REVISTA MEXICANA DE ANÁLISIS DE LA CONDUCTA 1998, 24, 311-337 NÚMERO 3 (DICIEMBRE)
MEXICAN JOURNAL OF BEHAVIOR ANALYSIS NUMBER 3 (DECEMBER)

# PRELIMINARY EVIDENCE FOR CROSS-MODAL GENERALIZED EQUIVALENCE CLASSES

## EVIDENCIA PRELIMINAR DE GENERALIZACIÓN ENTRE MODALIDADES DE EQUIVALENCIA ENTRE CLASES

**SCOTT D. LANE, ANDREW INNIS, JULIE K. CLOW, AND THOMAS S. CRITCHFIELD**[1]
AUBURN UNIVERSITY

## ABSTRACT

Previous studies have suggested that stimulus equivalence and stimulus generalization can interact to contribute to the formation of open-ended categories. The present study sought to extend that work by determining whether emergent stimulus classes involving cross-modal stimuli also can expand spontaneously via stimulus generalization. A pre-test and post-test of simple auditory discrimination suggested that participants (college students, N = 7) could discriminate among a range of tones used in the main study, and that patterns of discrimination were not globally altered by intervening phases of the study. Before beginning the main study, participants learned to use a rating procedure for categorizing sets of stimuli as class-consistent or inconsistent. After completing conditional discrimination training with new stimuli, the participants demonstrated the formation of cross modal equivalence classes involving both shapes and tones. Subsequently, the class-inclusion rating procedure was reinstituted, this time with cross-modal sets of stimuli drawn from the equivalence

[1] Scott D. Lane, Auburn University (now at Department of Psychiatry and Behavioral Sciences, The University of Texas Health Science Center at Houston); Andrew Innis, Auburn University; Julie K. Clow, Auburn University; Thomas S. Critchfield, Auburn University.

Portions of the data were presented at the 1996 Association for Behavior Analysis Conference in San Francisco. Manuscript preparation was partially supported by a training grant from the National Institute for Drug Abuse (DA07247) to the University of Texas Health Science Center at Houston. The authors thank Bill Dube, Lanny Fields, Kate Saunders, and Edelgard Wulfert for helpful suggestions regarding the manuscript.

Correspondence concerning this article should be addressed to Julie K. Clow, Department of Psychology, Auburn University, 226 Thach, Auburn, Alabama 36849. Electronic mail may be sent to clowjul@mail.auburn.edu.

classes. On some trials, novel tones were paired with shapes from the previously-demonstrated equivalence classes. The probability that these novel sets would be rated as class-consistent was a direct function of the degree of difference between the novel tone and the tone explicitly included in the equivalence class. The present results appear to show the formation of cross-modal generalized equivalence classes. Pending replication to resolve procedural ambiguities, these results add generality to prior work on generalization of stimulus equivalence classes, and thus support the role of operant processes in human category formation.

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

## RESUMEN

En estudios previos se ha sugerido que la equivalencia y la generalización de estímulos pueden interactuar para contribuir a la formación de categorías ilimitadas. En el presente estudio se buscó extender ese trabajo determinando si las clases de estímulo emergentes que involucran estímulos de diferentes modalidades pueden extenderse espontáneamente a través de la generalización de los estímulos. Una pre-prueba y una post-prueba de una discriminación auditiva simple sugirió que los participantes (estudiantes de licenciatura, N = 7) podrían discriminar entre un rango de tonos usados en el estudio principal y que los patrones de discriminación no fueron alterados globalmente por las fases intervinientes del estudio. Antes de empezar el estudio principal, los participantes aprendieron a usar un procedimiento de calificación para categorizar conjuntos de estímulos como consistentes e inconsistentes con la clase. Después de terminar el entrenamiento en discriminación condicional con estímulos nuevos, los participantes demostraron la formación de equivalencias de clases entre modalidades que incluyeron formas y tonos. Posteriormente, el procedimiento de calificación de inclusión en la clase, se reinstituyó, esta vez con conjuntos de estímulos de diferentes modalidades derivados de las clases equivalentes. En algunos ensayos, los estímulos nuevos se aparearon con formas de las clases equivalentes previamente demostradas. La probabilidad de que estos conjuntos nuevos se clasificaran como consistentes con la clase fue una función directa del grado de diferencia entre el tono nuevo y el tono explícitamente incluido en la clase equivalente. Estos resultados parecen mostrar la formación de generalización entre modalidades de equivalencia entre clases. Aparte de una replicación para resolver ambigüedades de procedimiento, estos resultados añaden generalidad al trabajo previo sobre generalización de los estímulos en equivalencia de clases y así apoya el papel de los procesos operantes en la formación de categorías en humanos.

Palabras clave: equivalencia del estímulo, generalización del estímulo, categorización, procedimiento de clasificación, presión del "raton" de computadora, estudiantes de licenciatura

Categorization, or classification, is the process of dividing stimuli into meaningful groups (e.g., Barsalou, 1992), and is evident both in differential labeling and in differential patterns of action with respect to stimuli. The means by which categorization is acquired remains a matter of substantial debate (e.g., Carpenter, Grossberg, & Reynolds, 1991; Fields, Reeve, Adams, & Verhave, 1991; Herrnstein, 1984; Saunders & Green, 1992; Sidman, 1994; Smith, 1995; Wasserman & Bhatt, 1992; Wright, 1992). Most theoretical accounts appear to agree, however, that category formation is both structurally and functionally influenced. That is, category formation can be guided both by the physical properties of stimuli, and the consequences for differentially associating them (e.g., Barsalou, 1992; Harnad, 1987; Herrnstein, 1990; Tversky, 1977; Vaughn, 1988).

The phenomenon known as stimulus equivalence (Sidman, 1994) may provide one example of a functional contribution to category formation. In laboratory equivalence procedures, the training of overlapping conditional discriminations leads to emergence of untrained stimulus relations, notable among which are transitive relations between stimuli with no explicit training history of association. The result is a group of physically dissimilar stimuli that become functionally interchangeable in that they have identical effects on behavior in appropriate contexts, and may even be identically labeled (e.g., Sidman, 1994).

Most important human categories are open-ended, that is, they incorporate a potentially infinite number of exemplars which, once the class-specific categorization repertoire has been acquired, can be identified without explicit training. Therefore, according to Harnard (1996), equivalence class formation alone cannot account for human category formation because of its reliance on a finite pool of stimuli. Yet several recent studies nicely illustrate a conjunction of structural and functional influences in category formation, and in the process suggest a possible response to Harnard's (1996) criticism.

Fields and his colleagues have shown that stimulus classes created through the functional processes of stimulus equivalence can expand, via stimulus generalization, to include new stimuli that have never been involved in explicit training (Adams, Fields, & Verhave, 1993; Fields et al., 1991; Fields, Adams, Brown, & Verhave, 1993; Fields, Adams, Buffington, Yang, & Verhave, 1996; Fields, Reeve, Adams, Brown, & Verhave, 1997). The general procedure of these studies was as follows. Differential reinforcement during conditional discrimination training was used to teach overlapping stimulus relations. The training stimuli then cohered into stimulus classes indicative of stimulus equivalence (as defined by Sidman, 1994), including the emergence of relations among stimuli related, during training, only through a common associate. Among these training stimuli was a horizontal line of intermediate length.

