials, the children next proceeded to listener training with mixed pairs.

Listener training in pairwise trials—Mixed pairs. The experimenter reorganized the stimuli into three new mixed stimulus pairs so that each zog stimulus had a new vek pair member. The stimuli remained in the mixed pairs arrangement for the remainder of the study during which they were designated for reporting purposes as Pair 1 (Zog 1/Vek 1), Pair 2 (Zog 2/Vek 2), or Pair 3 (Zog 3/Vek 3). The training procedure and the learning criterion for these new mixed pairs were the same as for initial pairs training.

Reinforcement reduction. The purpose of this stage was to maintain 100% correct responding in the absence of reinforcement—the conditions employed in all subsequent test procedures. The criterion was correct responding throughout one block of listener trials in the absence of reinforcement. If the criterion was not met in the first test block, a VR2 schedule was introduced until responding was 100% correct; following this, performance was again tested in the absence of reinforcement, and so on until the zero reinforcement criterion was met on Pair 1. The same reinforcement reduction procedures were next implemented for the Pair 2 and Pair 3 stimuli. In all other respects the procedure was identical to that of the previous training stages.

Tact tests. For 10 of the subjects, tact responses to each of the six arbitrary stimuli were next tested in the absence of reinforcement (Immediate tact test—see Figure 1, left-hand track). To examine the effects of presenting the tact test not just before but also after category testing, 4 of the children (MW, AJ, CD, and FLJ), chosen on a random basis, were given the tact test after all category testing had been completed (Deferred tact test—see Figure 1, right-hand track). To minimize the possibility of experimenter-cueing, Experimenter 2 was introduced at this stage; she was known to the children but was not acquainted with the specifics of the study. Experimenter 2 placed one of the stimulus pairs on the table, pointed to one of the shapes and said, ''What's this?'' The child was given approximately 4 s to respond, and, if no response was given, the question was repeated and a further 4-s response interval ensued. The next scheduled trial was then begun.

The Pair 1 stimuli were presented on four trials, each stimulus being targeted twice, once on the left and once on the right, in a prespecified quasi-random order. This was followed by four trials with Pair 2, and then with Pair 3, arranged in the same manner. This procedure was repeated until each stimulus had been targeted four times, making 24 trials in total. Eighteen or more correct responses over 24 trials (i.e., combining both categories) would be significantly different from chance ($p < .01$). A further criterion, ensuring that both categories (i.e., zog and vek respectively) have been fully established, is 9 out of 12 (75%) correct responses per common listener category ($p = .05$).

Novel function training: Pair 1. Clapping and waving were selected as the novel behaviors to be trained to Zog 1 and Vek 1. These are common behaviors in the repertoires of most children and are often under various forms of stimulus control. These novel relations were trained only with the Pair 1 stimuli. For each child, one of the behaviors (either clapping or waving) was randomly assigned to the Zog 1 stimulus and the other to Vek 1. The two Pair 1 stimuli were placed in front of the child; Experimenter 1 pointed to one of them and said, ''Look at this; it goes like this.'' The experimenter then performed the action for that particular stimulus (i.e., waving or clapping as appropriate). This was followed by, ''Can you show me how it goes?'' If the child responded correctly the experimenter delivered social praise or, otherwise, provided corrective feedback. Once the child had responded reliably to the above instruction across one block of trials, the experimenter shortened subsequent instructions to, ''Can you show me how this one goes?'' The scheduling of trials and mastery criterion were as for the Pair 1 listener behavior training.

Reinforcement reduction. The child's production of the novel functions with the Pair 1 stimuli was next tested in the absence of reinforcement. If performance was 100% correct over one eight-trial block, the child progressed to category transfer-of-function testing; otherwise, the reinforcement reduction procedure described for listener training was implemented.

Category transfer-of-function Test 1: Novel behavior production. This was completed in two stages. Step 1, conducted prior to each of the test sessions, ensured that all the trained relations were intact. Experimenter 1 conducted four listener test trials with each of the three stimulus pairs, followed by four test trials of the novel behavior (i.e., clap or wave) trained to Pair 1 (two trials with Zog 1 and two with Vek 1). The experimenter provided no reinforcement or corrective feedback during this stage; the criterion was 100% correct responses.

In Step 2, testing was conducted by Experimenter 2 who, following a prespecified randomized sequence, placed either Pair 2 or Pair 3 on the table in front of the child and, pointing to the scheduled target stimulus and having ensured that the child was attending said, ''Look at this; can you show me how this one goes?'' Trials were organized into blocks of eight and were counterbalanced within each block with the constraint that the same trial type did not occur twice in succession. When the child had responded, the next trial was begun and so on until each of the four stimuli had been targeted eight times, four times on the left and four times on the right, making a total of 32 test trials. If the child did not respond within 4 s, Experimenter 2 repeated the instruction and waited a further 4 s. If the child still did not respond, the trial was marked as incorrect, as were trials in which the child produced the wrong response. No reinforcement or corrective feedback was given during test sessions. Twenty-four correct responses over 32 trials (i.e., combining both categories—the overall mastery criterion) would be significantly different from chance ($p < .01$). A further criterion, ensuring that both categories (i.e., zog and vek respectively) have been fully established, is 12 out of 16 (75%) correct responses per common listener category ($p < .05$).

Category transfer-of-function Test 2: Novel behavior comprehension. All three of the stimulus pairs were used to test for the emergence of listener behavior (comprehension) to the novel behaviors that were trained only to the Pair 1 stimuli. Pair 1 was included in this test because only production, and not comprehension, of the novel behaviors had been trained directly to the Pair 1 stimuli. Prior to each of the test sessions, trained behaviors were reviewed (as in Step 1 of category transfer-of-function Test 1). Then, in each trial, Experimenter 2 presented one of the three stimulus pairs, in a prespecified random order. She asked the child, ''Can you give me the one that goes like this?'' and modeled either a clap or a wave gesture. The child was given a maximum of two opportunities per trial to respond, as in Test 1. Over 24 test trials, each of the six possible listener relations (i.e., see a clap [wave] —select an arbitrary object from Pair 1, 2, or 3) was tested four times, counterbalanced across trials, with the target object presented twice on the left and twice on the right. The trials were otherwise conducted in the same manner as for category transfer-of-function Test 1. Eighteen correct responses over 24 trials (i.e., combining both categories) would be significantly different from chance ($p < .01$). A further criterion, ensuring that both categories (i.e., zog and vek respectively) have been fully established, is 9 out of 12 (75%) correct responses per common listener category ($p = .05$).

Of the 10 subjects who were given the tact test immediately after listener training, 4 (RH, CM, FJ, and MH) were selected on a random basis to progress to the category match-to-sample test; Set 1 procedures ended at this point for the remaining 6 (see Figure 1, left-hand track). Of the 4 subjects who had not yet received the tact test (see Figure 1, right-hand track), 3 of them (MW, AJ, and FLJ) progressed to the category match-to-sample test, following which they were given the tact test. The remaining subject, CD, who was due to leave the nursery shortly, proceeded directly to the tact test after which his participation in the study ended.

Category match-to-sample Test 1. Prior to the test trials with the arbitrary stimuli, it was necessary, as in other test phases, to ensure that the participants could respond appropriately to the experimental instructions. For this reason, Experimenter 2 first conducted the category match-to-sample procedure with the familiar objects. Experimenter 2 placed the stimuli on the table in two rows of three, in randomized predetermined positions. She then picked up the prespecified target stimulus (either a hat or a cup) and said to the child, ''Look at this; can you give Teddy the others?'' Experimenter 2 then waited for the child to complete his or her selections from the five remaining stimuli. All six stimuli then were removed from the table and repositioned for the subsequent trial, and so on, until each hat and each cup had served as the sample stimulus once (i.e., six trials). The criterion was correct sorting of the hats and cups by giving the other two hats when a hat was the sample, or the other two cups when a cup was the sample, in all six trials. If any of the children did not pass all six trials, training trials were conducted (see Lowe et al., 2002);

otherwise, the children progressed to the arbitrary stimulus category match-to-sample Test 1.

The procedure for the arbitrary match-to-sample test was the same as for the everyday stimuli except that the six arbitrary stimuli were employed; each of these served as the sample three times (18 trials in all). Trials were classified as correct if, when presented with a vek [zog] stimulus as the sample, the child selected the remaining two veks [zogs] from the five-stimulus array; all other responses, such as giving more or less than the required two stimuli, and/or giving stimuli from the other designated category, were classified as incorrect. Six correct sorts over 18 trials (i.e., combining both categories) would be significantly different from chance ($p < .01$). A further criterion, ensuring that both categories (i.e., zog and vek respectively) have been fully established, is four out of nine correct sorts per common listener category ($p < .01$).

