*THE MERGER AND DEVELOPMENT OF EQUIVALENCE
CLASSES BY UNREINFORCED CONDITIONAL
SELECTION OF COMPARISON STIMULI*

RICHARD R. SAUNDERS, KATHRYN J. SAUNDERS,
KIMBERLY C. KIRBY, AND JOSEPH E. SPRADLIN

BUREAU OF CHILD RESEARCH, UNIVERSITY OF KANSAS

Three experiments assessed the likelihood that subjects with histories of equivalence class development would respond conditionally on new discriminations in the absence of differential consequences for responses. In the first two experiments, two groups of subjects with different experimental histories, but whose performances showed four equivalence classes, responded on trials without explicit reinforcement involving samples from two of the classes and comparisons from the other two classes, in a two-choice matching-to-sample format. Subjects consistently selected a particular comparison in the presence of a particular sample. Subsequent tests showed the emergence of equivalence relations between stimuli from classes linked by the unreinforced conditional selections. Subsequently, in Experiment II, the subjects' responses in the conditional selection trials were reinforced if the selection was reversed from that made previously. Although reversed selection was maintained, 2 of the 3 subjects continued to perform on equivalence relation trials according to their original unreinforced selections. In the third experiment, these 2 subjects responded on a series of conditional discriminations involving three new pairs of sample stimuli and one new pair of comparison stimuli. No explicit reinforcement followed responses on any trial in this experiment. Subsequent tests for equivalence between sample stimuli revealed the development of two equivalence classes.

Key words: stimulus classes, equivalence relations, unreinforced conditional selection, matching-to-sample, conditional discrimination, mentally retarded

The study of stimulus classes usually involves teaching a subject a series of conditional discriminations and then testing to determine whether new conditional discriminations emerge without direct training. Often two-choice conditional discriminations are used in these studies (Devany, Hayes, & Nelson, 1986; Dixon, 1978; Dixon & Spradlin, 1976; Dube, McIlvane, MacKay, & Stoddard, 1987; Saunders, Wachter, & Spradlin, 1988; Spradlin, Cotter, & Baxley, 1973; Spradlin & Dixon, 1976; Spradlin & Saunders, 1986; Wetherby, Karlan, & Spradlin, 1983).

A simple stimulus-class study could begin by teaching a two-choice conditional discrimination; that is, the subject is taught that when stimulus A_1 occurs as the sample, stimulus B_1 is the correct comparison (denoted as A_1B_1);

when A_2 is the sample, B_2 is the correct comparison (A_2B_2). The subject might also be taught that when A_1 is the sample, C_1 is the correct comparison (A_1C_1), and that when A_2 is the sample, C_2 is the correct comparison (A_2C_2). Following this training, tests for the development of stimulus classes can be conducted. In these trials, when C_1 and C_2 are the comparisons, if C_1 is selected when B_1 is the sample and C_2 is selected when B_2 is the sample, the performance may indicate the development of classes. Similarly, selecting B_1 when C_1 is the sample and selecting B_2 when C_2 is the sample are further indications of class development. Performances showing these relations are indicative of a particular type of stimulus class that is labeled an equivalence class (Sidman & Tailby, 1982).

The ease with which human subjects can be taught two-choice discriminations varies considerably across subjects, with some subjects taking many sessions to learn the first conditional discrimination (Saunders et al., 1988; Wetherby et al., 1983). However, after one or two conditional discriminations have been taught to the subject, subsequent two-choice

Preparation of this article was supported in part by Federal NICHD grants 5-P30HD02528 and 1-P01HD18955 to the Bureau of Child Research and grant 5-T32HD07173 to the Department of Human Development and Family Life, University of Kansas.

Requests for reprints should be sent to Richard R. Saunders, Bureau of Child Research, P.O. Box 738, Parsons, Kansas 67357.

conditional discriminations are usually learned rapidly. In fact, a subject who has been taught a number of two-choice conditional discriminations may respond correctly on every trial of a new conditional discrimination subsequent to the first trial, regardless of whether the response on the first trial was correct or incorrect (across subjects, first-trial performance is at the level of chance) (Wetherby *et al.*, 1983).

Dixon (1977) demonstrated that if retarded youths are taught in repeated trials to select a specific stimulus (B_1) when B_1 and B_2 are presented as comparisons and A_1 is presented as the sample, many will select B_2 when presented A_2 as the sample in the first trial of this type. That is, many human subjects respond in a conditional manner even though their immediate experimental history required only simple discrimination learning.

Dixon's findings and the finding that human subjects who have been taught a number of two-choice discriminations respond at chance (across subjects) on the first trial of a new discrimination and subsequently perform correctly suggest that events in a single trial may be sufficient to produce consistent subsequent responding. If the first response is followed by reinforcement, no further errors are likely to occur. In the absence of a consequence for the first response, subsequent responding also may be conditional and consistent with the sample-comparison relation shown on the first trial.

Recently, we trained a mentally retarded adult with no prior laboratory history on a series of two-choice conditional discriminations as follows: A_1B_1 , A_2B_2 , then B_1C_1 , B_2C_2 , and then C_1D_1 , C_2D_2 . After the subject's performance was accurate and stable in sessions containing a mix of all the trained relations, unreinforced probe trials for the derived relations A_1D_1 and A_2D_2 were interspersed among a mixture of the reinforced training trials. While maintaining correct performance on the training trials, the subject consistently responded on the probes by selecting comparison stimulus D_1 in the presence of sample stimulus A_2 and comparison stimulus D_2 in the presence of sample stimulus A_1 . Clearly the subject was responding conditionally, but not as expected. These results strongly suggest that subjects who have been taught a series of conditional discriminations will respond conditionally when presented previously unre-

lated samples and comparisons, even in the absence of explicit reinforcement. The three experiments presented below address this issue.

If subjects demonstrate conditional responding in the absence of explicit reinforcement, do these conditional relations function in the same way as the reinforced relations of a conditional discrimination? Sidman, Kirk, and Willson-Morris (1985) reported that by training the linking relations between two small stimulus classes, one larger class emerged. To do this, they taught subjects to select a stimulus from one class when it was presented as a comparison stimulus in the presence of a sample that was a stimulus from the other class. The relations were trained between three pairs of three-member classes, resulting in performances that showed the emergence of three six-member classes. In that study, all of the linking relations were trained with explicit reinforcement of comparison selection. Would this merger of classes be likely to occur if the linking relations were not trained? If subjects acquired four equivalence classes via two-choice discriminations and samples then were presented from two classes and comparisons presented from the two unrelated classes, would the subjects respond conditionally in the absence of reinforcement? If they do respond conditionally, will the four classes be merged into two larger classes on the basis of such conditional responding?

To answer these questions, an experiment was conducted with 2 subjects who had extensive histories of conditional discrimination learning and of showing equivalence classes by their performances on probes. The results of that experiment led to two additional experiments, one with naive subjects who were given similar but briefer histories, and a follow-up experiment with 2 of the subjects who had participated in the second experiment.

EXPERIMENT I

METHOD

Subjects

Two mentally retarded adults who had served in an earlier experiment (Subjects RF and TB in Saunders *et al.*, 1988) participated in this experiment. These subjects had histories of performing conditional discriminations

Table 1
Subject Characteristics.

