EVALUATING MERGER AND INTERSECTION OF EQUIVALENCE CLASSES WITH ONE MEMBER IN COMMON

HARRY A. MACKAY[1,2], KRISTA M. WILKINSON[1,3], COLLEEN FARRELL[4], AND RICHARD W. SERNA[1]

[1]SHRIVER CENTER, UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
[2]NORTHEASTERN UNIVERSITY
[3]EMERSON COLLEGE
[4]BLOOMFIELD (CT) PUBLIC SCHOOLS

Sidman (1994) noted that the existence of a member that is common to more than one class may produce either class merger (union) or class intersection. A multiple-selection, matching-to-sample test was developed to examine the conditions under which these outcomes occur. Test trials each required three conditional discriminations involving selection or rejection of comparison stimuli under control of samples representing two categories. Test results obtained from an initial group of typical adults using familiar stimuli (DOG and BIRD, pictures of dogs and birds and relevant printed breed names (e.g., DALMATIAN, RETRIEVER) showed the conditional stimulus control best described as intersection. For example, the word DALMATIAN provided the context for selecting the dalmatian but not the retriever picture. However, these results may have depended on the participants' verbal history as English speakers. Would conditional-discrimination training with overlapping sets of laboratory-generated stimuli also result in intersection? Naïve typical adults were assigned to one of three different training conditions. Like the participants tested with familiar stimuli, these participants demonstrated highly reliable test outcomes best described as showing class intersection, regardless of training condition. These findings begin to elucidate the necessary and sufficient conditions for establishing complex category-like classes of stimuli.

Key words: equivalence classes, merger, intersection, contextual control, semantic classes, computer mouse click, adult humans

––––––––––––––––––––––

Sidman's analyses of verbal classes and stimulus equivalence (e.g., Sidman, 1986, 1994) fundamentally altered research and theory concerning the stimulus control of behavior. In typical research on equivalence (e.g., Sidman, 1971), matching-to-sample (conditional discrimination) procedures are used to establish a subset of the possible relations between physically dissimilar stimuli, like pictures and their printed and spoken names. Then, without further training, matching-to-sample tests are used to examine the mutual substitutability of the stimuli, thus assessing the emergence of new relations between the stimuli. The emergence of these untrained relations has been likened to the development of symbolic relations (e.g.,

Mackay & Sidman, 1984; Sidman, 1994) and may be related to phenomena studied by cognitive psychologists (e.g., Nelson, 1985; Waxman, 1991). This general notion has provoked considerable discussion and research (e.g., Branch, 1994; Fields, Reeve, Adams, & Verhave, 1991; Stemmer, 1990; Wasserman & Devolder, 1993; Wilkinson & McIlvane, 2001). It also figures prominently in the Special Issue of this journal concerned with categorization and concept learning (Critchfield, Galizio, & Zentall [Eds.], 2002)

Little has been done to evaluate directly whether classes established via the laboratory methods and procedures that yield equivalences possess the same properties as categories or semantic classes acquired extraexperimentally. However, examination of this issue is essential. As Catania (1996, p. 277) noted, ''no elaboration of artificial contingencies to account for natural phenomena is likely to be convincing without evidence that such contingencies are also likely to operate in natural environments…'' Likewise, Galizio (1996, p. 288) stated that ''a key test area [is]… research that directly extends [equivalence] analysis to the study of natural language classes and the phenomena associated with them.'' The cur-
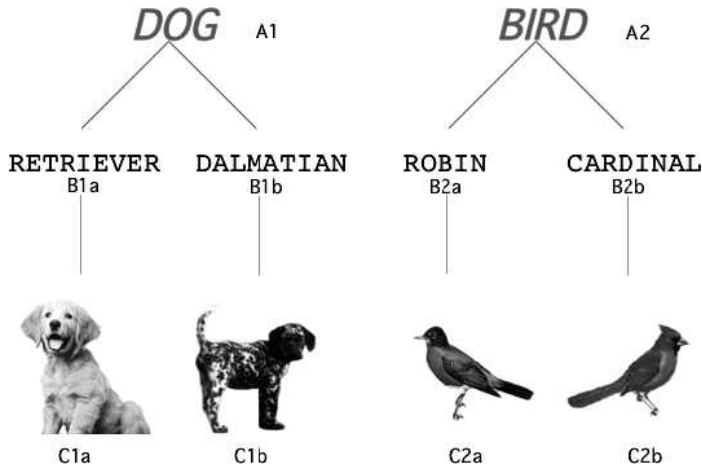
Fig. 1. Familiar categories and stimuli used in preliminary research. Pictures were in natural colors. As stimuli, the words DOG and BIRD were blue and the others black. The alphanumeric labels identify particular stimuli and did not appear in stimulus displays.

rent research initiated such an extension. First, familiar categories that represented extraexperimental semantic classes were examined using matching-to-sample procedures typical of experiments on equivalence and a new test procedure that permits selections of multiple comparison stimuli on the same trial, thus simulating a sorting task (see Wilkinson & Rosenquist, 2006). Novel categories then were taught to adult participants and examined using the same procedures. Did these novel categories show the structural properties and constraints observed in the familiar extraexperimental semantic categories?

Figure 1 shows members of two familiar categories (dog and bird) used in computer-presented tasks described later. The four pictures are exemplars of the two categories and each is related to a printed word so that there are one-to-one relations between the words and the pictured exemplars. Within each category two words (as experimental stimuli, printed in black), RETRIEVER and DALMATIAN for the dog category, ROBIN and CARDINAL for the bird category, are related to the respective single words DOG and BIRD (as stimuli, printed in blue and larger than the breed names). Consideration of the membership and structure of the categories illustrated in Figure 1 suggests that the members of each category may form two classes (potentially equivalence classes, see

below) linked by a common superordinate member, DOG in one case, BIRD in the other.

The experiments described here aimed to examine the functional significance of the sharing of class members. As Sidman (1994) has noted, the existence of a member that is common to more than one class may produce either class merger (union) or class intersection. In the case of merger, possession of the common member results in the combination of two or more classes into one large class. In the case of intersection, the classes remain separate and class membership changes across different circumstances, that is, class size remains limited and class membership depends upon contextual control. These possibilities raise considerations relevant to Catania's (1996) and Galizio's (1996) challenges cited earlier, concerning contingencies in the natural environment that may yield the highly flexible repertoires involved. For example, the speaker of English can determine whether the words DALMATIAN and RETRIEVER are equivalent, and therefore substitutable for one another at one time (e.g., when instructed to ''Name dogs'' or ''Name something that barks'') but not at another (e.g., when instructed to ''Point to the Dalmatian'' or ''Point to the one with spots,'' where either the written word or the picture is appropriate). What processes of acquisition might yield such flexibility? Research directed at answering

such questions and considerations will enrich the scope and impact of the stimulus equivalence approach. The analyses reported here serve that purpose.

Sidman, Kirk, and Willson-Morris (1985) examined conditions that resulted in mergers of classes of equivalent stimuli. They first established two groups of three classes, each class consisting of three stimuli. Conditional discrimination training then was used to combine these classes into three 6-stimulus equivalence classes. In the training, a stimulus from one of the classes to be combined (e.g., E1 of class D1E1F1) served as the sample, and a stimulus from the other class (e.g., C1 of class A1B1C1) served as the correct comparison. After this limited training, subjects matched any two of the six stimuli, thus demonstrating class mergers. Saunders, Saunders, Kirby and Spradlin (1988) also described the occurrence of class mergers but without the direct training used by Sidman and his colleagues. However, these subjects had histories with two-choice conditional discrimination procedures that led to reliable, though unreinforced, conditional responding when new stimulus pairings were presented. The class memberships following merger were consistent with the untrained conditional responding that occurred. Lane, Clow, Innis, and Critchfield (1998) used educationally relevant stimuli in their study that demonstrated the applied significance of procedures that yield merger of classes. They established the four 3-stimulus classes A–O–''vowel'' and D–V–''consonant'', and then E–U–''vowel'' and K–T–''consonant''. The outcomes of test trials (e.g., with A as sample select E as correct comparison, with D as sample select K as correct comparison) then showed that the sets of stimuli that shared the same auditory stimulus merged to form a five-stimulus vowel class (A–O–E–U–''vowel'') and a five-stimulus consonant class (D–V–K–T–''consonant''). In other research, Fields and his colleagues have examined mergers of equivalence classes with perceptual classes (e.g., lines of similar lengths; Fields, Reeve, Adams, Brown, & Verhave, 1997), and of two perceptual classes (e.g., Fields, Matneja, Varelas, Belanich, Fitzer, & Shamoun, 2002). These authors suggest that ''establishment of some conditional discriminations between members of two classes may lead to class merger, regardless of class type'' (p. 288) and

that class mergers may underlie the development of open-ended classes of infinite size.

