JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1998, **70,** 87–102 NUMBER 1 (JULY)

STIMULUS EQUIVALENCE: EFFECTS OF A DEFAULT-RESPONSE OPTION ON EMERGENCE OF UNTRAINED STIMULUS RELATIONS

ANDREW INNIS, SCOTT D. LANE,
ERIN R. MILLER, AND THOMAS S. CRITCHFIELD

AUBURN UNIVERSITY

Default-response options, intended to measure uncertainty, sometimes are included in discrete-choice measurement in an attempt to clarify stimulus control of remaining response options. Recent experiments have used a default-response procedure to investigate emergent stimulus relations, but no study to date has compared effects of different default-response procedures on emergence of the untrained relations that define stimulus equivalence. Five groups of college undergraduates (each $n = 16$) completed a conditional discrimination training procedure to instate the stimulus relations prerequisite to three three-member equivalence classes; a training review intermingling all of the explicitly trained relations; and tests for emergent relations. The groups differed in terms of (a) presence versus absence of a "none" option during emergent relations tests and (b) the amount of experience with "catch trials" in which "none" was the correct selection. Stimulus equivalence was demonstrated in 94% of subjects in a control group who were trained and tested without the "none" response option and without catch trials and in 41% of subjects in the "none" groups. Among subjects in the "none" groups who failed to demonstrate equivalence initially, 95% did so when retested under control-group conditions. Across "none" groups, probability of equivalence class formation was positively correlated with amount of experimental experience with catch trials in preliminary training and equivalence testing. Among the emergent relations defining stimulus equivalence, reflexivity was most often precluded by the "none" option, although there was evidence of group differences in relation specificity. These results suggest that a default-response option can interfere with the formation of emergent relations, and that the effects are contextually sensitive. Although there may be advantages to employing default-response procedures in studies of emergent stimulus relations, the responses they control should be viewed as behavior under specific stimulus control rather than a generic expression of uncertainty.

Key words: conditional discrimination, stimulus equivalence, default-response option, uncertainty, points, computer mouse click, college students

Stimulus equivalence procedures are notable for promoting the emergence of numerous untrained, arbitrary relations from the training of only a few. Much of the stimulus equivalence literature is devoted to understanding when and how these emergent relations occur, and the range of methods applied to this endeavor has grown in recent years. Although the primary focus remains on accuracy (i.e., class consistency) of performance in match-to-sample tasks, equivalence experiments increasingly include measures of other outcomes, such as latency of match-to-sample responses (Bentall, Dickins, & Fox, 1993; Spencer & Chase, 1996), self-reports about match-to-sample responses (Lane & Critchfield, 1996; Wulfert, Dougher, & Greenway, 1991), or transfer of function among stimuli involved in the match-to-sample procedures (Dougher, 1994).

One recent, but largely unexplored, innovation in research on emergent stimulus relations is to include a default-response option within the match-to-sample procedure used to test for emergent relations. Match-to-sample tasks require subjects to select from among a few stimuli, and weak stimulus control is manifested by inconsistent selection patterns. Taken at face value, default-response procedures could be said to allow subjects to answer with the equivalent of "don't know" or "none of the above." A guiding assumption in the design of such procedures is that, by strongly controlling behavior that otherwise would be weakly determined by experimental manipulations, the default-re-

We thank Carol Pilgrim for a helpful experimental design suggestion. We are also indebted to two anonymous reviewers, whose comments had a major impact on the final version of this paper. Portions of the data were presented at meetings of the Association for Behavior Analysis (1996) and the Southeastern Association for Behavior Analysis (1997).

Address correspondence to T. Critchfield, Department of Psychology, Illinois State University, Normal, Illinois 61790.

sponse option can remove noise from the measurement system and thus clarify the control of behavior by other stimuli (e.g., Bickel, Oliveto, Kamien, Higgins, & Hughes, 1993; Fields, Reeve, Adams, Brown, & Verhave, 1997; Shields, Smith, & Washburn, 1997; cf. Dube & McIlvane, 1996). Any tool that clarifies stimulus control should be welcome in the arsenal of methods used to examine stimulus equivalence.

