JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1998, **70,** 267–279 NUMBER 3 (NOVEMBER)

GENERALIZATION OF CROSS-MODAL STIMULUS EQUIVALENCE CLASSES: OPERANT PROCESSES AS COMPONENTS IN HUMAN CATEGORY FORMATION

SCOTT D. LANE, JULIE K. CLOW, ANDREW INNIS, AND THOMAS S. CRITCHFIELD

UNIVERSITY OF TEXAS HEALTH SCIENCE CENTER–HOUSTON AND AUBURN UNIVERSITY

This study employed a stimulus-class rating procedure to explore whether stimulus equivalence and stimulus generalization can combine to promote the formation of open-ended categories incorporating cross-modal stimuli. A pretest of simple auditory discrimination indicated that subjects (college students) could discriminate among a range of tones used in the main study. Before beginning the main study, 10 subjects learned to use a rating procedure for categorizing sets of stimuli as class consistent or class inconsistent. After completing conditional discrimination training with new stimuli (shapes and tones), the subjects demonstrated the formation of cross-modal equivalence classes. Subsequently, the class-inclusion rating procedure was reinstituted, this time with cross-modal sets of stimuli drawn from the equivalence classes. On some occasions, the tones of the equivalence classes were replaced by novel tones. The probability that these novel sets would be rated as class consistent was generally a function of the auditory distance between the novel tone and the tone that was explicitly included in the equivalence class. These data extend prior work on generalization of equivalence classes, and support the role of operant processes in human category formation.

Key words: stimulus equivalence, stimulus generalization, categorization, rating procedure, computer mouse click, college students

Categorization is of interest to researchers in a variety of fields, including animal cognition (Herrnstein, 1984; Wasserman, Kiedinger, & Bhatt, 1988; Wright, 1992), human cognition (Smith, 1995), neuroscience (Carpenter, Grossberg, & Reynolds, 1991), and the experimental analysis of human behavior (Fields, Reeve, Adams, & Verhave, 1991; Saunders & Green, 1992; Sidman, 1994). Definitions of the term *category* vary somewhat across research domains. In studies of stimulus equivalence, the term may refer to a phenomenon in which each stimulus in a given group evokes a specified response (verbal or nonverbal) (Sidman, 1994, pp. 416–417). For example, human categories often consist of written words, spoken words, and their referent objects (real or pictorial). These can be established via matching-to-sample (MTS) training and equivalence class formation (Stromer & Mackay, 1992). However, some categories also exist as a result of generalization across stimuli with common attributes (Herrnstein, 1990). Differing descriptions of what constitutes a category may reflect the possibility that there exists more than one type.

A variety of models attempt to explain how organisms come to group stimuli into meaningful categories. For example, structural accounts emphasize the physical properties of the stimuli (Harnad, 1987; Tversky, 1977); connectionist accounts address the networks of neural connections that conjoin when organisms categorize stimuli (Carpenter & Grossberg, 1990); and functional accounts, such as those derived from the experimental analysis of behavior, focus on the consequences for differentially associating stimuli (Fields, Reeve, Adams, Brown, & Verhave, 1997; Fields et al., 1991; Herrnstein, 1990; Vaughan, 1988; Wasserman & Bhatt, 1992). Most models, however, imply that categorization has both structural and functional properties.

Stimulus equivalence is an important component in functional accounts of categoriza-

This research was conceived and initially conducted while the first author was at Auburn University. Preparation of this manuscript was supported in part by a fellowship grant from the National Institutes of Health (DA07247) and the University of Texas Health Science Center at Houston. We thank Lanny Fields, William Dube, and William McIlvane for helpful suggestions regarding the manuscript.

Correspondence and reprint requests should be addressed to the first author at The University of Texas Health Science Center at Houston, Department of Psychiatry and Behavioral Sciences, 1300 Moursund, Houston, Texas 77030 (E-mail: slane@msi13.msi.uth.tmc.edu).

tion in humans. In a typical stimulus equivalence procedure, stimuli from a finite training set become interchangeable members of a class, although only a subset of relations between those stimuli are directly trained (Saunders & Green, 1992; Sidman & Tailby, 1982). Those not directly trained are described as emergent relations. It is important to note that equivalence classes are created by differential consequences for selecting stimuli conditional on the presence of other stimuli, rather than physical similarities between those stimuli (Sidman, 1994).

It is widely assumed that many naturally occurring categories are open-ended, in that they incorporate a potentially infinite number of nonidentical exemplars (Fields et al., 1991, 1997; Herrnstein, 1984). Any plausible account of categorization, therefore, must explain the reliable categorization of novel instances. Thus, stimulus equivalence alone probably cannot provide a comprehensive account of category formation because equivalence classes, as traditionally defined, encompass a finite set of stimuli with the requisite shared history of conditional discrimination (e.g., see Harnad, 1996).

Fields and colleagues have argued that equivalence, in combination with stimulus generalization, possibly along multiple stimulus dimensions, might account for category inclusion of novel exemplars (e.g., Fields et al., 1991, 1997). At least three studies have provided support for this view (Adams, Fields, & Verhave, 1993; Fields et al., 1991, 1997). These studies began by establishing two equivalence classes, each composed of nonsense three-letter words, and either a short line (Class 1) or a long line (Class 2). Subsequent generalization tests were conducted in the context of MTS trials involving the same nonsense words, plus lines that were both longer and shorter than those used in training. Class inclusion followed an orderly generalization gradient, in that inclusion of the novel stimuli into Class 1 or Class 2 was a function of the degree of similarity (line length) to the trained stimulus (Adams et al., 1993; Fields et al., 1991, 1997). The authors termed this outcome a *generalized equivalence class* because class membership was not restricted to the stimuli used in training. The generalized equivalence class provides a closer union with structuralist models (Harnad, 1987, 1996; Smith, 1995) than would be possible based on stimulus equivalence alone.