Although Sidman (1994) discussed the topic of class intersection, little is known about the behavioral processes that yield that outcome. Some laboratory research has demonstrated the conditional control of equivalence relations that is a basis for inferring intersection: The same stimuli are shown to belong to one class in one context and to another class in a second context. In many of these studies, the different contexts, and related contingencies, were cued by stimuli presented throughout training trials (e.g., display colors, Wulfert & Hayes, 1988, Griffee & Dougher, 2002; Hayes, Kohlenberg & Hayes, 1991; tones, Bush, Sidman & DeRose, 1989; forms, Gatch & Osborne, 1989; Kennedy & Laitinen, 1988; Rehfeldt, 2003; Serna, 1987). Lynch and Green (1991) used dictated nonsense syllables that accompanied only the sample stimuli for training trials. The contextual control exercised by these auditory stimuli then was transferred to arbitrary visual forms in further training. Sidman (1994, pp. 516–524) describes experiments by Fucini (1982) in which the contextual cues controlling the occurrence of merger and intersection were the particular negative comparison stimuli presented on a given trial.

The current research extends the earlier analyses in a new direction; the critical contextual stimuli were different members of the same class presented as samples on different trials. To illustrate using the potential stimuli shown in Figure 1, the words DOG and RETRIEVER (or DALMATIAN) would be presented as samples on trials in which both dog pictures appeared among the comparisons. We were interested specifically in distinguishing whether the establishment of two equivalence classes that included a shared member (the word DOG, for example) resulted in merger or intersection. To do so, we developed a new procedure that provides a test that distinguishes these two potential outcomes. The test procedure is a variant of matching-to-sample tasks that allows more than a single selection to be made on each trial, (i.e., *a multiple-selection matching-to-sample task*; see Wilkinson & Rosenquist, 2006). While this resembles sorting procedures used in studies with tabletop methods (e.g., Cowley, Green, & Brownling-McMorrow, 1992; Lowe,
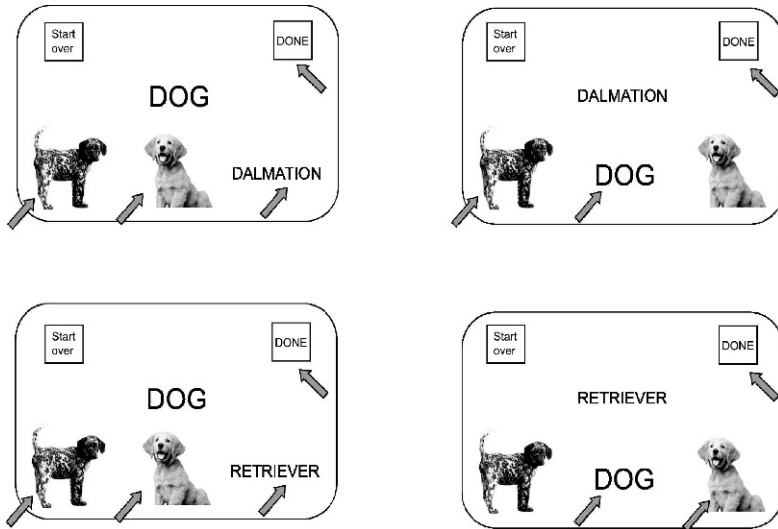
Fig. 2. Displays for test trials involving one, two, and three selections.

Horne, Harris, & Randle, 2002; Pilgrim & Galizio, 1996) these instances involved simultaneous presentations of all training stimuli for choice. In the current research, the procedures were computer based and subsets of three stimuli were presented for choice (see Figure 2 and Appendix A). One, two, or all three could be selected before touching a key labeled DONE to end the trial.

Figure 1 demonstrates how members of two familiar categories can be classified using typical equivalence notation. Figure 2 presents our multiple-selection matching-to-sample task in which more than one of the comparison stimuli could be correct. There are clear expectations about the outcomes of tests in which the stimuli illustrated in Figure 1 (and related text) are presented to English speakers in matching-to-sample tasks. For example, any English speaker given the trials illustrated in Figure 2 would be expected to select all three comparisons when the sample is DOG (left panels), but only two comparisons when the sample is DALMATIAN or RETRIEVER (right panels).

Considered broadly, the trials illustrated in Figure 2 exemplify trial types that might yield empirical tests that distinguish between class merger and intersection. Each row shows a pair of trials involving the same four stimuli but the one serving as the sample differs. (There are four comparable trial types for the bird category.) A participant's selections on

these trial types (and others involving two C stimuli from the same category as comparisons; see below) may answer the question: Does the sample stimulus for a particular trial exert differential instructional control (Cumming & Berryman, 1965) determining which stimuli are treated as equivalent? In other words, do different members of the same class exercise conditional control of equivalence? The trials with a B sample and both C stimuli as comparisons (panels at right in Fig. 2) are of particular interest because the participant's selections on these trials (Appendix A, Trials 9, 14, 19, 28) distinguish class merger from intersection. Selection of only the C stimulus that had been related directly to the B sample by training or extraexperimental history suggests intersection whereas selection of both C stimuli suggests their interchangeability and is consistent with merger of the classes.

Occurrence of the expected performances with the familiar, extraexperimental stimuli would help to validate that our procedures can reveal outcomes consistent with intersection. We conducted the relevant research (unpublished) with 5 young adult participants, using the dog and bird stimuli illustrated in Figure 1. These participants did just as expected. First, they were given matching tasks that demonstrated stimulus relations (baseline) consistent with English. For example, in the A–B matching task, all five selected RETRIEVER rather than CARDINAL given the sample

Table 1

Stimuli presented on multiple-selection test trials.

Trial #	Sample	Comparisons		
8	A1	C1a	B1b	C1b
12	A2	B2b	C2a	C2b
24	A1	C1a	B1a	C1b
31	A2	C2b	C2a	B2a
9	B2a	C2b*	C2a	A2
14	B1b	C1b	A1	C1a*
19	B2b	A2	C2a*	C2b
28	B1a	C1a	A1	C1b*
4	B1a	C1b*	C1a	B1a
10	B2b	C2b	C2a*	B2b
22	B1b	C1a*	B1b	C1b
23	B2a	C2a	B2a	C2b*

* suggests intersection of classes if not selected

DOG; in the B–A task, they selected DOG given the sample RETRIEVER; in the B–C task they selected the retriever picture given RETRIEVER as sample. Additional performances demonstrated the expected C–B, A–C, and C–A relations. Then, in the multiple-selection task, the patterns of selection for all 5 participants were identical and as anticipated.

These highly reliable results for the particular test trials in which the comparisons included both C stimuli from different subclasses in the same category (i.e., C1a and C1b or C2a and C2b) may be presented with reference to Table 1 which lists these trials. For a single participant performance may be described simply as follows: The stimuli marked by an asterisk never were selected, whereas the unmarked stimuli always were selected. Across the 5 participants, then, the same statement describes performances on a total of 60 test trials, each involving three conditional discriminations.

More specifically, the upper section of Table 1 shows the trials illustrated at left in Figure 2 (see also Appendix A, Trials 8, 12, 24, 31). On all four of the trials with DOG (A1) and BIRD (A2) as samples each participant selected both related C stimuli (dog and bird pictures respectively), thus confirming performance with the AC relations but in the multiple-selection task rather than standard matching. The third comparison for these trials, a B stimulus, also was selected. In contrast, for the trials shown in the middle and lower sections of Table 1 with a B stimulus as the sample (e.g., RETRIEVER, DALMA-

TIAN; see Figure 2; also Appendix A, Trials 9, 14, 19, 28 and 4, 10, 22, 23) selections among the same C stimuli were conditional upon the samples presented. The picture in the BC relation that had been established outside the experimental context (e.g., C1b is the match to B1b on trial 14; Table 1 middle section) always was selected. However, the stimulus (e.g., C1a on trial 14) whose selection on each trial would have suggested class merger, was never selected. The performances involving the BC relations thus were evaluated on eight trials (four per category) and provided within-subject replications using stimuli from the two different categories.

Note that different comparison stimuli were presented in addition to the C stimuli for the trials shown in the middle and bottom sections of Table 1. For the trials in the middle section, the A stimuli presented were from the same class as the concurrent samples. For the trials shown at bottom, the third comparison was the B stimulus physically identical to the concurrent sample. Selections of these stimuli always occurred, thus confirming performances already demonstrated in the standard matching tasks.