Subject	Age	Length of institution-alization	Sex	Measured I.Q.	Diagnosis
TB	25 years	16 years	M	62	Mild mental retardation, seizures, visual, speech, motor, and behavior dysfunction
RF	18 years	6 years	M	63	Mild mental retardation related to "oddities of movement" and speech abnormalities
LM	26 years, 7 months	4 months	M	63	Mild mental retardation, communication delays, moderate unilateral hearing loss, and behavioral impairment
SD	21 years, 3 months	9 years, 4 months	M	63	Mild to moderate mental retardation, speech dysfunction, and behavioral impairment
JP	14 years, 7 months	7 months	M	65	Mild mental retardation, mild unilateral hearing loss, speech, visual/spatial, and behavioral impairments

(273 and 245 sessions, involving 20 and 21 trained conditional discriminations, respectively), including discriminations with both visual stimuli and auditory stimuli as sample stimuli and visual stimuli as the comparison stimuli. The age, sex, measured intelligence, and diagnosis of each subject (all were verbal) are shown in Table 1.

Before beginning the experiment, all of the possible untrained relations involving stimuli lettered A through H shown in Figure 1 had been tested (but never reinforced), and two eight-member classes were demonstrated. The relations I_1L_1 and I_2L_2 and their reciprocals had been tested (but never reinforced) following the training shown with the solid arrows. These subjects' performances strongly suggested that two additional four-member classes were also formed. (Additional tests would have been necessary for conclusive evidence of the two equivalence classes.)

Apparatus and General Procedures

Subjects were seated at a large table that supported a stimulus display box, shown in Figure 2, that housed the computer equipment used to control the stimulus displays, response consequences, and so on. The equipment was the same as described in Saunders et al. (1988). The face of the box was 120 cm wide by 61 cm tall. A horizontal row of three display windows measuring 5 by 5 cm each, located 2.5 cm apart, was cut into the wall 105 cm above the floor. During the experimental sessions, an Apple® monochrome monitor was positioned so that when stimuli were displayed on

the screen, they were visible and centered in each window. Stimuli were displayed on the monitor screen using Apple's high-resolution graphics. Plastic, spring-loaded buttons, were mounted one under each window. Each button had an exposed diameter of 2.5 cm and was recessed 1 cm. To the subject's lower left was a plastic box into which pennies or tokens could be dispensed through a 5 by 5 cm opening in the front wall of the display box.

The general procedures for each session were the same as described in Saunders et al. (1988). To begin each session, and in each subsequent trial of a session, the sample stimulus was displayed in the center window. A press of the button under this window resulted in the concurrent display of two comparison stimuli, one to the left of the sample and one to the right. Further presses to the center button had no programmed consequences. A press of the button under either comparison window resulted in various programmed consequences, the temporary removal of all displays for 2 s, and the presentation of the next sample stimulus.

During training, "correct" selections of the comparison stimuli resulted in a computer-generated auditory jingle of approximately 1-s duration that imitated the musical phrase, "ta da, ta daa," that often prompts fans at sporting events to yell "charge." The delivery of a penny as described above accompanied the jingle. "Incorrect" responses resulted in a buzzer-like sound of approximately 1-s duration. On probe trials, to be described later, and on any training trials with no programmed consequences, selections of comparison stimuli resulted only in

Fig. 1. The trained (solid arrows) and probed (dashed arrows) relations among the eight stimuli in each of Classes 1 (A_1 – H_1) and 2 (A_2 – H_2), and the four stimuli in each of the probable Classes 3 (I_1 – L_1) and 4 (I_2 – L_2), that were part of Subjects RF and TB's experimental histories before the start of Experiment I. The numbers above the solid training arrows refer to the order of the two phases in which these relations were trained in an earlier experiment. The subscript associated with each letter designating a stimulus indicates the arbitrary sample-comparison assignment made by the experimenters. Thus, in training, a comparison stimulus with the subscript "1" was the correct comparison in the presence of a sample stimulus with the subscript "1"; similarly, comparisons with the subscript "2" were correct for samples with the subscript "2."

Fig. 2. A drawing of the apparatus mounted on a table. Stimuli are shown visible in the stimulus-presentation windows cut into the display wall, with the response buttons mounted beneath the windows and reinforcer aperture and cup to the extreme left. The computer and feeder were positioned behind the wall and were not visible when subjects were seated.

the removal of all displays for 2 s and advancement to the next trial.

Each session in the experiment consisted of 32 trials, regardless of the type of session conducted. In each phase of the experiment, a criterion of either one session of 100% correct or "expected" performance or two consecutive sessions of 90% or higher was required to permit advancement to the next phase or stage within a phase. Subjects were required to maintain performances on baseline trials of 90% or higher during probe sessions; otherwise, retraining to 90% or higher was instituted. When sessions involved unreinforced

conditional selection trials (to be described below), if the subject's overall performance met criterion in one session, a second session was required to ensure consistency of the unreinforced selection. In sessions in which no explicit programmed consequences occurred following responses, the subjects were paid at the end of the session, receiving one penny for each correct response on a training trial and one penny noncontingently for every other type of trial in the session. Subjects were paid for their performances in a lump sum and were not informed of the method of computation. When the subjects exchanged the pennies for larger coins, they could observe and often commented upon totals of less than 32 cents. Thus, subjects were aware of "imperfect" performances, but received no direct information regarding which responses had not produced a penny. Questions regarding payment were answered with comments to the effect that the "computer keeps track of the ones you get right and tells me how much to pay you." No subject disputed the computer's computation.

Sessions were conducted each weekday un-

less the subject was ill, away from the facility, or had a temporary conflict. Usually, two to five sessions were conducted each day, depending upon available time for the subject and experimenter. The interval between sessions averaged 5 min.

RESULTS

Phase 1: Tests for Unreinforced Conditional Selection

Initially, sessions were conducted in which two stimuli, A_1 and A_2 from the first two classes, served as sample stimuli, and two stimuli from the third and fourth classes, I_1 and I_2 , served as comparisons on eight of the 32 trials. These trials are hereinafter labeled "unreinforced conditional selection trials" (see the dotted lines in Figure 3). The remaining 24 trials were the baseline training trials AB, AC, AD, and AE for each of the two classes (arrows marked "1" in Figure 3). These training trials had been used in the first phase of training in the earlier experiment that led to the development of the eight-member classes. Selections of comparison stimuli had no programmed consequences on any trial of either type, except advancement to the next trial following a 2-s intertrial interval.

In this phase, subjects maintained accurate responding on the former baseline trials and consistently (always) selected a specific (the same) comparison in the presence of a particular sample on the eight unreinforced conditional selection trials (see Figure 3). In other words, they responded conditionally in the absence of explicit differential reinforcement. Subject RF made the selections as A_1I_2 and A_2I_1 . Subject TB made the selection as A_1I_1 and A_2I_2 .

