JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2000, **74,** 101–114 NUMBER 1 (JULY)

EQUIVALENCE CLASSES IN INDIVIDUALS WITH MINIMAL VERBAL REPERTOIRES

D. CARR, K. M. WILKINSON,
D. BLACKMAN, AND W. J. MCILVANE

WELSH CENTRE FOR LEARNING DISABILITIES,
E. K. SHRIVER CENTER AND NORTHEASTERN UNIVERSITY,
AND UNIVERSITY OF WALES CARDIFF

Studies from two different laboratories tested for equivalence classes in individuals with severe mental retardation and minimal verbal repertoires. In the first study, 3 individuals learned several matching-to-sample performances: matching picture comparison stimuli to dictated-word sample stimuli (AB), matching those same pictures to printed letter samples (CB), and also matching the pictures to nonrepresentative forms (DB). On subsequent tests, all individuals immediately displayed Emergent Relations AC, AD, BC, BD, CD, and DC, together constituting a positive demonstration of equivalence (as defined by Sidman). The second study obtained a positive equivalence test outcome in 1 of 2 individuals with similarly minimal verbal repertoires. Taken together, these studies call into question previous assertions that equivalence classes are demonstrable only in individuals with well-developed language repertoires.

Key words: conditional discrimination, stimulus equivalence, verbal behavior, touchscreen press, humans

---

Among the most controversial issues in contemporary behavior analysis is the role of the verbal repertoire in the formation of equivalence classes. Horne and Lowe (1996) have argued, for example, that equivalence classes depend upon a process termed *naming,* said to emerge as one acquires language; we shall term this the *naming hypothesis.* Hayes (1991) has also theorized that equivalence classes may be related to language learning, a broader *language hypothesis.* In his account, such learning experiences instantiate the contingencies necessary to generate arbitrarily applicable relational responding, one type of which is shown by positive outcomes on equivalence tests. By contrast, Sidman (1994) has suggested that stimulus equivalence may be a fundamental behavioral process that is generated by the reinforcement contingency and not reducible to other processes. In Sidman's view, language learning may depend in part on the emergence of equivalence relations.

Horne and Lowe (1996) supported their arguments with data that showed that negative equivalence class outcomes in young children were replaced by positive outcomes when children were taught to name the stimuli (e.g., Dugdale & Lowe, 1990). Hayes' group has reported positive equivalence outcomes in young children who have demonstrable language and negative outcomes in children who seemingly do not (Devany, Hayes, & Nelson, 1986). The difficulty in demonstrating equivalence classes in nonhumans can also be interpreted as providing support for both the Horne and Lowe and the Hayes (1991) positions.

In their recent commentary on the Horne and Lowe (1996) article, however, McIlvane and Dube (1996) questioned existing interpretations of the negative equivalence outcomes in nonverbal participants and the facilitative effects of teaching naming (see also Dube & McIlvane, 1996). Concerning the negative outcomes, they suggested that researchers more fully consider the history of stimulus control research with nonhumans and young children. A much-replicated finding is that there is often a mismatch between the stimuli that the experimenter wants to control behavior and the stimuli that actually

Study 1 was supported by a grant from the National Institutes of Child Health and Human Development (HD25995). Study 2 was part of a PhD dissertation submitted by D. Carr to the School of Psychology, University of Wales College, Cardiff. First and second authorship is equal. We thank Bill Dube for comments on an earlier version of this manuscript.

Reprint requests may be sent to W. J. McIlvane, E. K. Shriver Center, Psychological Sciences Division, 200 Trapelo Road, Waltham, Massachusetts 02452 (E-mail: wmcilvane@shriver.org) or D. Carr, Welsh Centre for Learning Disabilities ARU, Meridian Court, North Rd., Cardiff CF4 3BL, Wales (E-mail: debcarr@writeme.com).

gain control of behavior (see Lionello & Urcuioli, 1998, for a recent example). McIlvane, Serna, Dube, and Stromer (in press) subsequently used the term *stimulus control topography coherence* to describe the situation in which the experimenter-specified stimuli do gain control, and they argued that coherence was necessary for a truly valid test of the equivalence potential of any research participant.

McIlvane and Dube (1996) went on to argue that teaching research participants to name the stimuli may encourage coherence of participant- and experimenter-defined topographies, for example, by encouraging prerequisite sample/positive-stimulus (S+) controlling relations (Sidman, 1987) or verifying the successive and simultaneous discriminations that are prerequisite to a successful equivalence test outcome (cf. Saunders & Spradlin, 1990). Notably, Schusterman and Kastak's (1993) demonstration of stimulus equivalence in a sea lion was accomplished with a training regimen that featured a large number of different stimuli and stimulus relations, including reversing sample and comparison functions during baseline training. Providing multiple exemplars is a well-established technique for encouraging learners to attend to experimenter- or teacher-defined aspects of the task being learned (e.g., Engelmann & Carnine, 1982).

What pattern of data would be needed to support the position that naming skills per se may not provide the critical behavioral prerequisites for positive equivalence outcomes? Schusterman and Kastak's (1993) demonstration has had relatively little impact on those who advocate a central role for naming or some other aspect of the language repertoire in stimulus equivalence. Horne and Lowe (1996), for example, questioned whether the positive outcomes with the sea lion came about via the same behavioral histories and processes as those responsible for equivalence in humans. Indeed, Horne and Lowe suggested that equivalence in humans may come about via a complex set of interconnected behavioral repertoires involving words heard, words spoken (overtly or covertly), and a variety of other stimulus functions associated with the environmental events to which words refer. In light of their proposal, it is of great interest to test for stimulus equivalence in humans who have never developed the capacity for speech. Such a deficit effectively eliminates the route by which the naming process operates, and equivalence test outcomes should be negative.

The studies reported here reflect the efforts of two independent laboratories to assess the merits of naming theory. In both studies, the intent was specifically to test Horne and Lowe's (1996) assertion that children who had poorly developed language and communication skills would not readily show stimulus equivalence. In their words, "one of the most fruitful ways to test [the naming hypothesis] is with young children who have not yet learned to name or in whom naming skills are not yet well established" (p. 224). In the studies reported here, research participants were older individuals with significant developmental disabilities who displayed virtually no functional spoken language. Their repertoires as speakers were meager, typically consisting principally of manual signs (see below). Their repertoires as listeners were comparably meager. Of greatest significance, the participants appeared to lack the critical prerequisites for equivalence classes that were articulated by Horne and Lowe. Thus, they provided an opportunity for a reasonably strong test of the naming hypothesis: These participants should not display positive equivalence class test outcomes.

