Journal of the
Experimental Analysis of Behavior

JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2014, **101,** 38–50 NUMBER 1 (JANUARY)

MERGING SEPARATELY ESTABLISHED STIMULUS CLASSES WITH OUTCOME-SPECIFIC REINFORCEMENT

CAMMARIE JOHNSON[1], OLGA MELESHKEVICH[2], AND WILLIAM V. DUBE[3]

[1]WESTERN NEW ENGLAND UNIVERSITY, NEW ENGLAND CENTER FOR CHILDREN
[2]NEW ENGLAND CENTER FOR CHILDREN
[3]UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL, WESTERN NEW ENGLAND UNIVERSITY

This study extended previous research on equivalence relations established with outcome-specific reinforcers to include the merger of separately established stimulus classes. Participants were four adults. Conditional discriminations AC and BC were trained first. Correct selections of C1 (C2, or C3) in the presence of A1 or B1 (A2 or B2, or A3 or B3) were followed by red (blue, or white) tokens; tokens were exchanged for value added to three participant-selected gift cards. Outcomes on equivalence tests for three-member classes ABC were positive. DF and EF were trained with the same reinforcing consequences, and tests were positive for three-member classes DEF. Results of class merger tests with combinations of stimuli from the ABC and DEF classes (AD, FB, etc.) were immediately positive for two participants, demonstrating six-member classes ABCDEF with reinforcers as nodes. Merger tests for a third participant were initially negative but became positive after brief exposure to unreinforced probe trials with reinforcers as comparison stimuli. Following class merger, tests for matching the reinforcers to samples and comparisons were also positive. Class-merger test results were negative for a fourth participant. The results provide the first demonstration of eight-member equivalence classes including two outcome-specific conditioned reinforcing stimuli.

Key words: Stimulus equivalence, class merger, conditional discrimination, match to sample, outcome-specific reinforcement, mouse click, humans

Research in stimulus equivalence examines the emergence of relational stimulus control that is not the result of direct training. Much of equivalence research has used matching-to-sample procedures in which a baseline of two or more conditional discriminations is first established by direct training, followed by tests for other, untrained conditional discriminations among the stimuli (Sidman, 1994). By analogy to the mathematical definition of equivalence, Sidman and Tailby (1982) defined the behavioral requisites for equivalence relations in matching to sample as reflexivity (e.g., train AB, test AA and BB; the first and second letters refer to sets of stimuli that serve as samples and comparisons, respectively), symmetry (train AB, test BA), transitivity (train AB and BC, test AC), and combined symmetry and transitivity (train AB and BC, test CA). Positive results on such tests document equivalence classes in which the class members are mutually substitutable within the training context. The baseline discriminations in such an arrangement share a set of common stimulus elements (the B stimuli in the example above) termed a *node* by Fields, Verhave, and Fath (1984).

The great majority of equivalence research has been conducted with *nonspecific* reinforcement procedures in which the same reinforcing consequence follows all correct comparison selections. In an alternative procedure, sometimes termed *outcome-specific* reinforcement, different reinforcers follow responses controlled by stimuli from different classes. For example, during AB baseline training, selecting comparison B1 on trials with sample A1 might be followed by reinforcer R1 (e.g., potato chip), and selecting B2 on other trials with sample A2 followed by a different reinforcer, R2 (e.g., sip of juice). Previous research has shown that outcome-specific reinforcement procedures may lead to the inclusion of reinforcing stimuli in equivalence classes, and the results are

William Dube acknowledges partial support for manuscript preparation from the Commonwealth Medicine Division, University of Massachusetts Medical School. We thank Nicole Campbell and Elizabeth Morrison for assistance in data collection, and Greg Hanley, Amanda Karsten, and Jason Bourret for comments on the manuscript. This study was completed in partial fulfillment of the requirements for a doctorate of philosophy in behavior analysis in the College of Arts and Sciences of Western New England University for the first author.

Olga Meleshkevich is now at The Durham Center for Education, Institute of Professional Practice, Fitchburg, MA.

Address correspondence to Cammarie Johnson, New England Center for Children, 33 Turnpike Road, Southborough, Massachusetts, 01772 (e-mail: cjohnson@necc.org).

doi: 10.1002/jeab.61

consistent with an interpretation that the reinforcers may function as nodes (e.g., Dube, McIlvane, Maguire, Mackay, & Stoddard, 1989; Schenk, 1994).

Since the early demonstrations of equivalence relations in the 1970s, the experimental analysis of stimulus-equivalence classes has documented increasingly complex experimental paradigms. Three levels of complexity can be seen in the progression from:

(1) *Class formation.* Equivalence classes documented according to the criteria of Sidman and Tailby (1982), as in the example above and variations (e.g., train AB and AC, test BC and CB). Examples with outcome-specific reinforcers as nodes include Dube and McIlvane (1995) and Schenk (1994, Exp. 2). For example, in Schenk, children were given identity matching training AA, BB, CC, and DD with outcome-specific conditioned reinforcers: a red (or blue, counterbalanced across subjects) bead following correct responses to A1, B1, C1, or D1, and a blue (or red) bead following correct responses to A2, B2, C2, or D2. Accumulated beads were periodically exchanged for prizes. Subsequent tests for emergent arbitrary matching AB, BC, DA, and so forth documented ABCD equivalence classes for six of eight children.

(2) *Class expansion.* In a more complex paradigm, previously established classes are expanded to include new stimuli. Lazar, Davis-Lang, & Sanchez (1984) provides an example with nonspecific reinforcement. Participants were first trained to perform AD and DC matching and then emergent AC and CA matching documented the formation of ACD equivalence classes. Then, after ED was trained directly, emergent CE, EC, AE, and EA performances showed expansion of the classes to include the E stimuli. Examples with outcome-specific reinforcement include Dube et al. (1989), Goyos (2000), and Schenk (1994, Exp. 1). For example, in both Dube et al. and Schenk, ABC classes were established by training AB and BC with outcome-specific reinforcement. After exposure to DD identity matching with the same reinforcers, class expansion to include the D stimuli was documented by emergent matching AD, DA, BD, and so forth.