Phase 2: Tests for Equivalence Classes

Next, sessions were conducted that incorporated tests for equivalence relations between members of the eight-member classes and members of the four-member classes. In these sessions, eight probe trials for a pair of equivalence relations (e.g., the EL relations, E_1L_1 and E_2L_2 ; see Figure 3) replaced eight of the baseline training trials of the preceding baseline. Therefore, these sessions consisted of 16 trials involving the AB, AC, AD, and AE relations, eight unreinforced conditional selection trials (AI), and eight probe trials for equivalence relations. Each of the pairs of

Fig. 3. The unreinforced conditional selection relation (dotted line) and the subsequent probed relations for class merger (dashed lines) in Experiment I.

equivalence relations, AL, BL, CL, DL, EL and LB (shown as dashed lines in Figure 3) were tested separately in one or more sessions. These sessions were alternated with one or more sessions without probes (i.e., Phase 1 sessions). The exact order of sessions is shown in Table 2.

Table 2 also shows the subjects' performances on probe trials. Both subjects' performances showed equivalence relations in accordance with the between-class unreinforced conditional selection. In other words, subjects' performances showed the development of two large equivalence classes. These classes were directly shown to be at least nine members each, and may be inferred to be 12 members each if the F, G, and H stimuli (that were members of the preexperimental eight-member classes) are included. For Subject RF, Class 1 merged with Class 4 and 2 with 3. For Subject TB, Classes 1 and 3 merged, as did 2 and 4.

DISCUSSION

The results showed a general tendency by experienced subjects to make unreinforced conditional selections of comparisons to samples, and to have performances on equivalence tests that show the merger of existing equivalence classes into larger classes based upon these unreinforced conditional selections. In

Table 2

Session sequence and probe performance (number equivalent of eight) in Experiment I.

Subject	Session type	Number correct
TB	Session without probes	
	Session without probes	
	Probe BL relations	8
	Probe BL relations	8
	Session without probes	
	Probe CL relations	8
	Session without probes	
	Probe DL relations	8
	Session without probes	
	Probe EL selection	8
	Session without probes	
	Probe AL relations	8
	Session without probes	
	Probe LB relations	8
RF	Session without probes	
	Session without probes	
	Session without probes	
	Probe BL relations	8
	Session without probes	
	Session without probes	
	Session without probes	
	Probe CL relations	8
	Session without probes	
	Probe DL relations	8
	Session without probes	
	Probe AL relations	8
	Session without probes	
	Probe LB relations	8
	Session without probes	
	Probe EL relations	8

other words, class merger was based on a conditional discrimination that was not trained directly.

Sidman *et al.* (1985) reported the merger of six small stimulus classes into three larger classes by directly training the linking relations. Thus, the stimuli through which the merger occurred—the nodal stimuli—were related through training. Nodal stimuli are those stimuli that relate two other stimuli through a trained relation with each (Fields, Verhave, & Fath, 1984). Subsequent to this training, tests for equivalence relations between stimuli from the two smaller classes included tests across two, three, and four nodes. In the present experiment, the linking stimuli, A and I, were related through unreinforced conditional selection. The subsequent tests for equivalence relations included tests (AL) across three linking stimuli (I, J, and K) and tests (EL, BL,

CL, DL, and LB) involving four linking stimuli (A, I, J, and K). Thus, the present study replicates that of Sidman *et al.* (1985), but without training the class-merging relations with explicit reinforcement procedures (*i.e.*, without a training node).

The results of Experiment I demonstrate that an extensive history of two-choice conditional discrimination training and tests for equivalence relations in a two-choice conditional discrimination test is sufficient to produce performances that show the development of equivalence relations based on conditional, but unreinforced, selection of comparison stimuli. In this case, both the sample stimuli and comparison stimuli were members of existing classes (or what appeared to be classes). Experiment II was conducted to replicate the results of Experiment I with four definitive classes and to determine whether the subjects would show these performances following only the minimally necessary training to produce the four classes upon which the subsequent tests would be based.

EXPERIMENT II

METHOD

This experiment had 10 phases as shown in Table 3, ultimately involving 16 visual stimuli (Figure 4). The first six phases used matching-to-sample training procedures and two-choice conditional discrimination probes to produce performances that showed the emergence of four four-member equivalence classes and the maintenance of the performances in sessions without explicit trial-by-trial reinforcement. Establishing the discriminations involved training subjects to select each comparison stimulus (the training node) in the presence of each of several different samples. This method has been referred to as a multiple-sample single-comparison procedure (Saunders *et al.*, 1988).

Phases 7 and 8 replicated the previous experiment by permitting the unreinforced conditional selection of a comparison stimulus from each of two of the equivalence classes in the presence of each of two sample stimuli that were members of the two remaining classes, respectively, and then testing for equivalence relations among the samples "linked" by the selections. Phases 9 and 10 were used to reinforce responses on conditional selection trials

Table 3

Number of sessions to criterion in each phase of Experiment II.

Phases	Subjects		
	LM	SD	JP
1: Identity match	1	1	1
2: Discrimination training for Classes 1 and 2	21	15	35 ^a
3: Probes for equivalence	Not applicable		
4: Discrimination training for Classes 3 and 4	29 ^b	13	27
5: Probes for equivalence	Not applicable		
6: Establishing concurrent performance of probes in Classes 1, 2, 3, and 4	4	21 ^b	9
7: Unreinforced conditional selection	3	3	3
8: Probes for class merger	Not applicable		
9: Reversed conditional selection	14	38	23
10: Probes for remerger	Not applicable		

^a Subject who required postsession payment contingency.

^b Both subjects required 11 review sessions following a 2-month break.

that were the reverse of that performed when the selections were unreinforced in Phase 7, and to test again for equivalence relations among the samples, as in Phase 8.

Thus, the 10 phases of the experiment permitted (a) a determination of whether subjects with a brief history of forming equivalence classes would make consistent unreinforced conditional selections of comparisons to samples; (b) whether, following such selections, performances would show equivalence relations among comparisons from different classes, joined only by the unreinforced conditional selections; and (c) whether the selections could be reversed through direct training methods and whether, subsequently, the equivalence relations would be shown to be reversed.

The apparatus and general procedures, including criteria for phase change, were the same as in Experiment I, except as noted below.

PROCEDURES AND RESULTS

Subjects

Two retarded young adults and one retarded adolescent participated in the experiment. The chronological age, sex, intelligence quotient, and diagnosis of each (LM, SD, and JP) are shown in Table 1. All tests scores were current within 2 years and were obtained with

EXPERIMENT II

Phase 1 Training and Tests for Equivalence Relations

Phase 2 Training and Tests for Equivalence Relations

Fig. 4. The trained relations (solid lines) and probed one-node equivalence relations (dashed lines) in Phases 2 through 6 of Experiment II.

a Wechsler test of intelligence. All subjects exhibited functional speech and were experimentally naive to our laboratory and the procedures used.

Phase 1: Pretraining with Identity Matching

Each subject received one session of 32 trials of identity matching with 10 of the 16 visual stimuli used in the experiment. All 3 subjects performed at 100% correct in the first session with contingent reinforcement, and with the verbal instruction to "push this (the middle button) and now push the one that goes with the one in the middle," on the first trial of the session.