STUDY 1

This study was conducted at the Shriver Center, which conducts extensive research on discrimination learning in children and adults with developmental limitations. In past research, the Shriver laboratories occasionally tried to demonstrate equivalence classes in individuals who lacked any speaking and listening repertoires (briefly described by McIlvane & Dube, 1996). Those studies failed because the participants could not be taught the requisite conditional discrimination baselines despite extensive training; no equivalence tests could be conducted. With individuals with better developed repertoires, however, success on equivalence class tests has been routine (e.g., Dube, McIlvane, Mackay, & Stoddard, 1987; Dube, McIlvane, Maguire, Mackay, & Stoddard, 1989). Prior to the present study, however, there had been only very limited equivalence work with individuals who might be characterized as "minimally verbal." The research of this type that was available had other aims and used nonstandard procedures, leading to results that were somewhat difficult to interpret (e.g., McIlvane & Stoddard, 1985). The present study initiated conventional stimulus equivalence work with individuals from this population.

Method

Participants

Participants were 3 individuals with severe mental retardation. DJB and IVB were male, and JRV was female. They were 19, 21, and 15 years of age, respectively, at the time of their participation. They achieved age-equivalent scores of 2 years 3 months, 2 years 1 month, and 2 years 3 months, respectively, on the Peabody Picture Vocabulary Test—Revised (PPVT-R; Dunn & Dunn, 1981). The PPVT-R surveys the repertoire of the listener for extant relations between dictated words and representative line drawings of familiar objects or events. As the PPVT-R score indicates, the participants entered the study able to match a small number of pictures to dictated names. None of the 3 individuals achieved a basal score, however, and their scores may overestimate their listening repertoires relative to the sample on which the test was normed.

It is important to note that none of the participants had significant oral naming skills. This was assessed by administration of the Gardner Expressive One-Word Picture Vocabulary Test—Revised (EOWPVT-R; Gardner, 1990), in which single line drawings are presented and participants are asked to name each picture. JRV and IVB were untestable with this instrument (scoring <2 years); DJB had enough vocalizations to receive an age-equivalent score of 2 years 1 month (with no basal score).

To give them some medium for expressing wants and needs, DJB and JRV had been taught a limited repertoire of manual signs. DJB and IVB had also been provided with a picture-based augmentative communication board. These signs and pictures most often functioned as mands (i.e., to request reinforcers). DJB was retested on the EOWPVT-R using signs rather than spoken words as his expressive medium. His age-equivalent score increased slightly to 2 years 5 months.

All participants were students at behaviorally oriented schools for persons with intellectual disabilities, and they had received extensive special education throughout their lives. Although the source of their meager communication repertoires was not documented, it is reasonable to conclude that extant skills were the products of direct teaching. None of the participants had any reading skills.

Apparatus

Sessions consisting of 42 to 60 discrimination trials were conducted two to four times per week in a quiet room at the participant's school. Experimental stimuli were presented by a Macintosh Plus® computer adapted for the research (Dube, 1991). Its touch-sensitive screen (19 cm by 14 cm) displayed black-and-white stimuli on a white background. Participants responded by touching the stimuli. Reinforcers were tokens, which were exchanged after each session for food items, drinks, or preferred activities.

Figure 1 shows the matching-to-sample stimuli used in this study. Stimuli used in preliminary training included dictated English words (*dog, cake,* and *shirt*) (Set A), line drawings of corresponding pictures (Set B), and printed letters of the alphabet (D, C, and S). Subsequent tests for equivalence introduced nonrepresentative forms (Set D), which were recognizable as "picture-like" (Wilkinson & McIlvane, 1994), but they did not appear to the experimenters to be similar in form to any of the other visual stimuli.

Procedure

Preexperimental assessment and training. Participants began their participation with a standard battery of assessments that evaluated extant discrimination skills (see Dube, Iennaco, & McIlvane, 1993, for a description). This quasi-programmed assessment provided training on basic discrimination skills, including various matching-to-sample performances. Following this training, all participants were able to match at least three pictures to dictated English names. They had mastered generalized identity matching of those pictures and also of nonrepresentative forms.

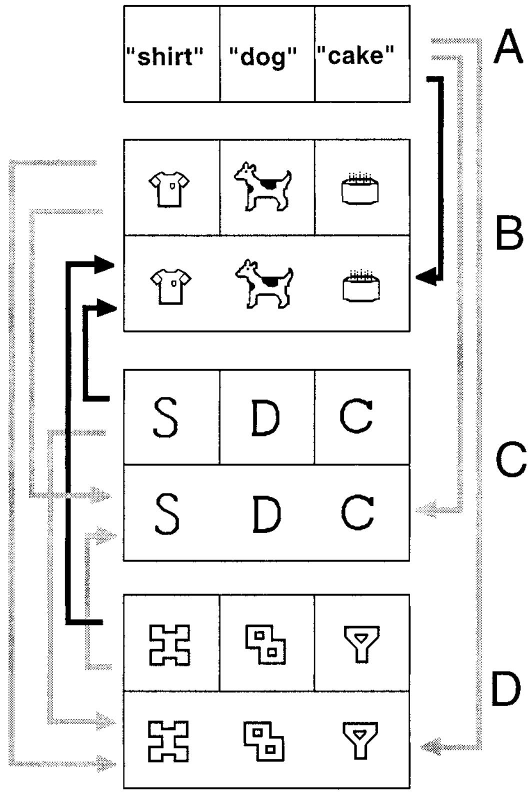

Fig. 1. Stimulus relations trained and tested in Experiment 1. Arrows point from sample stimuli to comparison stimuli. Black arrows indicate baseline (trained) relations, and gray arrows indicate potentially emergent relations.