Subjects who passed the tact test (immediate or deferred) and category match-to-sample Test 1 (MH, FLJ), ended their participation in the Set 1 procedures. Any subject who passed the tact test but failed category match-to-sample Test 1 proceeded to category match-to-sample Test 2 (FJ). Of the 4 subjects who failed the tact test, Subject AJ left the nursery at this point (see Figure 1) and the remaining 3 (RH, CM, and MW) were next given tact training.

Category match-to-sample Test 2. The procedure was the same as for Test 1 except that the Test 2 task instruction, "What's this? Can you give Teddy the others?" required the subject to tact the sample before selecting the comparisons. Six trials, with a different sample stimulus in each, were first conducted with the hat and cup stimuli. When performance was 100% correct in trials with the everyday objects, 18 test trials were conducted with the six arbitrary stimuli.

Repeat of category match-to-sample Test 1. Irrespective of their performance in Test 2, and to control for repeated testing, subjects next participated in a further category match-to-sample Test 1. Set 1 procedures ended for Subject FJ at this point.

Tact training. For all three mixed pairs, subjects (RH, CM, and MW) were trained to produce the tacts ("zog"/"vek") in accordance with the previously trained listener relations. In each of the first block of trials the experimenter placed one pair of stimuli in their pre-specified positions in front of the child and as he pointed to each stimulus said, "Look at this, it's a zog [vek], can you say zog [vek]?" In subsequent trial blocks the child was asked for each stimulus, "What's this?" If the child produced the correct vocal response, the experimenter said, "Yes, clever girl [boy]. It is a zog [vek]." If the child produced an inaccurate response or remained silent, the experimenter pointed to the appropriate stimulus once more and provided the additional prompt, "This is a zog [vek]. Can you say it?" Trials were otherwise scheduled as for the mixed pairs listener training and to the same pairwise mastery criteria.

Repeat category transfer-of-function tests. The repeats of Test 1 (novel behavior production) and Test 2 (novel behavior comprehension) were conducted in the same way and to the same criteria as for the first presentation of these tests. One subject (CM) left the nursery after completing these tests.

Repeat category match-to-sample tests. For the 2 remaining subjects (RH and MW) the repeats of the category sorting tests were conducted in the same way and to the same criteria as on their first presentation. This completed training and testing with the Set 1 arbitrary stimuli.

Arbitrary stimuli: Set 2. Three subjects (SO, PW, and CG) participated in listener training for an additional set of arbitrary stimuli, Set 2. From the pool of arbitrary stimuli each child was randomly assigned six wooden shapes that had not been included in their previous training; as for Set 1, the listener stimuli were the auditory stimuli /zog/ and /vek/. The Set 2 procedures employed for these subjects were the same as those implemented for their respective Set 1 stimuli except that the everyday objects phases conducted for Set 1 were omitted and, following listener training, no novel function training was conducted with any of the Set 2 stimuli. All 3 subjects were given the tact test with the Set 2 stimuli immediately after listener training. The Pair 1 stimuli of Set 1 were next presented and performance of the novel (clap/wave) behaviors that had previously been trained only to these stimuli during Set 1 procedures was retested in the absence of reinforcement on one eight-trial block. If performance was still at criterion, subjects proceeded directly to the

Set 2 transfer tests. If not, these behaviors were retrained in accordance with the Set 1 procedures and mastery criteria. The Set 2 category transfer-of-function Test 1 measured generalization of the Set 1 novel behaviors to each of the six Set 2 stimuli. Thirty-two correct responses over 48 trials (i.e., combining both categories) would be significantly different from chance ($p < .01$). A further criterion, ensuring that both categories (i.e., zog and vek respectively) have been fully established, is 18 out of 24 (75%) correct responses per common listener category ($p < .01$). The category transfer-of-function Test 2 measured transfer to the Set 2 stimuli of comprehension of the novel behaviors established with the Pair 1, Set 1 stimuli. The mastery criteria for the Set 2 transfer-of-function Test 2 were the same as for the Set 1 stimuli.

Interobserver reliability. An independent observer scored all trials in a randomly selected 25% of all training sessions; interobserver agreement on these trials was 97.8%. Similarly, all test trials were scored; interobserver agreement on these trials was 100%. The independent observer reported no discrepancies between the scheduled and implemented procedures.

RESULTS

Everyday Objects

Listener overtraining, and category transfer-of-function testing. All 14 children completed the pairwise listener overtraining to criterion in an average of 27.2 trials (range 24–48). Most produced the conventional functions trained with the Pair 1 objects to criterion in one trial block; exceptions were Subject FJ who required four blocks and Subjects MH, SH, and AJ, who required two. All children subsequently passed the category transfer-of-function test with the Pair 2 and Pair 3 objects.

Arbitrary Stimuli Set 1

Subjects' data shown in Figures 2 and 3 are grouped according to their performances in the tact test and category tests. Figure 2 shows the listener training, tact test and category test performances for the 8 subjects who passed the tact test given immediately after listener training and went on to pass the categorization test (i.e., ''namers/categorizers''). Figure 3 (top row) shows comparable data for 2 subjects who were given, and passed, the tact test *after* succeeding on the category tests (top row). Next shown are the data for subjects (''listeners/noncategorizers'') who, in addition to the category tests failed the immediate tact test (middle rows) or the deferred tact test (bottom row). In Figures 2 and 3, the individual category mastery criterion for each test is shown as a horizontal dashed line.

Listener training in pairwise trials: Initial pairs, mixed pairs, and reinforcement reduction. All subjects completed listener training; the average number of trials to reach the 100% reinforcement criterion on the Initial Pairs 1, 2, and 3 was 53.7 (range 16–152), 46.3 (range 16–168), and 45.1 (range 16–208), respectively. Subjects SO and FLJ met the criterion in the minimum number of training trials (16) on all pairs and 4 subjects (BB, MH, SH, and CG) did so on two of the pairs but required one further eight-trial block for the third. The mean number of trials to criterion on Mixed Pairs 1, 2, and 3, at 100% reinforcement, was 27.4 (range 16–104), 44.0 (range 16–208), and 36.6 (range 16–152), respectively. Eight subjects (BB, FJ, MH, SO, MD, CD, FLJ, and CG) met the criterion for all three mixed pairs in the minimum total number of 48 trials. A mean number of 9.7, 10.3, and 10.3 trials were required for all subjects to meet the mixed pairs zero reinforcement criterion for Pairs 1, 2, and 3, respectively. Nine subjects met the mixed pairs zero reinforcement criterion in the minimum of one eight-trial block per pair; the remaining 5 subjects required two or three blocks on one or more of the pairs.

Tact test. Eight of the 10 subjects who were given the tact test immediately after listener training passed the test (see Figure 2); Subjects RH and CM failed (see Figure 3, middle panel). Of the 4 subjects who were tested after they completed all category testing, 2 passed (see Figure 3, top panel); Subjects MW and AJ failed (see Figure 3, bottom panel). Table 2 shows each subject's verbal responses to each stimulus in the tact test. Of the 10 subjects who passed the tact test, only Subject CG made an error. Of the 4 subjects who failed, 2 (RH and CM) scored zero mostly due to their making no verbal responses, whereas the remaining 2 (MW and AJ) scored well below chance, either not responding at all, or producing familiar object names or wrongly applied experimental tacts.