Two recent studies of emergent arbitrary relations have employed a default-response option. In a test of relational frame theory, Roche and Barnes (1996) sometimes included a response option that was designated by a question mark, and told subjects to choose this option ''if you think that none of the available choices are correct'' (p. 462). Emergent relations predicted from relational frame theory apparently were not disrupted by the presence of this response option, and 6 of 10 subjects chose the question mark on trials that contained no match as defined by the theory. In a study of equivalence class expansion through primary generalization, which involved tests for matching class members to novel exemplars defined by physical similarity to same-class members, Fields and colleagues sometimes included a response option labeled ''neither'' (Fields et al., 1997). Some stimuli that were treated as class consistent in a traditional procedure (excluding the ''neither'' option) were treated as class inconsistent with the default-response option present. In both studies just described, conclusions about the robustness of emergent stimulus control were based partly on the default-response data.

Default responses are not unique to studies of equivalence, and studies from other research areas indicate that (a) default-response options can alter the distribution of responses made to other options, but (b) the frequency and effects of default responses depend on a variety of factors (e.g., Bickel et al., 1993; Critchfield, 1996; Johansen, Gips, & Rich, 1993; Shields et al., 1997; Sudman, Bradburn, & Schwartz, 1996; Thomson, 1920). It seems likely, therefore, that default-response procedures will be most informative to the study of emergent stimulus relations when their operating characteristics have been investigated in the proper context (e.g., see R. Saunders, 1996). The equivalence studies by Roche and Barnes (1996) and Fields et al. (1997) represent interesting applications of default-response technology. But it remains to be seen how variations in default-response procedures affect performance during equivalence tests, and how variations in equivalence procedures modulate the impact of default options on equivalence class formation. No single study can meet this agenda, but the present report describes an attempt to identify some of the relevant variables. Five groups of subjects completed conditional discrimination training and equivalence testing that were identical except for the presence or absence of a default-response option and the amount of experience that might help to define the function of the default option. The resulting groups included one that was trained and tested traditionally and four others that completed equivalence testing with a default-response option (labeled ''none'') available.

METHOD

Subjects and Apparatus

Undergraduate students ($N = 80$), recruited from psychology classes, participated in exchange for bonus course credit. Participation time was earned according to contingencies described below. Time of participation was converted to course credit according to various policies established by course instructors, typically within the constraint that bonus credit not exceed 2% of the course total. Subjects worked alone in a small room, responding with a mouse to figures presented on a color monitor. An IBM-compatible personal computer, housed in an adjacent room, was used to control session events and collect the data, using custom programs written in Microsoft QuickBasic®.

Procedure

Stimuli. The experimental contingencies divided nine stimuli, each red and about 2 cm square, into three mutually exclusive three-member classes, arbitrarily defined in the sense that there were no intended physical or thematic similarities among members of a class. Figure 1 shows the stimuli and classes.

Reinforcer. A deception was used to establish points as reinforcers. Subjects were told at the start of participation that each correct re-

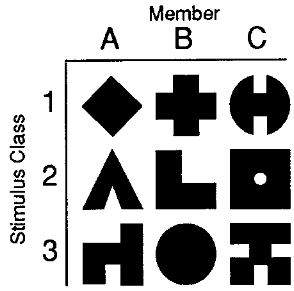

Fig. 1. Stimuli used in the experiment.

sponse earned 20 s of participation time, and that the amount of total participation time documented for extra credit purposes depended entirely on session performance. The research protocol actually required that (a) all subjects receive credit at least commensurate with time actually spent in the experiment; and (b) any subject who earned fewer seconds than actually spent in the experiment be debriefed immediately. Otherwise, subjects were debriefed by mail at the end of the academic quarter. For all subjects, time earned at least matched actual time of participation. Thus, the deception was not revealed until after all subjects had completed the experiment.