Finally, although not shown in Table 1, the results of the trials that involved assessment of relations between the B stimuli (e.g., B1a and B1b) and between the C stimuli (e.g., C1a and C1b) were consistent with the preceding outcomes. For example, on both trials in which the two B stimuli from the same category (e.g., ROBIN and CARDINAL) were comparisons with one as the sample (e.g., ROBIN), the other (i.e. CARDINAL) was not selected as a comparison. Likewise, the two C stimuli (pictures of the dogs and birds) were not matched with one another on two trials in which one served as the sample and the other was one of the comparisons presented. On all remaining trials, opportunities to demonstrate baseline relations were provided. Performance on these trials was perfect.

The importance of the procedures described in this introduction lies in the validation of our method. For these familiar stimuli, the training presumably occurred extraexperimentally during language acquisition long before the current experiments began. Nevertheless, the test outcomes are clear. The participants' linguistic histories included contingencies that established two dog classes that

intersect and two bird classes that intersect. Although a stimulus was held in common by these pairs of classes, the classes did not merge. Our experimental question could now be asked: Will the outcomes for overlapping equivalence classes established within the laboratory also be consistent with the intersection of the classes, rather than their merger? That is, will training provided only in experimental contexts with stimuli that participants have never seen immediately yield intersecting classes with characteristics like those involving the familiar stimuli? The answer to this question is crucial if one of the aims of the study of experimentally induced equivalence classes is to clarify acquisition processes that might produce readily the highly flexible repertoire seen in natural semantic categories.

A secondary, exploratory question in the present research asked whether the particular type of training given might affect the final outcome of the training. Thus, two different arrangements of the components of training were used with separate sets of arbitrary stimuli to reflect differing ways in which semantic categories may be established. The participants in experimental Condition A and its systematic replication with different stimuli and participants in Condition B, received linear series training, whereas the participants in experimental Condition C received many-to-one (comparison-as-node) training (Saunders & Green, 1999). These methods often differ in the effectiveness with which they yield equivalence classes (e.g., Arntzen & Holth, 2000). The interest here is whether outcomes like those illustrated with the familiar dog and bird stimuli will result from both training methods.

For example, both the linear (Figure 3, left panel) and comparison-as-node (right panel) methods potentially establish two overlapping, three-stimulus classes (e.g., A1, B1a, C1a and A1, B1b, C1b) like those in the DOG (or BIRD) category in Figure 1, but in different ways. As marked by the solid arrows in Figure 3, the linear method trains the four baseline relations, A1–B1a, B1a–C1a, A1–B1b, B1b–C1b; the comparison-as-node method trains the four baseline relations, A1–C1a, B1a–C1a, A1–C1b, B1b–C1b.

For both types of training, the broken arrows point to the C stimuli (e.g., C1b on a trial with B1a as sample) that could function as negative comparisons during BC training and
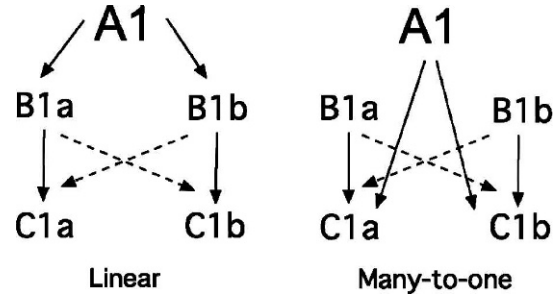

Fig. 3. Schematic outlines for linear and many-to-one training methods. Solid arrows point from sample to positive comparison stimuli. Broken arrows point toward negative comparison stimuli.

provide the bases for membership of each C stimulus in only one class, just as with the familiar dog and bird examples. The response on each multiple-selection test trial with a B stimulus as sample then would be selection of a single C stimulus and evidence of intersecting classes. The conditional discrimination training and test procedures (either method) in which each trial has only one correct comparison may favor that same outcome.

In contrast, the two methods of establishing the baseline prerequisites of the overlapping classes could generate other outcomes. The comparison-as-node training outlined in Figure 3 (right panel) provides one illustration and, notably, involves sources of stimulus control absent from linear training. The training of the AC and BC relations potentially results in the C stimuli serving as nodes that supply the basis for membership of the B stimuli in a single class with the A and both C stimuli. That would occur, for example, if BC training establishes only sample–S+, "select," relations leaving the C stimuli without the S− functions described above (i.e., the broken lines would be absent from Figure 3; see Johnson & Sidman, 1993). Such an outcome differs from that yielded by the familiar dog and bird stimuli. However, performance on the multiple-selection test would distinguish whether one large class or two separate intersecting classes had been established by the training. (Note that analogous considerations apply also in the case of linear series training and other possible comparison-as-node training procedures e.g., with B stimuli as comparisons. These other possibilities will not be discussed here.)

Catania's (1996) and Galizio's (1996) challenges cited earlier concerning the contingencies that operate in the natural environment provide the general rationale for the research described here. The outcomes and analyses will enrich the scope and impact of the stimulus equivalence approach.

METHOD

Participants

Sixteen experimentally naïve young adults were recruited. Most were current undergraduate students and received course credit following their participation. The others were volunteers who had a Bachelor's level of education and participated during work hours. Three different groups of participants received tasks with the stimuli for Conditions A (n = 5), B (n = 5), and C (n = 6) described below.

Apparatus

For Condition A an Imac (Apple ©) computer fitted with a MicroTouch GoldStar touch-sensitive screen (MicroTouch Systems Inc. ®) controlled stimulus presentations and other experimental events and recorded responses automatically. For Conditions B and C a McIntosh G3 computer was used instead and participants responded using a mouse.

Stimuli

The individual stimuli for Conditions A, B, and C are shown in Figure 4. Categories are identified using the same alphanumeric system as for Figure 1. For Condition A (trigram shared), the stimuli in Set A are the word-like printed trigrams SIG and BEZ (A1 and A2 respectively, the numbers suggesting potential memberships in two categories analogous to the categories dog and bird). Within each of the categories, the Set B stimuli were two graphic symbols (lexigrams) and the Set C stimuli were two colored, animal-like stimuli called Fribbles (Tarr, M. http://stims.cnbc.cmu.edu/Image%20Databases/TarrLab/). The lexigrams to be related by direct training to stimulus A1 will be called B1a (left leg of category 1) and B1b (right leg, category 1). The lower case letters reflect potential memberships in two pairs of equivalence classes to be established by training. Likewise, the two lexigrams related to A2 will be called B2a and B2b

(left and right legs respectively of category 2). Similarly, the Fribbles will be called C1a (left leg, category 1) and C1b (right leg, category 1), and C2a and C2b.

Though not evident in Figure 4, the body shape and color of the Fribbles were the same within each category but differed across the two categories (blue for those in the SIG category, red for the BEZ category) in ways that are analogous to the similarities and differences among the Set C (picture) stimuli in the familiar dog and bird categories. Figure 4 also illustrates components of lexigrams that were the same for the SIG and BEZ categories, yellow circles and triangles respectively (shown as gray). This is analogous to the use of blue print for only Set B of the familiar stimuli.

The stimuli for Conditions B and C are shown in the lower panel of Figure 4. The same Fribbles as in Condition A remained as the Set C stimuli. However, contrasting with Condition A, the Set A stimuli were two lexigrams rather than word-like trigrams, four of which now served as the Set B stimuli. Also, the Set B and Set C stimuli all were black on white backgrounds rather than partly colored. Each of these conditions will be referred to as lexigram-shared.

The number and roles of the lexigrams and trigrams used in training were changed in order to gain new data concerning the use of different types of stimuli at different points in the sequence of training tasks that establish the prerequisites for equivalence class formation. Whereas all the stimuli used here are unfamiliar, some research, which is directed primarily at assessing the role of naming in facilitating class formation (e.g., Holth & Arntzen, 1998; Mandell, 1997), has used mixtures of arbitrary unfamiliar stimuli and familiar stimuli.