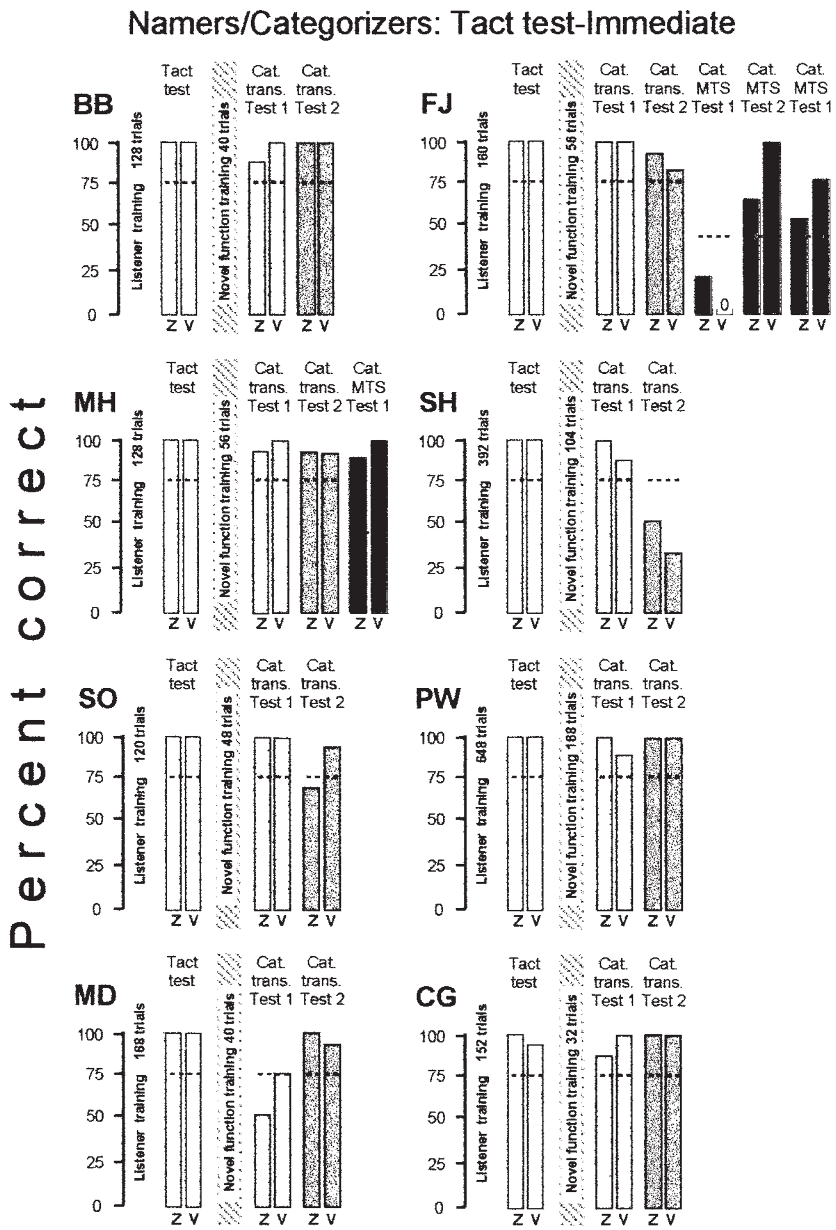

Fig. 2. Performances in each experimental phase of the 8 subjects who passed the arbitrary stimulus categorization tests and the tact test given immediately following listener training with the arbitrary stimuli (namers/categorizers). Numbers of arbitrary stimulus listener training trials to criterion are shown in the first column, followed in the next pair of columns (unfilled) by percent correct "zog" and "vek" responses in the tact test. Numbers of Pair 1 novel-function training trials to criterion are shown in the third column, followed in the next two column pairs by performance in the category transfer Test 1 (light gray) and the category transfer Test 2 (dark gray), respectively. For Subjects FJ and MH, performance in the category match-to-sample Test 1 is next shown (black) and, for FJ, performance on Test 2 and a repeat of Test 1 are also shown in the final pair of columns. The individual category mastery criterion on test trials is shown as a horizontal broken line.

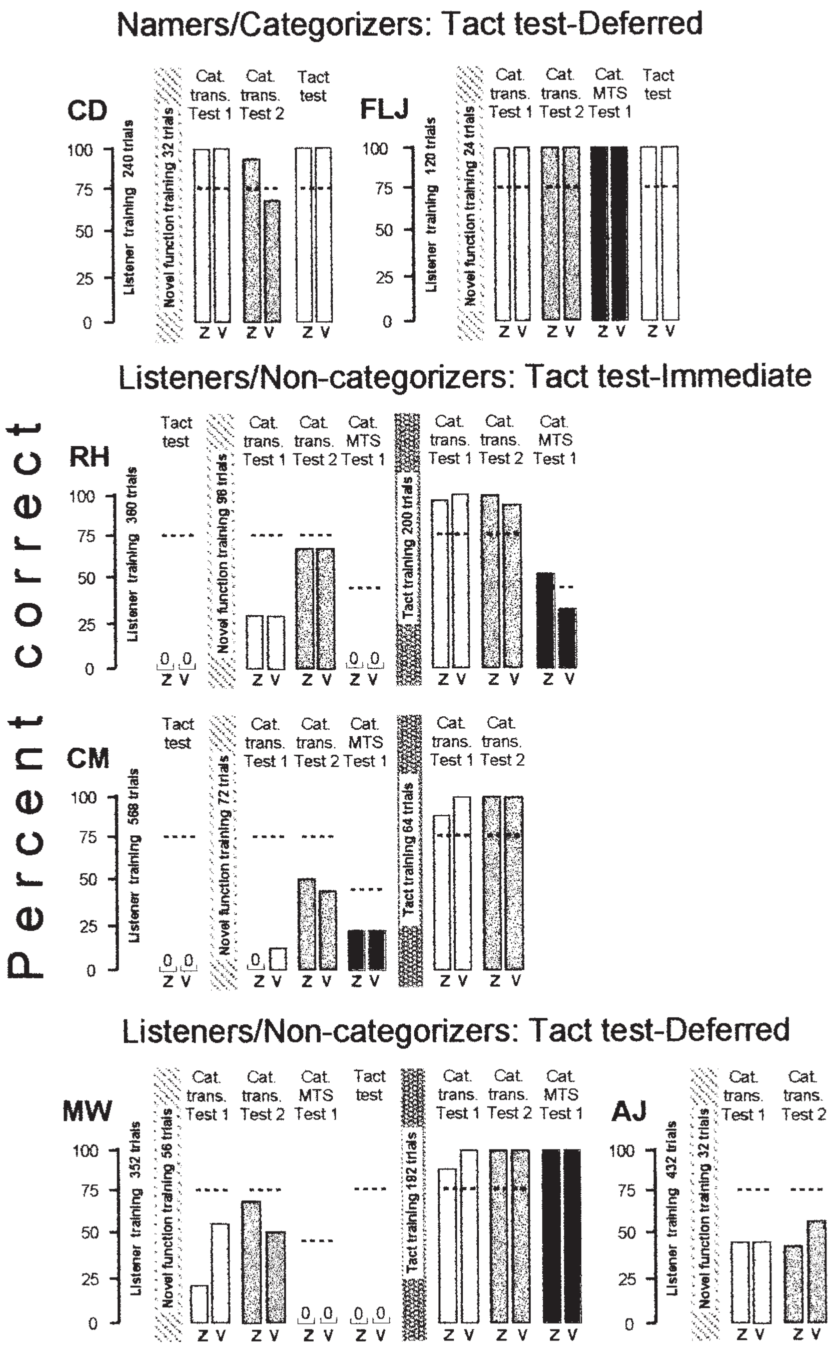

Fig. 3. As for Figure 1, except that the first 2 subjects were namers/categorizers who passed the tact test given at the end of category testing (deferred). Next shown are the performances of the listeners/non-categorizers who failed either the immediate (Subjects RH and CM) or the deferred (Subjects MW and AJ) tact test and, except for Subject AJ, were given tact training followed by repeat category tests.

Novel function training: Pair 1. The mean number of trials required to learn the novel behaviors for the Pair 1 stimuli under continuous reinforcement was 51.4 (range 16–152). When the scheduled reinforcement was reduced to zero, 9 subjects met the 100% correct criterion in the minimum of one eight-trial block, while the remainder required one further test block. Though subjects were not required to verbalize during novel function training trials, 3 subjects (SO, MD, and FLJ) said ''zog'' or ''vek'' (appropriately) to the target stimulus in at least one of these trials. Subject FLJ also described the relation between the Zog 1 stimulus and the trained waving response (see Table 2).

Category transfer-of-function Test 1: Novel behavior production. Of the 10 subjects who passed the tact test, all met both the overall and the individual category mastery criteria with the exception of MD who failed to reach the overall criterion and the individual criterion with the zog stimuli (see Figure 2 and Figure 3, top panel). These children produced no overt stimulus names during the test with the exceptions of Subject SO who correctly named Zog 2 in the second trial of the transfer test and Subject FLJ who before she clapped her hands said, ''The vek goes like this'' and, on another trial, ''The big one [zog] goes like this'' before she waved (see Table 2). All 4 subjects (RH, CM, MW, and AJ) who failed the tact test also failed transfer Test 1 (see Figure 3, middle and bottom panels); none of them produced any verbalizations during the test.