Phases 2 through 6: Establishing Four Equivalence Classes

The training for the first two of four classes was conducted in Phase 2 via a multiple-sample single-comparison training procedure (see Figure 4). Thus, each discrimination involved a pair of samples (e.g., B_1 and B_2 shown in Figure 4) and two comparisons (A_1 and A_2) and was trained in separate sessions first. After all three discriminations had been trained independently, they were trained concurrently in the same sessions. Subjects were instructed on the first two trials of the first discrimination, "When the spool comes up, you push the button under the flag. When the bow comes up you push the button under the apple." No explicit instructions on how to respond were given to any subject throughout the remainder of the experiment. When a subject's performance on the mix of the three discriminations met criterion, sessions of the same mix were conducted, but with eight trials unreinforced (no sound or penny delivered). When performance met criterion on these intermittent reinforcement sessions, the same mix of trials was used in sessions conducted with no trial-by-trial reinforcement (no sound or penny delivered for any response). The subjects were paid 32 cents at the beginning of each session during sessions with no trial-by-trial feedback.

Table 3 shows that the performances reached the criterion for exposure to probe sessions within 15 to 35 sessions. In addition to the training described above for this phase, subjects required refresher or review training on some of the sample stimulus pairs or additional practice under intermittent reinforcement conditions prior to maintaining their performance without differential consequences. The greater number of sessions required by JP was due to three failed attempts to maintain the performance of the mixture of three conditional discriminations under conditions of no feedback and with payment before the session. After each failed attempt, the mixed baseline was brought to criterion under reinforcement conditions, followed by the no-feedback condition. The fourth exposure to the no-feedback condition was preceded by instructions that the computer was keeping track of his performance and he would be paid at the end of the session. At the end of these sessions, JP was paid a penny for each correct trial, but was not given specific feedback on errors. The sub-

ject's performance met criterion under these conditions.

Following this training, a series of sessions to probe for one-node equivalence relations was conducted (dashed lines in Figure 4) in Phase 3. The composition of probe trials for equivalence is shown in Table 4. Each probe session for equivalence contained a sample of four probes from Class 1 and four from Class 2. On all probe trials, the comparison stimuli were presented as stimulus pairs, based on their order of introduction into the experiment (e.g., if B_1 was the correct comparison stimulus, B_2 was the distractor stimulus). The probe trials were interspersed among the original training trials unsystematically, except that probe trials were always separated by one or more training trials. Probe sessions were alternated with sessions of training trials until all possible equivalence relations in each of the two directions of each relation (e.g., B_1C_1 and C_1B_1) were tested at least once (which required three probe sessions) and until performance on the probes was stable. Each day's sequence of sessions was begun with a training trial session. If performance was 100%, a probe session was conducted next. If performance on the training trial session was less than 100%, additional training trial sessions were conducted until performance reached two consecutive sessions at 90% correct or better or one at 100% correct; a probe session then followed.

Responses on probe trials that indicated the formation of the predicted stimulus classes were considered correct. Table 5 shows that all subjects were correct in eight of eight one-node equivalence probes for all three sessions containing probes. Thus, two four-member stimulus classes emerged: Class 1, consisting of A_1 , B_1 , C_1 , and D_1 ; and Class 2, consisting of A_2 , B_2 , C_2 , and D_2 . In addition, no subject's performance was ever below 96% correct on training trials in any session.

Following the emergence of equivalence relations in the first two classes, eight new stimuli were used in training (Phase 4) and testing (Phase 5) procedures identical to those above to produce a third and fourth equivalence class. However, after a subject met criterion on each new pair of samples, review sessions involving a mix of the discriminations for Classes 1 and 2 were conducted. No subject performed below 96.9% in any of these review sessions. Figure 4 and Tables 3, 4, and 5 show the trained and

Table 4
Equivalence tests in Experiment II.

Phase 3 probes (one-node)						Phase 5 probes (one-node)					
Class 1			Class 2			Class 3			Class 4		
Sample	Equiva- lent com- parison	Dis- tractor	Sample	Equiva- lent com- parison	Dis- tractor	Sample	Equiva- lent com- parison	Dis- tractor	Sam- ple	Equiva- lent com- parison	Dis- tractor
B ₁	C ₁	C ₂	B ₂	C ₂	C ₁	F ₁	G ₁	G ₂	F ₂	G ₂	G ₁
C ₁	B ₁	B ₂	C ₂	B ₂	B ₁	G ₁	F ₁	F ₂	G ₂	F ₂	F ₁
B ₁	D ₁	D ₂	B ₂	D ₂	D ₁	F ₁	H ₁	H ₂	F ₂	H ₂	H ₁
D ₁	B ₁	B ₂	D ₂	B ₂	B ₁	H ₁	F ₁	F ₂	H ₂	F ₂	F ₁
C ₁	D ₁	D ₂	C ₂	D ₂	D ₁	G ₁	H ₁	H ₂	G ₂	H ₂	H ₁
D ₁	C ₁	C ₂	D ₂	C ₂	C ₁	H ₁	G ₁	G ₂	H ₂	G ₂	G ₁
Phases 8 and 10 probes (linked by one or three stimuli)						Phases 8 and 10 probes (one-node review)					
Class 1 and 3 merger			Class 2 and 4 merger			Class 1 and 3			Class 2 and 4		
Sample	Equiva- lent com- parison	Dis- tractor	Sample	Equiva- lent com- parison	Dis- tractor	Sample	Equiva- lent com- parison	Dis- tractor	Sample	Equiva- lent com- parison	Dis- tractor
B ₁	G ₁	G ₂	B ₂	F ₂	F ₁	B ₁	C ₁	C ₂	B ₂	C ₂	C ₁
C ₁	F ₁	F ₂	B ₂	H ₂	H ₁	B ₁	D ₁	D ₂			
C ₁	H ₁	H ₂	C ₂	G ₂	G ₁	G ₁	F ₁	F ₂	G ₂	F ₂	F ₁
D ₁	G ₁	G ₂	D ₂	F ₂	F ₁	H ₁	F ₁	F ₂			
D ₁	H ₁	H ₂	F ₂	C ₂	C ₁						
F ₁	B ₁	B ₂	G ₂	B ₂	B ₁						
F ₁	D ₁	D ₂	G ₂	D ₂	D ₁						
G ₁	C ₁	C ₂	H ₂	C ₂	C ₁						
H ₁	B ₁	B ₂	H ₂	D ₂	D ₁						

tested relations, sessions to criterion, probe trials, and results, respectively.

In Phase 6, four training stages were used to produce the maintenance of performances in Phases 2 through 5 concurrently without programmed consequences for responses: (a) sessions containing an equal number of training trials from each of the four classes with programmed consequences, (b) sessions containing these same trials with programmed consequences on 24 of the 32 trials (intermittent reinforcement), (c) sessions containing these trials with no programmed consequences, and (d) eight one-node equivalence probe trials (two from each class) interspersed among 24 training trials from the four classes, with no programmed consequences for either type of trial. Probe sessions were alternated with sessions without probes until a total of three probe sessions had been conducted with no programmed consequences on individual trials. Subjects LM and SD were paid 32 cents at the beginning of each session; Subject JP

was paid after the sessions as described for Experiment I.