Subsequent generalization testing, in match-to-sample format, showed that novel stimuli (lines both longer and shorter than the training line) sometimes were matched to the other (non-line) training stimuli (e.g., see Fields et al., 1991). Thus, classes had formed that reflected both functional and structural influences. The authors labeled these potentially open-ended classes as *generalized equivalence classes* (Fields et al., 1991).

Generalized equivalence classes provide evidence that stimulus classes built around equivalence processes need not be limited to stimuli with which an individual has prior experience. Because naturally occurring categories tend to be large, diverse, and multi-faceted, however, much remains to be learned before stimulus equivalence can contribute to a general-purpose account of human category formation. An important feature of many human concepts, for example, is that they are cross-modal. A boy's concept of *dog* may encompass written and spoken versions of the word "dog" (plus "chien" if he is bilingual), pictures of dogs, the sounds of dogs barking, the feel of a dog's fur, and so forth. So far, it is not known whether generalized equivalence classes can extend across different stimulus modalities.

The present study sought evidence for generalized equivalence classes incorporating both auditory and visual stimuli. Stimulus classes were created through conditional discrimination training involving overlapping stimulus relations among two visual stimuli and a tone. Of interest was whether tones other than the training tone subsequently would be treated as related to the visual stimuli. In the studies by Fields and colleagues, these extended stimulus relations were demonstrated through traditional match-to-sample tests. In the present study, we probed for stimulus generalization effects using a new procedure in which participants rated groups of stimuli as class-consistent or class-inconsistent. These groups could include only class-consistent members, members from different classes, or members from a single class plus a novel tone that was physically similar to the one used in training. To facilitate interpretation of the ratings, we provided preliminary training to help establish, in advance, the "operating characteristics" of the rating repertoire (e.g., Critchfield, Tucker, & Vuchinich, in press; Saunders, 1996).

## METHOD

### *Participants and Apparatus*

Seven college-student volunteers reporting no prior experience in conditional discrimination experiments completed a study that was described in the informed consent agreement as concerned with "reasoning in symbolic logic." One additional volunteer was excluded because preliminary testing revealed poor discrimination among the tones to be used in the main study. For participating, participants received extra credit in undergraduate psychology courses. The informed consent agreement indicated that the amount of time for which course credit could be received was linked to performance in the study (details of these instructions are described below). Participants participated for 8 to 10 hours total, distributed, in visits lasting about 2 hours, over a period of 4 to 14 days, as was convenient to their schedules.

Participants 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, a speaker (taken from a Craig Model JS8633 portable compact disc player), and a fluorescent lamp. They performed the MTS task by using the mouse to move a cursor to appropriate locations on the video screen. The speaker, positioned on the table next to the video monitor, was used to present auditory stimuli. An IBM-compatible microcomputer located in an adjacent room was used to control experimental procedures, present stimuli, and collect data. Tones were created using a Soundblaster ® Sound Card installed on the PC. All operations were controlled by a custom program written in Microsoft QuickBasic ®.

*Instructions*

No spoken instructions were provided during the experiment, but throughout the procedures, a variety of printed messages appeared on the participant's screen. These are described in conjunction with the relevant procedures below. In general, these instructions were designed to (a) describe mechanical selected aspects of the procedure (for example, to initiate each trial by mouse clicking in the sample box); and (b) in conjunction with the informed consent agreement, establish "seconds" as reinforcers.

*"Seconds" as Reinforcers*

The informed consent agreement indicated that the time for which participants would receive course credit was limited to the seconds earned as consequences in the experimental task. The informed consent stated the following:

> "You will work alone in an office-sized room, where you will view information on a computer screen and solve problems by pressing buttons on a mouse. You will be asked to solve a series of brief problems during each visit. You will not be given extensive instructions but rather will be asked to learn from experience as you work...Each time you solve a problem correctly, you will earn points worth seconds of participation time. The participation time documented for your instructor is measured in terms of these points, not in terms of how much time you spend in the lab. The more accurately you work, the

quicker you will accrue participation points. If you make few mistakes, you could earn 10 hours of participation points in as little as about 8 hours. If you make many mistakes, you could work in the study for 1 hour but earn fewer participation points. For example, if you worked for 10 hours, but earned 8.5 hours in points, you would receive documentation of working for 8.5 hours. In the past, most participants in our research have earned points roughly commensurate with their actual involvement in the study. But your point earnings depend on your performance, so we can make no guarantees about what you will actually earn."

The research protocol actually required that participants receive credit for all time spent in the study (10 hours maximum), regardless of performance. Participants were to be immediately debriefed should seconds accumulated at the end of the study fall short of the amount of time spent participating, but all participants earned the maximum possible extra credit. Thus, all were debriefed by mail at the end of the academic quarter.

*Phase 1: Auditory Discrimination Pretest*

Table 1 summarizes some key features of this and all subsequent phases. Phase 1 was designed to determine whether participants could distinguish 9 different tones from one another. During each trial in this phase, two 2-s tones were presented separated by a 1-s pause. On the first five trials only, the message, "Are the following two tones same or different?" appeared on screen just before the first tone began. Participants responded by clicking a mouse inside one of two horizontally adjacent boxes labeled "same" or "different." A single session included 3 presentations of each of the 81 possible binary combinations of 9 different tones. Thus, the two tones were identical on 27 (12.5%) of the 243 trials. The 9 tones, shown in Table 2, were spaced apart roughly equidistant along the musical scale, with middle C (264 Hz) as the third highest tone. Two of these tones were used as stimuli in the match to sample (MTS) training and equivalence testing in Phases 4 and 5. The other seven stimuli, as well as the trained tones, were used in equivalence generalization tests in Phase 6 (details described below).

*Phase 2: Preliminary Training of the Class-Inclusion Rating Procedure*

This phase was designed to mimic MTS training and class-inclusion rating procedure that would be used in Phases 4, 5, and 6. Two classes of stimuli were used in training, consisting of the capital letters A, B, and C and numerals 1, 2, and 3, all printed on screen in green. The A and 1 stimuli always served as samples, and the remaining stimuli were used as comparison stimuli. Thus, for example, when A was the sample, B and 2 were the comparison stimuli, and when 1 was the sample, C and 3 were the comparison stimuli. Under the assumption that the two classes already existed in each participant's repertoire, these stimuli were chosen to facilitate the rapid mastery and high levels of accuracy throughout training and testing. Appendix A shows the stimulus arrangements for each relation that was trained and tested. The following instructions were shown on the participant's screen prior to each session:

> At the beginning of each trial four boxes will appear on the screen, one on top and three below. The top box will have a figure in it. Moving the cursor into the upper box and pressing the mouse button will make three figures 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.