(3) *Class merger.* A higher level of complexity is shown with the merger of separately established equivalence classes. An example with nonspecific reinforcement is Sidman, Kirk, & Willson-Morris (1985, Exp. 2). AB and AC training generated three-member equivalence classes ABC, and DE and DF training generated three-member classes DEF. After EC training with one member of each class, subsequent tests (AD, BE, FC, etc.) documented merger of the three-member classes into six-member classes ABCDEF.

The research literature does not include studies examining class merger with outcome-specific reinforcement and reinforcers as nodes. A study by Minster, Jones, Elliffe, and Muthukumaraswamy (2006, Exp. 2) reported "interclass matching" results that may resemble class merger in some ways but nevertheless seem distinct from merger. College students were trained to perform two conditional discriminations (AB, CB) with a novel procedure that included both nonspecific and outcome-specific reinforcement. There were four comparison stimuli on each baseline trial. Two sets of the sample–comparison relations (A1–B1, C1–B1; and A2–B2, C2–B2) were trained with outcome-specific contingencies (R1 or R2, respectively), and the other two (A3–B3, C3–B3; and A4–B4, C4–B4) with nonspecific contingencies (R3 for all). Initial tests with standard procedures documented the formation of ABC equivalence classes. In subsequent interclass matching tests with sample stimuli that had the reinforcement history with R3 (classes 3 and 4), a nonstandard procedure was used; the comparison stimulus from the same equivalence class as the sample stimulus was absent from the comparison array. For example, on an AB interclass test trial with sample A3, the comparison array was B1, B2, and B4 (i.e., equivalence class member B3 was missing). On trials like these, five of six participants selected the comparison that had previously been related to R3 (B4 in the example). At first, these results might appear to be a merger of the A3–B3–C3 and A4–B4–C4 classes with R3 as node. However, an interpretation of class-3 and -4 merger seems problematic because the interclass test context was inconsistent with those classes. On both

baseline and ABC equivalence-class test trials, matching any class-3 stimulus to a class-4 stimulus was an error and in conflict with the reinforcement contingencies that defined the A3–B3–C3 and A4–B4–C4 classes. The interclass test-trial context was designed to be one in which the original A3–B3–C3 and A4–B4–C4 classes were impossible to demonstrate, and thus it seems inappropriate to characterize the result as a merger of those classes within that context. An alternative interpretation to merger is that the original and interclass test contexts included different but incompatible sets of equivalence relations.

One goal of the present experiment was to examine the potential for class merger via outcome-specific reinforcers in a context that is consistent with the baseline training contingencies, with procedures like those of Sidman et al. (1985). As in Sidman et al., three-member ABC and DEF classes were established separately, but with outcome-specific reinforcers. In contrast to Sidman et al., stimuli from the two 3-member classes were never presented together on training trials prior to the class-merger tests. Positive outcomes on class-merger test trials (AD, BE, FC, etc.) would indicate that the reinforcers functioned as nodes for the merger.

A secondary goal of the present experiment was to test for the inclusion of two types of outcome-specific reinforcers within each equivalence class. The reinforcer delivery procedure was similar to that of Maki, Overmier, Delos, and Gutmann (1995, Exp. 2 and 3) and others. The immediate consequences for correct responses were presentations of colored tokens, with a different color for each potential stimulus class. Participants placed the tokens into bowls of the same color. A distinctive logo for a participant-selected commercial establishment (e.g., Papa Gino's, Walmart, etc.) was attached to each bowl. After sessions, the different color tokens were exchanged for value added to gift cards for these establishments. Following the class-merger tests for six-member equivalence classes described above, participants were given test trials in which digital pictures of the tokens and gift-card logos were presented as samples or comparisons. Positive results on such tests would provide the first experimental demonstration of eight-member equivalence classes that each included two conditioned outcome-specific reinforcers.

Method

Participants

Participants were recruited through personal contacts based on their schedule availability, and the first four who volunteered were included in the experiment. Participants were three adults and one adolescent, all with no known clinical conditions. Rich was a 52-year-old man with some college education employed as a supervisor for group homes. Bob was a 21-year-old man with a high school education employed at a fast food restaurant. Mark was a 19-year-old undergraduate student. Emily was a 14-year-old high school student. None of the participants reported any formal training relevant to stimulus control research.

Setting and Materials

Runs of sessions, typically lasting 1–2 hrs and consisting of 9–19 sessions, were conducted up to 3 days per week in a quiet room where the participant worked or lived. Pretraining materials included a laptop and mouse, small metal bucket, and pennies or paperclips. Experimental materials included the laptop; lined paper and pencils; a video camera and tripod; red, blue, and white bowls (14 cm x 6 cm), each with the logo or name of a specific commercial establishment attached to it; a box of 4-cm red, white, or blue poker chips; and printed data sheets for the experimenter's trial-by-trial record of responses. The bowls were placed in unsystematic locations to the side of the computer.

Experimental stimuli were presented by Microsoft™ PowerPoint software, and the participant used the mouse to select stimuli. At least half of all training sessions were videotaped by a camera oriented towards the computer screen and area where tokens were delivered.

Stimuli and Trial Configuration

Figure 1 shows the 18 black abstract forms that served as the experimental stimuli (taken from Dube & Hiris, 1999). Each was approximately 3 cm^2 when presented on the computer screen. The alphanumeric stimulus designations in the figure were not available to participants.