General Procedure and Overview

For all standard matching-to-sample trials (training and test), the display consisted of five "keys," 5-cm by 5-cm squares on a white background without visibly marked boundaries. One key was centered on the computer screen and the others occupied the corners. The sample stimulus appeared in the center key. The participant had to make an observing response to the sample (i.e., touch it or use the mouse to position the cursor on the sample
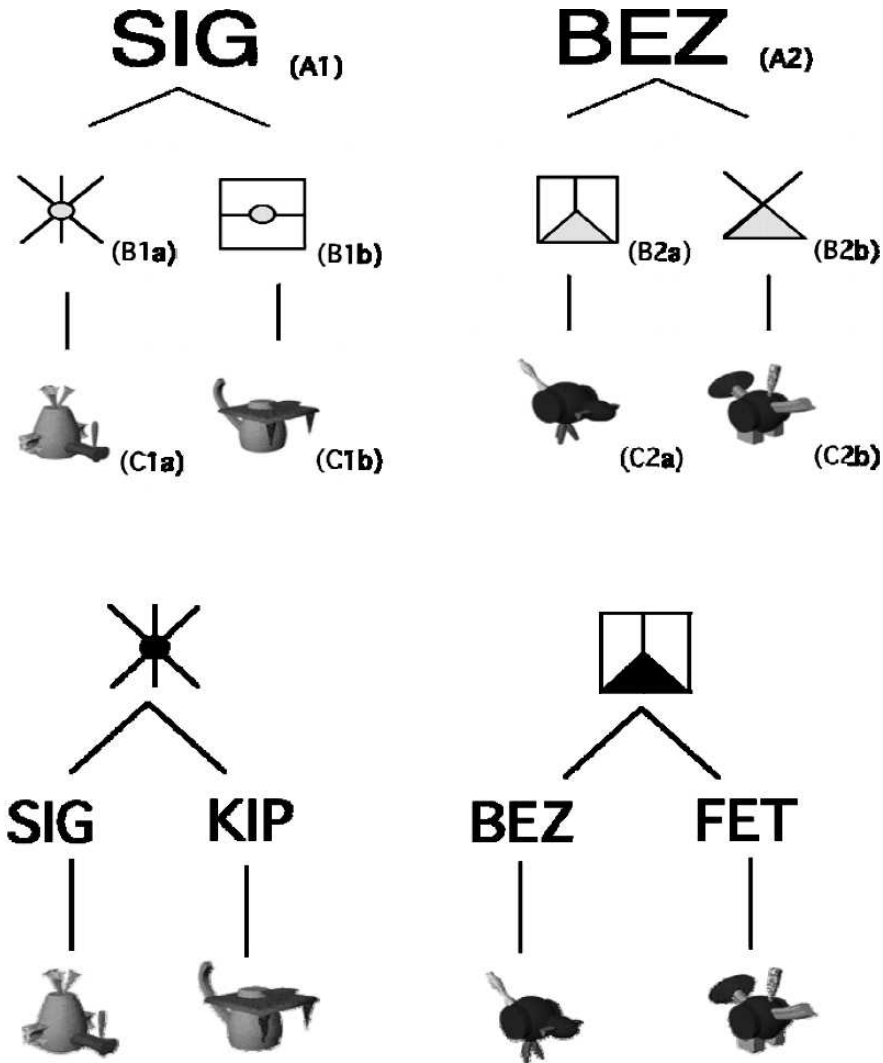
Fig. 4. Categories and stimuli used in Conditions A (upper panels), B, and C (lower panels). The trigrams for Condition A were black, the pictograms were black lines with yellow circles and triangles. The Fribbles were colored (See text for viewing location). The pictograms and trigrams for conditions B and C were black, as shown in the lower panels. Alphanumeric labels added only for stimuli shown in upper panels.

before pressing the button on the mouse (clicking) which resulted in the presentation of comparison stimuli on the corner keys. The participant then touched or used a mouse click to indicate selection of a comparison.

The multiple-selection matching-to-sample task was the new test procedure developed for this research. In it the sample appeared in a box at the top of the screen. After the participant responded to the sample, a pool of three comparison stimuli appeared on the bottom half of the screen, in a separate area surrounded by an outline. Selection of one, two, or all three of these stimuli then was permitted. For all procedures, the sample remained visible throughout the trial.

Each participant received standard matching-to-sample training, review, and testing tasks and then the multiple-selection task for the experimental conditions was assigned. On sessions that involved training and review tasks, feedback was scheduled to occur on all or on designated trials. Chimes sounded if the participant selected a correct comparison

Table 2

Order of training and testing phases in Conditions A and B.

Session characteristics and sequence	# of trials	% reinforced
Identity matching, set A and set B stimuli	20	100%
Identity matching, set C stimuli	24	100%
Train BC relations	36	100%
Review BC relations	16	100%
Criterion review	16	100%
CB symmetry probes in BC matching baseline	28	75%
Train AB relations	24	100%
Review AB relations	16	100%
BA symmetry probes in AB matching baseline	28	75%
Mixed AB and BC relations	16	100%
AC transitivity probes in AB, BC matching baseline	32	75%
CA equivalence probes in AB, BC matching baseline	32	75%

(defined by the particular task) on a trial in which reinforcement was scheduled. The 1.5-s intertrial interval then began. If no reinforcement was scheduled or if an incorrect selection occurred the intertrial interval began without chimes. Contingencies for testing sessions are described later.

All participants in all conditions began with identity matching-to-sample tasks (see Table 2), first using the stimuli in Sets A and B for a particular condition (20 trials), and then the stimuli in Set C (24 trials) for that same condition. For A and B stimuli, trials were two-choice matching-to-sample because there were only two possible comparisons for each set (S− stimuli were always drawn from the same stimulus set. For example, when the S+ on a trial was A1, then the S− was A2). Minimal verbal instructions (''Touch it'' or ''Click on it'') were given while the sample was present on the first of these trials. For the Set C stimuli, trials involved three comparisons. All correct matches performed during these segments were reinforced. Participants' accurate performances in these assessments showed that they discriminated the stimuli without difficulty. For this reason, further data will not be presented below.

Condition A: Training and Test Procedures

The participants in Condition A received training on the BC and then the AB relations[1] between the experimental stimuli presented (See Figure 4, upper panels). After completing this modified form of linear training (Saunders & Green, 1999) (note that the BC relations were trained before the AB relations), tests were given to assess all potential emergent relations (CB, symmetry; BA, symmetry; AC, transitivity; CA, symmetry as well as transitivity, often termed equivalence). Table 2 lists successive components of the teaching/ testing sequence and shows the type and minimum number of trials presented, and the reinforcement schedule. After completing the training and testing outlined in Table 2, participants received the multiple selection test.

BC training and CB tests. After identity matching assessment, training began for the four 1-to-1 BC relations (B1a–C1a, B1b–C1b, B2a–C2a, B2b–C2b). This training, which required no verbal instructions, was conducted across three blocks of trials within a single session. The first block included four trials. In each trial, a B sample appeared in the center of the screen. After the response to the sample occurred, the experimenter-designated S+ from Set C appeared in one of the four corners of the screen, in essence providing a ''single-choice'' matching task in which each sample and its corresponding correct comparison were paired. Each of the four B samples and its corresponding C comparison appeared on one trial of this initial block. In the second block of trials (16 trials), the B stimuli were samples and all four Set C stimuli appeared as

---

[1] Initially (Condition A), BC training was given on one day, AB training on the next. On the second day, participants were asked whether they wished to review (16 trials) the BC relations trained earlier. Three of the 5 opted for the review; the others did not and one of them completed all training and testing efficiently in one session. Subsequently, this efficient procedure was used for participants in Conditions B and C. No performance differences were found related to whether review of the BC relations was conducted.

comparisons on every trial, one in each corner of the computer screen. On each trial, the S+ was highlighted (prompted) by a black outline. Finally, in the third block of trials this outline prompt was removed, thus requiring the participant to select comparisons conditionally upon the sample. The block consisted of 16 trials, 4 with each of the B stimuli as sample. Two different C stimuli were the comparisons on all 16 trials. Eight trials involved presentation of both stimuli from the same category (e.g., C1a and C1b): On the other eight, one stimulus from each category was presented (e.g., C1a and C2b). If more than one error occurred in this last block, training began again with the initial block of single-choice trials. This part of training ended after a maximum of 60 trials. After accuracy of performance reached at least 90% on the final block of training, 16 trials were presented to review a mixture of all the BC relations (four trials for each). After accuracy of at least 90% occurred in the review, a probe session examined the emergence of CB matching, to demonstrate symmetry of the trained relations.

CB symmetry tests. In these tests, participants received 12 CB ''probe'' trials interspersed among 16 BC baseline trials. No reinforcer was scheduled for 8 of the 12 probe trials (the first three and five additional trials selected unsystematically). Reinforcers were available on the remaining four probes following selections consistent with the BA and CB relations. Reinforcers also were available on 12 of the 16 baseline trials; the four trials without reinforcement were selected unsystematically. The participant was informed of this change in reinforcement schedule at the beginning of the session by the following on-screen message:

> Hi there. This session will have a slightly different format. Up until now, the computer has ''chimed'' whenever you responded correctly. In this session, the computer will still ''chime'' every now and then, but not on every single correct trial. Some trials, the computer will just move along even if you responded correctly.