Table 1. Summary of experimental phases. See text, Table 2, and Figure 2 for descriptions of the stimuli

| Phase | Description | Stimuli | Probabilit of Feedback | Class-Inclusion Ratings? | Trials per Session | Sessions |
|---|---|---|---|---|---|---|
| 1 | Auditory discrimination test | Tones 1-9 | 0 | No | 243 | 1 |
| 2a | Preliminary training: MTS training | A, B, C<br>1, 2, 3<br>A, B, C and 1, 2, 3 | 1.0<br>1.0<br>1.0 | No<br>No<br>No | 12<br>12<br>24 | 2 at 100%<br>2 at 100%<br>2 > 95% |
| 2b | Preliminary training: Equivalence trials | A, B, C and 1, 2, 3 | 1.0 | No | 26 | 2 > 95% |
| 2c | Preliminary training: Class inclusion procedure | A, B, C, b & 1, 2, 3, | 1.0 | Yes | 84 | 2 |
| 3 | Baseline class-inclusio test | Tone 1-9*, B1, B2, C1, C2 | 0 | Yes | 16 | 1 |
| 4 | Conditional discrimination training: | A1, B1, C1<br>A2, B2, C2<br>A1, B2, C1 & A2, B2, C2<br>A1, B1, C1 & A2, B2, C2 | 1.0<br>1.0<br>1.0<br>0.4 | No<br>No<br>No<br>No | 12<br>12<br>24<br>24 | 2 > 91%<br>2 > 91%<br>2 > 95%<br>2 > 95% |
| 5 | Equivalence test | A1, B1, C1 & A2, B2, C2 | 0 | No | 52 | 2 |
| 6 | Class inclusion test | Tones 1-9*, B1, B2, C1, C2 | 0 | Yes | 136 | 3 |
| 7 | Equivalence test | A1, B1, C1 & A2, B2, C2 | 0 | No | 52 | 1 |
| 8 | Auditory discrimination test | Tones 1-9 | 0 | No | 243 | 1 |

*Tone 3 = A1, Tone 7 = A2

Table 2. Summary of the auditory stimuli.

| Tone | Hertz |
|---|---|
| 1 | 132 |
| 2 | 198 |
| 3 (A1)[a] | 264 |
| 4 | 352 |
| 5 | 528 |
| 6 | 792 |
| 7 (A2)[b] | 1056 |
| 8 | 1408 |
| 9 | 2112 |

Note: Distance between each of the tones is either a fourth or a fifth interval based on a diatonic scale (white piano keys only).
[a] "Middle C"
[b] A "C" two octaves above Middle C.

The left column of panels in Figure 1 shows some key elements of the screen during MTS trials. Participants clicked the empty sample box to start each trial. A sample stimulus appeared, accompanied by three comparison stimuli, located in three horizontally aligned comparison boxes below the sample box, with location counterbalanced across trials. A click in any of the comparison boxes registered a response. Each correct MTS selection was worth 20 s of extra-credit time, and feedback, indicating the number of seconds earned (0 or 20), normally followed every trial. A "one-to-many" training procedure (Saunders, Saunders, Williams, & Spradlin, 1993) was used to teach conditional discriminations prerequisite to the formation of the two stimulus classes. Participants progressed through training by scoring 100% correct on two consecutive sessions of (a) 12 trials involving the A-B and 1-2 relations; (b) 12 trials involving the A-C and 1-3 relations; and (c) mixed training review involving 12 trials each of A-B/1-2 and A-C/1-3 relations, intermingled within sessions. Feedback followed every trial until the mixed tone review was completed, whereupon the probability of feedback was reduced to 0.40 and the mixed training review was repeated. Following a session, the number of seconds earned for that session was shown on the monitor for 5 s.

Next, sessions mimicking equivalence tests were conducted using all of the possible trained and emergent relations derivable from the two stimulus sets. Trial format was identical to that used during training, using the stimulus configurations shown in Appendix A. Unlike in traditional equivalence tests, responses during these test sessions were followed by feedback, to facilitate accuracy.

The class-inclusion rating procedure was introduced next. In addition to the six stimuli used in the training and equivalence testing, the stimuli included two new shapes: 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). Assuming that participants classified accordingly, these novel stimuli would provide participants with an experimental history of classifying stimulus groups including members that were not part of training sets.

Sessions consisted of 84 trials long, and presented most of the possible combinations of the eight stimuli described above. Each participant completed two sessions. On each trial, a set of two or three stimuli was displayed with members of the set horizontally arranged in the middle of the screen. There were four types of sets: (a) class-consistent, trained (e.g., B A C); (b) class-inconsistent, trained (e.g., C 3 2); (c) class-consistent, novel (e.g., A C b or 3:1); and (d) class-inconsistent, novel (e.g., A C : or 3 b 1). Thus, this phase provided no experience with using physical similarity as a basis for including novel stimuli among equivalence-based stimulus sets. It allowed, but did not require, the more general experience of evaluating stimulus sets that included a potentially class-consistent member that had not been part of conditional discrimination training.

The right column of panels in Figure 1 shows some key elements of the display during the class-inclusion rating procedure. Each participant read the following instructions on the computer screen before beginning this phase:

> 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 onto one of the blue arrows and pressing the red mouse button. Each press will be counted in the seconds-earned box above the arrow. The number of 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.

On the first five trials per session only, the message, "Do these go together?" was 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 rectangle, the confidence bar, about 1 cm high and 12 cm long, and bisected by 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. Each click on the "Yes" box filled one-eighteenth of the confidence bar (18 responses maximum). The first click on the "No" box filled the confidence bar, and each subsequent response cleared one eighteenth of the bar (18 responses maximum). The initial click of either box cleared the alternative from the screen.

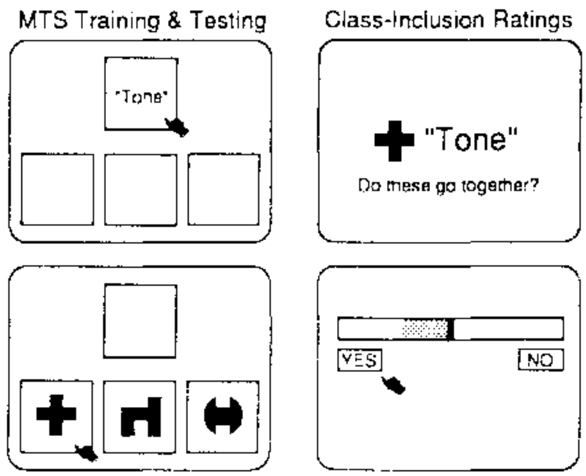


Figure 1. The participant's display during MTS training (left column of panels) and during the class inclusion rating procedure (right column of panels). See text for details.

Feedback and point consequences, described below, were established to teach participants to move the cursor into the "Yes" box and respond at a high rate when the stimulus set was class-consistent, and to move the cursor into the "No" box and respond at a low rate, or not at all, when the stimulus set was class-inconsistent. For example, to indicate maximal certainty that the stimuli in the set were class-consistent, a participant could respond 18 times in the "Yes" box. To indicate a lower degree of certainty that the stimuli were class-consistent, a participant could respond only a few times in the "Yes" box. To indicate maximal certainty that the stimuli were class-inconsistent, a participant could respond just once in the "No" box. To indicate a lower degree of certainty that the stimuli were class inconsistent, a participant 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. The word "CORRECT" or "INCORRECT," as appropriate to the outcome of the trial, was printed at the top of the screen. On the first five trials of each session only, the message, "Pressing repeatedly may or may not affect how many seconds you earn or lose" appeared at the bottom of the feedback screen. In the center of the screen was the message "TIME EARNED=X seconds" (for YES ratings about class-consistent stimulus sets, X=19 minus number of responses) or "TIME LOST=-X seconds." (for NO ratings about class-consistent stimulus sets, X=number of responses; for YES ratings about class-inconsistent stimulus sets, X=19 minus number of responses).