The concept of the generalized equivalence class is useful because it supports the role of operant principles in category formation. However, naturally occurring categories are complex psychological phenomena, and existing studies in support of generalized equivalence classes are limited in number and scope. One source of complexity is that natural categories tend to be cross-modal. A boy's concept (category) of dog, for example, may encompass written and spoken versions of the word *dog* (or *chien* if he is bilingual), pictures of dogs, the sounds of dogs barking, the feel of a dog's fur, and so forth. Although conventional equivalence classes can be cross-modal (Bush, 1993), it is not known whether generalized equivalence classes can extend across different stimulus modalities.

In the present study, we investigated generalized equivalence in cross-modal stimulus classes. In addition, we sought to evaluate the inclusion of novel stimuli into existing classes differently than in previous studies. In the studies of Fields et al. (1991, 1997), Fields, Adams, Brown, and Verhave (1993), and Adams et al. (1993), generalized equivalence classes were measured through subjects' selections on MTS tests that presented novel stimuli. Consistent selection of a novel comparison stimulus in the presence of a given sample stimulus indicated inclusion of that stimulus in the same class as the sample. The critical test trials were thus presented in a context (MTS) in which (a) there was a history of a correct match on every trial, and (b) the correct match was always a single comparison stimulus. In the present study we sought to broaden that context, and in doing so capture more of the complexity of naturally occurring categories.

The present approach was based loosely on one used in a previous study of self-reports about emergent relations (Lane & Critchfield, 1996). A critical feature of that procedure was preliminary training in which contingencies of point reinforcement and punishment were used to calibrate self-reports and subsequent confidence ratings (e.g., see Critchfield, Tucker, & Vuchinich, in press; Saunders & Green, 1996) with respect to well-defined events such as selections on

MTS trials. In the present study, subjects who had demonstrated the formation of equivalence classes by conventional means used a previously trained rating procedure to identify groups of stimuli as class consistent or class inconsistent. These groups sometimes included a probe stimulus that was physically similar to another stimulus in the same class, thus providing an opportunity to measure generalization. Results analogous to those of Fields and colleagues would support previously reported effects and expand the context in which they can occur.

METHOD

Subjects

Ten undergraduates without prior experience in conditional discrimination experiments completed an experiment described in the informed consent agreement as focusing on ''reasoning in symbolic logic.'' One additional volunteer was excluded after his baseline performances suggested preexisting knowledge of class structure. Subjects participated for 8 to 10 hr, over a period of 4 to 8 days. Each visit to the laboratory lasted 1.5 to 2 hr. All subjects were students enrolled in an undergraduate psychology course. Participating in the study was one of several ways to fulfill a project requirement worth 50 of 575 total course points. In the present case, students were told that 10 hr of participation, plus attendance at a 1-hr group debriefing session at the end of the academic term, could earn full credit for the project. The informed consent agreement indicated that volunteers would gain or lose participation time, in seconds, depending on responses made during the study. The experimental protocol actually required that subjects receive credit for all time spent in the study (10 hr maximum), regardless of performance. Subjects were debriefed immediately if the number of seconds accumulated at the end of the study fell short of the amount of time spent participating, but all subjects earned the maximum possible extra credit.

Apparatus

Each subject worked alone in a small room containing a table, a chair, a video monitor, a mouse with the left button marked by a red sticker, and a speaker placed on the table next to the monitor. They performed the MTS task by using the mouse to move a cursor to appropriate locations on the video screen. An IBM-compatible microcomputer in an adjacent room, equipped with a SoundBlaster® sound card, ran custom programs written in MicroSoft QuickBasic® to control experimental procedures, present stimuli, and collect data.

Procedure

Each subject completed eight phases. Table 1 outlines these phases, showing the number of trials per session, number of sessions, and mastery criterion for each phase. Phase 1 employed a simple discrete-trials auditory discrimination procedure to estimate the extent to which tones used in later phases could be distinguished from one another. Phase 2 employed different stimuli from the main experiment, and provided training in the use of a class-inclusion rating procedure that, when reinstated in later phases, provided the primary data for the study. Using this same procedure, Phase 3 provided a baseline assessment of class-inclusion ratings made about the stimuli used in the main experiment prior to their involvement in conditional discrimination training or equivalence testing. Phase 4 provided conditional discrimination training prerequisite to the formation of two three-member equivalence classes, each composed of one auditory and two visual stimuli. Phase 5 tested for the emergence of the equivalence classes predicted from the Phase 4 training.

Phase 6 used the class-inclusion rating procedure to test for generalization of the emergent relations to physically similar stimuli. The novel stimuli used in this procedure varied in terms of their similarity to the auditory member of the previously established equivalence classes. In Phase 7, the emergent relations tests of Phase 5 were repeated to ensure that the stimulus classes remained intact. In Phase 8, the auditory discrimination procedures of Phase 1 were repeated to ensure that the prior phases had not globally disrupted subjects' discrimination of the tones.

Phase 1: Auditory discrimination pretest. Each trial of a successive discrimination task began with the message, ''Are the following two tones the same or different?'' printed on the screen, followed, 0.5 s later, by two 2-s tones

Table 1

Summary of experimental phases. See text, Table 2, and Figure 2 for descriptions of the stimuli.