Category transfer-of-function Test 2: Novel behavior comprehension. Of the 10 subjects who passed the tact test, all met the overall mastery criterion, with the exception of SH. Seven subjects met the individual criterion on both categories, 2 subjects (SO and CD) met the individual mastery criterion on one category, and Subject SH failed to do so on both (see Figure 2 and Figure 3, top panel). Only Subject BB was heard to name any of the stimuli during the test and she named each zog stimulus ''zoggy'' before she selected it. All 4 children who failed the tact test also failed transfer Test 2 (see Figure 3, middle and bottom panels); none of them produced any on-task verbal behavior.

In sum, all 10 children who passed the tact test met the overall mastery criterion on one or both of the category transfer-of-function tests—9 passed Test 1 (MD failed) and 9 passed Test 2 (SH failed)—whereas all 4 children who failed the tact test failed both transfer tests.

Category match-to-sample tests. Of the 3 subjects who passed the tact test and were given category match-to-sample tests, all passed; 2 (MH and FLJ) succeeded in Test 1 whereas Subject FJ failed Test 1 but passed Test 2 and then also succeeded in a repeat of Test 1 (see Figure 2 and Figure 3, top panel). The 4 subjects (RH, CM, MW, and AJ) who failed the tact test were also given the category match-to-sample Test 1 which they all failed (see Figure 3, middle and bottom panels).

Tact training. Subjects RH, CM, and MW, who failed the tact test, learned the tact relations for Pair 1, Pair 2, and Pair 3 of the Set 1 stimuli in a mean of 61.3 (range 24–88), 48.0 (range 16–64), and 42.7 (range 24–56) trials, respectively.

Repeat category tests. Figure 3 (middle and bottom panels) shows that following tact training, all 3 subjects passed the repeat of the category transfer-of-function Tests 1 and 2. Two subjects (RH and MW) were next given a repeat of the category match-to-sample Test 1. Subject MW performed correctly in all 18 trials. Subject RH passed the overall sorting criterion; he produced a total of 8 out of 18 correct sorts ($p < 0.0002$) but did not meet the individual category mastery criterion on the vek stimuli.

Set 2 Stimuli: Auditory Listener Training

As shown in Figure 4, the 3 subjects (SO, PW, and CG) learned the Set 2 listener behaviors in a mean of 165.3 trials (range 120–232). All 3 passed the tact test without error, except for CG who made one error. All 3 subjects also passed both category transfer-of-function Tests 1 and 2.

DISCUSSION

Common Naming and Categorization

All three previous studies in this series (Horne et al., 2004; Lowe et al., 2002; 2005) have shown that common naming of arbitrary stimuli is a very effective means of establishing stimulus classes in young children. The present results are best viewed in the context of the earlier experiments and, to facilitate this, we

Table 2

The task-relevant verbal responses of each subject (presented in order of his or her age) to individual stimuli during the tact test, Pair 1 novel function training (Novel funct. train.), and the category transfer Tests 1 and 2 (cat. trans.). When an utterance occurred more than once, the appropriate number is indicated (e.g., zog × 3); no responses in the tact test are shown as NR. Data are for the Set 1 stimulus set except where indicated as Set 2.

Subject	Stage/phase	Target Stimulus	Verbalization
RH	Tact test	Zog 1, 3	NR ×4
		Zog 2	NR ×3; uhm
		Vek 1, 2	NR ×4
		Vek 3	Uhm-vvv; NR ×3
BB	Tact test	Vek 1, 2	vek ×3; vek- round ×1
		Vek 3	vek × 3; vek- me
		All Zogs	correct tact responses
	Cat. trans. Test 2	All Zogs	zoggy
CM	Tact test	All Zogs / Veks	NR (except uhm ×1 for Zog 2 & Vek 2)
	Novel funct. train.	Zog 1	can fly (makes aircraft noises)
		Vek 1	that's not egg
	Tact training	Zog 2	ponytail ×5
FJ	Tact test	All	correct tact responses
MH	Tact test	All	correct tact responses
SH	Tact test	All	correct tact responses
SO	Tact test	All	correct tact responses
	Novel funct. train.	Zog 1	zog ×7
		Vek 1	vek ×4
	Cat. trans. Test 1	Zog 2	zog
MW	Tact test	Zog 1	NR ×4
		Zog 2	NR ×2; game; fruit
		Zog 3	cunch; cadyn; NR; game
		Vek 1, 2	NR ×4
		Vek 3	Cunch; cadyn; NR; farmer
AJ	Tact test	Zog 1	gee (waves); this (points to stimulus and waves); Christmas tree; bek
		Zog 2	bek; NR (but he claps); gog; bek
		Zog 3	NR; bek ×3
		Vek 1	gek; NR (but he claps); goes round (as he holds the stimulus); NR
		Vek 2	gog ×2; bek; NR
		Vek 3	gog; bek (and he claps); a little man; bek
PW	Tact test	All	correct tact responses
MD	Tact test	All	correct tact responses
	Novel funct. train.	Vek 1	vek
CD	Tact test	All	correct tact responses
FLJ	Novel funct. train.	Zog 1	Zoggy, the one that goes like that (waves) is like a windmill. That looks like rain doesn't it? (pointing to Vek 1)
	Cat. trans. Test 1	Pair 2	The vek goes like this (claps). The big one goes like this (waves).
	Tact test	All	correct tact responses
CG	Tact test	All Zogs	correct tact responses
		Vek 1, 2	correct tact responses
		Vek 3	vek ×3; zog
SO-Set 2	Tact test	All	correct tact responses
PW-Set 2	Tact test	All	correct tact responses
CG-Set 2	Tact test	All Zogs	correct tact responses
		Vek 4, 5	correct tact responses
		Vek 6	vek ×3; zog

have summarized the findings from all four studies in Table 3. Although the present study set out to train common listener relations, in the process of its doing so, 10 of the 14 children also learned to common tact the Set 1 stimuli, that is, they learned to name them appropriately as either ''zog'' or ''vek''. All 10 showed transfer of function. Taking these

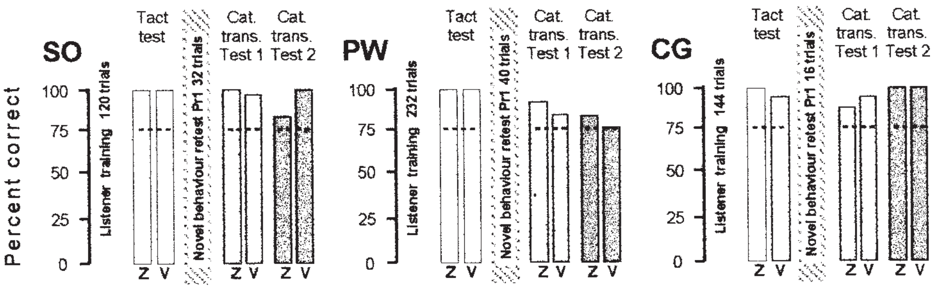

Fig. 4. For 3 subjects, training and test performances with the second stimulus set (Set 2). (Details as in Figure 2.)

results together with those of Lowe et al. (2005), of the 19 children who learned the common names during training of either common listener or common tact relations, all went on to pass tests of transfer of function, showing that two three-member stimulus classes had been established (see Table 3).

In addition, the 3 subjects in the present study given listener training with the Set 2 stimuli all learned the common names and succeeded on the category transfer tests, showing that function is easily transferred via naming, and that six-member stimulus classes had been established. This again corroborates the findings of Lowe et al. (2005) showing transfer of function within six- and nine-member classes. Naming clearly is highly effective in bringing about transfer of function in children from 1 to 4 years old.

Having learned to name the Set 1 stimuli and pass the category transfer tests, 3 children (FJ, MH, and FLJ) in the present study were next given the category match-to-sample test, which they also passed (either Test 1 or 2). Combining data from all four studies in the series, Table 3 shows that of the 20 children who learned to name the stimuli during training of either common listener or common tact relations, all 20 succeeded on match-to-sample testing. Common naming also readily established two six-member, and two nine-member, stimulus classes (Lowe et al., 2002; Lowe et al., 2005). Simply providing a common name for as many as nine arbitrary stimuli per class is clearly a remarkably effective means of ensuring children's success on category match-to-sample tests.

Across the four studies, a total of 32 subjects who named the stimuli were given either category transfer-of-function or category match-to-sample testing, or both. All 32 succeeded in categorizing. For young children aged from 1 to 4 years, these are the most extensive data and largest arbitrary stimulus classes ever achieved with a rigorous training and category-testing methodology. Such is the effectiveness of naming in establishing these classes that we have little doubt that the six- and nine-member stimulus classes shown in this series of studies could be readily extended and much larger stimulus classes demonstrated in children of this age. The precise scope and limits of name-driven categorization are matters for future research.