Prior to participating in the present study, DJB and JRV had served in a study of controlling relations in conditional discrimination. That study did not address equivalence classes and has been reported separately (Serna, Wilkinson, & McIlvane, 1998). During training for that study, both participants mastered a three-choice arbitrary matching-to-sample task in which letters of the alphabet served as samples and line drawings served as comparisons. Sample stimuli appeared in the center of the computer screen. Following a touch to the sample, comparison stimuli were displayed in three of the four corners of the screen. Correct sample–comparison relations were defined as D–dog, C–cake, and S–shirt. Such initial consonant–picture relations were unknown and essentially arbitrary for these nonreading individuals. (These performances were selected because they might have some later educational significance for the participants.) Although IVB did not serve in the prior study, his baseline training was similar in all major respects to that received by the others.

The auditory-visual (AB) matching relations were entry skills. The visual-visual baseline (CB) was taught via by the sample stimulus-control shaping method reported by Zygmont, Lazar, Dube, and McIlvane (1992). Briefly, this procedure began with identity matching with the Set B stimuli and then transformed the samples into the Set C stimuli by gradual changes over a series of trials. Potentially relevant to the present study was the subsequent use of the ''blank comparison'' procedure, which explicitly verified both sample-S+ and sample/negative-stimulus (S−) control in the baseline arbitrary matching relations. In this procedure, participants view one or more comparison forms and a black square. If the sample and one of the comparison forms match, the participant selects the form; if not, he or she selects the blank (see Serna et al., 1998, for a description of the earlier study and also McIlvane et al., 1987, for an extended discussion of controlling relations and their measurement).

Initial tests for emergent relations. The baseline for the present work was standard three-comparison matching to sample (i.e., no blank comparison trials were included in this phase of testing). Intertrial intervals were 1.5 s. The relevant skills were matching (a) dictated word samples (A) with picture comparisons (B) (AB matching) and (b) printed letter samples (C) with the same picture comparisons (CB matching). Auditory-visual and visual-visual performances were presented in separate trial blocks. Correct selections were followed by reinforcers on 75% of baseline trials. Initial tests evaluated whether the participants might prove to be capable of BC matching (a symmetry test) and AC matching (a transitivity test, provided the BC matching was exhibited). Because the A stimuli were auditory, BA and CA performances could not be tested with conventional matching-to-sample procedures. Thus, the baseline at this point permitted tests of emergent relations but not a comprehensive test for Sidman equivalence (i.e., as defined by Sidman & Tailby, 1982).

Every test session evaluated one potentially emergent relation (either BC or AC); the symmetry test was conducted first. Six probe trials were interspersed among baseline trials, and each relation was evaluated in two different sessions (a total of 12 probe trials for each relation). The baseline reinforcement schedule remained at 75%. Unlike many equivalence studies, the tests for emergence were not conducted in extinction. Rather, we used a modified version of the method reported by McIlvane and Stoddard (1985), which was designed specifically for testing low-functioning participants. No reinforcers followed selections on the first two probe trials of each type. However, reinforcers did follow two of the four subsequent probe-trial selections.

Stimulus equivalence: Training and tests. To develop a baseline that would permit Sidman equivalence tests, all participants were taught via sample stimulus-control shaping procedures to match nonrepresentative form samples (Set D, Figure 1) with picture comparisons (DB matching). Typical shaping sessions included 48 to 60 trials. This training established the basis for the following derived relations: BD (symmetry), AD (transitivity), CD (equivalence), and DC (equivalence); tests were conducted in this order. Positive outcomes on these tests for emergent relations would constitute a fully adequate demonstration of Sidman equivalence. Probe procedures were identical to those used during the initial tests for emergent relations.

Results and Discussion

Acquisition of Baseline Performances

To establish CB and DB matching, respectively, required the following number of shaping sessions: DJB: 80 and 5; IVB: 16 and 8; JRV: 4 and 4. Mastery criterion was one full session at 95% accuracy, and all participants maintained that criterion when baseline trial types were intermixed and, with one exception, throughout subsequent testing.

Tests for Emergent Relations

Twelve-trial tests for emergent AC, BC, AD, BD, CD, and DC relations were presented, for a total of 72 probe trials. DJB's performance was consistent with equivalence relations on 71 trials, and JRV's performance was consistent on all 72 trials. Thus, both participants demonstrated highly positive equivalence test outcomes, even though the first two probe trials for each relation were unreinforced. IVB's performance was consistent with equivalence on 68 of 72 trials (94% consistent). Three of the four inconsistencies occurred on BC matching probes during a single session in which his baseline performance also fell (35 of 42 baseline trials correct); notably, the initial BC probe responses were consistent and the deterioration occurred later in the session. Upon recovery of baseline performance (41 of 42 correct) his BC probe performance recovered as well (five of six correct). In all other sessions, the participants maintained criterion baseline accuracy.

It is difficult to reconcile these data with Horne and Lowe's (1996) naming hypothesis, given that none of the participants had appreciable oral naming skills. Further discussion of these data and their implications will be deferred until Study 2 is presented.

STUDY 2

This study was conducted at the School of Psychology, University of Wales College in Cardiff, Wales. Two male adolescents participated. Both had diagnoses of severe mental retardation and autism. One had a verbal repertoire similar to the participants of the preceding study. The other had somewhat more advanced oral skills, including a highly developed echoic repertoire. Study of such an individual might prove to be informative, given the prominent role of echoics in Horne and Lowe's (1996) formulation. The study was a systematic replication of Sidman (1971), except that all experimental stimuli were visual.

Method

Participants

BN was 13 years old at the time of the study. He had no speaking repertoire, but could imitate certain animal sounds. His most recent language assessment placed him at 2 years 0 months on Derbyshire Language Scheme assessment (Knowles & Masidlover, 1982). Because he did not speak, his language therapy focused on manual signs (Makaton system; Grove & Walker, 1990). In language training sessions, he had learned a large number of signs, but rarely used them outside the structured therapy sessions. Like many children with autism, BN's listening skills were severely limited. For example, he

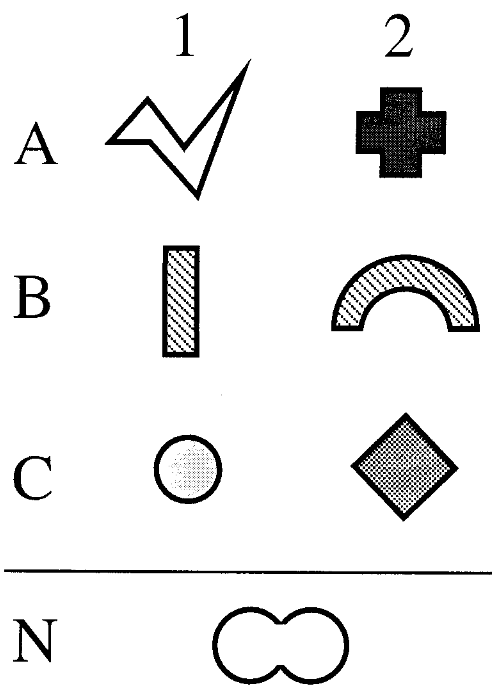

Fig. 2. Form stimuli used in Study 2. Different colors are suggested by different shades of gray.

did not respond to spoken requests unless the supplemental visual cues were given.