*Phase 3: Baseline Class-Inclusion Test*

Phase 3 began with one baseline session of an equivalence generalization test using the class-inclusion rating procedure. The procedures were identical to those of Phase 2 except that no feedback followed any trial, and the stimuli included all of the tones listed in Table 2 and the shapes shown Figure 2, configured as specified in Appendix B. This phase was used to ascertain whether any participants had any systematic patterns of responding to cross modal stimulus combinations prior to class training. Because participants had not yet been exposed to any of these stimuli, no systematic rating patterns were expected.

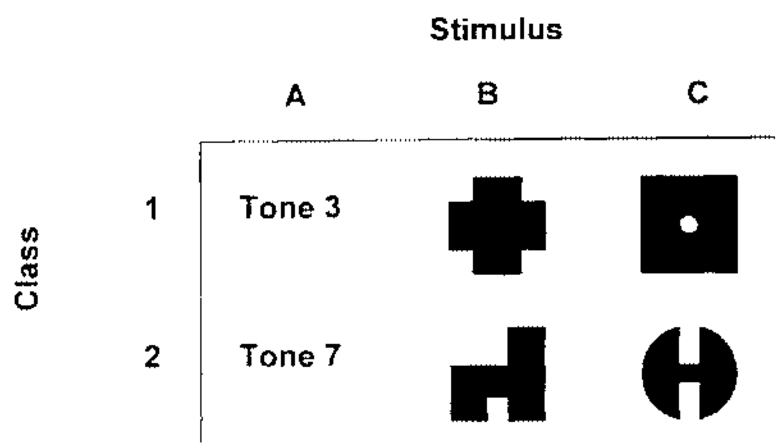


Figure 2. Stimuli used in the experiment.

*Phase 4: Conditional Discrimination Training*

A "one-to-many" training procedure (Saunders, Saunders, Williams, & Spradlin, 1993) was used to teach conditional discriminations prerequisite to the formation of two three-member, cross-modal equivalence classes using the stimuli shown in Figure 2 (A1-B1-C1 and A2-B2-C2). The A stimuli (tones) always served as the samples, and the B or C stimuli (shapes printed in red) served as correct or incorrect comparisons depending on the sample that was present on a given trial (see Appendix A). The auditory stimulus used as the A1 sample was a middle C, and the A2 sample was a C two octaves lower than the middle C (see Table 2).

The left column of panels in Figure 1 shows some key elements of the display during MTS trials. Instructions were those provided at the start of conditional discrimination training in Phase 2. Participants clicked the empty sample box to start each trial. A sample tone then sounded for 2 s, followed by the presentation of three visual comparison stimuli, located in three horizontally aligned comparison boxes below the sample box, with location counterbalanced across trials. Each correct MTS selection was worth 20 s of extra-credit time, and feedback initially followed every trial. After a participant scored 100% correct on two consecutive 12 trial sessions involving the A-B relations, and 100% on two consecutive sessions involving A-C relations, the A-B and A-C relations were mixed into single 24 trial sessions. The probability of feedback for mixed sessions began at 1.0 and, after two consecutive sessions at 100% accuracy, was decreased to 0.40 until the same criterion was met. Following each session, the total number of seconds earned for that session was displayed on the monitor for five seconds.

*Phase 5: Equivalence Tests*

This phase tested all the possible trained and derived relations from the two three-member stimulus sets (see participant instructions under Phase 2). Stimulus arrangements for Phase 5 test trials are shown in Appendix A. Each session was 52 trials long, and presented 12 training trials (6 each of A-B and A-C), 18 reflexivity trials (6 each of A-A, B-B, and C-C), 12 symmetry trials (6 each of B-A and C-A), and 12 equivalence, or combined symmetry-transitivity, trials (6 each of 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.

Trial format was similar to that used during conditional discrimination training, with modifications as appropriate to the stimuli. When reflexivity trials involved auditory stimuli (A1 or A2), participants mouse-clicked 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. As each tone played, a yellow asterisk appeared in one of the three comparison boxes, with location counterbalanced across trials. The asterisks remained present until participants made a selection by clicking the mouse in one of the corresponding comparison boxes. On symmetry test trials, sample stimuli were visual shapes, and the same procedure of presenting auditory comparison stimuli was used. Equivalence (combined) test trials presented only visual stimuli, and typical MTS procedures were used (as in Phase 2).

*Phase 6: Post-Equivalence Class-Inclusion Ratings*

This phase employed the class-inclusion rating procedure introduced in Phase 2 (Figure 1, right column of panels) and provided the primary data of the experiment. Instructions were the same as those given prior to the preliminary class-inclusion ratings of Phase 2. On each trial, a stimulus set was presented for 2 s, accompanied, on the first five trials only, by the printed message "Do these go together?" Participants labeled each set as class consistent ("Yes") or class inconsistent ("No"), and recorded a confidence rating as described previously. No feedback about earnings or performance followed any trial in this phase.

Appendix C lists the stimulus sets, which could include both tones and shapes. The tones included those used in conditional discrimination training and equivalence testing of Phases 4 and 5, plus probe tones (Table 2) from among those presented during the auditory discrimination pre-test in Phase 1. The shapes were those used in Phases 4 and 5 (Figure 2). Sessions were 136 trials long, and presented most possible combinations of 2 and 3 stimuli from the 13 stimuli (6 stimuli from training and 7 novel probe tones) used during this phase. On the 80 trials including only training stimuli, 40 sets were class-consistent, and 40 were class-inconsistent. Some sets included only visual stimuli. The remaining 56 trials presented visual stimuli used in training along with one of the 7 novel probe tones. On trials presenting two visual stimuli with a probe tone, the visual stimuli were always from the same class (e.g., B1 C1 rather than B2 C1) to insure that the auditory tone had to be considered in order to make a decision regarding class inclusion.

*Phases 7 & 8: Equivalence and Auditory Discrimination Post-Tests*

During Phase 7, the 52-trial equivalence test session from Phase 5 was repeated to ensure that exposure to the variety of stimulus sets during Phase 6 had not disrupted the equivalence classes evaluated during Phase 5. During Phase 8, the auditory discrimination test from Phase 1 was repeated to determine whether Phase 6 had altered the participants' discrimination among the tones.

*Dependent Measures*

On all MTS training and equivalence test trials, performances were measured by calculating the percent of experimentally designated class-consistent comparison selections (i.e., percent accurate). Class-inclusion ratings were summarized in two dependent variables. The first was the percent of stimulus sets chosen as class consistent (e.g., percent selections of the "YES" report box). The second was the magnitude of the confidence rating. Consistent with the contingencies established in Phase 2, response rates on YES and NO ratings represented measures of participants' "confidence" in the discrimination of a set as class-consistent or -inconsistent.

## RESULTS

*Auditory Discrimination Pre test and Post-Test*

Both prior to, and following, the main experiment, participants nearly always identified successively presented identical tones as "same" and nonidentical tones as "different." Overall accuracy was 98% or above on the auditory discrimination tests for all participants except S137, who scored 92% and 93% on the pre test and post-test, respectively. Discrimination patterns thus appeared not to have been globally altered by the experience of participating in the main study.