Table 3 shows the number of sessions in Phase 6, including probe sessions, and Table 5 shows the results of these probes. Because of experimenter error, Subject LM received only one probe session in this phase (see Table 5). All subjects were correct on each probe trial and training trial performances were maintained.

Phases 7 and 8: Unreinforced Conditional Selection and Probes for Equivalence (Class Merger)

Subjects were exposed to sessions in Phase 7 in which the sample stimuli were two samples (F₁ and F₂) from Classes 3 and 4, and the comparisons were the comparison stimuli (A₁ and A₂) from Classes 1 and 2 (dotted lines in Figure 5). Sessions were conducted with only these stimuli, one half of the trials with each sample stimulus presented in unsystematic order. When the subjects' conditional selections

Table 5

Performance on probes for equivalence expressed as number equivalent per session (of eight probe trials in Phases 3, 5 and 6; of six merger and two review probe trials in Phases 8 and 10).

		Subjects											
		LM			SD						JP		
		Probe sessions											
Phase	Probes	1	2	3	1	2	3	4	5	6	1	2	3
3	8 one-node	8	8	8	8	8	8				8	8	8
5	8 one-node	8	8	8	8	8	8				8	8	8
6	8 one-node	8			8	8	8				8	8	8
8	6 merger	6	5	6	6	6	6				6	6	6
	2 review	2	2	2	2	2	2				2	2	2
10 ^a	6 merger	0	0	0	0	0	2	1	0	1	3	4	0
	2 review	2	2	2	2	2	2	2	2	2	1	1	1

^a Equivalent responses are those in line with the reversed conditional selection.

were consistent, sessions were conducted with eight of these conditional selection trials interspersed among 24 training trials drawn equally from the four classes. No programmed consequences followed responses on any trial in Phase 7.

All subjects immediately performed on the unreinforced conditional selection trials in a conditional and consistent pattern. Each subject selected comparison stimulus A_1 in the presence of sample stimulus F_1 and selected A_2 in the presence of F_2 . After two sessions of 100% consistent selection by Subjects LM and SD and two sessions at 96.9% by JP, all subjects performed at 100% consistency on the first session with conditional selection trials interspersed among training trials.

In Phase 8, tests were conducted for the emergence of equivalence relations between the samples of the two classes "linked" by the conditional selections. These relations are shown as dashed lines in Figure 5; the composition of the trials that tested for these relations is shown in Table 4. Probe sessions contained 20 training trials from the four classes, four conditional selection trials, two one-node probe trials to test for the maintenance of the four original classes, and six probes of relations linked by one and three stimuli for the merger of classes.

Table 5 shows that performances in probe sessions showed the development of equivalence relations between sample stimuli from different classes. The performances on one- and three-node equivalence probes were always correct with respect to the consistent un-

reinforced selection, except for one trial in the second probe session for Subject LM. The accurate performances on the one-node review probes indicated that the classes that emerged in Phases 3 and 5 remained intact during this phase. Thus, the performances of all subjects showed the merger of Class 1 with Class 3 and Class 2 with Class 4. As during previous probe sequences, baseline performance was maintained at very high levels of accuracy.

Phases 9 and 10: Reversed Conditional Selection and Equivalence Probes (for Remerger of Classes)

In Phase 9, sessions were conducted that were identical to the unreinforced conditional selection sessions in Phase 7 except that responses contrary to the previous unreinforced conditional selection were reinforced (e.g., selecting A_1 in the presence of F_2). In these sessions, the buzzer sound occurred following selections of the type made in Phase 7 and the jingle and penny delivery followed selections that were reversed. When performance met criterion, sessions with eight reversed selection trials and six training trials from each of the four original classes, and with no programmed consequences for any response, were conducted. If criterion performance was not achieved with these sessions, other sessions were conducted, the type of which was determined by the particular performance of each subject. For example, if performance of the former training trials was disrupted by the inclusion of the reversed selection trials, sessions with only the training trials were conducted with

programmed consequences. Eventually, each subject's performance was required to meet criterion without trial-by-trial reinforcement, with training trials and reversed selection trials intermixed, prior to Phase 10. In these criterion sessions, correctness on the reversed selection trials was defined as "reversed from the selection when unreinforced." These trials had to be performed at 100% correct, and training trials had to be performed at 100% for one session or at 90% for two consecutive sessions, to meet the criterion for probing. Near the end of Phase 8, the experimenters learned that Subject SD would leave shortly for a vacation. Therefore, his final 56 sessions, including all of Phases 9 and 10, were conducted across 5 days.

Subject LM achieved 100% reversal of the conditional selection in isolation in two sessions and SD did so in three. However, when the eight conditional selection trials were first embedded in the training trial baseline under no trial-by-trial reinforcement conditions, both subjects reverted to their original unreinforced selection. It required 14 total sessions for LM and 38 total sessions for SD to reach criterion for first probing (Table 3).

Subject JP showed 100% reversal of the conditional selection in two sessions. His subsequent performance differed from the other subjects in that the reversal was maintained at levels of from 37.5% to 100% correct in four sessions in which the eight conditional selection trials were intermixed with all other trained relations without trial-by-trial reinforcement. For this subject, errors occurred on some reversed conditional selection trials and on previously trained relations involving the reversed conditional selection samples (F_1 and F_2). For JP, additional training included six sessions in which only the relations F_1E_1 and F_2E_2 were trained. Criterion for probing in Phase 10 was reached in 23 sessions (Table 3).

In Phase 10, probe sessions identical to those in Phase 8 were conducted. However, the class merger relations shown as probe trials would be responded to in reverse from those responses observed in Phase 8 if the new relations were based on the reversed conditional selection. If this occurred, Class 1 would then be merged with Class 4 and Class 2 with Class 3.

For Subjects LM and SD, performance on training trials was never below 91.7%. Re-

EXPERIMENT II

Tests for Equivalence Relations in Phase 4 Following Unreinforced Conditional Selection

Fig. 5. The unreinforced conditional selection relation (dotted line) and the subsequently probed equivalence relations (class merger) linked by one or three stimuli in Phases 7 and 8 of Experiment II. This schematic also shows the reversed conditional selection relation (dotted line) and subsequent probes (dashed lines) used in Phases 9 and 10 of Experiment II.

sponses on "review" probes for the original smaller classes were always correct. Subject LM responded to all six merger probes in all three probe sessions just as he had prior to the reversal of the selection, yielding no correct responses with respect to the reversed selection (Table 5). Subject SD responded incorrectly (with respect to the reversed selection) to all probes in the first two probe sessions, but responded correctly to two of the six merger probes in the third session. Therefore, additional probe sessions were conducted. SD was correct only twice in 18 opportunities over three additional sessions.

Subject JP's baseline performance fell to 83.3% correct in the first probe session; 15 sessions were required to reestablish the criterion performance before the next probe session could be given. Baseline performance during the final four probe sessions ranged from 91.7% to 100%. Responses to the review probes were always 50% correct, with Class 1 or 2 probes always correct and Class 3 or 4 probes

always incorrect. Performance on the class merger probes varied from zero to four correct (of six) with respect to the reversed selection, with three correct in each of the last two sessions (Table 5). Further analysis revealed that all 13 "correct" responses, based on the reversed selection, occurred on probe trials involving the F stimuli (the reversed conditional selection samples). These trials were the tests for class merger with one linking stimulus (A_1 or A_2). On the remaining 16 class merger probes not involving the F stimuli (the tests with three linking stimuli) and one trial involving the F stimuli, the subject responded according to the previous selection. All of the one-node review probes for Classes 3 and 4 involved the F stimuli, and all produced "correct" responses with respect to the reversed selection, indicative of a disruption of the original third and fourth small classes.