HF was 14 years old at the time of the study. His speaking repertoire was assessed with the Reynell Developmental Language Scales (Reynell & Huntley, 1985), and he obtained an age-equivalent score of 3 years 1 month. Although HF could speak, his speech consisted mainly of echoics, either immediate or delayed. Nonechoic speech was often nonsensical (e.g., "Put batteries in the window"), although he could sometimes complete sentences correctly if he had heard them before ("The school bus is . . . [white]"). His functional speech was comprised mainly of a few stock phrases (mands) that communicated basic wants and needs.

Apparatus

Participant and experimenter were seated opposite each other at a desk in a small, quiet room. The desk was empty, except for the experimental materials. Matching-to-sample stimuli were six physically dissimilar forms drawn on cards, differing in both form and color, as shown in Figure 2. These stimuli were used to constitute two sets of three stimuli each (A1B1C1 and A2B2C2) that might become two three-member equivalence classes.

Procedure

Sessions were conducted three to four times per week during school hours. Typical baseline sessions consisted of two 10-trial blocks presented in immediate succession. However, only one block was given if the children were inattentive or uncooperative on a given day. All test sessions consisted of 24 trials.

Participant BN. For BN, every training session began with a simultaneously spoken and signed instruction "Work first, then TV," an often-used cue that indicated the availability of television after a therapy session. Training and testing proceeded through the nine phases described in Table 1. On every trial, the sample stimulus was placed directly in front of BN, and two comparisons were placed adjacent to each other at the opposite edge of the desk. The left-right position of the S+ comparison varied unsystematically from trial to trial.

AB matching was established during Phases 1 through 3. On the early trials of Phase 1, the experimenter placed BN's index finger on the sample card (A1) while saying "That one and . . .," and then guided his finger to the positive comparison (B1), saying "that one." Manual guidance was used for several trials until BN independently touched the sample with only the experimenter's verbal prompt. If BN's subsequent comparison selection was correct, the experimenter said "that one!" and simultaneously delivered a single piece of candy. If BN's selection was incorrect, however, the experimenter said "Uh oh!" and no candy was delivered. Criterion for completing this phase and all subsequent training phases was 90% correct for a block of 10 trials. Procedures during Phase 2 were virtually identical to Phase 1. Conditional discrimination was required in Phase 3, during which the trial types from Phases 1 and 2 were intermixed. During this phase also, the reinforcement schedule was changed from continuous reinforcement to variable-ratio (VR) 3.

The first test for derived relations was presented in Phase 4, a BA symmetry test. Symmetry test trials were interspersed among AB

Table 1

Sequence of training and testing for Participant BN.

Phase	Sample	Comparisons	Trials per block
1. Teach A1:B1	A1	B1, B2	10
2. Teach A2:B2	A2	B2, B1	10
3. Teach Mixed AB baseline	A1	B1, B2	10
	A2	B2, B1	
4. Test BA symmetry (2 blocks)	B1	A1, A2	10 symmetry probes
	B2	A2, A1	+ 20 baseline trials
5. Teach A1:C1	A1	C1, C2	10
6. Teach A2:C2	A2	C2, C1	10
7. Teach mixed AC baseline	A1	C1, C2	10
	A2	C2, C1	
8. Test CA symmetry (2 blocks)	C1	A1, A2	10 symmetry probes
	C2	A2, A1	+ 20 baseline trials
9. Test BC and CB, combined test for symmetry	B1	C1, C2	
and transitivity (4 blocks)	B2	C2, C1	8 equivalence probes
	C1	B1, B2	+ 16 baseline trials
	C2	B2, B1	

baseline trials, and no reinforcers followed any selection on a test trial. During Phases 5 through 8, AC matching was established in the same manner as AB matching. Given the AB and AC matching baselines, tests for symmetry (BA and CA) and equivalence (BC and CB) became possible. These were presented in Phase 9. Test trials were interspersed among baseline trials (AB and AC on a VR 3 schedule); no differential consequences followed test trials.

Participant HF. During initial training, the procedures were virtually identical to those used with BN. The training outcome was different, however (see Results). That outcome led to the modified training regimen presented in Table 2. Briefly, the modified program substituted a blank card for the negative comparison on a proportion of trials ranging from 80% to 0%. This procedure was designed to minimize errors and to encourage HF to observe the sample and the positive comparison before making his matching selection.

Table 2

Revised sequence of training and testing for Participant HF.

Phase	Stage	% with blank	% with S−	Baseline schedule
Training	1	80	20	CRF
A1B1 and A2B2	2	70	30	CRF
	3	60	40	CRF
	4	50	50	CRF
	5	40	60	CRF
	6	30	70	CRF
	7	20	80	CRF
	8	0	100	VR 3
Symmetry test				
B1A1 and B2A2	9	0	100	VR 3
Training				
A1C1 and A2C2	10–17	As in Stages 1–8		
Symmetry test				
C1A1 and C2A2	18	As in Stage 9		
Transitivity and equivalence test				
B1C1 and B2C2	19	As in Stage 9		
C1B1 and C2B2				

During Stages 1 through 7, 20-trial blocks were programmed, and A1B1 and A2B2 trials were presented in an unsystematic order. Every correct response was followed by a reinforcer. The criterion for advancement from one stage to the next was 90% correct performance on the last 10 trials of a block. As Table 2 shows, the number of trials presenting the negative comparison instead of the blank card increased progressively. At Stage 8, all trials presented a negative comparison stimulus. In addition, the reinforcement schedule was changed; only every third response on average was followed by a reinforcer. The schedule change was made to prepare for B1A1 and B2A2 symmetry tests (Stage 9); no reinforcers followed any symmetry test-trial selection. During Stages 10 through 17, the A1C1 and A2C2 relations were taught via methods like those used in Stages 1 through 8. C1A1 and C2A2 symmetry test trials were introduced in Stage 18. Transitivity and equivalence tests were introduced in Stage 19. As with BN, all test trials were interspersed among baseline trials, and no reinforcers followed any test-trial selection.