Common Listener Relations and Categorization

The key purpose of the present study was not to investigate whether children who have learned common naming show category transfer of function but rather, to determine whether transfer can be shown by those who do not have these naming repertoires. The procedure was successful in yielding 4 such children: RH, CM, MW, and AJ. All initially learned the common listener relations with the Set 1 stimuli but did not produce the corresponding common names in the tact test. All 4 children failed not only both category transfer-of-function tests but also the category match-to-sample tests. This is entirely consistent with the Horne et al. (2004) study in which 7 of the children who were given common listener training as here, did not produce common names. All 7 failed the category match-to-sample tests: the 1 subject (LN) in the study who did produce the common names and completed the categorization testing, succeeded. It should be noted that, although more children (i.e., 10 out of 14) produced common names following listener training in the present study compared to those in the Horne et al. study (i.e., 2 out

Table 3

Summary data of the four studies in the series indicating, for each study, the trained relations, number of subjects, and the initial outcome following training, that is, the number of subjects who named (namers) or learned only listener relations (listeners). Also shown are the number of children who passed the category match-to-sample test (Cat. MTS), the Category transfer-of-function tests (Cat. Trans.), and the overall number who categorized on one or more of the tests (Cat.), for the 3-, 6-, and 9-member classes, respectively.

				3-member classes						6-member classes			9-member classes		
							After name training								
Study	Trained relations	No. Ss	Initial Outcome	Cat. MTS	Cat. Trans	Cat.	Cat. MTS	Cat. Trans	Cat.	Cat. MTS	Cat. Trans	Cat.	Cat. MTS	Cat. Trans	Cat.
Lowe et al (2002)	Tact	12	12 namers	12/12		12				2/2		2			
Lowe et al (2005)	Tact	9	9 namers	4/4	9/9	9				3/3	3/3	3	3/3	3/3	3
Horne et al (2004)	Listener	9	1* namer	1/1		1									
			7 listeners	0/7		0	5/7		5						
Present Study	Listener	14	10 namers	3/3	10/10	10					3/3	3			
			4 listeners	0/4	0/4	0	2/2	3/3	3						
Totals		44	**32 namers**	20/20	19/19	**32**									
			11 listeners	0/11	0/4	0	7/9	3/3	**8**	5/5	6/6	**8**	3/3	3/3	**3**

* An additional namer (HW) did not complete category testing.

of 9), the ages of the children differed across the two studies. Although age-variations may account for the differences in their naming outcomes, the relatively small numbers of subjects involved in the two studies preclude the possibility of confirming this by statistical analysis.

Table 3 shows that across the series of studies, 11 children initially learned common listener relations but not common names, and all 11 failed to categorize. This is in marked contrast to the findings from those children who learned the common names; all 32 of them succeeded in categorizing. This is good evidence in support of the hypothesis that naming may be necessary for the establishment of arbitrary stimulus classes.

From Common Listener Relations to Common Naming and Categorization

Further support for this hypothesis comes from those subjects who initially learned only listener relations and failed to pass the categorization tests but then were trained to produce common names for the stimuli. Three such children (RH, CM, and MW) in the present study were given tact training immediately following their failure on the category tests and all 3 passed both subsequent transfer-of-function tests. Two of these subjects (RH and MW) were given category match-to-sample Test 1, which both passed, though RH only met the overall and not the individual mastery criterion. These results are very similar to those of the Horne et al. (2004) study in which the 7 children who had learned common listener relations, but not common naming, failed the category match-to-sample tests and were then trained with the common names. Five of these 7 subjects then proceeded to pass the tests and 2 (Subjects CT and TB) failed. The high success rate following name training (i.e., 8 out of 10 subjects across the two studies) is further indication of the key role of naming in categorization of arbitrary stimuli.

In summary, of the 43 children in these four studies who completed category testing (either transfer of function, or category match-to-sample, or both), 42 of them learned the common names either after baseline tact or listener training or, failing this, following tact training later in the procedure (Subject AJ in the present study did not undergo tact training). All but 2 of the 42 children who had learned common names categorized the arbitrary stimuli on at least one test. There was not a single instance of a child passing any categorization test but failing to produce the class names. This again is further evidence that naming may be necessary for categorization. We return to the matter of test failures and what they tell us about the determinants of categorization in the next section.

The Naming Account

Horne and Lowe (1996) have proposed that naming is a powerful means of transferring novel behaviors across arbitrary stimuli that do not have any common distinguishing feature. For example, a child, having learned to name different objects as ''hat'', may also learn the conventional behavior of putting one such object on her head. Given that as she does so, and continues to name the object, the auditory stimulus (/hat/) she produces is likely to precede and become discriminative for the newly learned putting-on-hat behavior which, in the same manner, may then extend to the other objects that the child has learned to name ''hat'' (and see Lowe et al., 2005). Similarly, the 10 children in the present study who learned common naming during baseline common listener training may have continued, overtly or covertly, to name the Pair 1 stimuli during the subsequent novel behavior training trials, with the result that saying ''zog'', for example, may have become discriminative for clapping, and saying ''vek'' for waving. Indeed, 3 of these children (SO, MD, and FLJ) named aloud one or both stimuli during these trials and 1 (Subject FLJ) not only overtly named the stimulus but also correctly described the relation between the stimulus and the novel function, saying ''Zoggy, the one that goes like that [waves]''. Likewise, during the novel behavior training trials in the Lowe et al. (2005) transfer study, 4 of the 9 children who initially were trained with common names for the stimuli also named aloud the zog and/or vek stimulus before they respectively clapped or waved. The children in these studies were not required to say anything during novel function training trials. They were simply asked ''Can you show me how this one goes?'' However, the fact that over a third of them overtly named the Pair 1 stimuli may be a good indication of the type of covert naming that may have been occurring in the other children who also had learned the common names but did not name the stimuli out loud.

Given that these names were discriminative for the occurrence of the novel functions and that the child had learned to name the Pair 2 and Pair 3 stimuli correctly, it follows that when these stimuli were presented in transfer Test 1 trials, they would evoke the common names, either overtly or covertly, which would in turn occasion their respective functions (clap/wave). During transfer Test 1 in the present study, 2 subjects (SO and FLJ) named aloud the stimuli as ''zog'' or ''vek''. Subject FLJ also correctly described the relation between stimulus and function before she performed it (see Table 2), as did one of the subjects (BH) in the Lowe et al. (2005) study.

The naming that occurs during the novel function training, and the first transfer test that follows, provides the conditions whereby the novel behavior itself may become discriminative for the corresponding name. As during novel function training, the child learns to wave to one Pair 1 stimulus and clap to the other, she may not only say ''zog'' before, for example, clapping, but also afterwards. When this sequence is followed by a reinforcer, then not only may the name ''zog'' become discriminative for clapping, but clapping may now become discriminative for saying ''zog'', and the listener stimulus so produced (i.e., /zog/) should evoke the listener behavior of looking at any stimulus to which the child has previously oriented in response to that spoken word. So when the experimenter asks ''Can you give me the one that goes like this?'' the child may respond, either overtly or covertly, ''zog'' (as did Subject BB overtly in the present study—and see also Subject JJ in Lowe et al., 2005)—and then respond as a listener by orienting to and selecting any other stimulus that he or she has learned to name ''zog''. It also is possible that some of the children may have named the clap and wave behaviors they learned in Pair 1 novel function training (e.g., see Subject BH, Lowe et al., 2005) while they continued to name the wooden shapes as ''zog'' or ''vek''. Intraverbal name relations between, for example, saying ''clap'' and saying ''zog'', and vice versa, and between saying ''wave'' and saying ''vek'', and vice versa, may just as readily have specified all the behaviors tested in transfer Tests 1 and 2. These and other forms of verbal control that can facilitate transfer of function are described in more detail in Lowe et al. (2005).

According to the naming account, success in the category match-to-sample task also flows directly from the bidirectionality inherent in a common name relation. When a child is trained to name a variety of objects ''zog'' then, at one and the same time, she simultaneously may learn the corresponding listener behavior with respect to those objects; that is, to orient to and select them when she hears the spoken word /zog/ (see Horne & Lowe, 1996, p. 201, Lowe et al., 2002). By the time the children reached the category sorting tests in the present study, they produced little overt verbal behavior. In the Lowe et al. (2005) transfer-of-function study, however, 2 of the 9 children (Subjects JJ and EW) named the sample before sorting the remaining stimuli, and 1 of them (Subject EW) did so on all 18 test trials. Similar unprompted sample naming during category sorting trials was observed in several subjects in the Lowe et al. (2002) and Horne et al. (2004) experiments.