AB training and BA tests. Training of the AB relations was conducted in only two blocks of trials. The first block used a ''single-choice'' format like that used for BC teaching. After the participant completed the observing re-sponse to the A stimulus sample, the experimenter-designated S+ from Set B appeared in one of the four corners of the screen. The procedure sought to teach participants to match two of the B stimuli (B1a and B1b) to A1 as sample, and the other two (B2a and B2b) to A2 as sample. Each sample appeared on four trials on this initial block (total = eight trials), two for each AB relation.

Next, in the second phase of AB training, each A stimulus appeared as sample, with two comparisons from Set B. One comparison was the S+ (B1a or B1b, for sample A1, B2a or B2b for sample A2) and the other was an S− (B2a or B2b, for sample A1, B1a or B1b, for sample A2). This phase included a minimum of 16 trials and advanced to a review segment following accurate performance on 15 of the trials. As in the BC training, the program automatically scheduled return to the initial one-choice task following an error. A 16-trial review of the AB matching was conducted after the training was complete. During training and review, reinforcement followed all correct trials.

After the participant demonstrated at least 90% accuracy in the review, a probe session testing emergence of BA matching (symmetry of AB) was conducted. This probe session was like the CB symmetry probe session, using 12 BA trials interspersed among 16 trials of AB baseline, the same variable reinforcement schedule, and the same printed instructions concerning reinforcement at the beginning of the session.

Preparation for AC and CA tests. After completion of the BA symmetry testing, preparation was conducted to ready participants for the AC tests for transitivity and the CA tests for combined transitivity and symmetry (i.e., equivalence). This was done in a session in which trials of both AB and BC matching (eight trials each) were mixed together, thus establishing baseline within which AC and CA testing would be conducted.

AC and CA tests. Participants received two tests. In one, 16 AC probes were inserted into the mixed AB and BC matching-to-sample baseline and, in the other, 16 CA probes were inserted into the same baseline. In the AC test session, each of the sample stimuli (A1 and A2) was presented on eight trials on which the C-level stimuli served as comparisons in a carefully balanced fashion. For example, when

A1 was the sample, C1a and C1b served as S+ equally often (four trials). Furthermore, each of the C-level stimuli related to A2 served as S− on four trials, twice when C1a was S+ and twice when C1b was S+. These trials, and the comparable set for sample A2, were presented among the baseline trials in an unsystematic order. In the CA test session, each C stimulus was the sample equally often (four trials) so that the comparisons, A1 and A2, were S+ and S− equally often (eight trials). Reinforcement was available on the same schedule as for the CB probe sessions described earlier.

Multiple-selection matching-to-sample test. After participants completed the AC and CA tests, the 32-trial, multiple-selection task (see Appendix A) was presented. Recall that each trial had a single sample and three comparison stimuli and that the procedure allowed selection of one, two, or all three, comparison stimuli before touching DONE ended the trial (see Figure 2).

Before the multiple-selection task began the following printed instructions were presented on the computer screen and read by the experimenter:

> You will be presented with a single picture or written word (hence, sample stimulus) at the top of the screen. Use this stimulus as your model. After you've looked at this stimulus, click on it with your mouse. Three other pictures/written words will then appear at the bottom of the screen. Select all of the pictures that are EQUIVALENT to the sample stimulus. You can pick as many pictures/words as you think fit. If you make a mistake, select "START OVER". When you have selected as many stimuli as you would like, select "DONE".

Three times during the session, a subset of these instructions appeared on the screen as a reminder. Participants were assured that these instructions were programmed into the computer and did not reflect their performance. They also were instructed that no feedback would be provided after trials.

Condition B. The participants in Condition B were exposed to the same training and testing sequence as participants in Condition A (i.e., BC training and CB testing, then AB training and BA testing, followed by tests for emergent behavior) but with the stimuli depicted in the lower panels of Figure 4.

Condition C. The participants for Condition C were exposed to the same general proce-dures as those in Conditions A and B but the order of training and testing was changed. The stimuli for Condition C were the same as those for Condition B (lower panels, Figure 4) but the AC relations were trained first and the CA task then was used to assess symmetry. The BC relations were trained next and the CB task assessed symmetry before a mixture of the AC and BC tasks was scheduled. Finally, two sessions were devoted to tests for emergence of the AB and BA performances required to examine equivalence and a third involved the multiple-selection task. The numbers of trials given and the criteria used to evaluate perfor-mance were the same as before. This form of many-to-one or comparison–as-node training was used rather than the earlier modified linear training to help clarify whether such differences in methods of establishing equiva-lence classes might affect class merger or intersection.

RESULTS

Condition A: Training and Test Performances

Training was conducted initially over 3 days with the first 2 participants. Their performanc-es suggested that participants could complete the training/testing series more rapidly. Two new participants then took 2 days, and 1 took only a single session. The performance of each of the 5 participants quickly showed greater than 90% accuracy on each segment of the training that established the BC and AB relations.

Symmetry, transitivity, and equivalence tests. All participants demonstrated performances on symmetry and transitivity probes that were consistent with stimulus equivalence. The mean accuracy on BA, CB, AC, and CA probes was 98%, 100%, 95%, and 100%, respectively. Errors were rare on the baseline trials present-ed during these probe sessions.

Multiple-selection test performances. After dem-onstration in the matching-to-sample tasks that two equivalence classes had been established with each of the categories of stimuli illustrat-ed in the upper part of Figure 4, the multiple-selection task was conducted. The trials of particular interest are like those illustrated in Figure 2 and listed in Table 1. The pattern of performances shown by 4 of the 5 participants was identical to those produced with the familiar dog and bird stimuli. Immediately

after training these participants selected both C (Fribble) stimuli in the same category on trials when the sample was an A stimulus, SIG or BEZ (like the trials in the upper part of Table 1; eight opportunities), just as English speakers would select both dog pictures in the presence of the word DOG. The selection of members of both classes within the same trial was a new aspect of performance that occurred readily. In contrast, these participants' performances changed on trials when the different B stimuli (lexigrams) were samples (lower eight rows of Table 1). On these trials, they demonstrated the same kind of conditional control apparent in performances with the familiar stimuli. The Fribble that had been related to the sample by training was selected reliably (eight opportunities) whereas the Fribble in the other class (i.e., stimuli marked by * in Table 1) was not (eight opportunities). The performances of the remaining participant for all the trial types in the top eight rows of Table 1 were the same as for the other participants, both in this condition and with the familiar stimuli. However, selections on two trials early in the session did not show such consistency. On a trial with B1a as sample, both Set C comparisons were selected, the C1b selection being inconsistent with other aspects of performance. Additionally, on a trial with B2b as sample, C2a was selected, whereas C2b, the trained comparison, was not.

In general then, the A and B stimuli that were members of the same equivalence class functioned differently with respect to control of selections of the C stimuli. When the A stimuli that were common to both classes were samples, both C stimuli were appropriate selections based on the training of the BC and AB relations which gave rise to the AC relations. The new multiple-selection aspect of performance also derived from the training that established these relations and was enabled by the test instructions. In contrast, on trials when the B stimuli were samples the differential selections of only one of the C stimuli remained consistent with earlier BC training contingencies in which these stimuli functioned as positive and negative discriminative stimuli in the presence of the different samples during training. These observations demonstrate the contextual stimulus control that underlies application of the descriptive term *class intersection.* Moreover, they are

consistent with Sidman's (1994, p. 544) discussion suggesting that contextual control is required to resolve conflicts such as the present one which involves the contingencies that established the different discriminative functions of the C stimuli.

Clearly, the multiple-selection test provides a highly efficient method for assessing the stimulus control of performances generated under conditions that establish stimulus classes that contain a member in common and thus could merge to form a single large class or intersect and remain as separate classes. The results are highly unlikely to have occurred by chance. Recall that each of the 60 trials represented by Table 1 (12 trials by 5 participants) involved three conditional discriminations. For the trial types shown in the top four rows (like those at left in Figure 2), each participant selected all three possible stimuli on each of four trials, a total of 12 conditional selections, 8 of which were the selections of C stimuli listed in Table 1. (The remaining four selections were appropriate B stimuli.) The probability of that outcome being due to chance on a single trial is 0.125 ($p = 0.015$ for both trials of a particular type). Additionally, it is substantially less likely that the same outcome would occur by chance consistently for all 8 selections made across two sets of stimuli.