Tests for sample-S+ and sample-S− relations. Following the tests for emergent relations, both participants received follow-up tests that evaluated possible sample-S+ and sample-S− baseline relations. The method was similar to that reported by Stromer and Osborne (1982). Specifically, a single novel form was substituted for the S− on sample-S+ tests (see Figure 2); the novel form was substituted for the S+ on sample-S− tests. Baseline reinforcement procedures were like those used in the prior tests for emergence. Test blocks scheduled eight test sample-S+ or sample-S− trials interspersed among 16 baseline trials. Each test block was presented twice (i.e., 16 total test trials for each type of relation). No differential consequences followed test selections.

Results and Discussion

Figure 3 summarizes the major results of Study 2 and shows that both participants acquired baseline matching-to-sample performances. BN did so via the initial training program; scores in every training session were well above chance (i.e., 50%). The initial program did not succeed with HF, however. The star in Figure 3 gives the mean accuracy on AB conditional discrimination trials during the that program. Detailed initial program data, presented in Figure 4, show that HF never exceeded 70% correct whenever the AB conditional discrimination was required. Moreover, continued exposure to program conditions did not improve performance. For this reason, the subsequently successful revised training program was used.

On the tests for emergent performance (Figure 3), BN's BA symmetry test score was only intermediate. Performance on the CB symmetry test, however, was perfect. On the equivalence tests, BN's initial score was low, but subsequent scores improved to a high level despite the fact that test-trial performances produced no differential consequences. Thus, BN seemed to display the gradual emergence phenomenon that is frequently reported in stimulus equivalence research (Sidman, 1994).

Probe data from HF showed no evidence of emergent matching relations. With negative findings like this, one is led to ask whether additional exposure to the training and test conditions or procedure changes might produce a positive outcome. In follow-up work, reported by Carr (1997), HF was directly taught bidirectional performances (cf. Schusterman & Kastak, 1993) in an effort to encourage positive equivalence test outcomes. Although HF's test scores improved with this procedure, his performances were unstable and fell far short of a convincing demonstration of stimulus equivalence.

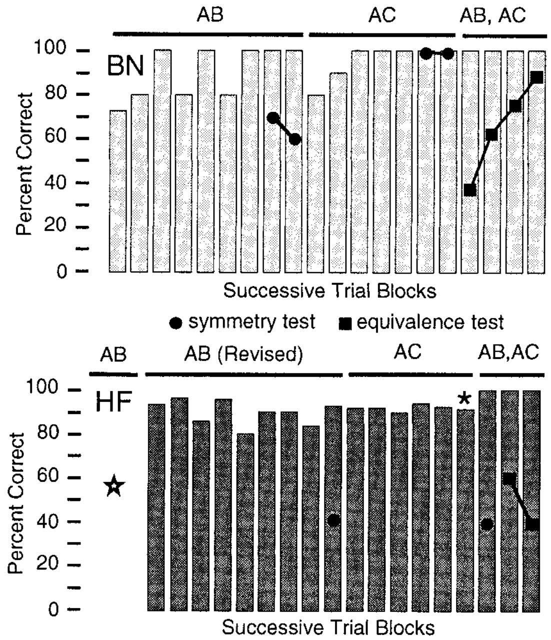

Fig. 3. Training and test results from Study 2. Histograms plot percentage of correct selections in successive 10-trial (BN) or 20-trial (HF) blocks of AB or AC baseline trials. Filled circles and squares indicate performance on tests for emergent relations. The star in the left portion of HF's plot indicates mean accuracy on conditional discrimination training trials that preceded those shown here (see text and Figure 4). The asterisk indicates a single 30-trial block that preceded the AC symmetry and BC and CB transitivity tests.

On tests for sample-S+ and sample-S− relations, Figure 5 shows that BN's probe selections remained consistent with established equivalence classes. By contrast, HF's performance was more difficult to interpret, largely because his matching-to-sample baseline deteriorated during the tests. Performance on trials that tested sample-S+ control was comparable to baseline levels; the score on the

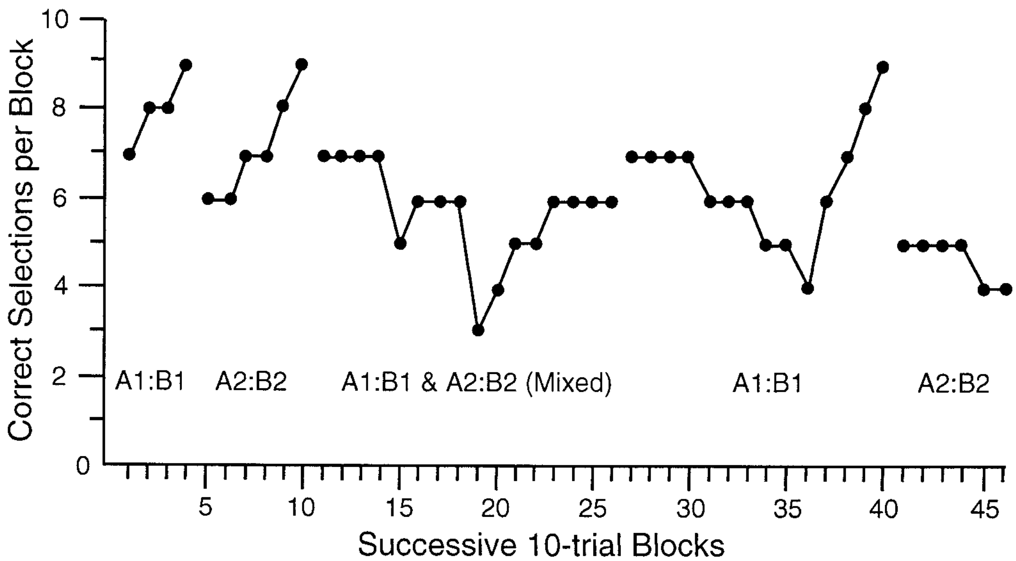

Fig. 4. Results of the original attempt to teach HF the AB conditional discrimination. The filled circles show accuracy scores on the performaces indicated immediately below them.

sample-S− test was lower and in the chance range of performance. Notably, the 3 participants in Study 1 all had demonstrated both sample-S+ and sample-S− relations with their baseline performances (i.e., in the earlier study by Serna et al., 1998, and in replicating those procedures with IVB). Thus, when data from both Study 1 and Study 2 are combined, the 4 individuals who showed positive equivalence class outcomes also showed both sample-S+ and sample-S− relations. By contrast, the single individual who had chance-level scores on equivalence tests had similar scores on tests for sample-S− relations (see Carrigan & Sidman, 1992, for a different perspective on this issue). Further research is needed to determine whether these findings are replicable, and if so to clarify their meaning.