Phase	Description	Probability of feedback	Class-inclusion ratings?	Trials per session	Mastery criteria sessions at % correct
1	Auditory discrimination test	0	No	243	1 > 85%
2	Preliminary training:				
2a	MTS training (A-B)	1.0	No	12	2 at 100%
	(A-C)	1.0	No	12	2 at 100%
	(A-B, A-C)	1.0	No	24	2 > 95%
2b	Equivalence tests	0	No	26	2 > 95%
2c	Class-inclusion test	1.0	Yes	84	2 (ave > 80%)
3	Baseline class-inclusion test	0	Yes	136	1
4	MTS training (A-B)	1.0	No	12	2 > 91%
	(A-C)	1.0	No	12	2 > 91%
	(A-B, A-C)	1.0	No	24	2 > 95%
	(A-B, A-C)	.4	No	24	2 > 95%
5	Equivalence tests	0	No	66	2 > 91%
6	Class-inclusion test	0	Yes	136	3 or 4[a]
7	Equivalence test	0	No	66	1[a]
8	Auditory discrimination test	0	No	243	1[a]

[a] No mastery criterion used.

separated by a 1-s pause. At the end of the second tone, two horizontally aligned boxes, labeled ''SAME'' and ''DIFFERENT,'' appeared in the middle of the screen. Clicking inside one of these boxes completed the trial. Subjects completed one 243-trial session incorporating three iterations of each of the 81 possible pairings of nine different tones. The tones were identical on 27 (12.5%) of the trials.

The tones, depicted in Table 2, were nine musical notes, each spaced approximately one half octave apart along the musical scale. Two of these tones were used as MTS training and equivalence test stimuli in Phases 4 and 5, and the other seven stimuli were used for the equivalence generalization tests in Phase 6.

Table 2

Summary of the auditory stimuli. The distance between two tones was approximately one half octave, that is, a fourth or fifth interval based on a diatonic scale (white piano keys only). Tone 3 (the A1 stimulus in the main phases of the experiment) was a middle C. Tone 7 (the A2 stimulus) was a C two octaves below middle C. Hertz values are based on standard note-to-frequency conversions rather than measurement of the actual tones used in the study.

Tone	Musical note	Hertz
1	C	132
2	G	198
3 (A1)	Middle C	264
4	F	352
5	C	528
6	G	792
7 (A2)	C	1056
8	F	1408
9	C	2112

Phase 2: Preliminary training of the class-inclusion rating procedure. This phase was designed to teach the rating procedure that would be used in later phases. Subjects first completed conditional discrimination training with two three-member classes of stimuli consisting of the capital letters A, B, and C and the numerals 1, 2, and 3, all printed on the screen in green. The stimuli were chosen to facilitate rapid mastery and high levels of accuracy throughout training and testing. Stimulus arrangements for the trained and tested relations are shown in Appendix A, and are analogous to those in Phases 4 and 5, the training and equivalence tests of the main part of the study.

When mastery on the conditional relations had been demonstrated, the class-inclusion rating procedure was introduced with the following on-screen instructions:

> During this session you will be presented with groups of figures. After you view the figures, you will have a chance to EARN or LOSE seconds by moving the cursor into one of the boxes and pressing the red mouse button. Each press will be counted in a seconds-earned box above the figures. The number of

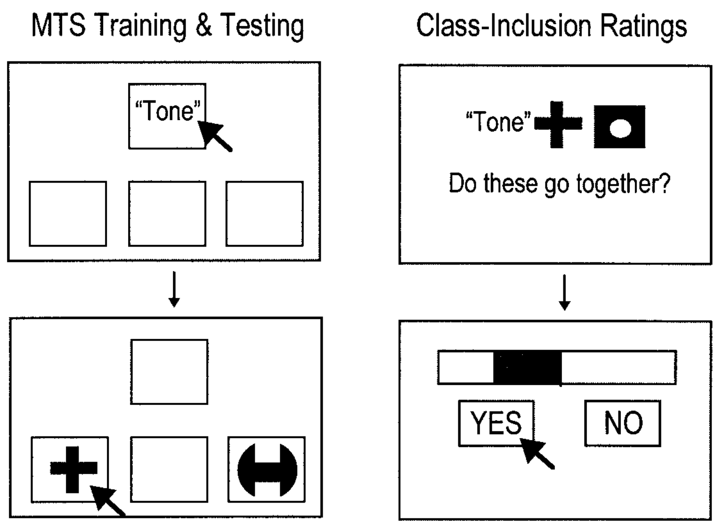

Fig. 1. Depiction of the visual display on the subject's monitor screen during the MTS training (left panels) and class-inclusion rating procedure (right panels). See text for details concerning the class-inclusion ratings.

> presses you make may or may not be related to the number of seconds you earn or lose. You may or may not get feedback about your performance this session.

Only visual stimuli were used in this phase. In addition to the six stimuli used in the training and equivalence testing, two additional stimuli were employed: the lower case letter b, and a vertical arrangement of two small circles (**:**). These stimuli were selected to be readily classified with one member of each "trained" stimulus class (e.g., b to B, and two circles to the numeral 2), thus providing subjects with an experimental history of classifying stimulus groups that included members that were not part of training sets.

Key features of the subject's display during the class-inclusion rating procedure are shown in the right panels of Figure 1 (which depicts the rating procedure with stimuli used during later phases of the study). On each trial, a set of two or three stimuli was displayed with members of the set horizontally aligned across the middle of the screen. There were four types of sets: (a) class consistent, trained (e.g., B A C; 1 2 3); (b) class inconsistent, trained (e.g., C 3 2; A B 3); (c) class consistent, novel (e.g., A C b; 3 **:** 1); and (d) class inconsistent, novel (e.g., A C **:**; 3 b 1).

Subjects completed two 84-trial sessions; each session presented most of the possible combinations of the eight stimuli described above. Each trial began with the presentation of a stimulus set accompanied, on the first five trials per session only, by the message, "Do these go together?" printed directly below the stimuli. A message at the bottom of the screen stated, "Click the red mouse button to continue." Clicking the red mouse button cleared the stimuli and replaced them with a confidence bar consisting of a rectangle about 1 cm high and 12 cm long, and bisected by a thin vertical line. Below the confidence bar were two smaller boxes, one labeled "YES," aligned under the left end of the confidence bar, and one labeled "NO," aligned under the right end of the confidence bar.