*Preliminary Training of the Class-Inclusion Rating Procedure*

Participants made few errors during MTS training and equivalence testing involving the A-B-C and 1-2-3 stimulus classes (Phase 2), as might be expected if the stimuli were part of a pre-existing class. The mean percent accurate was 97.34 (*S.E.* = 1.05) on MTS training trials, and 97.73 (*S.E.* = 1.51) on equivalence test trials. No participant required more than 3 sessions to demonstrate mastery (2 consecutive sessions at 100% accuracy) at any point in the training or testing.

Participants also made few errors on the class-inclusion ratings of preliminary training (mean accuracy = 96.29%, S.E. = 1.19). Table 3 summarizes these ratings. Class-inconsistent stimulus sets were routinely labeled as class-inconsistent, regardless of whether the sets included probe stimuli. The percentage of "Yes" selections ranged from 0 to 29, and mean response outputs ranged from 1.0 (the minimum) to 4.2 out of a maximum of 18. Class-consistent stimulus sets containing only training stimuli were routinely labeled as class consistent. The percentage of "Yes" selections was 90% or higher for all participants, and associated response rates were around 10 or higher, out of a maximum of 18, for 6 of the 7 participants. The exception was S138, who always made a single response whether selecting "Yes" or "No." When class-consistent training stimuli were accompanied by probe stimuli to which they were, presumably, thematically related, rating patterns changed little for 6 of the 7 participants. The exception was S136, whose percentage of "Yes" selections dropped from 94 to 58, and whose mean response output dropped from 14.9 to 5.6 out of a maximum of 18. Overall, by the end of preliminary training, all participants had demonstrated use of the class-inclusion rating system in a manner that was consistent with its intended purpose and with common-sense interpretations of stimulus groupings.

*Table 3. Class-inclusion ratings during preliminary training (phase 2), as a function of* whether the stimulus set was class-consistent or class-inconsistent, and whether it included training stimuli only (No Probe) or training stimuli plus a novel stimulus (probe). Percent "Yes" refers to stimulus set labelled as class-consistent. Responses refers to a rate-based "confidence rating." See text for details.

| | Participant | | | | | | |
|---|---|---|---|---|---|---|---|
| | 129 | 130 | 133 | 134 | 136 | 137 | 138 |
| *Percent "Yes"* | | | | | | | |
| Class-consistent | | | | | | | |
| No Probe | 100 | 98 | 98 | 96 | 94 | 90 | 96 |
| Probe | 100 | 96 | 92 | 92 | 58 | 83 | 92 |
| Class-inconsistente | | | | | | | |
| No Probe | 4 | 13 | 0 | 0 | 29 | 4 | 0 |
| Probe | 0 | 1 | 0 | 0 | 29 | 4 | 0 |
| *Responses* | | | | | | | |
| Class-consistent | | | | | | | |
| No Probe | 11.2 | 14.0 | 16.3 | 10.7 | 14.9 | 9.7 | 1.0 |
| Probe | 10.7 | 14.0 | 15.7 | 10.6 | 5.6 | 12.9 | 1.0 |
| Class-inconsistent | | | | | | | |
| No Probe | 1.7 | 3.1 | 1.1 | 1.2 | 2.6 | 4.2 | 1.0 |
| Probe | 1.7 | 3.1 | 1.1 | 1.0 | 1.4 | 3.2 | 1.0 |

### *Conditional Discrimination Training and Equivalence Tests*

Individual conditional discrimination training and equivalence test scores from the main part of the study are shown in the middle columns of Table 4.

Participants typically required few sessions to meet mastery criteria at each step of the training (Phase 4). Mean accuracy for all participants across all training sessions was 96.77% (*S.E.* = 0.86). The results from Phase 5 (right-hand columns of Table 4) suggested that three-member equivalence classes had formed for all 7 participants. Mean accuracy for all participants across all equivalence tests was 96.78% (*S.E.* = 1.21).

Table 4. Percent correct on class-inclusion pretest and posttests; and on condition discrimination training and equivalence testing of the main study. Values within a column for individual subjets show performance in successive sessions. In conditional Discrimination Training, AB = sessions with A-B relations, AC = sessions with A-C relations, and Mix = Mixed Training Review, sessions with both A-B and A-C relations.

| | Class-Inclusion Tests | | Cond. Discrimination Training | | | Equivalence Tests | |
|---|---|---|---|---|---|---|---|
| Subjec | Pre-test | Post-test | AB | AC | Mix | Main | Post-test |
| 129 | 57 | 91 | 100 | 83 | 96 | 94 | 92 |
| | | 66 | 100 | 100 | 100 | 92 | |
| | | 55 | | 92 | | | |
| 130 | a | 100 | 100 | 92 | 100 | 98 | 92 |
| | | 99 | 100 | 100 | 100 | 100 | |
| | | 98 | | | | | |
| 133 | 50 | 95 | 75 | 100 | 79 | 98 | 100 |
| | | 95 | 100 | 100 | 96 | 96 | |
| | | 98 | 100 | | 100 | | |
| 134 | 53 | 91 | 100 | 100 | 100 | 100 | 98 |
| | | 93 | 100 | 100 | 96 | 100 | |
| | | 98 | | | | | |
| 136 | 48 | 99 | 92 | 100 | 96 | 100 | 100 |
| | | 96 | 100 | 100 | 100 | 100 | |
| | | 99 | | | | | |
| 137 | 49 | 85 | 92 | 100 | 100 | 100 | 97 |
| | | 91 | 100 | 100 | 100 | 85 | |
| | | 96 | | | | | |
| 138 | 49 | 95 | 83 | 92 | 96 | 92 | 92 |
| | | 91 | 100 | 92 | 100 | 100 | |
| | | 93 | 100 | | | | |

[a] Data were lost due to equipment failure.

### *Class-Inclusion Ratings of the Main Study*

The leftmost columns in Table 4 show that the baseline accuracy of class-inclusion ratings about sets of stimuli from the main study, made prior to conditional discrimination training and equivalence testing, tended to be around 50% (chance). Following demonstration of equivalence, accuracy improved to over 90% for all participants except S129, whose post-equivalence accuracy initially was 92%, but dropped to 55% across two subsequent sessions.

Recall that the class-inclusion rating trials differed in terms of which tones were paired with visual stimuli from within a single equivalence class. Ratings indicating class consistency were expected to be most common, and of highest magnitude, for stimulus sets including a training tone and class-consistent visual stimuli. Ratings indicating class consistency were expected to become less common, and of lower magnitude, as probe tones became increasingly different from the training tone. The generalization gradients based on the training tones were expected to peak at different tones depending on the visual stimuli with which tones were paired.

Baseline class inclusion ratings (for economy of presentation, not shown) appeared to be unrelated to either the type of tone or the visual stimuli included in the stimulus set. The percentage of "YES," or class-consistent, ratings was low and variable, ranging unsystematically across the 9 tones from 11% to 36% (*S.E.* = 9.7 to 14.1). Importantly, there was no obvious evidence of differential ratings as a function of the visual stimuli with which tones were paired. For example, when the tone later used as stimulus A1 was paired with Class 1 stimuli, ratings were not significantly different than when the same tone was paired with Class 2 visual stimuli, $t(6) = .65$, $p = .54$. Similarly undifferentiated ratings occurred when the tone later used as stimulus A2 was paired with visual stimuli from Class 1 versus Class 2, $t(6) = 1.04$, $p = .34$. Such outcomes were anticipated given that participants had no training history of associating the tones and the visual stimuli.