Conditional discriminations were presented via a matching-to-sample procedure. Each trial began with presentation of a sample stimulus in the center of the lower part of the screen,

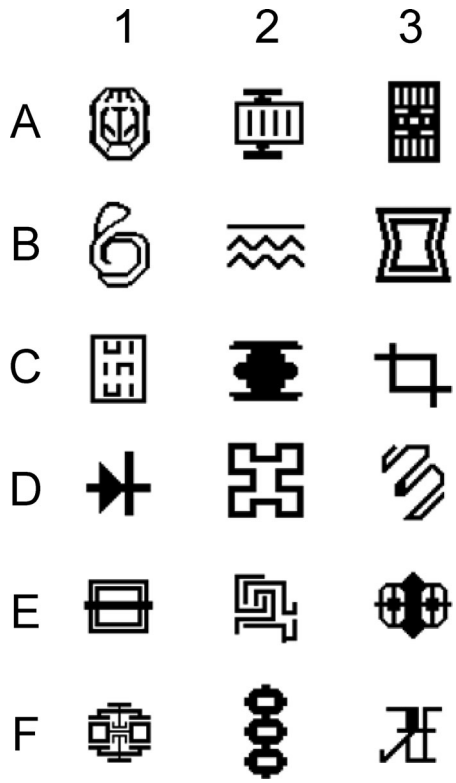

Fig. 1. Experimental stimuli and alphanumeric designations.

followed 0.5 s later by three comparison stimuli in the upper part of screen. A checkbox appeared under each comparison stimulus and a green button with a hyperlink arrow was located in the bottom right corner of the screen. Each PowerPoint slide presented one trial, and teaching sessions consisted of 54 trials (unless noted differently in the procedures) with each sample, comparison, and comparison/location combination appearing on an equal number of trials. The sequence of trials within a session was randomized using a MathCAD free distribution formula.

Procedure

Pretraining. Pretraining sessions introduced the matching-to-sample and delayed-cue procedures to be used in the experiment proper (details below). Matching to sample was introduced with thematically related pictures as stimuli, for example, comparisons school bus, computer, and volleyball ball selected conditionally upon samples ambulance, computer diskette, and soccer ball, respectively (Pilgrim, Jackson, & Galizio, 2000). At the beginning of the first pretraining session, the experimenter gave the following instructions:

> On the screen of the computer you will see four stimuli. One of them will appear on the lower part of the screen and three other stimuli will appear on the upper part of the screen. Put a checkmark into the checkbox under one stimulus. Sometimes I will give you a paperclip [or penny] that will be worth $0.0X at the end of the experiment. Please put the paperclip into the bucket first, and then go to the next page by clicking on the green button in the lower right corner of the screen.

The experimenter presented the paperclip (or penny) immediately following correct responses, and there were no differential consequences following incorrect responses. Clicking on the green button initiated a 0.5-s intertrial interval (ITI), followed by the next trial. The value of the paperclips varied across participants between $0.01 and $0.03. Pretraining continued to a criterion of 6/6 correct unprompted trials.

Overview of experimental phases. The teaching and testing sequences consisted of (a) teaching baseline conditional discriminations with outcome-specific reinforcement, (b) verifying high accuracy on baseline performances in extinction, and (c) testing emergent relations in extinction. The experiment consisted of three phases and advancing through the phases was dependent on the participant's performance. In Phase 1, AC and BC baseline training with outcome-specific reinforcement was followed by emergent-matching tests for ABC equivalence classes. Next, DF and EF training was followed by tests for DEF classes. In Phase 2, tests were conducted to determine if the separately established ABC and DEF equivalence classes would merge based on the shared outcome-specific reinforcing stimuli. If the outcome of these class-merger tests was positive, testing advanced to Phase 3 with tests for matching the reinforcing stimuli to the training stimuli. If the outcome of the initial set of class-merger tests was negative, the participant was given one stimulus–reinforcer matching test

(without differential consequences) and then the class-merger tests were repeated. If class-merger results were then positive, testing advanced to Phase 3.

Instructions to participants. Before the first Phase-1 session, participants were asked to name three gift cards they would like to receive for their participation, and told that money earned during sessions would be accrued on these cards. Because the participants understood verbally stated contingencies, we assumed that participant-selected gift cards and tokens exchanged for value added to those gift cards would function as reinforcers. No formal reinforcer assessments were conducted.

The instructions given before Phase-1 training were similar to those in pretraining, but reflected the changes in consequences:

> [As above] …checkbox under one stimulus. Sometimes I will give you a token, white, blue, or red. When you get a token, put it in the same-color container, white token in the white container, blue token in the blue container, and so forth. Each token will be equal to N cents towards the gift card associated with that container. Please look at which container corresponds with each gift card—white tokens are worth money on X gift card; blue tokens are worth money on Y gift card, and red tokens are worth money on Z gift card. For example, earning 100 red tokens will mean earning $N towards Z gift card.

When these instructions were given to a participant, N was replaced by the value of each token, and X, Y, and Z were replaced by the names of the selected gift cards (e.g., for Rich, "Papa Gino's," "Target," and "Walmart" were associated with white, blue, and red tokens, respectively).

Prior to baseline or emergent relation tests conducted in extinction, the participants were told, "This time there will not be any tokens, but next time tokens will be worth twice as much."

Consequences. Tokens of different colors were presented as differential and class-specific reinforcers, and will be designated as R: R1 (red tokens) followed correct responding to comparison stimuli C1 and F1, R2 (blue tokens) followed correct responding to C2 and F2, and R3 (white tokens) followed correct responding to C3 and F3. When the experimenter delivered a token by dropping it on the tabletop, the participant picked up the token and put it in the bowl of the same color, thus verifying observation and discrimination of the token color.

There were no differential programmed consequences following incorrect responses. After all trials, the participant initiated the 0.5-s ITI by clicking the green button on the computer screen.

The logos or names of the participant-selected gift cards were printed on paper and attached to the red, blue, and white bowls. The logo stimuli will be designated r1 (attached to the red bowl), r2 (blue bowl), and r3 (white bowl). These logo-to-bowl color assignments remained consistent for each participant during all experimental sessions.