A maximum of 4 s was allotted to complete each rating. The initial choice of either box cleared the alternative box from the screen. Each click of the YES box filled one 18th of the confidence bar (18 responses maximum). When the NO box was clicked, the first response filled the confidence bar, and each additional response cleared one 18th of the bar (18 responses maximum). The consequences of making ratings, described below, were designed to promote YES selections accompanied by high-rate responding when the stimulus set was class consistent and NO selections accompanied by low-rate responding when the stimulus set was class inconsistent. For example, to indicate a high degree of certainty that the stimuli in the set were class consistent, a subject could respond 18 times in the YES box. To indicate a lower degree of certainty that the stimuli were class consistent, a subject could respond only a few times in the YES box. To indicate a high degree of certainty that the stimuli were class inconsistent, a subject could respond just once in the NO box. To indicate a lower degree of certainty that the stimuli were class inconsistent, a subject would respond several times in the NO box.

After 4 s had elapsed or 18 responses had been made, whichever came first, the self-report screen was replaced immediately by a feedback screen indicating how many seconds had been gained or lost through the rating procedure. Printed at the top of the screen, in large font, was the word CORRECT or INCORRECT, as appropriate to the outcome of the trial. In the center of the screen, printed in a smaller font, was the message "TIME EARNED = X seconds" or "TIME LOST = −X seconds." The value of X was determined as follows. When the stimulus set was class consistent, a subject either earned seconds equal to the number of responses made in the YES box or lost seconds equal to 19 minus the number of responses made in the NO box. When the stimulus set was class inconsistent, a subject either lost seconds equal to the number of responses made in the YES box or gained seconds equal to 19 minus the number of responses made in the NO box.

On the first five trials of each session, the printed message, "Pressing repeatedly may or may not affect how many seconds you earn or lose," also appeared at the bottom of the feedback screen. This message was added to the procedure after pilot data suggested that, without the printed message, subjects had difficulty acquiring the high-rate response pattern for "yes" ratings.

Phase 3: Baseline class-inclusion test. Phase 3 consisted of one baseline session of an equivalence generalization test using the class-inclusion rating procedure. The procedure (Figure 1, right panels) was identical to Phase 2 except that no feedback followed any trial, and the stimuli were novel and included all the tones in Table 2 and the visual stimuli shown in Figure 2. The stimulus arrangements (shown in Appendix B) were also used in Phase 6 to obtain postequivalence generalization gradients. Because subjects had not yet been exposed to any of the visual stimuli, and because the tones had not yet been included in the experimental contingencies, no systematic rating patterns were expected.

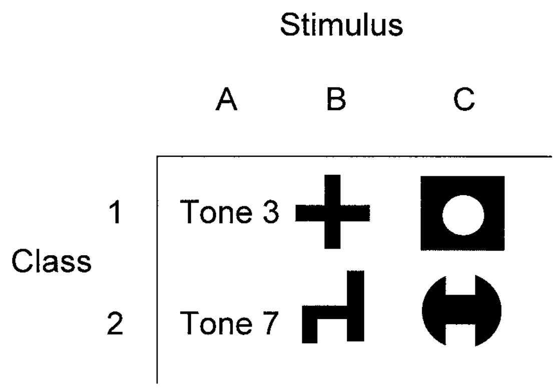

Fig. 2. The visual stimuli used in Phases 3 through 7.

Phase 4: Conditional discrimination training. Subjects were taught conditional discriminations prerequisite to the formation of two three-member cross-modal equivalence classes (A1B1C1 and A2B2C2). The basic training procedure is depicted in Figure 1 (left panels). At the start of the phase, the following instructions appeared on the subject's screen:

> At the beginning of each trial four boxes will appear on the screen, one on top and three below. Moving the cursor into the upper box and pressing the red mouse button will produce a tone, and then two figures will appear in the lower boxes. You may earn seconds of extra credit by selecting the correct figure from one of the lower boxes. To make a selection, place the cursor in the lower box you wish to select and press the red mouse button.

A "one-to-many" training procedure (Spradlin & Saunders, 1986) was used in which the A stimuli always served as the sample and the B or C stimuli were always comparisons. The A stimuli were Tones 3 and 7 (Table 2); B and C stimuli were arbitrary visual shapes displayed in red (Figure 2). To start each trial, subjects clicked the mouse in an empty sample box. This produced a 2-s sample tone followed immediately by the presentation of visual comparison stimuli in two of three horizontally aligned comparison boxes below the sample box, with locations counterbalanced within each session. Feedback followed every trial, and each correct MTS selection was worth 20 s of extra-credit time. Following each session, the total number of seconds earned in the session was displayed on the monitor for 5 s. Stimulus arrangements for these sessions are shown in Appendix A.

Initially, sessions consisted of 12 trials of A-B relations, six from each stimulus class. When two sessions were completed at a minimum accuracy of 92%, A-C relations were trained in the same format. Then the same A-B and A-C relations were intermixed in 24-trial sessions until two sessions were completed at a minimum accuracy of 95%.

Phase 5: Equivalence tests. This phase tested all possible trained and emergent relations. Stimulus arrangements for the Phase 5 test trials are shown in Appendix A. Each 66-trial session consisted of 12 training trials (A-B, A-C), 18 reflexivity trials (A-A, B-B, C-C), 12 symmetry trials (B-A, C-A), and 24 equivalence trials (combined symmetry and transitivity, B-C and C-B). No feedback was given. Following each session, a message reading "Session information withheld" was displayed on the monitor for 5 s.