Figures 3 and 4 show class-inclusion ratings made after equivalence classes had formed (Phase 6). The data in these figures combine ratings made when stimulus sets incorporated two stimuli (two visual stimuli, or one visual stimulus and one tone) with ratings made when stimulus sets incorporated three stimuli (two visual stimuli and one tone). Preliminary analyses showed no systematic rating differences as a function of whether the stimulus sets had two or three members.

Figure 3 shows the percentage of "Yes" selections for the nine tones as a function of whether they were paired with visual stimuli from Class 1 or Class 2. Along the abscissa of each panel, the training tones are designated with the data symbols associated with the two stimulus classes. "Yes" selections nearly always occurred when Class 1 visual stimuli were presented with Tone 3, and when Class 2 stimuli were presented with Tone 7. "Yes" selections almost never occurred when Class 1 stimuli were presented with Tone 7, and when Class 2 stimuli were presented with Tone 3. Thus, "Yes" selections reliably designated class-membership as expected based on the results of equivalence tests.

When probe tones accompanied the visual stimuli, the probability of "Yes" selections was positively related to the degree of similarity between the probe tone and the training tone associated with the visual stimuli. For example, only tones similar to Tone 3 tended to evoke "Yes" selections when presented in combination with Class 1 visual stimuli. Only tones similar to Tone 7 tended to evoke "Yes" selections when presented in combination with Class 2 visual stimuli. Across participants, the extent of this generalization ranged from pronounced (e.g., S138) to minimal (e.g., S134), but all participants produced at least 65% "Yes" responses to at least one probe tone for each class of visual stimuli.

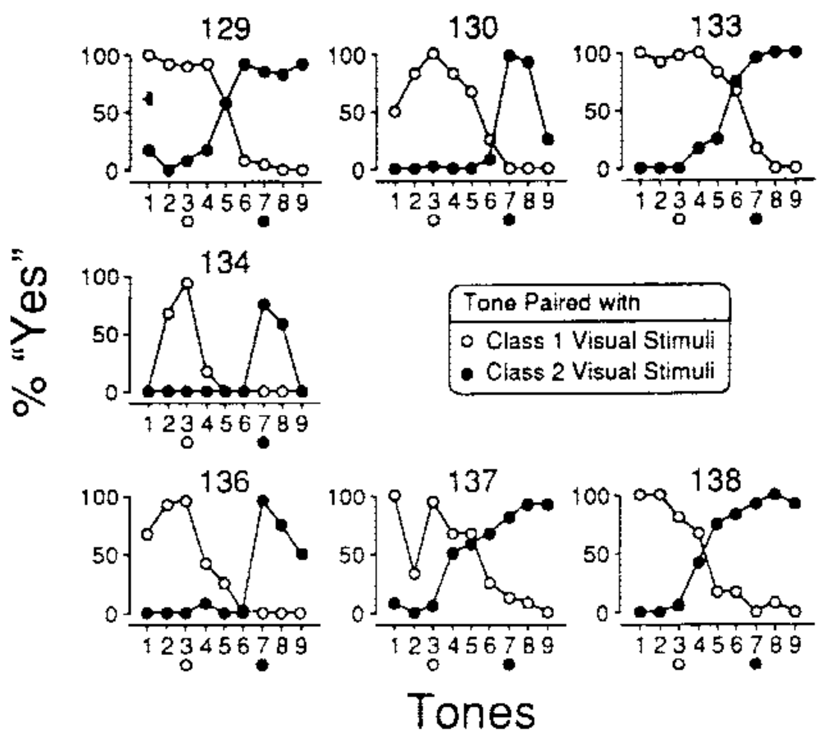


Figure 3. The percentage of "Yes" selections for the nine tones as a function of whether they were paired with visual stimuli from Class 1 or Class 2. Along the abscissa of each panel, the training tones are designated with the data symbols associated with the two stimulus classes.

Figure 4 shows the mean class-inclusion rating magnitudes for the nine tones as a function of whether they were paired with visual stimuli from Class 1 or Class 2. Results were similar to those shown in Figure 3, with two exceptions. First, S137 showed less class-related differentiation in rating magnitudes than in percentages of "Yes" responses. Second, consistent with his performance during the training of the class-inclusion rating procedure (Phase 2), S138 made a single response regardless of whether selecting "Yes" or "No."

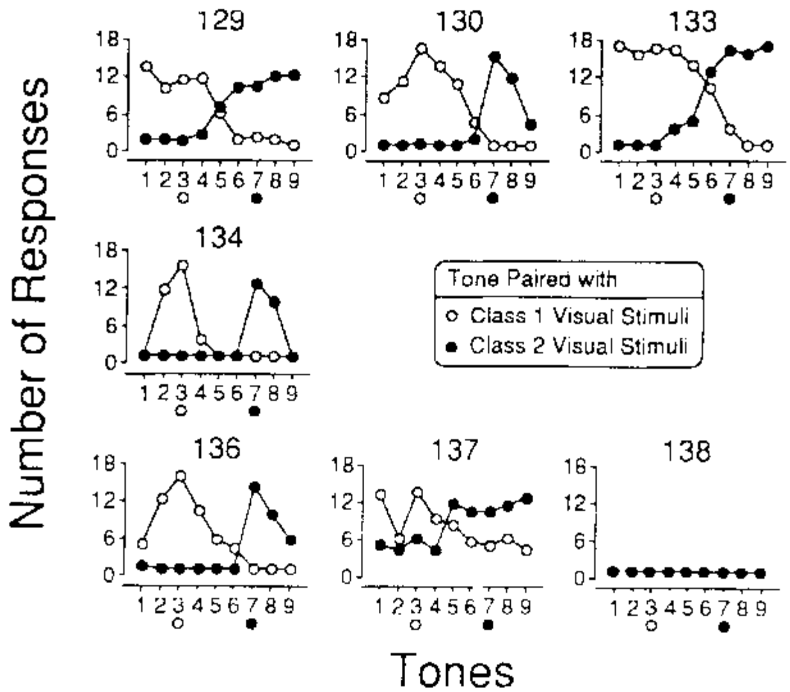


Figure 4. The mean class-inclusion rating magnitudes for the nine tones as a function of whether they were paired with visual stimuli from Class 1 or Class 2. Along the abscissa of each panel, the training tones are designated with the data symbols associated with the two stimulus classes.

### *Follow-up Equivalence Test*

Phase 7 repeated the equivalence tests of Phase 4 to determine whether equivalence classes remained intact following the class-inclusion rating procedure. The data in Table 4 support the conclusion that equivalence classes remained intact. Mean overall accuracy was 97% (*S.E.* = 1.36).

## DISCUSSION

The key results of the present study can be summarized as follows: Before starting the main study, it was demonstrated that the tones used for subsequent testing were all discriminable from one another. Participants also learned to use a rating procedure for categorizing sets of stimuli as class-consistent or inconsistent. After completing conditional discrimination training with new stimuli, the participants demonstrated the formation of two, three-member cross-modal equivalence classes involving both shapes and tones. Subsequently, the class-inclusion rating procedure was re-instituted, this time with sets of stimuli drawn from the equivalence classes. Occasionally, the tones of the equivalence classes were replaced by novel tones. The probability that these novel sets would be rated as class-consistent was a function of the auditory distance between the novel tone and the tone explicitly included in the equivalence class. For some participants, the test tones occasioned the same performances occasioned by test trials that contained tones that were class members. Therefore, the data for those participants support the view that the extension of equivalence classes by generalization reflects the merger of perceptual classes and equivalence classes, as noted by Fields et al. (1997).