Before beginning Phase 1, the average response rate from preliminary training was used to determine a token value for each participant that would approximate the then-current minimum wage. At the end of each session, the participant counted the numbers of red, blue, and white tokens earned and recorded this information on paper. At the end of each run, the experimenter recorded the total numbers of tokens and the amount of money that would be credited on each gift card. The cumulative total earned on each gift card was reviewed with the participant at the beginning of each run, and the amount to be added was reviewed at the end of each run.

Delayed-cue prompting procedure. A variation of the delayed-cue procedure (Handen & Zane, 1987; Touchette, 1971) was used to teach arbitrary matching-to-sample performances. The correct comparison stimulus blinked three times as a prompt. There was a delay interval between presentation of the comparison stimulus array and the prompt. This delay interval remained constant within each session. The delay was 0.5 s on the first teaching step, 3 s on the second step, and on the third and final step there was no prompt. From session to session, the teaching program advanced or backed up one step according to the performance criteria described below.

Performance criteria. The performance accuracy criterion was two or fewer errors per session and no more than one error for any

particular sample–comparison trial type (e.g., sample A1 on AC trials). This criterion was used (a) during delayed-cue training for advancing the teaching step, (b) as a measure of required accuracy for unprompted baseline sessions, and (c) as a measure of positive outcomes on test probe trials. During delayed-cue training, the criterion to back up to the previous teaching step was more than two errors with the same sample or four total errors. If this error criterion was met, the session was terminated and a session with the previous teaching step was presented.

General testing protocol. Probe trials for tested relations were interspersed among baseline trials. All trials within test sessions, both probe and baseline, were conducted with no programmed trial consequences (procedural extinction) and the token bowls were not present during tests. If at least half of the tests within a series of test sessions yielded positive outcomes, those with negative outcomes were repeated in order to evaluate delayed emergence, a relatively common observation in equivalence research (Sidman, 1994).

Each test run began with a baseline review session with reinforcement, repeated if necessary until performance met the accuracy criterion. If performance on baseline trials during test sessions did not meet the accuracy criterion, further testing was suspended until the participant demonstrated criterion performance of those baseline relations. Each test session was followed by a baseline review session with double reinforcement (tokens worth twice the usual amount). The double reinforcement contingency was used so that overall rate of reinforcement during runs that included extinction would be comparable to that of previous training runs.

Phase 1a: ABC Three-Member Equivalence Classes

Baseline training. A comparison-as-node teaching structure (Saunders & Green, 1999) was used. AC was trained first with the delayed-cue procedure and reinforcement procedures as described above (matching sample A1 to comparison C1 followed by R1, etc.). After acquisition of AC, BC was trained with similar contingencies (matching B1 to C1 followed by R1, etc.). These trained performances are indicated by the arrows on the left side of Figure 2.

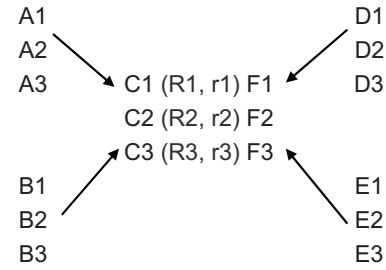

Fig. 2. Schematic showing the trained relations. R1, R2, R3 and r1, r2, r3 indicate the outcome-specific reinforcer tokens and associated gift-card logos, respectively.

After AC and BC training was completed, a session of AC/BC mixed baseline with these trial types interspersed was presented with reinforcement. If the accuracy criterion was not met, the discrimination(s) on which errors occurred were reviewed with a 3-s delayed cue before the mixed baseline session was repeated. When the accuracy criterion was met on the mixed baseline with reinforcement, a session was presented without programmed consequences (no token deliveries), followed by a session with double reinforcement. For Rich (who made many errors in training), the run following the introduction of the AC/BC mixed baseline began with reviews of AC and BC with the 3-s delayed cue until criterion performance was verified.

ABC stimulus equivalence tests. Combined symmetry and transitivity tests (AB and BA) were conducted first, followed by tests for symmetry (CA and CB, see thick dashed arrows on the left side of Fig. 3). Probe trials were

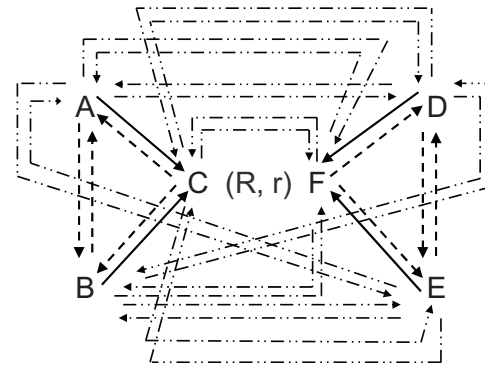

Fig. 3. Schematic showing the trained relations (solid arrows), ABC and DEF emergent relations tests from Phase 1 (thick dashed arrows), and ABCDEF merger tests (thin dashed arrows).

interspersed among AC/BC mixed baseline trials. Each session presented one test type (i.e., AB, BA, CB, or CA) and consisted of 18 probe trials interspersed among 36 baseline trials.

Phase 1b: DEF Three-Member Equivalence Classes

Training and testing paralleled that for the ABC classes. DF and EF conditional discriminations were trained with the delayed cue procedure and outcome-specific reinforcement contingencies (R1 followed matching D1 to F1, etc., see arrows on the right side of Fig. 2). Following acquisition, sessions of DF/EF mixed baseline were presented with reinforcement, and then in extinction. Probe trials tested for combined symmetry and transitivity (DE, ED) and symmetry (FD, FE, see thick dashed arrows on the right side of Fig. 3).