During the A-A reflexivity tests (auditory stimuli), subjects clicked the mouse in an empty sample box to produce a 2-s sample tone. Immediately afterward, two 2-s comparison tones followed, separated by a 1-s pause. A yellow asterisk appeared in one of the three comparison boxes during the sounding of each comparison tone. Across the session, each comparison box was used equally often. The asterisks remained present until the subject made a selection by clicking in a corresponding comparison box. The B-B and C-C reflexivity tests (visual stimuli only) were presented using typical MTS procedures. During symmetry tests, sample stimuli were visual shapes, and auditory comparison stimuli were presented as described on the A-A reflexivity tests. During equivalence (combined) tests (visual stimuli only), stimuli were presented using typical MTS procedures (as in Phase 2).

Phase 6: Postequivalence class-inclusion test. This phase provided the study's primary data. The class-inclusion rating procedure presented in Phase 3 was repeated using the stimuli from Phases 4 and 5 (see Figure 1, right panels). The stimulus sets sometimes included probe tones that varied in degree of similarity to the trained tones (see Table 2). On each trial, a two- or three-member stimulus set was presented for 2 s, accompanied, on the first five trials only, by the printed message "Do these go together?" Subjects labeled each set as class consistent or class inconsistent, and recorded a confidence rating as described previously. No feedback about earnings or performance followed any trial in this phase.

Depending on the amount of time available, subjects completed three (Subjects 315, 316, 317, and 322) or four (remaining subjects) sessions. Sessions were 136 trials long, and presented most possible combinations of two and three stimuli from the 13 stimuli used during this phase (six stimuli from training and seven novel probe tones). Eighty (59%) of the 136 stimulus sets shown in Appendix B incorporated only training stimuli; 40 of these sets were class consistent, and 40 were class inconsistent. The remaining 56 sets (41%) combined trained visual stimuli with one of the seven probe tones. When a set consisted of two visual stimuli and a probe tone, the visual stimuli were always from the same class so that ratings could not be based solely on the visual stimuli.

Phases 7 and 8: Equivalence and auditory discrimination posttests. Phases 7 and 8 replicated the equivalence test sessions of Phase 5 and the auditory discrimination test of Phase 1, respectively, to determine whether Phase 6 had altered these critical performances.

RESULTS

Auditory Discrimination Pretest and Posttest

Patterns of tone discrimination were not altered by the experience of participating in the main study. Across both auditory discrimination tests, overall accuracy was 97.2% (range, 91.75% to 100%).

Preliminary Training of the Class-Inclusion Rating Procedure

Subjects made few errors during MTS training and equivalence testing (Phase 2) with the A-B-C and 1-2-3 stimulus classes; all met the mastery criterion in three sessions or less. Overall mean accuracy was 94.4% during training and 99.2% on equivalence tests.

During the preliminary class-inclusion rat-

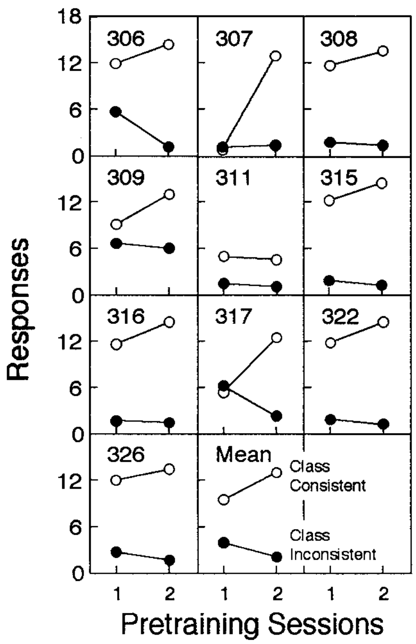

Fig. 3. Performance on the two preliminary class-inclusion test sessions (Phase 2, pretraining stimuli). The panels show class-inclusion rating magnitudes for individual subjects and the group mean (bottom, center). White circles show the mean number of responses per trial for class-consistent stimulus sets, and black circles show class-inconsistent sets. Higher scores indicate confidence that the sets were class consistent; lower scores indicate confidence that they were class inconsistent.

ing tests, subjects readily distinguished class-consistent from class-inconsistent sets (mean accuracy = 88.20%, *SE* = 3.21). Figure 3 summarizes the ratings. Typically, subjects made "yes" ratings and responded at a high rate when sets were class consistent and made "no" ratings and responded at a low rate when sets were class inconsistent.

Conditional Discrimination Training and Equivalence Tests

Individual conditional discrimination training (Phase 4) and equivalence test (Phase 5) performances were uniformly high. All subjects met mastery criteria across the stages of MTS training in three sessions or less. Mean accuracy was 90.40% (*SE* = 2.73) for AB and AC training sessions and 99.33% (*SE* = 0.36) for mixed training sessions. In equivalence tests (Phase 5), 9 of 10 subjects demonstrated near perfect class-consistent performances on all relation types and met the mastery criterion of two consecutive sessions at greater than 91% accuracy within three sessions. Mean accuracy for all 10 subjects was 93.69% (*SE* = 2.39) for reflexive relations, 94.84% (*SE* = 2.61) for symmetrical relations, and 89.34% (*SE* = 3.94) for combined (symmetry and transitivity) relations. Subject 316 failed initial equivalence tests, and twice (after Test Sessions 5 and 8) was returned to mixed training prior to retesting; subsequently, this subject met the mastery criterion in Sessions 9 and 10. The results from Phase 5 thus indicate that three-member equivalence classes had formed for all 10 subjects. These equivalence classes remained intact during the equivalence posttests (Phase 7); mean overall accuracy was 99.55% (*SE* = 0.26).