The present results also suggest that the merger of perceptual and equivalence classes is not hindered by the cross-modal relationships inherent in some equivalence classes. That is, generalized equivalence classes apparently can be created in cross-modal sets of stimuli. Cross-modal equivalence classes have been demonstrated before (Bush, 1993; Sidman & Tailby, 1982), and it seems reasonable to expect that these classes can expand based on primary generalization. Many natural human categories are both cross-modal (as evidenced by the relations between written words, spoken words, and non-language referents), and open-ended (incorporating numerous exemplars). A convincing account of categorization based on conditioning processes must be compatible with these observations.

The conclusions prompted by this study must be viewed as preliminary pending the resolution of three issues. First, a demonstration of generalized equivalence depends on an unequivocal demonstration of equivalence. Although the equivalence test performances of the present study were consistent with expected emergent relations, the study's design does not rule out the possibility that "correct" selections sometimes occurred on another basis. Specifically, stimulus configuration can guide performance on equivalence tests, independent of the relations specified by experimental contingencies (e.g., Harrison & Green, 1990; Carrigan & Sidman, 1992). Most MTS trials of the present study included three comparison stimuli drawn from two stimulus classes: a class-consistent stimulus and two class-inconsistent stimuli. On some trials, it would have been possible for participants to select the correct comparison stimulus without attending to the sample stimulus, thus producing partially spurious equivalence outcomes. For example, when B1 and C1 appeared together as comparisons, the remaining stimulus always was correct. Due to the way that stimulus-representation frequency was counterbalanced across trials (Appendix A), neither B1 nor C1 ever was correct when the other was present. It is worth noting that participants may prefer response strategies that apply consistently (Sidman, 1994), and such configural cues were not available on all trials. Nevertheless, a procedure employing only one comparison stimulus from each stimulus class would be required to resolve ambiguity about the validity of the equivalence classes.

A second source of ambiguity derives from the means by which generalization was assessed in the present study. Previous studies of generalized equivalence classes have employed MTS procedures during generalization tests. The present study used a class-inclusion rating procedure, and it remains unknown whether cross-modal, generalized equivalence classes would emerge in MTS-based procedures. Figure 5 suggests that the present results are at least superficially similar to those of previous studies using MTS procedures. The top panel shows data from Fields et al. (1991). The Y-axis shows the proportion of MTS choices as a function of the similarity between trained stimuli and probe stimuli. Trained stimuli are shown inside the hashed boxes; novel stimuli are data points not inside those lines. The bottom panel shows mean data derived from the upper panel of Figure 5, with left and right panels collapsed. The Y-axis shows proportion of "Yes" ratings as a function of the similarity between trained tones and probe tones. Trained stimuli are also shown inside the hashed boxes. Despite differences in stimuli, procedures, and dependent measures, the two gradients are quite similar. The concordance would be strengthened, however, if MTS-based generalization tests were conducted with the same participants as the class inclusion tests.

A third point for consideration is the possibility that our results are idiosyncratically linked to the preliminary training procedures. Presumably, everyday categories expand spontaneously through stimulus generalization. In the present study, preliminary training to instate the class-inclusion rating procedure involved reinforcement for grouping novel stimuli (the lower case letter b and the vertical arrangement of two circles) into an established class. It is possible that this pretraining may be one of the parameters that are required to obtain the present results, though this is not likely given the results of previous studies that did not utilize a pretraining protocol (Fields et al., 1991, 1993, 1996, 1997). Nevertheless, a replication of this study using different preliminary training procedures would be necessary to rule out this possibility.

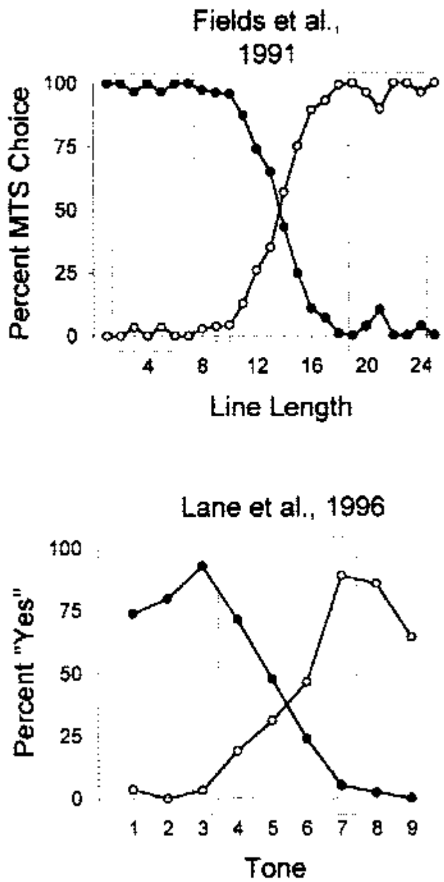


Figure 5. Comparison of data from Fields et al., 1991 (top panel) and data from the present study (bottom panel). Data for both studies are presented as group means. Solid circles and open circles represent Class1 and Class 2 stimuli, respectively, as a function of the physical dimensions of the stimuli (short to long line length and high to low tone pitch). Data points inside the dotted lines represent trained stimuli, points not inside the dotted line represent novel stimuli.

In summary, several recent studies have supported the concept of the generalized equivalence class as a means by which structural and functional processes may combine in the formation of stimulus classes that approximate simple categories (Adams et al., 1993; Fields et al., 1991, 1993, 1996, 1997). The present study extends that work by suggesting, for the first time, that generalized equivalence classes can extend across stimulus modalities, which is an important prerequisite for any phenomenon held to contribute to natural categorization. If confirmed by additional studies that resolve the design-related ambiguities identified here, the present study will help to buttress the plausibility of a conditioning-based account of categorization in all of its everyday complexity.

## REFERENCES

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

Barsalou, L. W. (1992). *Cognitive psychology: An overview for cognitive scientists*. Hillsdale, NJ: Erlbaum.

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

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.

Carrigan, P. F., Jr., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 58*, 459-504.

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.

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

Fields, L., Adams, B. J., Buffington, D. M., Yang, W., & Verhave, T. (1996). Response transfer between stimuli in generalized equivalence classes: A model for the establishment of natural kind and fuzzy superordinate categories. *The Psychological Record, 46*, 665-684.

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., Adams, B., & Verhave, T. (1991). Stimulus generalization and equivalence classes: A model for natural categories. *Journal of the*

*Experimental Analysis of Behavior, 55*, 305-312.

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

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

Harrison, R. J., & Green, G. (1990). Development of conditional and equivalence relations without differential consequences. *Journal of the Experimental Analysis of Behavior, 54*, 225-237.

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.

Saunders, K. J., Saunders, R. R., Williams, D. C., & Spradlin, J. S. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record, 43*, 725-744.

Saunders, R. R. (1996). From review to commentary on Roche and Barnes: Toward a better understanding of equivalence in the context of relational frame theory. *The Psychological Record, 46*, 477-487.

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

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. (Eds.), *Thinking: An invitation to cognitive science* (Vol. 3, 2nd Ed.). Cambridge, MA: MIT Press.

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

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.