Measures of categorization. A key issue for the present series of studies, and for the naming account, is how well they measure categorization and how the different measures do or do not covary. The results show that if subjects succeed on transfer-of-function Test 1, they also are highly likely to succeed on Test 2. The correlation was perfect for the 7 subjects who undertook both transfer tests in Lowe et al. (2005). In the present study only 1 (Subject MD) out of the 10 subjects who named, failed Test 1 outright but he passed Test 2; a different subject (SH) failed Test 2 but passed Test 1.

If subjects pass a transfer-of-function test, they also are highly likely to succeed on a category match-to-sample. For both the present study and Lowe et al. (2005) combined, there were 7 subjects who passed the transfer tests and then were given match-to-sample tests; all went on to pass one or other of the category match-to-sample tests.

The biggest discrepancy between the different categorization tests lies in performances on the category match-to-sample Tests 1 and 2. Over the four studies, 16 of 28 subjects who had learned to name the stimuli passed Test 1, in which they were asked only to look at the sample before they selected among the remaining stimuli; 10 of the 12 who failed Test 1 then succeeded when they were asked, in Test 2, to name the sample. The clearest comparison, however, comes from the two studies (Lowe et al., 2002; Lowe et al., 2005) in which the subjects were trained to produce common names for the stimuli at the start of the study (note that many subjects in the two listener studies, i.e. Horne et al. 2004 and the present one, had initially failed the tact and category tests before they were given tact training and repeat category tests). Of those 16 subjects then given the category match-to-sample tests, 10 succeeded on Test 1 and the remaining 6 on Test 2.

These failures to categorize on Test 1 are instructive about measures of categorization and arbitrary stimulus classes, and the theoretical approaches that underpin them. From the perspective of the naming account, such failures are entirely to be expected. For example, the different outcomes of the match-to-sample Tests 1 and 2 are precisely what one would predict from the naming account. When the subject is required to name the sample in Test 2, the naming helps to overcome competing sources of control over stimulus selections and ensures that selection is driven by the common name (and see Horne et al., 2004; Lowe et al., 2002).

The failure of the 2 subjects (CT and TB) in Horne et al. (2004) to pass either of the two match-to-sample tests even though they had been trained to produce common names for the stimuli, may also relate to problems in establishing verbal control. Before the experimental names were introduced, both these subjects had learned only listener behavior and had failed the category match-to-sample tests. There had been considerable opportunity for extraexperimental names and position preferences to become established (and there was direct evidence of both of these in the study) and to act as competing sources of stimulus control to the experimental names. Stimulus control, even by naming, is not a given but is a matter of experimental procedure interacting with human subjects' extraexperimental history over which researchers do not have control.

The differences between performances on the transfer Tests 1 and 2 may again relate to the experimental instructions. The instruction in Test 1 was the same as that used during novel function training (i.e., ''Can you show me how this one goes?'') whereas in Test 2 it was more novel (i.e., ''Can you give me the one that goes like this?''). How important these instructional differences are requires investigation.

It should be clear from the foregoing, however, that whether or not subjects show evidence of stimulus-class formation is not a straightforward issue. Young children come to the experiment with extensive verbal and nonverbal histories of behavior that may continue to determine how they respond to the experimental stimuli. The task for an experimental analysis is to deal with such interactions, particularly verbal control, in all their complexity.

When does listener behavior training give rise to naming? As we have observed previously (Horne & Lowe, 1996, p. 201; Lowe et al., 2002; Lowe et al., 2005), when young children are trained to tact they also learn at the same time the corresponding listener behavior, that is, they learn to name. As is clear from the data from the 7 subjects in Horne et al. (2004) and 4 in the present study (i.e., 11 out of 23), the converse is not true—children who learn listener relations do not invariably learn to tact. This is integral to the naming account (Horne & Lowe, 1996), according to which young children who learn listener behavior learn the corresponding speaker relation only if they echo, either overtly or covertly, the listener stimulus while at the same time looking at the object (referent). Why some children echo the listener stimulus in this way whereas others do not is difficult to determine. One possibility is that some subjects may begin the study with pre-existing individual names for the stimuli (see Dugdale & Lowe, 1990; Keller & Bucher, 1979). As listener training proceeds, their continued production of the pre-existing names may interfere with echoing the listener stimulus and learning the experimental tacts. Evidence of extraexperimental naming was observed in several of the subjects who failed to learn the experimental names during listener training in the Horne et al. (2004) and present study. Some children, however, produced no tact responses at all in the tact test—they (e.g., RH and CM in the present study and MJ in Horne et al.) simply said nothing when asked ''What's this?'' for each stimulus. Others (e.g., NW and RE in Horne et al., 2004) did produce the tacts ''zog'' or ''vek'', suggesting that they had at some point echoed the listener stimuli, but these vocal responses were often produced at random.

Does naming affect the learning of baseline relations? The 14 children in the present study learned the six listener relations in an average of 283.4 trials (range 120–648), but the average for the 10 children who named the stimuli was 225.6 trials compared to 428 trials for the 4 subjects who did not. The number of trials for those who named in this study is very similar to the 221 trials required for the 9 subjects who were trained to name the stimuli in the Lowe et al. (2005) study. It should be noted, however, that in the present experiment the average age of the 10 who named was 2 years 10 months (range 1 year 10 months to 4 years) compared to 2 years 4 months (range 1 year 10 months to 2 years 8 months) for the 4 who failed to name. But even when the age of subjects is taken into account, it seems to be the case that learning the baseline relations typically takes many fewer trials when the subjects produce common names than when they fail to do so. For example, of the 4 subjects in the present study who did not learn to name during listener training, 3 (Subjects RH, CM, and MW) had close age matches with 3 of the children (BB, FJ, and SO) who did name. Those who failed to name averaged 422 training trials to criterion compared to 136 for those who named. Comparing those who learned to name in the present study with those who were directly trained to do so in Lowe et al., there were 5 subjects (BB, FJ, SO, MD, and CD) here who were closely matched in age to 5 subjects (AF, LN, RC, CS, and CH) in Lowe et al. The 5 subjects in the present study required an average of 163 trials to reach criterion, which is little different from the average of 166 trials required by those in Lowe et al. Though this observation remains to be investigated systematically, it is consistent with the evidence from several other studies showing the facilitative effect of naming on learning (Birge, 1941; Eikeseth & Smith, 1992; Kendler, 1972; Pilgrim et al., 2000; K. Saunders, Saunders, Williams, & Spradlin, 1993).

Other Accounts of Categorizing

We have tried to show how the results of the present series of studies are consistent with the naming account and, indeed, were largely predicted by it. But it also is important to consider whether the findings would be predicted by other theories of categorization and whether they are consistent with those accounts. There are two key findings at issue:

(i) whereas all 32 subjects who initially learned common names succeeded on the categorization tests, the 11 subjects who failed to name also failed to categorize—when 10 of these latter subjects were trained to produce the common names, 8 of them categorized; and (ii) tact training gives rise to the corresponding listener relations but listener training does not invariably give rise to the corresponding tact.

Stimulus equivalence. The finding that only subjects who learned the common names passed the categorization tests is clearly not predicted by Sidman's (1994) theory of stimulus equivalence and, indeed, runs directly counter to it. According to that theory, naming is not necessary for categorization of arbitrary stimuli. It should not matter whether common listener or common speaker relations are trained—on subsequent testing in a match-to-sample (or transfer) procedure, the arbitrary stimuli should be partitioned into stimulus classes based upon either the common listener stimulus or the common speaker response (Sidman, 2000, p. 145; and see Horne et al., 2004). Moreover, the baseline procedures employed in these common tact and common listener studies entailed no positional differences in the stimuli across training and test trials of the kind that Sidman (1994, p. 527) has suggested may selectively disrupt transfer test performances following one-to-many match-to-sample discrimination training. It has been argued by some (e.g., R. Saunders et al., 1999) that a one-to-many match-to-sample procedure may fail to establish the predicted partition because the trained baselines may not establish all the prerequisite simultaneous and successive discriminations required for success in a subsequent match-to-sample test of stimulus equivalence. However, in the transfer-of-function categorization test employed here and in Lowe et al., (2005), these problems are avoided. The exact same stimulus pairings, in the same spatial arrangements, are given in training as in testing: all the prerequisite discriminations were shown to be in place. Regardless of whether the baseline training consists of pairwise listener training or pairwise tact training, the children could not progress to the pairwise transfer test unless they discriminated reliably between the two stimuli in each pair, in the absence of reinforcement. Nevertheless, there was a marked difference in category test outcomes related to whether just listener relations or naming had been learned: All subjects who categorized had also named, whereas all those who failed to name also failed to categorize. This is not a subtle difference or a matter of degree, but a major dichotomy. Along with other evidence such as the failures of nonverbal organisms to pass tests of stimulus equivalence (e.g., see Jitsumori et al., 2002; Lionello-DeNolf & Urcuioli, 2002; cf. Carr et al., 2000; Horne et al., 2004; Horne & Lowe, 1997; Kastak et al., 2001), this dichotomy poses major difficulties for equivalence theory.