DISCUSSION

The subjects in this experiment developed performances that demonstrated the development of four four-member equivalence classes. They required between 46 and 67 sessions to develop performances that met the criterion required for introduction of the unreinforced conditional selection trials. Thus, their experimental histories were briefer by about 200 sessions than those of the subjects in Experiment I. Nevertheless, like the subjects in Experiment I, they immediately and consistently selected a particular comparison stimulus in the presence of a particular sample stimulus. In Experiment I, unreinforced conditional selection trials were introduced by interspersing them among baseline training trials, with no differential consequences for any response. In Experiment II, the unreinforced conditional selection trials were presented in isolation, that is, comprising all 32 trials in the initial session and not interspersed among training trials. Despite the fact that in this procedure these trials are more easily discriminated as novel, and despite briefer experimental histories, the subjects responded conditionally, as they had when new reinforced discriminations were introduced in Phases 2 and 4.

Performances on the tests for equivalence relations across classes linked by the unreinforced conditional selection were unequivocal. Each subject consistently responded as correctly during Phase 8 tests as he had performed

on Phase 3 and Phase 5 tests, except for one incorrect response by Subject LM. Thus, the likelihood appears the same that equivalence relations will emerge when the relations are based on an unreinforced relation as when they are based on a reinforced relation, regardless of whether when the subject has a brief or an extensive history of conditional responding and equivalence class performance.

This demonstration is limited, however, to equivalence relations that show class merger. It does not indicate whether classes also can be expanded in this way rather than merged with other classes. These results do not permit an evaluation of the relative importance of the conditional discrimination history versus the history of performance that shows the emergence of equivalence classes. Even though these subjects had much briefer histories than the subjects in Experiment I, the number of trained conditional discriminations was six. It is not possible to assess whether the results could have been produced with fewer (e.g., four) discriminations.

If the tests for equivalence in Phases 3 and 5 had not been conducted, would the subjects have made the unreinforced conditional selections? If they did make the unreinforced conditional selections without a history of equivalence class performance, would they perform on the merger probes in a manner that showed equivalence? If subjects were to show the emergence of these large classes, without prior tests for the potential, smaller four-member classes, it may indicate that relatively large classes can emerge directly, without intermediate tests that might build the large class in stages. Conversely, it may indicate that the smaller classes emerged without testing, and were functional in the emergence of the relations across the three linking stimuli in the merger probes. Research on these possibilities has implications for applied and theoretical problems. For example, how many members of classes should be presented to students as conditional discriminations before using procedures that facilitate the students' performance with the stimuli as class members? Are teaching and testing in stages necessary when teaching large equivalence classes? In the emergence of verbal classes in the young child, do words that mean the same emerge as class members without overt requirements to use them equivalently?

In Phase 10, 2 subjects (LM and SD) retained the directionality of responding on the equivalence relations that had been established by the Phase 7 and 8 procedures, despite performing concurrently in the reversed direction on the reversed conditional selection trials. There are at least two interesting features of these results. First, both subjects (especially SD) required considerably more than the minimum number of sessions in Phase 9 to meet the criterion for probing. This indicates a rather surprising resistance to shift from an unreinforced selection to the reinforced reverse and to maintain the reverse in the context of Phase 2 and Phase 4 training trials. Second, the derived relations that indicated class merger in Phase 8 remained intact despite the ultimately reversed performance on the unreinforced conditional selection trials. Saunders et al. (1988) reported that performances showed equivalence class relations under conditions of no trial-by-trial reinforcement for as long as 5 months following the termination of an experiment. Coupled with the current findings, the implication is that stimulus classes, even those formed with untrained relations, can be extremely resistant to alteration.

What if, however, attempts to alter the composition of the classes were made before the tests for equivalence class emergence or class merger were conducted? In the preceding experiment, if Phase 8 had been omitted (no tests for merger conducted) and Phases 9 and 10 had been conducted, would performance on the reversed selection be so difficult to produce? Would the subsequent Phase 10 probes have revealed the merger of Class 1 with 4 and 2 with 3 (i.e., based on the reversed selection), and would these classes have the same durability as those that emerged in Phase 8? To understand how and when new stimuli become class members, we should investigate whether resistance to change of the unreinforced conditional selections is a function of permitting the derivative equivalence relations to emerge in overt performance.

Conversely, if the selections in Phase 7 had been reinforced, we would expect to see the same class merger performances as were observed with the unreinforced conditional selection. Would this pair of relations, trained with explicit reinforcement for comparison selection, be as difficult to reverse subsequently with similar reinforcement procedures? Are

relations acquired without explicit reinforcement more resistant to alteration than those performed as a function of reinforcement procedures? Is the strength of relations established with reinforcement procedures also a function of whether the relations have or have not yet been involved as links in the overt performance of equivalence relations? The implications of the answers to these questions for instruction in educational settings are exciting.

In Phase 10, Subject JP did not perform on either the one-node or the class merger equivalence tests in exactly the same manner shown by the other 2 subjects. Whereas they responded according to the direction of equivalence relations indicative of control by the original unreinforced conditional selection, JP did so only if the trial did not involve an F stimulus as either sample or comparison. Thus, responding by this subject was controlled by the presence or absence of the stimuli used as reversed conditional selection samples. The results do not permit a determination of whether two differently configured pairs of merged equivalence classes coexisted, each controlled by different samples on equivalence trials. However, the results replicate those of Subjects LM and SD in demonstrating that stimulus classes are not easily altered. Further, this performance demonstrates that the relations not involving the F stimuli directly (e.g., B_1H_1), were no longer dependent on the direction of the relations intermediate to the two stimuli in the relation. That is, although a relation such as B_1H_1 was originally derived from intermediate relations such as F_1A_1 and F_1E_1 , when the FA and FE relations were subsequently reversed the BH relation remained unaffected. This suggests that the linking relations are important in the development of derived relations, but that their alteration may not necessarily affect the derived relations later. This is to say that once an equivalence class emerges, other sources of stimulus control may alter performance on a subset of the relations without altering performance on the remainder of the relations in the class. This reminds us of rhubarb and tomatoes. As children, mostly under the control of the distinction between a main course and dessert, we inferred that rhubarb was a fruit and a tomato was a vegetable. Subsequent learning put these stimuli in their botanically correct categories, but no disruption to the memberships of peaches and ba-

EXPERIMENT III

Tests for Equivalence Relations
Following Unreinforced Conditional Selection of
Comparison Stimuli with Three Pairs of Samples

Fig. 6. The three unreinforced conditional selection relations (dotted lines) and the subsequently probed equivalence relations (dashed lines) in Experiment III.

nanas or to potatoes and celery occurred, despite the fact that their class membership may have been derived initially across relations involving rhubarb and tomatoes.