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.

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

| | | | Comparisons | | | Times Presented |
|---|---|---|---|---|---|---|
| Phase | Description | Sample | CO + | CO- | CO- | Per Session |
| 2 | Preliminary training | | | | | |
| | A-B | A | B | 2 | 3 | 6 |
| | | 1 | 2 | B | C | 6 |
| | A-C | A | C | 2 | 3 | 6 |
| | | 1 | 3 | B | C | 6 |
| | Mixed Review | Same as above | | | | 6 each (24) |
| | Equivalence tests | | | | | |
| | Trained | Same as above | | | | 1 each |
| | Reflexive | A | A | 1 | 2 | 1 |
| | | 1 | 1 | A | B | 1 |
| | | B | B | 1 | 2 | 1 |
| | | 2 | 2 | A | B | A |
| | | C | C | 2 | 3 | 1 |
| | | 3 | 3 | B | C | 1 |
| | Symmetrical | B | A | 1 | 2 | 2 |
| | | 2 | 1 | A | B | 2 |
| | | C | A | 2 | 3 | 2 |
| | | 3 | 2 | B | C | 2 |
| | Combined | B | C | 2 | 3 | 2 |
| | | 2 | 3 | B | C | 2 |
| | | C | B | 2 | 3 | 2 |
| | | 3 | 2 | B | C | 2 |
| 4 | MTS Training | | | | | |
| | A-B | A1 | B1 | B2 | C2 | 6 |
| | | A2 | B2 | B1 | C1 | 6 |
| | A-C | A1 | C1 | B2 | C2 | 6 |
| | | A2 | C2 | B1 | C1 | 6 |
| | Mixed Review | Same as above | | | | 6 each (24) |
| 5 | Equivalence Tests | | | | | |
| | Trained | Same as above | | | | 3 each (12) |
| | Reflexive | A1 | A1 | A2 | | 2 |
| | | A2 | A2 | A1 | | 2 |
| | | B1 | B1 | B2 | C2 | 2 |
| | | B2 | B2 | B1 | C1 | 2 |
| | | C1 | C1 | B2 | C2 | 2 |
| | | C2 | C2 | B1 | C1 | 2 |
| | Symmetrical | B1 | A1 | A2 | | 4 |
| | | B2 | A2 | A1 | | |
| | | C1 | A1 | A2 | | 4 |
| | | C2 | A2 | A1 | | 4 |
| | Combined | B1 | C1 | B2 | C2 | 3 |
| | | B2 | C2 | B1 | C1 | 3 |
| | | C1 | B1 | B2 | C2 | 3 |
| | | C2 | B2 | B1 | C1 | 3 |

Appendix B. Stimulus configurations for class-inclusion test trials of Phase 2 (preliminary training). Actual stimuli are shown. Stimuli appeared in the left-to-right sequences shown below. Each stimulus configuration occurred once per session.

| Sets Including Training Stimuli | | | | | | Sets Including Novel Shapes | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Class Consistent | | | Class Inconsistent | | | Class Consistent[a] | | | Class Inconsistent[a] | | |
| A | B | C | A | C | 3 | A | C | b | A | C | : |
| A | C | B | A | B | 2 | C | b | A | C | : | A |
| B | A | C | B | 3 | 1 | b | A | C | : | A | C |
| B | C | A | B | 1 | 3 | 1 | 3 | : | 1 | 3 | b |
| C | A | B | B | C | 1 | 3 | : | 1 | 3 | b | 1 |
| C | B | A | C | 2 | A | : | 1 | 3 | b | 1 | 3 |
| 1 | 2 | 3 | C | A | 1 | A | b | | 1 | b | |
| 1 | 3 | 2 | C | 2 | 3 | C | b | | b | 1 | |
| 2 | 1 | 3 | 1 | 3 | B | b | C | | 3 | b | |
| 2 | 3 | 1 | 1 | 3 | C | 1 | : | | : | A | |
| 3 | 1 | 2 | 1 | A | B | : | 1 | | C | : | |
| 3 | 2 | 1 | 2 | C | 2 | 3 | : | | : | C | |
| A | B | | 2 | A | 1 | | | | | | |
| A | C | | 2 | B | C | | | | | | |
| B | A | | 3 | C | 3 | | | | | | |
| B | C | | 3 | 2 | B | | | | | | |
| C | B | | 3 | C | 1 | | | | | | |
| C | A | | A | 1 | A | | | | | | |
| 1 | 2 | | A | 1 | | | | | | | |
| 1 | 3 | | A | 2 | | | | | | | |
| 2 | 1 | | B | 3 | | | | | | | |
| 2 | 3 | | B | 1 | | | | | | | |
| 3 | 2 | | B | 2 | | | | | | | |
| 3 | 1 | | C | 3 | | | | | | | |
| | | | C | 1 | | | | | | | |
| | | | C | 2 | | | | | | | |
| | | | 1 | 3 | | | | | | | |
| | | | 1 | A | | | | | | | |
| | | | 1 | B | | | | | | | |
| | | | 2 | C | | | | | | | |
| | | | 2 | A | | | | | | | |
| | | | 2 | B | | | | | | | |
| | | | 3 | C | | | | | | | |
| | | | 3 | A | | | | | | | |
| | | | 3 | B | | | | | | | |
| | | | | C | | | | | | | |

[a] Class consistentecy defined in terms of expected classification of novel stimuli based on typical pre experimental history. See text for details.

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

| Sets Including Training Stimuli Only | | | | | | | | Sets Including Probe Tones | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Class Consistent Sets | | | | Class Inconsistent Sets | | | | | | | | | | | | | |
| Auditory-Visual | N | Visual Only | N | Auditory-Visual | N | Visual Only | N | Visual Stimuli | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| B1 C1 A1 | 4 | B1 C1 | 2 | B1 C1 A2 | 4 | B1 C2 | 1 | B1 C1 | 1 | 1 | * | 1 | 1 | 1 | * | 1 | 1 |
| C1 B1 A1 | 4 | C1 B1 | 2 | C1 B1 A2 | 4 | C1 B2 | 1 | C1 B1 | 1 | 1 | * | 1 | 1 | 1 | * | 1 | 1 |
| B2 C2 A2 | 4 | B2 C2 | 2 | B2 C2 A1 | 4 | B2 C1 | 1 | B2 C2 | 1 | 1 | * | 1 | 1 | 1 | * | 1 | 1 |
| C2 B2 A2 | 4 | C2 B2 | 2 | C2 B2 A1 | 4 | C2 B1 | 1 | C2 B2 | 1 | 1 | * | 1 | 1 | 1 | * | 1 | 1 |
| B1 A1 | 4 | | | B1 A2 | 4 | B1 B2 | 1 | B1 | 1 | 1 | * | 1 | 1 | 1 | * | 1 | 1 |
| C1 A1 | 4 | | | C1 A2 | 4 | B2 B1 | 1 | C1 | 1 | 1 | * | 1 | 1 | 1 | * | 1 | 1 |
| B2 A2 | 4 | | | B2 A1 | 4 | C1 C2 | 1 | B2 | 1 | 1 | * | 1 | 1 | 1 | * | 1 | 1 |
| C2 A2 | 4 | | | C2 A1 | 4 | C2 C1 | 1 | C2 | 1 | 1 | * | 1 | 1 | 1 | * | 1 | 1 |