Trial format. During sessions, the subject's screen was black except for five boxes, each about 5 cm square, outlined in white. One box was located in the center of the screen, and the others appeared at the corners. At the start of each trial, a sample stimulus appeared in the center box. A mouse click on this stimulus produced comparison stimuli in three of the corner boxes. If the "none" option was available, the word "none" appeared in white, about 0.5 cm high, inside the fourth box. The location of the correct comparison stimulus, and of the "none" option, was counterbalanced within sessions. A mouse click inside one of the comparison boxes registered a response and produced feedback, if scheduled, followed immediately by the next trial.

Feedback. During training phases, subjects received feedback after each trial. Immediately after a comparison stimulus was selected, the stimuli and boxes were cleared from the screen, and the word "CORRECT" or "INCORRECT" appeared in letters about 3 cm high at the top of the screen. About 5 cm below, near the center of the screen, was the message "TIME EARNED = X SECONDS," with X = 20 for correct responses and 0 for incorrect responses. After 1 s, the screen cleared and the next trial began.

Instructions. The informed consent agreement, and text shown on the subject's screen prior to the start of the first session, stated the following:

> You will be asked to use a mouse to identify logical relationships between alphanumeric and geometric symbols on the screen of a computer. Correct selections will earn seconds of extra credit, which may or may not be shown on your computer screen. At the beginning of each trial five boxes will appear on the screen, one in the middle and four in the corners. A shape will appear in the middle box. Moving the mouse into the middle box and pressing the mouse button will produce shapes in either three or four of the corner boxes. You may earn seconds of extra credit by selecting the correct shape from one of the corner boxes. To make a selection, place the cursor in the corner box you wish to select and press the red mouse button. To begin this session, press the red mouse button now.

Experimental design. Subjects were assigned nonsystematically to one of five groups, each containing 16 subjects. For all groups, conditional discriminations first were taught separately and then were intermingled within sessions in a training review; emergent relations were subsequently tested during sessions with no feedback. Table 1 summarizes the key differences between groups. First, some groups participated in a preliminary phase in which the selection of the "none" option was reinforced on *catch trials,* which included no class-consistent selection. Second, the groups received different amounts of experience with catch trials following the conclusion of the preliminary training phase in which they were introduced. Specifically, both during the mixed training review and during equivalence testing, there were group differences in the proportion of catch trials encountered among the previously trained relations. This feature was included in the design under the assumption that ongoing experience with catch trials might help to determine the function of the "none" option. Finally, the "none" option could be present or absent during emergent relations testing.

Table 1

Procedure summary for the five groups.

Variable	Group				
	Control	No training	0%	25%	50%
Catch-trial pretraining with none option	No	No	Yes	Yes	Yes
Catch trials in mixed training review	No	No	No (0%)[a]	27%[a]	50%[a]
None option on AB and AC trials during mixed training review	No	No	No	Yes	Yes
Catch trials during equivalence testing	No	No	No (0%)[a]	20%[a]	50%[a]
None option on test trials	No	Yes	Yes	Yes	Yes

[a] Percentage of previously trained relations that were catch trials rather than A-B and A-C relations.

Table 1 illustrates that the between-groups manipulations represented packages of variables rather than a component analysis involving all possible combinations of these variables. Table 2 shows the sequence of phases for each group and the stimulus configurations employed in the trials of each phase.

Phases 1 and 2: Conditional discrimination training. Initial training was identical for all groups and incorporated relations necessary for three separate three-member stimulus classes. For each potential class, a single sample stimulus was paired, on different trials, with each of the other two stimuli in the class. Phases 1 and 2 employed identical procedures but used different stimuli. In Phase 1, all of the A-B relations (A1-B1, A2-B2, and A3-B3) were trained. In Phase 2, all of the A-C relations (A1-C1, A2-C2, and A3-C3) were trained. In both phases, sessions were 24 trials long, and the mastery criterion was completion of two consecutive sessions with accuracy of at least 92%. Relations from different stimulus classes alternated unsystematically across trials, with the caveat that sample-comparison pairs from a single class could appear on a maximum of three consecutive trials. In Phases 1 and 2, each comparison stimulus array included the correct match and two nonmatching comparison stimuli (Table 2). The fourth comparison box was blank. Feedback followed every trial.