Phase 2: ABCDEF Six-Member Equivalence Classes

Combined baseline. After the three-member equivalence tests were completed, a combined baseline with six interspersed trials each of AC, BC, DF, and EF was presented with reinforcement. As above, if the accuracy criterion was not met, the discrimination(s) with errors were reviewed with the 3-s delayed cue before the combined baseline session was repeated. For Rich (who made many errors during Phase 1), the run following the introduction of the combined baseline began with a 48-trial session that included 12 trials of each discrimination with the 3-s delayed cue, and criterion performance on each discrimination was required before the combined baseline was reviewed. When the accuracy criterion was met on the combined baseline with reinforcement, a session was presented without programmed consequences, followed by a session with double reinforcement.

Six-member stimulus equivalence tests. A total of 54 sample-comparison relations were tested in 18 test types: AD, AE, AF, BD, BE, BF, CD, CE, CF, DA, DB, DC, EA, EB, EC, FA, FB, FC (see thin dashed arrows in Fig. 3). In test sessions, 18 test trials of one type were interspersed among 36 combined baseline trials. The test series consisted of one session with each test type, presented in a random order across participants (see data presentation in Results for individual test sequences). If the majority of test results were positive, the tests with negative outcomes were repeated at this point. If the majority of test results were negative, the participant received one AR test session (described below) to probe the status of the reinforcers as class members, and then the six-member tests were repeated.

Phase 3: Reinforcer Matching Tests

Each participant received one or more of the following tests: AR, DR, FR, RD, RC, Ar, Cr, Er, rB, rF. These tests sampled relations between baseline stimuli and digital pictures of the tokens (R1, R2, R3), or baseline stimuli and graphic representations of the gift-card logos (r1, r2, r3). These 10 test types were selected from the 24 possible tests types such that reinforcers were presented as samples on some test types and as comparisons on others, and with baseline stimuli drawn equally from ABC and DEF classes. Two additional tests types, Rr and rR, tested relations between pairs of reinforcing stimuli. Test sessions for Bob, Mark, and Rich included 18 probe trials of one test type interspersed among 36 combined baseline trials (e.g., AR). For Emily the procedure was modified to evaluate more emergent relations per session, with nine trials for each of two different test types within one session (e.g., ARFR). As in earlier phases, criterion performance on the combined baseline was required before each test session.

Response Measurement and Interobserver Agreement

The experimenter manually recorded correct and incorrect responses on printed data sheets. The computer also recorded the participant's selections in PowerPoint format as a filled checkbox under one of the comparison stimuli. Interobserver agreement was assessed by an independent recorder who compared the PowerPoint permanent record with the experimenter's completed data sheet for 33% of all sessions for Bob and Mark, or 20% of teaching and testing sessions for Rich and Emily (i.e., excluding baseline reviews before or after emergent relation tests). Agreement scores were calculated by dividing the number of agreements by the total of agreements plus disagreements. The mean agreement score for each participant was greater than 0.99.

Procedural integrity measures on the independent variables not managed by the computer

were assessed for 20% of teaching sessions that had reinforcer deliveries. An independent observer watched videotaped sessions and recorded trial-by-trial data on appropriate delivery or nondelivery of tokens based on response (correct or incorrect), delivery of the appropriate class-assigned token, and immediacy of token delivery (within 1 s of participant response). Procedural integrity was calculated by dividing the number of trials with all measures completed as prescribed by the number of trials observed. Mean procedural integrity score for each participant was greater than 0.94.

Results

Pretraining

Bob, Emily, and Mark responded correctly on all prompted trials and all six unprompted trials in the first session. Rich made errors on unprompted trials in 15 sessions within the first run; he responded almost exclusively to one of the three comparisons. When new pretraining stimuli were introduced in a second run, Rich met a more stringent criterion of 26/27 correct in the third unprompted session.

Phase 1a: ABC Three-Member Equivalence Classes

Table 1 shows that all participants met the established accuracy criterion for AC, BC, and AC/BC mixed baseline. Bob, Mark, and Emily met the accuracy criterion for AC and BC within 162 trials or fewer, and for the AC/BC mixed baseline in 54 trials. Rich, however, required many additional trials to meet and maintain criterion performance on AC and to meet criterion on the AC/BC baseline. Bob, Emily, and Rich made no errors on their first AC/BC mixed baseline in extinction (not shown in Table 1). Mark made three errors, but then performed without error when AC/BC was repeated in extinction after a mixed baseline session with double reinforcement. All participants maintained criterion performance on unreinforced AC/BC trials interspersed with probe trials during ABC equivalence tests and the reinforced trials in AC/BC mixed baseline sessions with double reinforcement that followed each test.

Outcomes were positive for all participants on each of the four emergent relation tests. Performances consistent with equivalence classes were shown in 18/18 probe trials for each test type for three participants. Emily's responses were class-consistent on 18/18 probe trials for all tests except AB, which was 16/18 with the inconsistent responses on different trial types.

Phase 1b: DEF Three-Member Equivalence Classes

The bottom portion of Table 1 shows that all participants met the accuracy criterion for DF, EF and DF/EF mixed baseline. All except Rich met the criterion immediately on the DF/EF mixed baseline. Rich required retraining on DF and EF when performances deteriorated (see Table 1), typically following changes in session parameters such as mixing trial types or introducing extinction.

Table 2 shows that participants Bob, Emily, and Mark had immediately positive outcomes

Table 1

Number of Trials to Meet and Maintain Criterion Performance on Baseline Relations

Phase	Relation/s	Performance	Number of Trials			
			Bob	Emily	Mark	Rich
1a	AC	Criterion	108	162	108	281
		Retraining				135
	BC	Criterion	108	162	108	162
	AC/BC	Criterion	54	54	54	270
1b	DF	Criterion	108	162	108	162
		Retraining				108
	EF	Criterion	108	162	108	162
		Retraining				54
	DF/EF	Criterion	54	54	54	108
2	AC/BC/DF/EF	Criterion	54	54	54	270
	AC	Retraining				54

Note. Retraining was initiated when previous criterion performance was no longer maintained and continued until criterion performance was again demonstrated.