Furthermore, the same considerations arise for the trial types shown in the middle four rows of Table 1 (See examples at right in Figure 2.). Each of these four trials again involved a total of 12 conditional discriminations. However, each participant made two selections of stimuli (one C and one A related to the B sample) and one rejection (a C stimulus) per trial ($p = .015$ for both trials of each type). Similar performances were replicated by 4 participants for the four trials shown in the bottom four rows of Table 1; only 1 participant showed some inconsistency on these trial types.

Performance in the trials that examined the relations between B stimuli (e.g., B1a and B1b) and between the analogous C stimuli (not shown in Table 1) also were consistent with the preceding analysis and the results obtained with the familiar stimuli. In two multiple-selection trials, the samples were different B stimuli (one lexigram from each of the SIG and BEZ classes) and the stimuli presented for

selection included one identical to the sample and the B stimulus from the other equivalence class. On these trials, all participants responded only on the basis of identity and did not select the B stimulus from the other equivalence class. Four of the 5 participants also selected only on the basis of identity in two trials that provided opportunity to select both C stimuli.

During these sessions, performances on trials involving baseline tasks (identity matching, tasks trained directly, and their symmetrical counterparts) were virtually perfect across all 5 participants.

Condition B. Training and testing for 6 participants was completed in a single session lasting about 1 hr. Five showed greater than 90% accuracy in every segment of the training and testing. One participant was dropped after failing to show performance on CA trial types that was consistent with equivalence.

Symmetry, transitivity, and equivalence tests. For the 5 participants who demonstrated performances on probes that were consistent with stimulus equivalence mean accuracy on the CB, BA, AC, and CA probes was 98%, 100%, 99%, and 99%, respectively. Errors on the baseline trials presented during these probe sessions were rare.

Multiple-selection performances. The multiple-selection performances of all 5 of these participants closely resembled those produced using both the familiar and the Set A (trigram shared) stimuli. Both C stimuli always were selected when the sample was an A stimulus (top four trials listed in Table 1). In contrast, on both types of trials in which the B stimuli were samples (lower eight trials listed in Figure 1), the same kind of conditional control that was apparent with the familiar stimuli occurred reliably. The Fribble that had been related to the sample by training was selected whereas the Fribble in the other class (marked by * in Table 1) was not. Only one selection made by 1 participant was inconsistent with that description.

These results replicate closely those obtained with the Condition A (trigram shared) and the familiar stimuli, reducing much further the possibility that they were due to chance. The similarity in these outcomes again shows that performances on the multiple-selection test provide valid reflections of the difference between class intersection and class merger. In addition, it supports the conclusion that general outcomes were little affected by the procedural differences between the participants in Condition A that resulted from conducting training and testing within 1 day rather than across 2 days. Furthermore, the changes from Condition A in the color of stimuli and their particular roles during training made no difference in equivalence class formation. The data thus would not support the notion that critical or necessary functions (e.g., as prompts of common names for equivalence classes) may have been served by using word-like stimuli as potential superordinate stimuli held in common by the classes established in Condition A (cf., Arntzen, 2004).

Condition C. The AC and BC training and testing used for 5 of the 6 participants were completed in one session lasting about 1 hr. All 5 showed greater than 90% accuracy in every segment of training and testing, most performing virtually without error. However, 1 participant whose performance on transitivity tests did not permit the inference of that relational property was dropped.

Symmetry, transitivity, and equivalence tests. For the 5 participants who demonstrated performances on symmetry, transitivity, and combined tests that were consistent with stimulus equivalence, mean accuracy on the CA, CB, AB and BA probes was 99%, 99%, 100%, and 100%, respectively. Performance on the baseline trials presented during the probe sessions also was near perfect.

Multiple-selection performances. The multiple-selection test performances of all participants closely resembled those produced by the participants given the familiar stimuli. They also replicated systematically the results of Conditions A and B. With respect to the trials listed in Table 1, both C stimuli were selected when the sample was an A stimulus. In contrast, on both types of trials in which the B stimuli were samples, the conditional control apparent with the stimuli used in earlier Conditions again occurred reliably. Only one trial on which 1 participant did not select both C stimuli was inconsistent with that description. This replication indicates that the influence of chance on the outcomes is vanishingly small. Note that for this condition, the selections of C stimuli were a function of direct training: The AC relations did not

involve the property of transitivity as in Conditions A and B.

DISCUSSION

Overlapping classes of stimuli were established using laboratory procedures in order to produce and examine a class structure like that of simple, familiar categories established via extraexperimental experience (e.g., the dog and bird categories illustrated in Figure 1). The results suggest that the training, which was provided in three ways with sets of trigrams, lexigrams, and Fribbles, readily established paired classes of equivalent stimuli that had one stimulus in common just like the familiar categories. The differential effectiveness of linear series and many-to-one-training procedures that is sometimes reported (e.g., Arntzen & Holth, 2000) was not found. Also, performance was not a function of the different sets of stimuli used in Conditions A and B where the trigram presentations could have promoted, respectively, acquisition of common or separate names for the overlapping classes.

The multiple-selection task was developed to assess whether possession of a common member resulted in merger of two smaller classes into a single large class or, just as with the familiar stimuli, the conditional stimulus control that suggests class intersection (Sidman, 1994). Performances on the multiple-selection tests that involved the trigram, lexigram and Fribble stimuli were highly reliable, being replicated almost uniformly within and also across 15 participants regardless of training method. Each of these participants performed three conditional discriminations on each of 12 critical test trials by selecting or rejecting comparison stimuli under control of samples representing two different categories. The data suggest strongly that the relations among these experimental stimuli resembled the relations that exist among the stimuli in the familiar dog and bird classes shown in Figure 1. To illustrate, on trials involving presentation of the same four stimuli (one A, one B, and both Cs), when the A stimulus (e.g., a trigram analogous to DOG or BIRD) was the sample, participants selected both C stimuli (Fribbles analogous to dog and bird pictures respectively), thus suggesting that both were members of the same stimulus class in that context. In contrast, conditional control of the class membership of the C stimuli was demonstrated when the B stimuli (e.g., lexigrams or breed names) were samples. Under these conditions, only one C stimulus was selected. When the C stimuli were Fribbles, the only one selected had been related differentially to the current sample by training. For the familiar stimuli, the analogous breed name–picture relation reflected an unknown learning history that preceded experimental participation. These relations also were confirmed by performances on other multiple-selection trials involving the BC relations.

These data thus suggest that each of our procedural arrangements provided conditions in which stimuli from the same class exercised contextual control of changes in class membership. These changes occurred from one trial to the next as a function of the particular stimuli serving as samples. Notably too, after their use during training was changed (Condition A vs. Conditions B and C) the functions of the trigrams and lexigrams as contextual stimuli changed accordingly. In each case, the stimuli in the experimental classes functioned just like those in the familiar categories established extraexperimentally. The findings thus support the notion that the methods and procedures described here possess validity as means of studying categories just as proposed.

In accounting for the similarity of the outcomes obtained with the familiar stimuli and those of the three experimental conditions, it is likely that the characteristics of the stimuli play an important role. The general resemblance of the experimental stimuli to those in the world outside the laboratory may introduce influences that affect current behavior in the laboratory. Recall that all participants already were skilled language users and that the trigram stimuli have the formal characteristics of the familiar printed words: Indeed, participants referred to them in postexperimental interviews as ''words'' that had ''vowels'' in common. Furthermore, the Fribbles, which are designed to resemble animals (Williams, 1998), have physical characteristics (family resemblances) that may have occasioned the same treatment as the animal pictures for the familiar dog and bird categories. In general then, participants may have responded to the experimental stimuli in ways that reflected generalized repertoires devel-

oped during lengthy histories with similar stimuli in contexts outside the experimental situation. Systematic replications of the present research but with only arbitrary nonrepresentational stimuli would help to clarify the potential role played by the general resemblance of some of the current stimuli to familiar types of stimuli.

Little research has examined the influence that such general stimulus factors may have on performances involving equivalence classes. For some, the focus is on the effects of mixing stimuli that readily evoke names (e.g., pictures and line drawings of familiar objects, letter strings compatible with English) with arbitrary forms to examine the potential effects of naming (e.g., Bentall, Dickens & Fox, 1993; Mandell, 1997; Mandell & Sheen, 1994). Other experimental analyses concern the place during training that familiar stimuli are introduced (e.g., Arntzen, 2004; Holth & Arntzen, 1998). Practical and clinical as well as theoretical interests give direction to additional research with stimuli that evoke potentially disruptive emotional responses derived from cultural (e.g., McGlinchey & Keenan, 1997; Plaud, 1995; Plaud, Gaither, Franklin, Weller, & Barth, 1998) and personal (e.g., Keenan, McGlinchey, Fairhurst, & Dillenberger, 2000; Leslie, Tierney, Robinson, Keenan, & Watt, 1993) aspects of individuals' histories.