Class-Inclusion Ratings

Preequivalence class-inclusion ratings (Phase 3) showed no systematic patterns. All subjects rated the tones as class consistent on approximately 50% of the presentations; the range was 39% to 60%. Figure 4 shows the individual-subject class-inclusion rating gradients that followed equivalence testing (Phase 6). Typically, the highest percentage of "yes" ratings occurred at or near the training stimuli (A1 = Tone 3 and A2 = Tone 7), and each subject's two gradients intersected near the midpoint of the range of tones (Tone 5). Some subjects (Subjects 306, 307, 309, 326) showed sharp generalization gradients around the A1 and A2 tones, whereas others (Subjects 308, 311, 315, 317, and 322) showed more step-like functions with only two rating levels. These step-like functions were characterized by (a) levels near 100% on the trained tones (and those tones most similar to them) with class-consistent stimulus sets, and (b) levels near 0% on the trained tones (and those tones most similar to them) with class-inconsistent stimulus sets. Subject 317 showed a similar pattern, but levels on class-inconsistent sets were uniformly near 50%. In three instances, for Subjects 306, 307, and 316 (Tone A1 only), little to no generalization was evident. Specifically, few ratings of class-consistent sets occurred at the tails of the gradient (i.e., to the left of Tone 3, or A1), indicating that subjects did not treat these higher pitched tones as equivalent to the B1 and C1 visual stimuli. The two right panels of Figure 4 show mean number of responses made in conjunction with the class-

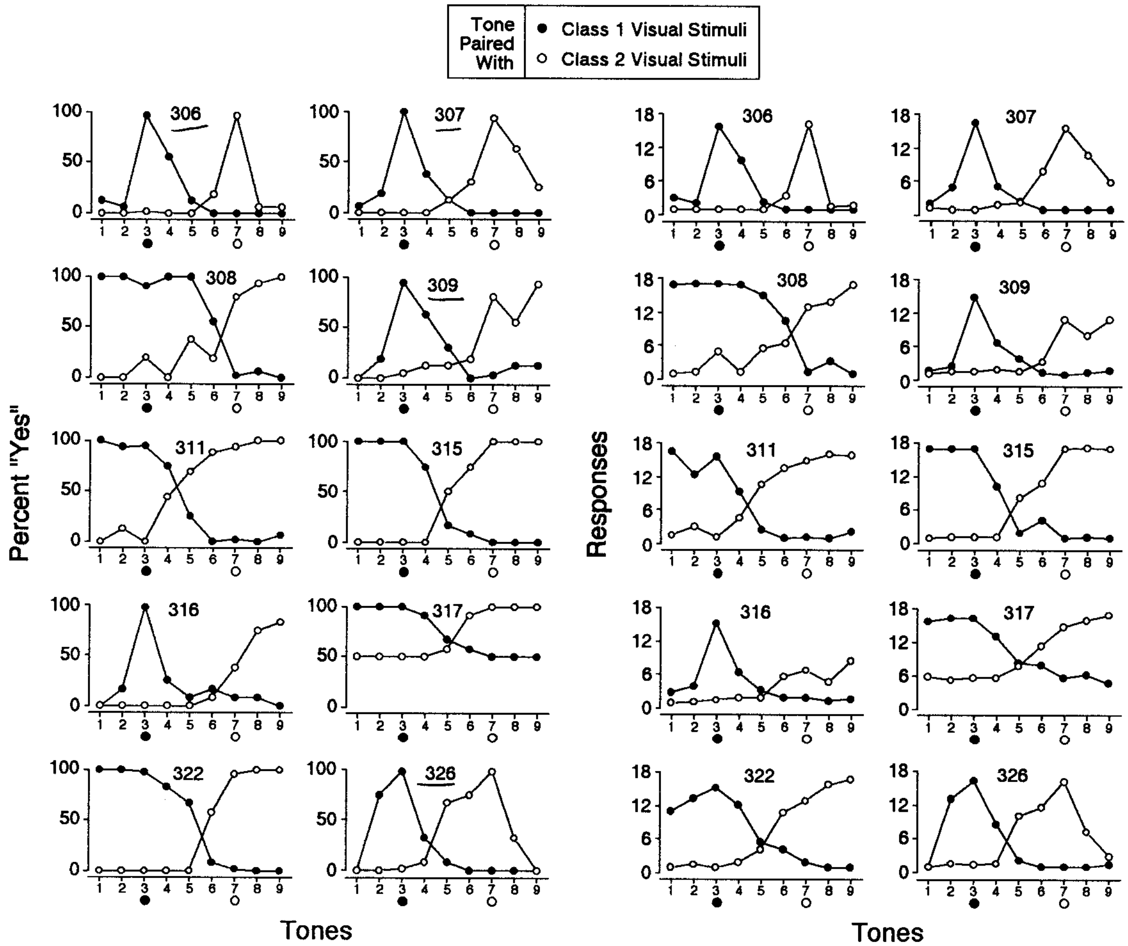

Fig. 4. Performances of all 10 subjects on the primary class-inclusion test sessions (Phase 6). The two left panels show the percentage of "yes" (class-consistent) responses as a function of the nine tones presented (Tone 3 = A1 trained stimulus, Tone 7 = A2 trained stimulus). Black and white symbols show, respectively, ratings when Class 1 and Class 2 visual stimuli were presented. The two right panels show the response magnitudes, indicating subjects' "confidence" during the class-inclusion ratings, plotted in the same format as the left panels. High ratings indicate confidence on class-consistent sets; low ratings indicate confidence on class-inconsistent sets.

inclusion ratings. These functions correspond quite consistently with percentage of "yes" reports; rating magnitudes were either peaked or step-like functions and were generally highest at or near the class-consistent trained tones and lowest at or near the class-inconsistent trained tones.