Results of Experiments I and II demonstrate that consistent conditional responding is likely when the stimuli are samples and comparisons from existing stimulus classes. Would subjects be as likely to respond conditionally in trials involving two novel sample stimuli and two novel comparison stimuli? If so, would they then respond conditionally on a series of such discriminations? If this too were likely, then tests for equivalence between these stimuli could be conducted as was done in Phase 3 of Experiment II. Experiment III was conducted to answer these questions and to determine whether the tests would show the development of equivalence classes based entirely on conditional, but untrained, relations.

EXPERIMENT III

METHOD AND RESULTS

The apparatus and general procedures were the same as described previously. The eight stimuli shown in Figure 6 were used in this experiment.

Subjects

The subjects in this experiment were Subjects LM and SD, who had participated in

Experiment II. Subject JP withdrew his consent to participate in further research at the conclusion of Experiment II.

Unreinforced Conditional Selection Sessions

As shown by the dotted lines in Figure 6, the subjects were presented with three new conditional discriminations, each involving a different pair of sample stimuli (M₁, M₂; N₁, N₂; and O₁, O₂) and the same pair of comparison stimuli (P₁, P₂). The presentation was conducted without verbal instructions and without programmed consequences on any trial. The subjects were paid 32 cents prior to each session. Otherwise the same procedures as described in Phase 2 of Experiment II were followed: each pair of samples was presented (or "trained without consequences") separately and then all six samples were presented in the same session in an unsystematic order. The performance criterion for the presentation of each new pair of sample stimuli and for presentation of the mixed sessions was two sessions in a row at 90% or better consistent selection (the direction of selection in both sessions had to remain constant). Two mixed sessions at 90% consistent or one mixed session at 100% was required for exposure to probe sessions.

Subject LM required few sessions for his performance to meet criterion for exposure to the probes for equivalence (Table 6). On the first pair of samples, his performance was initially unsystematic with respect to comparison selection, but in the second and third sessions he consistently selected stimulus P₁ in the presence of stimulus M₂ and P₂ in the presence of M₁. Subsequently, stimulus P₁ was selected in the presence of N₁ and O₁, and P₂ was selected in the presence of N₂ and O₂. In the first four sessions with mixed samples, Subject LM reversed his selections in the presence of the M stimuli, but on the fifth session was 100% consistent with his earlier selections.

Subject SD initially performed unsystematically on the first sample pair, but by the third session was consistently selecting a particular comparison in the presence of each sample. The selections of comparison stimuli with the second pair of samples became consistent in the fifth session with that pair. Selections with the third pair were consistent after the third trial in the first session with that pair. With all three pairs, SD selected the comparison

from Class 1 (Figure 6) in the presence of a sample from Class 1 and the Class 2 comparison in the presence of Class 2 samples. However, when all of the samples were presented in the same sessions, Subject SD performed on all trials involving the N samples by selecting comparison stimulus P_2 . Six sessions involving only the N stimuli as samples did not affect this new performance.

Probes for Equivalence Classes

When performance in the mixed arrangement met criterion, sessions were conducted consisting of 24 trials from the mixed-presentation sessions and eight trials that tested for equivalence relations among the sample stimuli (e.g., M_1N_1 and N_1O_1). The actual probe trials for each subject were determined by which comparison stimulus had been selected in the presence of each sample stimulus. Figure 6 shows the probes for equivalence (the dashed lines) as if both subjects had selected the comparison stimulus from Class 1 with the sample stimuli from Class 1. Table 7 shows the composition of the probe trials for each subject (the differences in number and type of trial between subjects are explained below). Probe sessions were alternated with sessions containing no probe trials.

During probe sessions, Subject LM's performance on the baseline conditional selection trials remained above 95% consistent with respect to his original conditional selection. The three responses on equivalence probe trials, during four probe sessions, that were inconsistent with the baseline trials all involved stimulus M_2 as either the sample, comparison, or distractor stimulus.

Subject SD's selections made in the presence of the N stimuli as samples precluded the inclusion of probe trials with these two stimuli as the comparison and distractor stimuli. For example, if stimulus M_1 were the sample on such a trial, then neither N_1 nor N_2 would be equivalent. Similarly, if M_2 were the sample stimulus, then both could be equivalent, because P_2 was selected in the presence of both. Therefore, the eight probe trials in each probe session included only the eight versions shown in Table 7. The probe sessions were conducted and Subject SD's performance showed equivalence on all 24 probe trials. Only one inconsistent choice was made on baseline trials in

Table 6

Trials to criterion for equivalence probes and performance on probes in Experiment III.

Presentations	Sessions to criterion	
	Subject LM	Subject SD
Sample pair M_1M_2	3	4
Sample pair N_1N_2	2	6
Sample pair O_1O_2	2	2
All samples mixed	5	13 ^a

Probes	Number equivalent per session					
	Subject LM			Subject SD		
8 one-node	6	8	8	7	8	8

^a Includes six review sessions with sample pair N_1N_2 .

these sessions. On the trials in which stimulus N_1 was the sample, SD selected the comparison stimulus related by conditional selection to comparison stimulus P_2 . Because these tests were combined tests for transitivity and symmetry, the results suggest that both N_1 and N_2 were symmetric with P_2 . Thus, unlike Subject LM whose performance demonstrated two four-member classes, SD's performance indicated a three-member and a five-member class.

DISCUSSION

The conditions sufficient to produce performances that show the emergence of equivalence relations have now been expanded well beyond those previously reported. The 2 subjects in Experiment III had performances that showed the emergence of equivalence classes even though the relations upon which the equivalence relations could be based were never directly trained with trial-by-trial reinforcement procedures. The selection of the comparison stimuli during the sessions in which each new pair of samples was presented, and when all the samples were presented in the same session, was clearly conditional. Specific sample stimuli controlled the selection of certain comparison stimuli. However, the control was not derived from specific reinforcing consequences for these selections. Presumably, the control was generalized as a function of the subjects' histories of contingent reinforcement for previous comparison selections in the presence of other samples. Moreover, those histories provided ample opportunity for the subjects to discriminate that the selection of only

Table 7
Equivalence tests in Experiment III.

	Class 1			Class 2		
	Sample	Equivalent comparison	Distractor	Sample	Equivalent comparison	Distractor
Subject LM	M ₁	N ₁	N ₂	M ₂	N ₂	N ₁
	N ₁	M ₁	M ₂	N ₂	M ₂	M ₁
	M ₁	O ₁	O ₂	M ₂	O ₂	O ₁
	O ₁	M ₁	M ₂	O ₂	M ₂	M ₁
	N ₁	O ₁	O ₂	N ₂	O ₂	O ₁
	O ₁	N ₁	N ₂	O ₂	N ₂	N ₁
Subject SD	N ₁	M ₁	M ₂	N ₂	M ₂	M ₁
	M ₁	O ₁	O ₂	M ₂	O ₂	O ₁
	O ₁	M ₁	M ₂	O ₂	M ₂	M ₁
	N ₁	O ₁	O ₂	N ₂	O ₂	O ₁

one comparison stimulus, and not both in the presence of a particular sample, was reinforced. Oddly, Subject SD eventually selected one comparison stimulus in the presence of both samples within the second pair.