While the preceding suggests the relevance of existing generalized repertoires for research on equivalence, the restricted size of the classes used in the current research (e.g., consisting of the three stimuli A1, B1a, and C1a) may be considered a limitation with regard to its importance for language and symbolic behavior. Harnad (1996), for example, commented on such restrictions in experiments on stimulus equivalence, referring to classes of few stimuli as specific, arbitrary ''associative clusters'' (p. 263) rather than broad abstractions of ''invariant properties of stimuli'' or ''perceptual relations'' that he views as the bases of natural language categories. However, as Galizio, Stewart and Pilgrim (2004, p. 254) state, ''It seems evident that equivalence classes would include not merely the trained stimuli, but through stimulus generalization would also include physically similar stimuli.'' Experimental demonstrations and analyses of generalized equivalence classes are described in reports of research by Fields and his colleagues (e.g., Fields, Adams, Brown, & Verhave, 1993; Fields & Reeve, 2000, 2001; Fields et al., 1997). Beyond that too, conditional control over the range of stimuli included in generalized equivalence classes has been produced and examined by others (Griffee & Dougher, 2002).

The research of Galizio et al. (2004) also examined another factor, typicality, which also may influence class size. In this respect, class membership may be based on multiple features of the stimuli while no given feature is necessary for membership. However, class members are regarded as more typical (even as prototypes) if they have more features in common than do less typical members of the class. The Fribbles used in the current research provide relevant examples. They were selected from ''species'' (Williams, 1998) with multiple exemplars constructed (with various combinations of body shape and color, ''head'', ''feet'', and ''ears'') to examine the role of such factors in categorization and object identification. Additional investigation thus would provide not only follow-up to the current research with predictably larger classes of stimuli of the kind sought by Harnad (1996) but also bring a behavior-analytic perspective together with the more cognitive perspective of investigators exemplified by Williams (1998).

The results of the tests showing the resemblances between the relations produced among the current experimental stimuli and the relations among members of the familiar hierarchical dog and bird categories suggest the relevance of the procedures described here to the production and analysis of such categories. In that respect, the illustrations in Figure 1 highlight the importance of two sets of relations between category members. One set involves relations between stimuli at the same level (e.g., between the words RETRIEVER and DALMATIAN or between their corresponding pictures). The other involves relations between the members of subordinate and superordinate levels of a category (e.g., relations between RETRIEVER or DALMATIAN and DOG exemplifying the B and A sets of stimuli respectively in the current research).

First, with regard to relations between members at the same level of a hierarchy, equivalence may be demonstrable under some circumstances but not others. As suggested earlier for example, retrievers and dalmatians may be

demonstrably equivalent to one another in the context of the printed word DOG or, of course, in other contexts such as ''animals that bark.'' On the other hand, these exemplars are not equivalent in the context established by the instruction ''Point to the dalmatian'' or other contexts such as ''which one has spots?'' In discussing such intralevel relations among category members, cognitive-developmental investigators (e.g., Waxman, 1991) suggest that a ''contrastive principle … maintains horizontal relations among classes at a given level of abstraction'' (p.110). Thus, a member of the dalmatian class cannot also be a member of the retriever class and the Fribbles shown in the upper part of Figure 4 are members of two mutually exclusive sets that belong to the SIG and BEZ classes, respectively. The current research, however, suggests an account of the contrastive memberships of the stimuli that provides an alternative to Waxman's (1991) appeal to a ''principle'' derived from logic (cf. Sidman, 1990). The relations among the stimuli in the research described here were based on the specific contingencies used in training the BC and AB (or AC) relations involved. These contingencies built the units of behavior for the integrated sets of conditional discriminations that form the overlapping pairs of classes for the categories illustrated. These contingencies established the functions of the Set C Fribble stimuli in the presence of stimuli from Sets A and B. More extensive, detailed functional analyses with categories that are larger than those used here are needed to elucidate the more general discriminative repertoires that Waxman (1991) terms a ''contrastive principle''.

With regard to relations among members of classes at different hierarchical levels, our methods and procedures suggest new considerations. The training conditions described here established the prerequisites for mutual substitutability of stimuli from Sets A, B, and C, which may be viewed as exemplars from different levels of a hierarchy. Further, the relations among these stimuli were shown to satisfy the three requirements for equivalence (reflexivity, symmetry, transitivity) as described by Sidman and Tailby (1982). Research in the area of cognitive development also has led investigators (e.g., Waxman, 1991) to comment upon transitive, (but not reflexive or symmetrical), relations between members at different levels of hierarchical categories. In general, one might say that if the members of Class Z are included in Class Y, and if the members of Class Y are included in Class X, then the members of Class Z are included in Class X. Such a statement with only slight modification would describe the outcomes of training for the present Conditions A, B, and C. The current research methods, then, are relevant to the study of class inclusion phenomena illustrated by statements like ''All retrievers are dogs but not all dogs are retrievers'' or ''All dogs are animals but not all animals are dogs'', the former virtually describing the multiple-selection test outcomes presented here and the latter suggesting extensions of the current research with additional stimuli.

While the preceding discussion reflects the use of the dog and bird stimuli as familiar categories it is important to note that the training used does not necessarily yield hierarchical relations and may produce other interesting forms of classification. For example, consider the following stimuli: The printed word BOW, the corresponding weapon, a picture of a similar weapon, a decorative knot with loops, and a picture of a similar knot. These stimuli clearly form two distinct classes, arguably equivalence classes, each including the common stimulus, BOW. Further, suppose that the training described here is used to establish the set of interrelated conditional discriminations resembling those in Condition A. That repertoire is likely to yield the same set of outcomes that were produced in Condition A. However, for a typical learner of English, the relations among the stimuli, trained and emergent, would not exemplify a hierarchy but instead two intersecting equivalence classes, each containing the same printed stimulus, BOW. In that regard, it may be important that the members of the respective dog and bird classes shown in Figure 1 share several features (e.g., General Shape, head shape etc. of dogs and birds respectively, cf. Rosch & Mervis, 1975) whereas such family resemblances are not shared by the stimuli in the BOW classes suggested. The present work may help point the way toward further research regarding such verbal classes and categories (cf. Mandell, 1997).

While the training described here yielded highly reliable immediate effects with the adult participants of the current research the procedures also will be useful in other respects with

other populations. For example, Wilkinson and Rosenquist (2006) suggest that the methods described here may be used to evaluate the semantic relations and structure of categories acquired by individuals with limited verbal skills. To illustrate, a child with autism might relate the word dog, written or dictated, to only a specific single dog (or picture). The multiple-selection test would allow presentation of probes to assess whether the specificity of performance reflected narrower than typical boundaries of the class dog or perhaps a difficulty with expressive vocabulary (e.g., the child selects more than one dog exemplar in matching tasks but uses the word dog to label just one). The results of the multiple-selection test thus may have direct implications for assessment of clients' appropriateness for alternative and augmentative communication training and the adequacy of particular prerequisite skills.

REFERENCES

Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record*, *54*, 275–291.

Arntzen, E., & Holth, P. (2000). Equivalence outcome in single subjects as a function of training structure. *The Psychological Record*, *50*, 603–608.

Bentall, R. P., Dickens, 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.

Branch, M. N. (1994). Stimulus generalization, stimulus equivalence, and response hierarchies. In S. C. Hayes, L. J. Hayes, M. Sato, & K. Ono (Eds.). *Behavior analyses of language and cognition* (pp. 51–70). Reno, NV: Context Press.

Bush, K. M., Sidman, M., & deRose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior*, *51*, 29–45.

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

Cowley, B. J., Green, G., & Braunling-McMorrow, D. (1992). Using stimulus equivalence procedures to teach name–face matching to adults with brain injuries. *Journal of Applied Behavior Analysis*, *25*, 461–475.

Critchfield, T., Galizio, M., & Zentall, T. (Eds.) (2002), Categorization and concept learning [Special Issue]. *Journal of the Experimental Analysis of Behavior*, *78*(3)

Cumming, W. W., & Berryman, R. (1965). The complex discriminated operant: Studies of Matching-to-sample and related problems. In D. Mostofsky (Ed.). *Stimulus generalization*. Stanford University Press, Stanford, CA.