Table 2

Number of Class-Consistent Responses on 18 Probe Trials in Phase-1b DEF Stimulus Equivalence Tests

Relations Tested	Participants			
	Bob	Emily	Mark	Rich
DE	18	18	18	12
				12
ED	18	18	18	12
				11
FD	18	18	18	18
FE	18	18	18	18
DE				12
				13
ED				15
				17
DE				17
				18
ED				18
				18

on each of the four emergent relations tests for DEF classes. Rich's responses were inconsistent with equivalence relations on the combined equivalence tests, which were repeated one time each before the symmetry tests. After positive outcomes on the symmetry tests, Rich's performance on combined equivalence tests was not consistent with equivalence relations until the DE and ED tests had been repeated several times (see Table 2). After performance consistent with equivalence relations was demonstrated, however, it was maintained with high accuracy for two consecutive sessions.

Phase 2: ABCDEF Six-Member Equivalence Classes

When AC/BC and DF/EF baselines were first combined, Table 1 shows that Bob, Emily, and Mark immediately demonstrated criterion performance with reinforcement; they also met the accuracy criterion subsequently in extinction (not shown in Table 1). Rich, however, required five combined baseline sessions (270 trials) and one session of retraining AC before meeting the criterion with reinforcement.

As shown in the upper panels of Figure 4, outcomes on the first set of class-merger equivalence tests were positive for Bob and Emily. After meeting the outcome criterion on the first test (FC), Bob did not reach criterion levels on the next three tests (BD, AF, DB), but all subsequent test types had positive outcomes, including repeated tests of those relations with initial negative outcomes (shown by gray bars in Fig. 4).

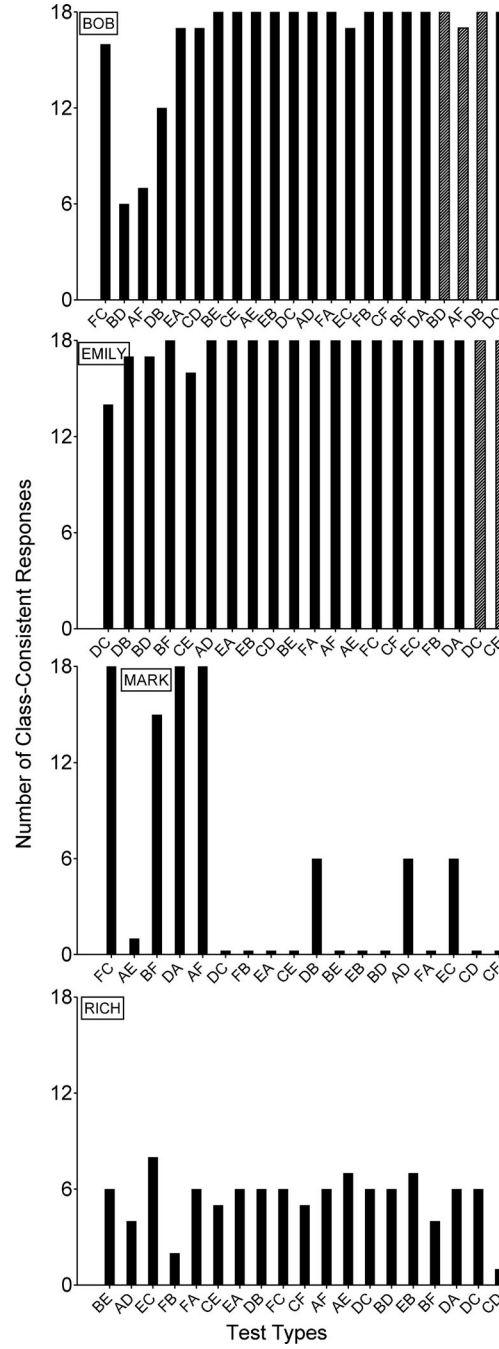

Fig. 4. Number of class-consistent responses on 18 probe trials in the initial set of Phase-2 class merger tests. Gray bars show tests repeated because initial results were negative.

Emily met the outcome criterion in the initial test session for 16 of the 18 test types. For two

early tests, DC and CE, the outcome criterion was not met until these tests were repeated at the end of the sequence.

Mark's results on the first set of class-merger tests were variable (third panel of Fig. 4). Outcomes were positive for only three tests near the beginning of the testing sequence (FC, DA, AF). On the tests with negative outcomes Mark consistently selected a specific comparison stimulus for each sample stimulus, indicating stimulus control by sample stimuli (sometimes termed "arbitrary assignment" or "unreinforced conditional selection"; Saunders, Saunders, Kirby, & Spradlin, 1988). These sample–comparison assignments, however, were inconsistent across test types and often inconsistent with the experimental stimulus classes (producing 0 class-consistent responses). The top panel of Figure 5 shows a negative outcome on the AR test, but then Mark's responses in the second set of merger tests became perfectly consistent with class merger on all test types.

Results for Rich were negative on every test type (lower panels of Figs. 4 and 5), including two AR tests. Rich typically selected the same comparison stimulus on all probe trials, indicating no stimulus control of comparison selections by the sample stimuli.