The stimulus equivalence account also does not predict that listener training should give rise to an emergent corresponding tact. For example, if the child is trained to select the zog stimulus when the auditory stimulus /zog/ is presented, this should result in the following partition members:

The auditory stimulus /zog/	The sight of the zog	The zog selection response

Clearly, there is no element in the listener partition that could generate the corresponding tact response ''zog''. Nevertheless, 11 of the 23 children in the present study and Horne et al. (2004) combined, learned the common tacts during listener training. It is not at all clear how any element of the partition described above, or any combination of these elements, could give rise to production of the tact ''zog''. For an account that originally promised to explain not just categorization but also verbal behavior itself (Sidman, 1994, pp. 336–352; and see Horne & Lowe, 1996, pp. 233–237), this is yet another serious flaw (and see Horne & Lowe, 1996, pp. 227–230; 233–239).

Relational frame theory. Insofar as it assigns no special role to naming in the establishment of stimulus classes, relational frame theory cannot account for the outcome in the present study—the observed dichotomy in categorizing between children who did and did not have common names. Indeed, on the basis of relational frame theory thus far articulated, it is not clear that we should even have found such differences in naming repertoires between the children. According to the relational frame account, when children of the same age as those in the present studies are trained to respond as either speakers or listeners with respect to a particular object, then a bidirectional (or ''frame of coordination'') relation should be established between the object and the vocal response (Lipkens, Hayes, & Hayes, 1993). Children of this age have extensive histories of learning relations of the form Object A–Name A, where they are presented with an object and are trained to tact it, and Name A–Object A, where they are presented with the name and are trained to select the object. When a new Object X–Name X relation is trained, the Name X–Object X relation emerges or is derived. Similarly, when a new Name Y–Object Y relation is trained, the Object Y–Name Y relation emerges. Given that by the age of 2 years, children have, on average, a 300-word speaker repertoire (Woodward, 2000, p. 82), then if a frame of coordination is indeed established during such name learning interludes, emergent relations such as those described above should have been readily apparent in the verbal performances of the 2- to 4-year-old children who participated in the present series of studies (and see Lipkens et al., 1993). Because, according to RFT, arbitrary applicable relational responding is a defining feature of verbal behavior, the fact that all of the subjects regularly conversed with us in well-formed sentences is a further indication that arbitrary applicable relational responding, including the frame of coordination, should have been well established in these children. By this account, training Name A–Object A (listener) relations should just as readily have given rise to corresponding Object A–Name A (tact) relations as training tact relations should have given rise to the corresponding listener relations.

But the data from the present series of studies show otherwise. Whereas it is true that all children trained to tact also showed corresponding listener behavior, for about half the children the converse did not hold. Moreover, the children who failed were not predominantly younger than those who passed, which suggests that limited development of the echoic repertoire per se was not the cause for their failure in the tact test (cf. Lipkens et al., 1993). There is clearly a fundamental asymmetry in the emergence of tacting and listener behavior, respectively, which is not predicted by the theory and, indeed, is inconsistent with the frame of coordination account (and see Baldwin, 1991; Benedict, 1979; Horne & Lowe, 1996, p. 232).

Neither does a relational frame account of the categorization findings, in terms of control by contextual stimuli, explain the differences between the naming and nonnaming subjects. For example, given that the same instructions (contextual cues) were employed in Pair 1 training and the transfer tests, the children, according to relational frame theory, should have passed the transfer tests irrespective of whether they were given common listener or common speaker training beforehand. There were other contextual cues that should also have established a frame of coordination: All the children were given pretraining with toy hats and cups, using exemplars that differed from each other in terms of shape and color, and using exactly the same task instructions (cues) as were later employed with the arbitrary stimuli. It may be assumed therefore that, as a result of pretraining, all the children should have had a strong ''experimental set'' to produce the tested relations (e.g., given training of the listener behaviors to the hats, they were tested for the corresponding tact ''hat'' and for transfer of the behavior of placing the hat on the head, and later, of sorting the hats into one category).

One might argue, perhaps, that the pretraining with familiar stimuli and the parity between the training and test instructions may not have been sufficient to instantiate the appropriate frame of coordination and that pretraining with two arbitrary stimulus sets might have been more effective in this regard. Even this appears not to guarantee success, however, as is shown by the performance of Subject LN in Horne et al. (2004). Following the hats and cups pretraining, listener training with her first arbitrary stimulus set resulted in this child sorting the stimuli correctly and passing the tact test. When she was given listener training on a second arbitrary stimulus set, however, using exactly the same listener stimuli (/zog/, /vek/) as were employed for Set 1, she failed the category match-to-sample test and the tact test, even following a second round of listener training, and repeated category and tact testing; indeed, she only passed the category test after she was trained to name the stimuli. This child not only had a prior history of responding correctly in the presence of the relevant contextual cues with the familiar hats and cups, but she also had a history, following Set 1 listener training, of producing all the emergent behaviors (tacting and sorting) with one arbitrary stimulus set before she failed to tact and sort the second set of arbitrary stimuli. These and other failures of the subjects who learned only listener behavior to pass either the transfer or the match-to-sample tests cannot be easily explained by relational frame theory.

Acquired equivalence. Whereas nonhumans routinely fail standard tests of stimulus equivalence, there is evidence of novel behavior transfer in pigeons from studies that have been designed to test a ''common coding'' or respondent (see Meehan, 1999) account of acquired equivalence. According to the common coding account, covert responses may mediate the transfer of a novel behavior from one arbitrary stimulus to another. Specifically, if the organism is trained to select C1 in the presence of A1, and again C1 in the presence of B1, a covert response ''c1'' is said to mediate selection of the C1 comparison given A1 or B1 as sample. Similarly, a covert response ''c2'' may mediate selection of C2, given A2 or B2 as sample. When a new behavior, selecting D1, is now trained to the sample A1, the latter continues to evoke the covert ''c1'' response that was established in the baseline training so that ''c1'' now also becomes discriminative for selection of D1. Similarly, when selecting D2 is next trained to the sample A2, a covert ''c2'' mediates the selection of D2. The common coding account of acquired equivalence predicts that the new behaviors, selection of D1 and D2, will transfer to samples B1 and B2, respectively, without training, by means of the corresponding covert ''c1'' and ''c2'' mediating responses. Further, it predicts that transfer will occur only when a many-to-one training paradigm, such as that just described, is employed to train the baseline discriminations. This is because the alternative one-to-many paradigm (e.g., if C1 select A1 and if C1 select B1) should not generate the common covert mediating responses required to produce transfer of the novel behaviors from one arbitrary sample stimulus to another. When a many-to-one baseline training structure has been employed, followed by reassignment training to one sample and testing for transfer of the novel relation to the other sample in the prospective functional class, pigeons usually showed the predicted pattern of transfer (Urcuioli, 1996; Wasserman & DeVolder, 1993; Zentall, 1996; but see Bhatt & Wasserman, 1989). In contrast, pigeons do not show novel behavior transfer in one-to-many procedures (Urcuioli, 1996; Zentall, 1996). In general, therefore, the pigeon data on transfer of function appear to be consistent with a common coding, or respondent, mediation account.

Direct comparison of such studies with those of human performance, however, is complicated by important procedural differences. The studies with nonhumans test only one direction of behavior transfer—the one consonant with the trained baseline relations—and not the opposite or symmetric counterpart that is a defining characteristic of arbitrary stimulus classes (Hayes & Hayes, 1992; Horne & Lowe, 1996; Sidman, 1994). Studies of acquired equivalence also have used reinforced test trials and often employ savings in learning as the main dependent measure of transfer, whereas for human participants, novel behavior transfer usually is measured directly, under extinction test conditions (e.g., Smeets et al., 1997). Although the small, averaged transfer effects in the pigeon studies are statistically significant, intersubject variability is nevertheless high, leading some to question whether the same process underlies pigeon performances in the transfer task as for humans (K. Saunders et al., 1996, pp. 102–104). It is interesting, nevertheless, to consider how the acquired equivalence account relates to transfer of function in humans, particularly in very young children. For example, the common tact training employed in Lowe et al. (2002) and in Lowe et al. (2005) may be viewed as a many-to-one paradigm, and the common listener training employed in Horne et al. (2004) and the present study as a one-to-many paradigm. This might suggest that the different outcomes between the speaker and listener studies could be explained by an acquired equivalence account. Close inspection of all the emergent outcomes, however, shows otherwise. Because the acquired equivalence account is essentially based on unilinear relations, it cannot explain the emergence of untrained relations that run in the opposite direction to those trained. So, for example, it cannot explain how speaker (many-to-one) training (e.g., given either A1 or B1 say ''zog'') can result in the untrained emergence of the corresponding listener behavior (i.e., if /zog/ select A1 or B1). And vice versa: It cannot explain how common listener training may give rise to the corresponding untrained common speaker behavior.