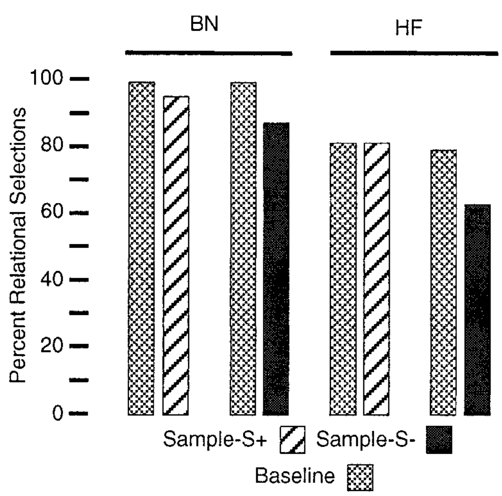

Fig. 5. Baseline and probe data from tests of sample-S+ and sample-S− relations.

GENERAL DISCUSSION

The present findings demonstrate equivalence classes in low-functioning individuals who lacked well-developed repertoires as either speaker or listener. Consider these participants' repertoires in relation to the processes that Horne and Lowe (1996) postulate undergird positive equivalence outcomes. For the 4 participants who had positive equivalence test outcomes, vocal echoics were not observed and had never been so at any point in their lives. Given the central role of echoics in the Horne and Lowe account, none of the participants should have shown a positive equivalence test outcome. What about the listening repertoires? Compared with a typically developing 2- or 3-year-old child, these participants' repertoires were meager indeed. It would be extremely difficult to apply Horne and Lowe's analysis to the verbal repertoires of these individuals.

A reasonable question is why our studies yielded positive results while those from certain other laboratories have been negative (e.g., Devany et al., 1986). Although we have no immediately compelling explanation, we speculate that differences in training procedures may have been an important variable. In Study 1, for example, the participants' auditory-visual baselines were entry skills, which probably reflected the effects of extensive prior teaching in special education settings. In addition, the critical visual-visual baselines were taught by careful, sometimes protracted, stimulus control shaping. In the language of McIlvane et al. (in press), such procedures might be expected to increase stimulus control topography coherence, that is, to encourage topographies that were consistent with positive test outcomes. It may not be coincidental that the only negative result came from Study 2, in which the teaching programs were somewhat less elaborate.

Proponents of the position that naming (or language, more generally) is essential for equivalence class formation will likely question whether the participants in our study truly lacked the crucial prerequisites. One might argue, for example, that some participants entered the study able to match certain pictured items to their spoken names or used a few signs to communicate basic wants. It is hard to argue, however, that the signing repertoires permitted naming in the sense meant by Horne and Lowe (1996). No overt signing was observed during the testing of any of the participants; proponents of the naming hypothesis must then argue that "covert signing" could have occurred. In Study 2, however, BN matched entirely arbitrary forms. Even if one accepted the possibility of covert signing, what signs would he have used? BN had no signs for basic shapes or colors, and thus would have to have developed purely arbitrary coding responses (see Carter & Werner, 1978, for a discussion of coding). Moreover, IVB had neither speech nor a signing repertoire. What was the source of his coding responses, if any?

Because the participants in these studies did have some rudimentary verbal skills, one might also ask whether they might actually have been more capable than their test results indicated. Because all of the participants in these studies were severely mentally retarded, they had been identified as candidates for special education early in life. We consider it especially noteworthy that a lifetime of formal and informal training had been sufficient to establish neither speech nor a signing repertoire that featured the full range of verbal functions (i.e., tacts, intraverbals, etc.). Contrast these limited achievements with the apparent facility with which the participants displayed equivalence classes. Thus, it seems inarguable that the data reported here must lead at minimum to modifications of proposals that well-developed language is central to stimulus equivalence. We turn next to considering the issue of whether equivalence classes (as defined by Sidman) might be demonstrable in the absence of any naming or language repertoire.

Ruling Out Naming or Language in Equivalence Class Formation in Human Participants

What pattern of data would be needed to conclude that naming or language was not in fact prerequisite for a display of equivalence classes? In the ideal case, the best procedures to test both hypotheses would be to study individuals who enter with no skills that even formally resembled communication. As noted by McIlvane and Dube (1996), meeting this test with human participants may be quite difficult. Neurological problems that render a human frankly noncommunicative despite years of specialized training are extremely hard to overcome. Such individuals may have profound disabilities in attending selectively to elements of complex stimuli, maintaining attending, and remembering recent events. Many fail to learn such basic skills as eating independently, finding their way in familiar environments, recognizing members of their families, and so forth. Learning anything resembling a conditional discrimination baseline (or even a simple discrimination for that matter) may be very difficult, even with the best teaching technology now available.