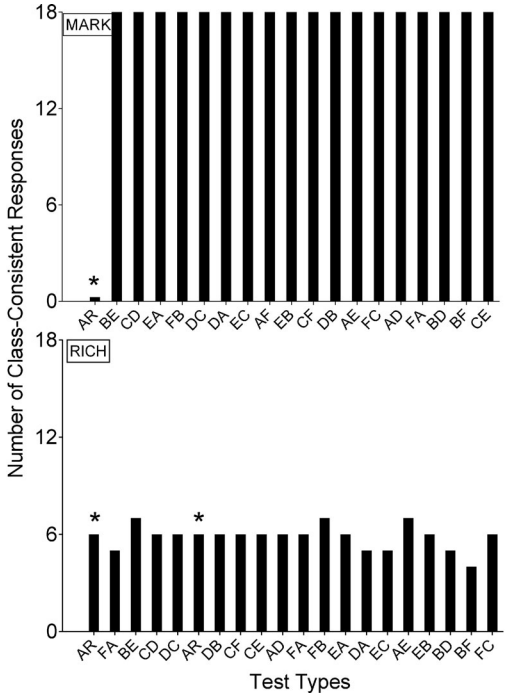

Fig. 5. Number of class-consistent responses on 18 probe trials in the second set of class merger tests that followed exposure to AR probes (shown by asterisks).

Phase 3: Reinforcer Matching Tests

Table 3 shows the results of reinforcer matching tests conducted after positive outcomes on class merger tests. Outcomes were consistently positive for Bob and Emily on these tests. Mark's results were also positive on his second AR test that followed the second set of class merger tests. Mark withdrew from the experiment before additional tests could be given.

Discussion

The results with three of the four participants documented the merger of separately established three-member equivalence classes ABC and DEF with outcome-specific reinforcers as nodes. Class-merger test results for two participants, Bob and Emily, were immediately positive for six-member ABCDEF classes. These participants were given additional reinforcer matching tests with stimuli from the six-member classes, pictures of the colored tokens, and pictures of the gift-card logos. The results of these tests were consistent with eight-member equivalence classes including the baseline sample and comparison stimuli and both of the outcome-specific conditioned reinforcing stimuli. These are the largest experimental equivalence classes with outcome-specific reinforcers demonstrated thus far. For a third

Table 3

Number of Class-Consistent Responses on 18 Probe Trials in Phase-3 Reinforcer Matching Tests

Relations Tested	Participants		
	Bob	Emily	Mark
AR	18		18
ARFR		18	
DR	18		
RDRC		18	
Ar	18		
ErCr		18	
rBrF		18	
rR	18		
rRRr		17	

Note. Mark withdrew from the experiment after one test in Phase 3.

participant, Mark, six-member class merger test results were initially negative but became positive after exposure to probe trials with one set of experimental stimuli as samples and pictures of the tokens as comparisons. Additional test results were consistent with seven-member equivalence classes; Mark withdrew from the experiment before receiving tests for eight-member classes. For the fourth participant, Rich, results of the initial ABC and DEF three-member equivalence class tests were positive, but repeated class merger and AR reinforcer matching tests were always negative.

One question that may arise concerning the immediately positive results with Bob and Emily is whether they are better interpreted as the merger of two three-member classes, or if the D, F, and E stimuli had been added to the ABC classes during the DF and EF training and thus better characterized as class expansion. McIlvane and Dube (1990) proposed that this type of question cannot be answered satisfactorily within behavior analysis at the level of overt operant behavior because the tests themselves may set the occasion for the emergence of the equivalence relations. (For one alternative approach, see electrophysiological data on neural activity correlates of equivalence class formation prior to vs. following behavioral tests for emergent relations in Haimson, Wilkinson, Rosenquist, Ouimet, & McIlvane, 2009.) Terms like "equivalence class," "expansion," and "merger" are expressions of the experimenter's inferences about stimulus control based on observing the outcomes of particular training and testing procedures. Procedurally, the sequence of events and behavior in the present experiment was class merger. Other sequences of events may produce other outcomes, but the present procedures provide the framework for the interpretation of the present results.

The question about class merger versus class expansion seems unlikely to arise with Mark's results because he failed the first series of class merger tests. Those initial test results documented two independent three-member classes, ABC and DEF. Further, the negative results for his first AR test indicated that the outcome-specific reinforcers were not equivalent to the baseline sample and comparison stimuli at that point in the experiment, at least not within a matching-to-sample context. These results are consistent with previous research in which failure of outcome-specific reinforcers to function as nodes was accompanied by failure to function as samples or comparisons on matching tests (Dube & McIlvane, 1995, Exp. 2; Joseph, Overmier, & Thompson, 1997; Schenk, 1994, Exp. 2). One possible explanation for these failures (and those with Rich, discussed below) is that the operative reinforcing stimulus was something shared by all of the consequences, such as the movements of the experimenter as she delivered tokens, the sound of the tokens dropping on the table top, the equivalent monetary value of the tokens, and so forth. In such a case the functional reinforcer would be nonspecific and equivalence relations would not be predicted because they would conflict with the baseline reinforcement contingencies.

Mark's exposure to the AR probes apparently facilitated class merger, even though trial-by-trial consequences were not provided during the test and his responses were inconsistent with the experimental equivalence classes. The sequence of events suggests that equivalence relations with the outcome-specific reinforcers were established during the reinforced baseline review session that followed the initial AR test. Positive outcomes on the subsequent second set of class merger tests indicated that the reinforcers then functioned as nodes, and results of the second AR test showed that the pictures of tokens were substitutable for other class members in matching to sample. It is not clear why exposure to the initial AR test, in which the R stimuli appeared for the first time within matching-to-sample trials, had a facilitative effect. One possibility is that seeing the tokens presented as comparisons encouraged him to observe and attend for the first time (in the baseline review session that followed) to the specific stimulus sequences of samples, comparisons, and token colors that distinguished the classes.

Mark's results are analogous in some ways to those of Minster et al. (2006). In both studies, results of initial tests indicated that there were no equivalence relations between the reinforcing stimuli and the baseline samples and comparisons (this is sometimes metaphorically referred to as "dropped out" of the equivalence classes; Sidman, 2000). This was shown in the present study by negative results on the first set of merger tests, and in Minster et al. by positive results on the initial equivalence tests with class-3 and -4 stimuli. Then, following a change in the testing context, stimulus–reinforcer equivalence relations were

demonstrated. In the present experiment the changed testing context was one in which the reinforcers had appeared as matching-to-sample stimuli, and in Minster et al. it was one in which behavior consistent with the original classes was no longer possible.