There also are major difficulties in accounting for the categorizing outcomes. Given common tact training (many-to-one), it cannot explain the selection of a B1 comparison given A1 as sample (and vice versa) in a category match-to-sample test given that the response ''zog'' (whether overt or covert) has never been established as a discriminative stimulus for the selection of B1 given A1, or of A1 given B1. Whereas the acquired equivalence account might appear to explain a positive outcome on transfer Test 1 following tact training (many-to-one)—insofar as a covert ''zog'' response may mediate production of the appropriate clap/wave response to the Pair 2 and Pair 3 test stimuli—it cannot explain success on transfer Test 2 in which, no matter how many times the child says ''zog'' (covertly or overtly), this will not enable him or her to respond to the sight of the experimenter's clap/wave by selecting the correct stimulus (i.e., zog or vek). Once again, the target response runs in the direction opposite to that trained and so it cannot occur via a covert mediating response such as is postulated by the theory. The acquired equivalence account cannot, therefore, explain most of the significant emergent outcomes in the four studies.

Conclusion

The findings of the present study, together with those of the others in the series, have major implications for theoretical and practical approaches to categorization. The key theoretical question that concerns many is whether language is necessary for the establishment of classes of arbitrary stimuli. It should be noted, however, that no experiment or even series of experiments can prove conclusively that language is necessary, as this has all the scientific problems associated with trying to prove the negative, such as, for example, the proposition that there are no big green martians in the universe or, as here, that there are no animal species, or individuals, without language from this or any other planet who can categorize arbitrary stimuli. What science can do, however, is *disprove* the proposition. It is a strength of the naming account that all that is required to show that naming is not necessary for categorization of arbitrary stimuli is to produce clear examples of organisms that cannot name the stimuli, but nevertheless categorize. This was the main theoretical thrust of the present series of studies, which succeeded in yielding subjects in considerable numbers who either did or did not have common names for the experimental stimuli. If there had been little difference in categorization outcomes between these two groups of children then this would have shown that naming is not necessary for categorization success. However, the fact was that, in experiment after experiment, not only were there differences between these groups, but there was a complete dichotomy—the only children who categorized the stimuli were those who learned the common names; if children did not name then they did not categorize. Far from disproving the naming hypothesis, this series of studies produces the strongest evidence yet that naming is indeed necessary. This conclusion is further supported by evidence from a range of studies that have failed to observe arbitrary stimulus classes in nonverbal organisms (K. Saunders et al., 1996; Urcuioli, 1996; Zentall, 1996; and see Horne & Lowe, 1996, 1997).

It is important to note, however, that whatever conclusions are derived about the necessity of naming, the present series of studies shows that at the very least, it has profound facilitative effects on categorization. This is again consistent with findings from studies with other groups of subjects that show how verbal interventions enhance learning (Pilgrim et al., 2000). The present results do indicate, however, that if we wish to teach normally developing young children to categorize their environment, it would be better to avoid training procedures, such as those used in the present study and Horne et al. (2004), that do not promote common naming. This applies particularly to nonverbal match-to-sample procedures, which frequently have been shown to be ineffective even with human adult subjects (Innis et al., 1998; R. Saunders et al., 1999). These data tell us instead that what we need to do is to train common names for the stimuli that are to be categorized. There are of course complexities underlying this simple proposition, some of which have been encountered in both the present study and others in the series. It is these complexities that should be the focus of future research.

It may be fruitful, of course, for some researchers to continue the quest for the Holy Grail of nonverbally governed categorization of arbitrary stimuli. If Sidman's (1994) theory is correct and equivalence is a given, then yet further studies with different animal species might yield results more favorable to the theory. Alternatively, it may be the case that equivalence is a given only for the human species, so that research might be most productively focused there. Some attempts along these lines have been made with mentally handicapped subjects but because the histories of these individuals, mainly adults or teenagers, have been so complex and their verbal repertoires so uncertain, they have not provided a satisfactory basis on which to resolve key theoretical issues (see Carr et al., 2000; Horne et al., 2004; cf. O'Donnell & Saunders, 2003). By far the most promising test might be provided by studies of preverbal subjects, that is, children even younger than those in the present studies. If such infants could learn the baseline relations, without a naming intervention, and pass the categorization tests, then the naming hypothesis would be disproved. The naming account, with support from the results from the present studies, indicates that this would not happen, but it should be tested.

An alternative route for future research may be to concentrate directly on the extraordinary behavioral repertoires that we term language or verbal behavior and the issue of how it transforms human learning. Skinner (1957, 1969) clearly recognized the importance of this behavior and had begun to establish a theoretical framework for language research. The naming account (Horne & Lowe, 1996, 1997; Lowe & Horne, 1996), in conjunction with the present series of experimental tests, is an attempt to advance this theory and to establish a coherent empirical base for its further development.

REFERENCES

Auguston, 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.

Barnes, D., Browne, M., Smeets, P., & Roche, B. (1995). A transfer of functions and a conditional transfer of functions through equivalence relations in three- to six-year-old children. *The Psychological Record, 45,* 405–430.

Barnes, D., & Keenan, M. (1993). A transfer of functions through derived arbitrary and non-arbitrary stimulus relations. *Journal of the Experimental Analysis of Behavior, 59,* 61–81.

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

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language, 6,* 183–200.

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.

Bhatt, R. S., & Wasserman, E. A. (1989). Secondary generalization and categorization in pigeons. *Journal of the Experimental Analysis of Behavior, 52,* 213–225.

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

Bones, R., Keenan, M., Askin, D., Adams, L., Taylor, D., & Nicholas, O. (2001). The effects of response topography on functional equivalence class formation. *The Psychological Record, 51,* 89–110.

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.

Dougher, M. J., Auguston, M. R., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior, 62,* 331–352.

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

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.

Fields, L., Reeve, K. F., Rosen, D., Varelas, A., Adams, B. J., Belanich, J., & Hobbie, S. A. (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., 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.

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

Greenway, D. E., Dougher, M. J., & Markham, M. R. (1995). S+/S− reversal procedures may not result in functional equivalence. *Experimental Analysis of Human Behavior Bulletin*, *13*, 16–18.

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

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, cognition, 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.

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

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.

Jitsumori, M., Siemann, M., Lehr, M., & Delius, J. D. (2002). A new approach to the formation of equivalence classes in pigeons. *Journal of the Experimental Analysis of Behavior*, *78*, 397–408.

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.

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

Kohlenberg, B. S., Hayes, S. C., & Hayes, L. J. (1991). The transfer of contextual control over equivalence classes through equivalence classes: A possible model of social stereotyping. *Journal of the Experimental Analysis of Behavior*, *56*, 505–518.

Lazar, R., & Kotlarchyk, B. J. (1986). Second-order control of sequence-class equivalences in children. *Behavioral Processes*, *13*, 205–215.

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.

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.

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

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

Meehan, E. F. (1999). Class-consistent differential reinforcement and stimulus class formation in pigeons. *Journal of the Experimental Analysis of Behavior*, *72*, 97–115.

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., 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.

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). New York: 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.

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

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

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., 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.

Sidman, M., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior, 52,* 261–274.

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.

Smeets, P. M., Barnes, D., & Roche, B. (1997). Functional equivalence in children: Derived stimulus-response and stimulus-stimulus relations. *Journal of Experimental Child Psychology, 66,* 1–17.

Steele, D. L., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior, 56,* 519–555.

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

Wasserman, E. A., & DeVolder, C. L. (1993). Similarity- and non-similarity based conceptualization in children and pigeons. *Psychological Record, 43,* 779–794.

Woodward, A. L. (2000). Constraining the problem space in early word learning. In R. M. Golinkoff et al. (Eds.), *Becoming a word learner* (pp. 81–114). New York: Oxford University Press.

Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50,* 125–141.

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: November 16, 2004
Final acceptance: October 6, 2005