Fields, L., Adams, B. J., Brown, J. L., & Verhave, T. (1993). The generalization of emergent relations in equivalence classes: Stimulus substitutability. *The Psychological Record*, *43*, 235–254.

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

Fields, L., & Reeve, F. K. (2000). Synthesizing equivalence classes and natural categories from perceptual and relational classes. In J. C. Leslie & D. Blackman (Eds.). *Experimental and applied analysis of human behavior* (pp. 59–84). Reno, NV: Context Press.

Fields, L., & Reeve, K. F. (2001). A methodological integration of generalized equivalence classes, natural categories, and crossmodal perception. *The Psychological Record*, *51*, 67–87.

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

Fields, L., Reeve, K. F., Adams, B. J., & Verhave, T. (1991). The generalization of equivalence relations: A model for natural categories. *Journal of the Experimental Analysis of Behavior*, *55*, 303–312.

Fucini, A. (1982). *Stimulus control of class membership*. Unpublished doctoral dissertation, Northeastern University, Boston, MA.

Galizio, M. (1996). Methodological issues in the study of naming. *Journal of the Experimental Analysis of Behavior*, *65*, 286–288.

Galizio, M., Stewart, K. L., & Pilgrim, C. (2004). Typicality effects in contingency-shaped generalized equivalence classes. *Journal of the Experimental Analysis of Behavior*, *82*, 253–273.

Gatch, M. B., & Osborne, J. G. (1989). Transfer of contextual stimulus function via equivalence class formation. *Journal of the Experimental Analysis of Behavior*, *51*, 369–378.

Griffee, K., & Dougher, M. J. (2002). Contextual control of stimulus generalization and stimulus equivalence in hierarchical categorization. *Journal of the Experimental Analysis of Behavior*, *78*, 433–448.

Harnad, S. (1996). Experimental analysis of naming behavior cannot explain naming capacity. *Journal of the Experimental Analysis of Behavior*, *65*, 262–264.

Hayes, S. C., Kohlenberg, B. S., & Hayes, L. J. (1991). The transfer of specific and general consequential functions through simple and conditional relations. *Journal of the Experimental Analysis of Behavior*, *56*, 119–137.

Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record*, *48*, 81–110.

Johnson, C., & Sidman, M. (1993). Conditional discrimination and equivalence relations: control by negative stimuli. *Journal of the Experimental Analysis of Behavior*, *59*, 333–347.

Keenan, M., McGlinchey, A., Fairhurst, C., & Dillenberger, K. (2000). Accuracy of disclosure and contextual control in child abuse: Procedures within the stimulus equivalence paradigm. *Behavior and Social Issues*, *10*, 1–17.

Kennedy, C. H., & Laitinen, R. (1988). Second order conditional control of symmetric and transitive stimulus relations: The influence of order effects. *The Psychological Record*, *38*, 437–446.

Lane, S. D., Clow, J. K., Innis, A., & Critchfield, T. S. (1998). Generalization of cross-modal stimulus equivalence classes: Operant processes as components in human category formation. *Journal of the Experimental Analysis of Behavior*, *70*, 267–277.

Leslie, J. C., Tierney, K. J., Robinson, C. P., Keenan, M., & Watt, A. (1993). Differences between clinically anxious and non-anxious subjects in a stimulus equivalence training task involving threat words. *The Psychological Record*, *43*, 153–161.

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

Lynch, D. C., & Green, G. (1991). Development and crossmodal transfer of contextual control of emergent stimulus relations. *Journal of the Experimental Analysis of Behavior*, *56*, 139–154.

Mackay, H. A., & Sidman, M. (1984). Teaching new behavior via equivalence relations. In P. H. Brooks, R. Sperber, & C. McCauley (Eds.). *Learning and cognition in the mentally retarded*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mandell, C. (1997). Stimulus equivalence and meaning: The influence of verbal behavior. In C. Mandell, & A. McCabe (Eds.). *The problem of meaning: Behavioral and cognitive perspectives*. Amsterdam: Elsevier Science.

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

McGlinchey, A., & Keenan, M. (1997). Stimulus equivalence and social categorization in Northern Ireland. *Behavior and Social Issues*, *7*, 113–128.

Nelson, K. (1985). *Making sense: The acquisition of shared meaning*. Orlando, FL: Academic Press.

Pilgrim, C., & Galizio, M. (1996). A class of correlations or a correlation of classes. In T. R. Zentall, & P. M. Smeets (Eds.). *Stimulus class formation in humans and animals* (pp. 173–197). Amsterdam: Elsevier.

Plaud, J. J. (1995). The formation of stimulus equivalences: Fear-relevant versus fear-irrelevant stimulus classes. *The Psychological Record*, *45*, 207–222.

Plaud, J. J., Gaither, G. A., Franklin, M., Weller, L. A., & Barth, J. (1998). The effects of sexually explicit words on the formation of stimulus equivalence classes. *The Psychological Record*, *48*, 63–79.

Rehfeldt, R. A. (2003). Establishing contextual control over generalized equivalence relations. *The Psychological Record*, *53*, 415–428.

Rosch, E. H., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

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

Saunders, R. R., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1988). The merger and development of equivalence classes by unreinforced conditional selection of comparison stimuli. *Journal of the Experimental Analysis of Behavior*, *50*, 145–162.

Serna, R. W. (1987). An investigation of the five-term contingency and the conditional control of equivalence relations. *Dissertation Abstracts International*, *48*, 2126B. (Orer No. DA8719451).

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

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thomson, & M. D. Zeiler (Eds.). *Analysis and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Erlbaum.

Sidman, M. (1990). Equivalence relations: Where do they come from? In D. E. Blackman, & H. Lejeune (Eds.). *Behavior analysis in theory and practice: Contributions and controversies* (pp. 93–114). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sidman, M. (1994). *Equivalence relations: 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.

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

Stemmer, N. (1990). Skinner's Verbal Behavior: Chomski's review, and mentalism. *Journal of the Experimental Analysis of Behavior*, *54*, 307–315.

Tarr, M. Retrieved 9/20/2010. *http://stims.cnbc.cmu.edu/ Image%20Databases/TarrLab/)*.

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

Waxman, S. R. (1991). Convergences between semantic and conceptual organization in the preschool years. In S. A. Gelman, & J. P. Byrnes (Eds.). *Perspectives on language and thought: Interrelations in development* (pp. 107–145). Cambridge: Cambridge University Press.

Wilkinson, K. M., & McIlvane, W. J. (2001). Methods of studying symbolic behavior and category formation: Contributions of stimulus equivalence research. *Developmental Review*, *21*, 355–374.

Wilkinson, K. M., & Rosenquist, C. (2006). Demonstration of a method for assessing semantic organization and category membership in individuals with autism spectrum disorders and receptive vocabulary limitations. *Augmentative and Alternative Communication*, *22*, 242–257.

Williams, P. (1998). Representational organization of multiple exemplars of object categories. Retrieved 8/21/2010, citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.5.8336.pdf

Wulfert, E., & Hayes, S. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior*, *50*, 125–144.

APPENDIX

List of stimuli presented on all trials of multiple-selection test.

Trial	Sample	Comparison stimuli			Trial	Sample	Comparison stimuli		
1	C1b	C2a	C1b	C2b	2	A2	A1	A2	B1b
3	A2	C2b	B2b	C1a	4	B1a	C1b	C1a	B1a
5	A1	A1	B2b	A2	6	C1a	C2b	C1a	C2a
7	A2	C2a	B1b	B2b	8	A1	C1a	B1b	C1b
9	B2a	C2b	C2a	A2	10	B2b	C2b	C2a	B2b
11	A1	B1b	C1a	B2b	12	A2	B2b	C2a	C2b
13	C2b	C1a	C1b	C2b	14	B1b	C1b	A1	C1a
15	A1	B2a	C1b	B1a	16	A2	C1b	B2a	C2a
17	A1	B1a	C1a	C2a	18	B1a	C1a	B1a	B2b
19	B2b	A2	C2a	C2b	20	A2	A1	A2	B1b
21	B1b	B1a	B2a	B1b	22	B1b	C1a	B1b	C1b
23	B2a	C2a	B2a	C2b	24	A1	C1a	B1a	C1b
25	A2	B1a	C2b	B2a	26	A1	C1b	B1b	C2b
27	B2b	B1a	B2a	B2b	28	B1a	C1a	A1	C1b
29	A1	A1	A2	B2a	30	B2a	B2a	B1b	B1a
31	A2	C2b	C2a	B2a	32	C1a	C1b	C1a	C2a