Rich's results were consistently negative on both sets of class merger tests and the interspersed AR tests. As noted above, such occasional failures of outcome-specific reinforcers to function as nodes have been reported in previous studies (Dube & McIlvane, 1995; Joseph et al., 1997; Schenk, 1994). Rich made many more errors than the other participants throughout baseline training, and one likely possibility is that whatever unknown forms of stimulus control produced those errors also occurred in test sessions. As noted in the Results section, one characteristic of this competing stimulus control was a failure of conditional control by sample stimuli, beginning in the first pretraining session and persisting through the test sessions. To determine whether class merger was demonstrable with this participant under any circumstances, he was trained to perform FC matching (after the tests shown in Fig. 5 were completed), a replication of the Sidman et al. (1985) procedure. A third set of class merger tests followed this training and results were positive on all test types, documenting six-member ABCDEF equivalence classes with conventional procedures that did not depend upon outcome-specific reinforcers as nodes. Results on several subsequent reinforcer-matching tests remained at chance levels, however, indicating that the reinforcers were still not equivalence class members.

The present results, showing that separately established equivalence classes may merge because of common outcomes, may have implications for clinical behavior analysis. Guinther and Dougher (2011) wrote that the clinical relevance of equivalence research is most apparent in demonstrations of the transfer of functions among stimuli that are members of the same equivalence classes. For example, Dougher, Auguston, Markham, Greenway, and Wulfurt (1994) showed that acquisition and extinction of a conditioned fear response, measured by skin conductance levels after Pavlovian conditioning with electric shock, transferred among equivalence class members. Similar results have been reported for conditioned sexual arousal (Roche & Barnes, 1997), operant shock avoidance (Augustson & Dougher, 1997), and a range of other stimulus functions (for a brief review see Guinther & Dougher, 2011). For almost all individuals, the environment includes equivalence classes whose members may, under certain circumstances, elicit emotions or evoke emotion-related operant behavior. An interpretive account of the present results suggests one way in which an individual's status quo with respect to such classes may be upset. If specific consequences for behavior controlled by stimuli from an equivalence class were to change, such that they became the same specific consequences as those for another previously independent equivalence class, there would be the potential for class merger. Subsequently, members of one of these classes may unexpectedly elicit or evoke emotional responses via transfer of function, possibly in ways that have clinical implications. Future investigations may lead to the development of laboratory models and translational research to study this possibility.

References

Auguston, E. M., & Dougher, M. J. (1997). The transfer of avoidance evoking functions through stimulus equivalence classes. *Journal of Behavior Therapy and Experimental Psychiatry, 28*, 181–191.

Dougher, M. J., Auguston, E., Markham, M. R., Greenway, D. E., & Wulfurt, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior, 62*, 331–351.

Dube, W. V., & Hiris, E. J. (1999). MTS v 11.6.7 [Computer software documentation]. Waltham, MA: E.K. Shriver Center.

Dube, W. V., & McIlvane, W. J. (1995). Stimulus–reinforcer relations and emergent matching to sample. *The Psychological Record, 45*, 591–612.

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

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.

Goyos, C. (2000). Equivalence class formation via common reinforcers among preschool children. *The Psychological Record, 50*, 629–654.

Guinther, P. M., & Dougher, M. J. (2011). From behavioral research to clinical therapy. In G. J. Madden (Ed.), *APA handbooks in psychology. APA handbook of behavior analysis, Vol. 2: Translating principles into practice* (pp. 3–32). Washington, DC: American Psychological Association.

Haimson, B., Wilkinson, K. M., Rosenquist, C., Ouimet, C., & McIlvane, W. J. (2009). Electrophysiological correlates

of stimulus equivalence processes. *Journal of the Experimental Analysis of Behavior, 92*, 245–256.

Handen, B. L., & Zane, T. (1987). Delayed prompting: A review of procedural variations and results. *Research in Developmental Disabilities, 8*, 307–330.

Joseph, B., Overmier, J. B., & Thompson, T. (1997). Food- and nonfood-related differential outcomes in equivalence learning by adults with Prader-Willi syndrome. *American Journal on Mental Retardation, 101*, 374–386.

Lazar, R. M., Davis-Lang, D., & Sanchez, L. (1984). The formation of visual stimulus equivalences in children. *Journal of the Experimental Analysis of Behavior, 41*, 251–266.

Maki, P., Overmier, J. B., Delos, S., & Gutmann, A. J. (1995). Expectancies as factors influencing conditional discrimination performance of children. *The Psychological Record, 45*, 45–71.

McIlvane, W. J., & Dube, W. V. (1990). Do stimulus classes exist before they are tested? *The Analysis of Verbal Behavior, 8*, 13–17.

Minster, S. T., Jones, M., Elliffe, D., & Muthukumaraswamy, S. D. (2006). Stimulus equivalence: Testing Sidman's (2000) theory. *Journal of the Experimental Analysis of Behavior, 85*, 371–391.

Pilgrim, C., Jackson, J., & Galizio, M. (2000). Acquisition of arbitrary conditional discriminations by young normally developing children. *Journal of the Experimental Analysis of Behavior, 73*, 177–193.

Roche, B., & Barnes, D. (1997). A transformation of respondently conditioned stimulus function in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior, 67*, 275–301.

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

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

Schenk, J. J. (1994). Emergent relations of equivalence generated by outcome-specific consequences in conditional discrimination. *The Psychological Record, 44*, 537–558.

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

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*, 127–146.

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.

Touchette, P. E. (1971). Transfer of stimulus control: Measuring the moment of transfer. *Journal of the Experimental Analysis of Behavior, 15*, 347–354.

Received: August 21, 2013
Final Acceptance: October 24, 2013