DISCUSSION

This study supports and extends recent work by Fields et al. (1991, 1993, 1997), and Adams et al. (1993), showing that novel stimuli may enter, via generalization, into previously established equivalence classes strictly on the basis of physical similarity to members of those classes. These previous findings are extended in two ways. First, unlike previous studies, each subject's inclusion of novel stimuli (tones) into existing equivalence classes was not measured using a MTS testing format. Rather, using a class-inclusion rating system, subjects responded differently to groups of stimuli that were class consistent than they did to groups that were class inconsistent. The novel tones were more likely to be rated as class consistent when they were closer in frequency to those explicitly included in the equivalence class (e.g., the spread of effect typically centered around the auditory stimuli that were equivalence class members). Second, the study demonstrates generalization of class membership in cross-modal sets of stimuli. Most human categories, at least those manifested linguistically, are cross-modal, incorporating written words, spoken words, and nonlanguage visual and auditory stimuli. A convincing account of categorization based on conditioning processes should be compatible with this observation. It is conceivable, for example, based on the present results, that a category like BIRD could be formed initially through stimulus equivalence involving a limited number of exemplars involving actual birds, printed words, spoken words, and pictures. The category then could expand rapidly via generalization of the original class members to novel, but similar, stimuli (like pictures of other birds).

Cross-modal equivalence classes have been demonstrated previously (Bush, 1993; Sidman & Tailby, 1982), and the fact that stimulus generalization could have an impact on class expansion may seem unsurprising. New psychological principles are not required to explain the present results, and may not be required to account for many observed instances of complex categorization. In fact, though sometimes associated with linguistic ability, the performances observed in the present study need not be considered uniquely human. Data from the animal cognition literature suggest that pigeons are able not only to categorize stimuli into functional classes but also to extend those categorizations to novel stimuli. In some cases, the novel stimuli were treated as equivalent based on perceptual similarity. In other cases, they were treated as such based on a conceptual relationship such as different (oddity), or people, flowers, and chairs (Astley & Wasserman, 1998; Urcuioli, 1996; Urcuioli, DeMarse, & Lionello, 1998; Wasserman & Bhatt, 1992; Wright, 1992; Zentall, Sherburne, & Urcuioli, 1993).

Although linguistic abilities need not be invoked to account for these data, it may be interesting to speculate on the role verbal behavior played in the present study. One possibility is that subjects established verbal categories that mediated the division of stimuli into classes (for example, based on the pitch of the training tones these categories could have been ''high'' and ''low''). Subsequently, during the presentation of novel tones, responding would be analogous to a signal-detection task in which ''yes'' or ''no'' responses were made based on the subject's decision criterion located along an auditory frequency distribution (see Dougherty & Wixted, 1996). Tones perceived as being to the right of the criterion would be partitioned into the ''high'' category; those to the left of the criterion would fall into the ''low'' category. Individual differences in the shape of the class-inclusion rating gradients (Figure 4) could be attributed to different decision criteria across subjects. Although we did not collect data to corroborate or refute this possibility, such a hypothesis is testable. One interesting strategy would be the inclusion of a pretraining phase in which subjects' decision criteria for classifying tones were manipulated via reinforcement contingencies (see Gescheider, 1985). Such a manipulation would allow investigators to make specific predictions about how tones would be categorized during test phases and could provide a type of preliminary model in which to study individual differences in category formation.

REFERENCES

Adams, B. J., Fields, L., & Verhave, T. (1993). Formation of generalized equivalence classes. *The Psychological Record, 43,* 553–566.

Astley, S. L., & Wasserman, E. A. (1998). Novelty and functional equivalence in superordinate categorization by pigeons. *Animal Learning & Behavior, 26,* 125–138.

Bush, K. M. (1993). Stimulus equivalence and cross-modal transfer. *The Psychological Record, 43,* 567–584.

Carpenter, G. A., & Grossberg, S. (1990). Neural dynamics of category learning and recognition: Structural invariants, reinforcement, and evoked potentials. In M. L. Commons, R. J. Herrnstein, S. M. Kosslyn, & D. B. Mumford (Eds.), *Quantitative analyses of behavior: Computational and clinical approaches to pattern recognition and concept formation* (Vol. 9, pp. 23–49). Hillsdale, NJ: Erlbaum.

Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. In G. A. Carpenter & S. Grossberg (Eds.), *Pattern recognition by self-organizing neural networks* (pp. 211–236). Cambridge, MA: MIT Press.

Critchfield, T. S., Tucker, J. A., & Vuchinich, R. E. (in press). Self-report methods. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior.* New York: Plenum.

Dougherty, D. H., & Wixted, J. T. (1996). Detecting a nonevent: Delayed presence-versus-absence discrimination in pigeons. *Journal of the Experimental Analysis of Behavior, 65,* 81–92.

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

Fields, L., Reeve, K. F., Adams, B. J., Brown, J. B., & 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). Stimulus generalization and equivalence classes: A model for natural categories. *Journal of the Experimental Analysis of Behavior, 55,* 305–312.

Gescheider, G. A. (1985). *Psychophysics: Method, theory, and application.* Hillsdale, NJ: Erlbaum.

Harnad, S. (1987). The induction and representation of categories. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 535–565). New York: Cambridge University Press.

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

Herrnstein, R. (1984). Objects, categories, and discriminative stimuli. In H. L. Roitblatt, T. G. Beaver, & H. S. Terrace (Eds.), *Animal cognition* (pp. 233–261). Hillsdale, NJ: Erlbaum.

Herrnstein, R. (1990). Levels of stimulus control: A functional approach. *Cognition, 37,* 133–166.

Lane, S. D., & Critchfield, T. S. (1996). Verbal self-reports of emergent relations in a stimulus equivalence procedure. *Journal of the Experimental Analysis of Behavior, 65,* 355–374.

Saunders, R. R., & Green, G. (1992). The nonequivalence of behavioral and mathematical equivalence. *Journal of the Experimental Analysis of Behavior, 57,* 227–241.

Saunders, R. R., & Green, G. (1996). Naming is not (necessary for) stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65,* 312–314.

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.

Smith, E. E. (1995). Concepts and categorization. In E. E. Smith & D. N. Osherson (Eds.), *Thinking: An invitation to cognitive science* (Vol. 3, 2nd ed., pp. 3–33). Cambridge, MA: MIT Press.