One possible route to working with nonverbal individuals is to follow the lead of Devany et al. (1986). They studied very young children who were developmentally delayed. Presumably, it should be possible to find children with reasonable learning potential who have not yet begun to speak, as they seem to have done. Many children with autism, for example, meet this criterion. However, Devany et al. did not report any formalized assessment of their participants' listening repertoires, so we cannot evaluate their actual capabilities. We do know, however, that all of their nonverbal children acquired arbitrary visual-visual matching-to-sample baselines via a conventional three-step training procedure (Guess & Baer, 1973; cf. Dixon, 1977). That method presents two invariant sample–comparison pairs successively in a simple simultaneous discrimination format and then intermixes trial types, requiring a conditional discrimination. Although the three-step method sometimes succeeds in teaching participants with developmental limitations, as it did with BN in Study 2, the typical outcome is a failure of the type seen with HF (e.g., Augustson & Dougher, 1991). Indeed, much research has been undertaken to find alternative methods for teaching conditional discriminations to such individuals (e.g., Saunders & Spradlin, 1990; Zygmont et al., 1992). The fact that all of Devany et al.'s children learned via the three-step method, indicating substantial learning potential even though they were all nonverbal, suggests a highly unusual, insufficiently characterized population. Perhaps that is why no direct replication of their work has ever been reported, despite the study's many citations and its prominent role in the history of stimulus equivalence research.

Yet another possibility is to consider studies of typically developing preverbal infants. Observations of the onset of early language consistently show that typically developing children begin to understand words as linked with objects or other environmental events (social routines such as ''bye-bye'') at about 9 months of age (Bates, 1979; Fenson et al., 1993). Thus, at least some components of the listener repertoire emerge before the first birthday. However, children most often do not begin to produce words until later (about 12 to 15 months), and do not appear to understand the symbolic potential of these words until perhaps as late as their second birthday (Bates, 1979). Thus, toddlers between 1 and 2 years old might provide a natural opportunity to explore equivalence class formation in individuals with limited verbal capabilities.

Extension of extant equivalence class methodology to a toddler population would be difficult but not impossible. There is no logical reason why equivalence methodologies might not be adapted to the kinds of procedures that have been used successfully with children of this age. For instance, Werker, Cohen, Lloyd, Casasola, and Stager (1998) have successfully implemented preferential looking procedures as a tool for assessing learning of conditional word–picture relations in infants as young as 8 months. Such methodologies offer potentially promising avenues for adapting equivalence tests, and could be implemented readily by any suitably prepared infant-study laboratory.

Other Methods for Falsifying Language-Related Hypotheses

Equivalence researchers who are interested in issues related to language appear to be on the horns of a dilemma. Despite our evidence to the contrary, it is still possible to argue that older individuals with developmental disabilities and severely limited language have more experiences with symbolic relations than is adequately revealed by tests like the PPVT-R. One possibility would be to invest further in studies with frankly nonverbal participants. Prior work in other aspects of relational learning has had positive results (e.g., McIlvane & Stoddard, 1981), but the necessary training is arduous, technically demanding, and somewhat expensive. Until such studies are undertaken or until methods are developed to allow study of preverbal infants, language-related hypotheses may be effectively unfalsifiable in work with human participants.

Clearly, more work is necessary with nonhumans who might have equivalence potential. Unfortunately, Schusterman's studies of marine mammals are necessarily limited by access to only a few animals. An intensive effort is needed with nonhuman primates, especially those that have not been part of research programs that sought to develop human-like speaking and listening repertoires (e.g., Savage-Rumbaugh, 1986). For example, such work may be possible with the Capuchin monkey (*Cebus apella*) (D'Amato, Salmon, Loukas, & Tomie, 1985).

Another possible route is to intensify efforts in the domain of connectionist modeling of stimulus equivalence (see Donahoe & Palmer, 1994, for a discussion of the role of such models in behavior analysis). By themselves, connectionist models will not convince everyone. Nevertheless, we suspect that development of high-quality biologically plausible models would make it somewhat harder to maintain the position that equivalence arises only from language learning. Should positive results emerge from studies of nonhumans and biologically plausible connectionist models, the counterargument that these entities acquire equivalence classes in a fundamentally different way (cf. Horne & Lowe, 1996, pp. 224, 233) would not be parsimonious and would be hard to defend on intellectual grounds.

Study of Equivalence in Humans with Minimal Verbal Repertoires

Although study of humans with developmental disabilities and limited verbal repertoires may not fully falsify the language hypothesis, we believe that there are still compelling reasons for continued study of this population. First, it would be of benefit to continue to accumulate empirical demonstrations like the one reported here. If methods can be developed to routinely demonstrate positive outcomes in progressively less verbal individuals, one might indeed learn something useful about the relation between equivalence classes and language repertoires. For example, if equivalence is a basic process underlying language, as Sidman (1994) has suggested, it seems reasonable to ask whether the number and complexity of classes that can be developed might be related in some fundamental way to the richness and complexity of verbal repertoires.

Second, such studies might help to address Sidman's (e.g., 1994, pp. 338, 445) suggestions that equivalence classes reflect a fundamental behavioral process associated with contingencies of reinforcement. One logical outcome of this suggestion is that equivalence classes should be demonstrable whenever appropriate matching-to-sample baselines are established, provided that there is no interfering experimental artifact (Sidman, 1994). We make a similar assertion when we suggest that positive equivalence outcomes result when there is coherence between participant- and experimenter-specified stimulus control topographies (McIlvane et al., in press). Taken in the abstract, if there is coherence, equivalence outcomes must be positive (provided no other artifact is present). Similarly, if there is no coherence, equivalence outcomes may be negative. Individuals with limited behavioral development may be the population of choice for evaluating the merits of our topography coherence hypothesis. Stimulus control analyses with this population frequently reveal differences between participant- and experimenter-specified topographies (cf. McIlvane, 1992; Stoddard & McIlvane, 1989). As methods of stimulus control analysis and stimulus control shaping are improved, it should be increasingly possible to detect lack of topography coherence and to encourage coherence by careful teaching.

There may be hints in the present findings about methods to encourage topography coherence. DJB, IVB, and JRV showed astonishingly rapid, errorless positive equivalence outcomes despite the fact that their measured mental ages were among the lowest studied thus far. For example, there was no gradual emergence. Notably, all 3 participants acquired the relevant arbitrary visual-visual matching-to-sample baselines via stimulus control shaping, a technique that has long been used to direct attending to relevant stimulus differences. Also, the blank comparison baseline directly verified the sample-S+ relations that are necessary for a positive outcome on a conventional equivalence test (cf. Sidman, 1987). Finally, the auditory-visual baselines were established merely by reinforcing relevant entry performances. As mentioned earlier, relating the letters to these familiar stimuli might well have been a variable in the positive outcomes that were obtained.