These results have some bearing on the questions raised earlier with respect to experimental history, because they demonstrate that the unreinforced conditional selection can occur with stimuli that are not yet members of an equivalence class. Although the subjects had performances that showed equivalence classes, these performances involved other stimuli.

Regardless of how the subjects' histories and the conditions in Experiment III interacted to produce the conditional selections, the result was that tests for equivalence among the samples revealed the emergence of equivalence relations. Thus, the results indicate that in addition to class merger, the development of existing classes through unreinforced conditional selection is likely. More important will be the demonstration that the emergence of an equivalence class based entirely on relations that were not explicitly trained can be replicated with subjects with no laboratory history of equivalence class development. If such results are possible, how brief a history of reinforced conditional responding will be necessary to make probable the unreinforced conditional selections through which the classes develop?

GENERAL DISCUSSION

Unreinforced, generalized responding is not in itself unusual. In the current series of experiments, however, the relations that estab-

lished the basis for the derived relations were also generalized. This finding is unusual, but important. When Fields *et al.* (1984) referred to the exponential increase in new relations when class size is increased "... metaphorically as the behavioral analogue of a cognitive breeder reactor ..." (p. 148), they were describing the growth as a function of the increasing number of relations trained with trial-by-trial reinforcement procedures. The current results show that the increase in new relations is also a function of increases in class size through untrained relations. Thus, a history of equivalence class emergence alone appears sufficient to produce class merger and class development when the subject is exposed to conditional discrimination problems of a similar type. Therefore, if we are to continue to use the term "node" in our vocabulary on equivalence classes, it should now be redefined as "a stimulus that is related to more than one other stimulus." That such relations be trained is not a necessary condition.

It may be that the present results are dependent on the two-choice (and therefore, forced choice) nature of the procedures used. With apparatus that would permit and procedures that employed more choices (specifically the opportunity to select *no* stimulus), the results might be different. Under those conditions, subjects might not make consistent unreinforced conditional selections or might not perform consistent with class formation. Because learning opportunities outside the laboratory are replete with both two-choice and multiple-choice conditional discriminations, and because these discriminations are important in human development and functioning,

an extension of the current series of experiments will be interesting and important. It will be equally informative to learn whether the long-term retention of relations is affected also by whether or not the relations were developed in forced, two-choice procedures.

Across the three experiments, subjects performed on training trials, conditional selections, and probes without explicit trial-by-trial reinforcement. What differences in performance resulted as a function of noncontingent payment before each session versus partially contingent payment (based on accuracy on training trials only) after the session? It may be the case that payment after a session affects future performance as a function of whether 32 cents or less than 32 cents is paid. This may be particularly likely given the multiple sessions conducted per day. Because all conditional selection trials are paid regardless of performance, feedback to the subject would be that performance was entirely accurate upon receipt of all 32 cents. Yet, 2 subjects performed similarly in their second session of this type, but had already been paid before rather than after the first session. Although the data from these experiments are insufficient to resolve this question, it is likely that the answer will be that some subjects will require payment after, with perhaps some portion contingent (for motivational reasons), but that the timing of the payment is not a major issue in the performances of the type demonstrated here. (See Saunders et al., 1988, for a discussion of the effects of probing in an explicitly reinforced baseline and the potential reinforcing effects of trial completion in sessions without explicit trial-by-trial reinforcement.)

We suspect, however, that class size and the number of classes previously formed are the most critical variables in whether unreinforced conditional selection occurs, whether equivalence classes will emerge that are based upon unreinforced conditional selection, and whether equivalence classes are stable over long periods of time. As Saunders et al. (1988) suggested, the maintenance of classes over extended periods of time may be directly a function of class size. In large classes, membership of any one stimulus becomes multiply determined as a function of its relation to each of the other stimuli in the class. If over time some of the relations are weakened (some stimuli no longer control selection of other stimuli), these relations may be recoverable by performance on a

series of other relations from which the weakened relation can be rederived. The smaller the size of equivalence classes, the fewer the number of other relations from which a weakened relation can be rederived. However, the finding that some relations cannot be performed after some period of time suggests that perhaps, although all the relations meet the test for equivalence, some are nevertheless different than other relations along some dimension. This suggestion is not unlike that of Fields et al. (1984) that associative distance, defined as the number of training nodes in a potentially transitive relation, might affect the probability that a particular relation will emerge. Although, subsequent research (e.g., Saunders et al., 1988) and the findings of Experiment I indicate that the emergence of equivalence relations is not affected by the number of connecting nodes, it is unknown whether this variable is involved in differential performance after lengthy time intervals.

How do we account for performances after lengthy intervals in which the relations are intact and stable, particularly with such complex arrays of stimuli? If Sidman and Tailby's (1982) analysis is correct, then no matter how large the array of stimuli in an emerged class, performance on the discriminations is reduced functionally to performance on relations that are entirely substitutable for one another, because all of the samples and comparisons in a class are equivalent. Four classes, as in Experiment II, can be conceptualized as four sets of substitutable relations, thus, functionally reducible to only four different relations. By making a consistent unreinforced selection that merges the classes, the four are reduced to two. Responding correctly on a large number of unrelated conditional discriminations is difficult; explicitly taught or merely inferred equivalence relations among the stimuli make the array parsimonious. Some of the experiments proposed earlier in this paper address this possibility. In particular, experiments that investigate whether consistent unreinforced conditional selection will be as likely to occur when the involved stimulus arrays are not or have not emerged as classes will be relevant.

REFERENCES

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

- Dixon, L. S. (1977). The nature of control by spoken words over visual stimulus selection. *Journal of the Experimental Analysis of Behavior*, **27**, 433-442.
- Dixon, M. H. (1978). Teaching conceptual classes with receptive label training. *Acta Symbolica*, **9**, 17-35.
- Dixon, M. H., & Spradlin, J. (1976). Establishing stimulus equivalences among retarded adolescents. *Journal of Experimental Child Psychology*, **21**, 144-164.
- Dube, W. V., McIlvane, W. J., MacKay, H. A., & Stoddard, L. T. (1987). Stimulus class membership established via stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior*, **47**, 159-175.
- Fields, L., Verhave, T., & Fath, S. (1984). Stimulus equivalence and transitive associations: A methodological analysis. *Journal of the Experimental Analysis of Behavior*, **42**, 143-157.
- Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior*, **49**, 95-115.
- Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior*, **43**, 21-42.
- Sidman, M., & 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.
- Spradlin, J. E., Cotter, V. W., & Baxley, N. (1973). Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. *American Journal of Mental Deficiency*, **77**, 556-566.
- Spradlin, J. E., & Dixon, M. H. (1976). Establishing conditional discriminations without direct training: Stimulus classes and labels. *American Journal of Mental Deficiency*, **80**, 555-561.
- 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.
- Wetherby, B., Karlan, G. R., & Spradlin, J. E. (1983). The development of derived stimulus relations through training in arbitrary-matching sequences. *Journal of the Experimental Analysis of Behavior*, **40**, 69-78.

Received December 22, 1987

Final acceptance May 15, 1988