Spradlin, J. E., & Saunders, R. R. (1986). The development of stimulus classes using match-to-sample procedures: Sample classification versus comparison classification. *Analysis and Intervention in Developmental Disabilities, 6,* 41–58.

Stromer, R., & Mackay, H. A. (1992). Spelling and emergent picture-printed word relations established with delayed identity matching to complex samples. *Journal of Applied Behavior Analysis, 25,* 893–904.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

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

Urcuioli, P. J., DeMarse, T. B., & Lionello, K. M. (1998). Transfer of performance to new comparison choices following differential outcome matching-to-sample. *Animal Learning & Behavior, 26,* 139–153.

Vaughan, W., Jr. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 36–42.

Wasserman, E. A., & Bhatt, R. S. (1992). Conceptualization of natural and artificial stimuli by pigeons. In W. K. Honig & J. G. Fetterman (Eds.), *Cognitive aspects of stimulus control* (pp. 203–224). Hillsdale, NJ: Erlbaum.

Wasserman, E. A., Kiedinger, R. E., & Bhatt, R. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 3,* 235–246.

Wright, A. A. (1992). The study of animal cognitive processes. In W. K. Honig & J. G. Fetterman (Eds.), *Cognitive aspects of stimulus control* (pp. 225–242). Hillsdale, NJ: Erlbaum.

Zentall, T. R., Sherburne, L. M., & Urcuioli, P. J. (1993). Common coding by pigeons in a many-to-one delayed matching task as evidenced by facilitation and interference effects. *Animal Learning & Behavior, 21,* 233–237.

Received August 26, 1997
Final acceptance June 17, 1998

APPENDIX A

Stimulus arrangements for Phases 2 (excluding class-inclusion ratings), 4, and 5. The location of the comparison stimulus was counterbalanced within sessions. CO+ = correct comparison stimulus. CO− = incorrect comparison stimulus. Actual stimuli are listed for Phase 2. Phase 4 stimuli are keyed to Figure 2.

Phase	Description	Sample	Comparisons CO+	Comparisons CO−	Times presented per session
2a	Preliminary training				
	A-B	A	B	2	6
		1	2	B	6
	A-C	A	C	3	6
		1	3	C	6
	Mixed review	Same as above			6 each
2b	Equivalence tests				
	Trained	Same as above			1 each
	Reflexive	A	A	1	1
		1	1	A	1
		B	B	2	1
		2	2	B	1
		C	C	3	1
		3	3	C	1
	Symmetrical	B	A	1	2
		2	1	A	2
		C	A	1	2
		3	1	A	2
	Combined	B	C	3	2
		2	3	C	2
		C	B	2	2
		3	2	B	2
4	MTS training				
	A-B	A1	B1	B2	6
		A2	B2	B1	6
	A-C	A1	C1	C2	6
		A2	C2	C1	6
	Mixed review	Same as above			6 each
5	Equivalence tests				
	Trained	Same as above			3 each
	Reflexive	A1	A1	A2	3
		A2	A2	A1	3
		B1	B1	B2	3
		B2	B2	B1	3
		C1	C1	C2	3
		C2	C2	C1	3
	Symmetrical	B1	A1	A2	3
		B2	A2	A1	3
		C1	A1	A2	3
		C2	A2	A1	3
	Combined	B1	C1	C2	6
		B2	C2	C1	6
		C1	B1	B2	6
		C2	B2	B1	6

APPENDIX B

Stimulus configurations, and frequency of presentation in each session (n), for class-inclusion test trials of the main experiment (Phases 3 and 6). Stimuli are those shown in Figure 2. Visual stimuli appeared in the left-to-right sequences shown below; a tone, if part of the array, was played simultaneously. A1 = Tone 3. A2 = Tone 7.

Sets including training stimuli only								Sets including probe tones									
Class-consistent sets				Class-inconsistent sets					n: Paired with tone								
Auditory-visual	n	Visual only	n	Auditory-visual	n	Visual only	n	Visual stimuli	1	2	3[a]	4	5	6	7[a]	8	9
B1 C1 A1	4	B1 C1	2	B2 C2 A1	4	B1 B2	1	B1 C1	1	1		1	1	1		1	1
C1 B1 A1	4	C1 B1	2	C2 B2 A1	4	B2 B1	1	C1 B1	1	1		1	1	1		1	1
B2 C2 A2	4	B2 C2	2	B1 C1 A2	4	C1 C2	1	B2 C2	1	1		1	1	1		1	1
C2 B2 A2	4	C2 B2	2	C1 B1 A2	4	C2 C1	1	C2 B2	1	1		1	1	1		1	1
B1 A1	4			B2 A1	4	B1 C2	1	B1	1	1		1	1	1		1	1
C1 A1	4			C2 A1	4	C2 B1	1	C1	1	1		1	1	1		1	1
B2 A2	4			B1 A2	4	C1 B2	1	B2	1	1		1	1	1		1	1
C2 A2	4			C1 A2	4	B2 C1	1	C2	1	1		1	1	1		1	1

[a] A1 (Tone 3) and A2 (Tone 7) were training tones.

1998 GUEST REVIEWERS

JEAB thanks the following individuals for serving as guest reviewers during the period September 1, 1997 to August 31, 1998:

B. Alsop
D. Barnes
D. Cerutti
P. Chase
G. Collier
A. R. Delameter
J. Dinsmoor
M. Domjan
D. Eckerman
D. Elliffe
R. Foltin
S. Fowler
J. Gibbon
B. Hienz
J. Hinson
S. Hursh
P. Killeen
K. A. Lattal
P. S. Lawrence
S. Lea
J. Myerson
D. Navarick
A. Neuringer
M. Pelaez-Nogueras
C. Pilgrim
K. Sawrey
R. Serna
A. Silberberg
J. Spradlin
D. Stafford
D. Stephens
B. Williams
D. Williams
T. Zarcone
T. Zentall