We conclude by giving perhaps the most compelling rationale for continuing study of equivalence classes in individuals who function at low behavioral levels. There is an immense practical need to develop more effective methods for establishing nonvocal communication in individuals who cannot or do not learn to speak. Communication boards (like that used by DJB) require skills that are much like those studied in laboratory research (e.g., scanning stimulus arrays, pointing accurately, and perhaps most important, learning relations between the forms on the board and corresponding events in world). An agenda that seeks reliable equivalence learning in this population will simultaneously contribute science and technology on which to base more effective communication training methods.

REFERENCES

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

Bates, E. (1979). *The emergence of symbols: Cognition and communication in infancy.* New York: Academic Press.

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

Carrigan, P. F., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 58,* 183–204.

Carter, D. E., & Werner, T. J. (1978). Complex learning and information processing by pigeons: A critical analysis. *Journal of the Experimental Analysis of Behavior, 29,* 565–601.

D'Amato, M. R., Salmon, D., Loukas, E., & Tomie, A. (1985). Symmetry and transitivity of conditional relations in monkeys (*Cebus apella*) and pigeons (*Columba livia*). *Journal of the Experimental Analysis of Behavior, 44,* 35–47.

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

Dixon, L. S. (1977). The nature of control by spoken words over visual stimulus selection. *Journal of the Experimental Analysis of Behavior, 27,* 433–442.

Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior.* Boston: Allyn & Bacon.

Dube, W. V. (1991). Computer software for stimulus control research with Macintosh computers. *Experimental Analysis of Human Behavior Bulletin, 9,* 28–30.

Dube, W. V., Iennaco, F. M., & McIlvane, W. J. (1993). Generalized identity matching to sample of two-dimensional forms in individuals with intellectual disabilities. *Research in Developmental Disabilities, 14,* 457–477.

Dube, W. V., & McIlvane, W. J. (1996). Some implications of a stimulus control topography analysis for emergent stimulus classes. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 197–218). North Holland: Elsevier.

Dube, W. V., McIlvane, W. J., Mackay, H. A., & Stoddard, L. T. (1987). Stimulus class membership established via stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior, 47,* 159–175.

Dube, W. V., McIlvane, W. J., Maguire, R. W., Mackay, H. A., & Stoddard, L. T. (1989). Stimulus class formation and stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior, 51,* 65–76.

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

Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test—Revised.* Circle Pines, MN: American Guidance Service.

Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications.* New York: Irvington.

Fenson, L., Dale, P. A., Reznick, J. S., Bates, E., Thal, D., & Pethick, S. J. (1993). *MacArthur communicative development inventories.* San Diego: Singular Publication Group.

Gardner, M. F. (1990). *Expressive one-word picture vocabulary test—Revised.* Novato, CA: Academic Therapy Publications.

Grove, N., & Walker, M. (1990). The Makaton vocabulary: Using manual signs and graphic symbols to develop interpersonal communication. *Augmentative and Alternative Communication, 6,* 15–28.

Guess, D., & Baer, D. M. (1973). An analysis of individual differences in generalization between receptive and productive language in retarded children. *Journal of Applied Behavior Analysis, 6,* 311–329.

Hayes, S. C. (1991). A relational control theory of stimulus equivalence. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 19–41). Reno, NV: Context Press.

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.

Knowles, K., & Masidlover, N. (1982). *Derbyshire language scheme.* Derby, UK: Derbyshire County Council.

Lionello, K. M., & Urcuioli, P. J. (1998). Control by sample location in pigeons' matching to sample. *Journal of the Experimental Analysis of Behavior, 70,* 235–251.

McIlvane, W. J. (1992). Stimulus control analysis and nonverbal instructional methods for people with intellectual disabilities. In N. W. Bray (Ed.), *International review of research in mental retardation* (Vol. 18, pp. 55–109). San Diego: Academic Press.

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

McIlvane, W. J., Kledaras, J. B., Munson, L. C., King, K. A., de Rose, J. C., & Stoddard, L. T. (1987). Controlling relations in conditional discrimination and matching by exclusion. *Journal of the Experimental Analysis of Behavior, 48,* 187–208.

McIlvane, W. J., Serna, R. W., Dube, W. V., & Stromer, R. (in press). Stimulus control topography coherence and stimulus equivalence: Reconciling test outcomes with theory. In J. Leslie & D. E. Blackman (Eds.), *Issues in experimental and applied analyses of human behavior.* Reno, NV: Context Press.

McIlvane, W. J., & Stoddard, L. T. (1981). Acquisition of matching-to-sample performances in severe retardation: Learning by exclusion. *Journal of Mental Deficiency Research, 25,* 33–48.

McIlvane, W. J., & Stoddard, L. T. (1985). Complex stimulus relations and exclusion in severe mental retardation. *Analysis and Intervention in Developmental Disabilities, 5,* 307–321.

Reynell, J. K., & Huntley, M. (1985). *Reynell developmental language scales* (2nd ed.). Los Angeles: Western Psychological Services.

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

Savage-Rumbaugh, E. S. (1986). *Ape language: From conditioned response to symbol*. New York: Columbia University Press.

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

Serna, R. W., Wilkinson, K. M., & McIlvane, W. J. (1998). Blank comparison assessment of stimulus-stimulus relations in individuals with mental retardation: A methodological note. *American Journal on Mental Retardation, 103,* 60–74.

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

Sidman, M. (1987). Two choices are not enough. *Behavior Analysis, 22,* 11–18.

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

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

Stoddard, L. T., & McIlvane, W. J. (1989). Generalization after intradimensional discrimination training in 2-year-old children. *Journal of Experimental Child Psychology, 47,* 324–334.

Stromer, R., & Osborne, J. G. (1982). Control of adolescents' arbitrary matching-to-sample relations. *Journal of the Experimental Analysis of Behavior, 37,* 329–348.

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

Wilkinson, K. W., & McIlvane, W. J. (1994). Stimulus organization and learning by exclusion: A preliminary experimental analysis. *Experimental Analysis of Human Behavior Bulletin, 12,* 17–20.

Zygmont, D. M., Lazar, R. M., Dube, W. V., & McIlvane, W. J. (1992). Teaching arbitrary matching via stimulus control shaping to young children and mentally retarded individuals: A methodological note. *Journal of the Experimental Analysis of Behavior, 57,* 109–117.

Received March 3, 1999
Final acceptance April 17, 2000