Phase 3: Catch-trial training. In Phase 3, subjects learned to use the ''none'' option. Phase 3 was omitted for the control group, which never encountered the default-response option, and for the no-training group, which first encountered the default-response option at the start of equivalence testing. Sessions included 24 catch trials in which there was no correct response option according to the contingencies established in Phases 1 and 2. The sample stimuli were the same as used in Phases 1 and 2, and the comparison stimuli all were from different classes than the sample (Table 2), making ''none'' the correct response option on all trials. Feedback followed every trial. The mastery criterion was the same as in Phases 1 and 2.

Phase 4: Mixed training review. Previously learned trial types were intermingled within sessions (12 A-B and 12 A-C trials, plus catch trials for some groups). The number of catch trials per session varied across groups, and was defined as a proportion of total trials: 0 (zero trials) for the control, no-training, and 0% groups; .27 (nine trials) for the 25% group; and .50 (24 trials) for the 50% group. The ''none'' option was present on every trial for the 0%, 25%, and 50% groups, and was not present for the control and no-training groups. Feedback followed every trial. Mastery was defined as accuracy of at least 96% on each of two consecutive sessions.

Phase 5: Emergent relations testing. No feedback followed responses on any trial in this phase. Sessions always included 96 emergent relations tests: 24 for reflexivity, 24 for symmetry, and 48 involving the previously unpaired B and C members of each stimulus class. In the present training format, B-C and C-B relations are considered to be combined tests for symmetry and transitivity (R. Saunders & Green, 1992), because neither stimulus has previously served as a sample stimulus. We will refer to these hereafter as *combined tests.*

Intermingled among the novel trials were tests for the maintenance of previously trained relations. Among these were 12 A-B and 12 A-C tests, plus catch trials for some

Table 2

Trial configurations and number of trials of each type per session, by group, for each of the experimental phases. Stimulus locations on the subject's screen were counterbalanced within sessions. See text for details. (Co+ = class-consistent selection; Co− = class-inconsistent selection)

		Stimuli					Groups				
Phase	Description	Sample	Co+	Co−	Co−	Co−	Control	No training	0%	25%	50%
1	A-B training						24	24	24	24	24
		A1	B1	B2	B3		8	8	8	8	8
		A2	B2	B1	B3		8	8	8	8	8
		A3	B3	B1	B2		8	8	8	8	8
2	A-C training						24	24	24	24	24
		A1	C1	C2	C3		8	8	8	8	8
		A2	C2	C1	C3		8	8	8	8	8
		A3	C3	C1	C2		8	8	8	8	8
3	Catch-trial training						0	0	24	24	24
		A1	N	B2	C2	C3	0	0	8	8	8
		A2	N	B1	B3	C3	0	0	8	8	8
		A3	N	B2	C1	C2	0	0	8	8	8
4	Mixed training review						24	24	24	33	48
		A1	B1	B2	B3	N[a]	4	4	4	4	4
		A2	B2	B1	B3	N[a]	4	4	4	4	4
		A3	B3	B1	B2	N[a]	4	4	4	4	4
		A1	C1	C2	C3	N[a]	4	4	4	4	4
		A2	C2	C1	C3	N[a]	4	4	4	4	4
		A3	C3	C1	C2	N[a]	4	4	4	4	4
		A1	N	B2	C2	C3	0	0	0	3	8
		A2	N	B1	B3	C3	0	0	0	3	8
		A3	N	B2	C1	C2	0	0	0	3	8
5	Equivalence tests						120	120	120	126	144
	Trained relations						24	24	24	24	24
		A1	B1	B2	B3	N[b]	4	4	4	4	4
		A2	B2	B1	B3	N[b]	4	4	4	4	4
		A3	B3	B1	B2	N[b]	4	4	4	4	4
		A1	C1	C2	C3	N[b]	4	4	4	4	4
		A2	C2	C1	C3	N[b]	4	4	4	4	4
		A3	C3	C1	C2	N[b]	4	4	4	4	4
	Catch trials						0	0	0	6	24
		A1	N	B2	C2	C3	0	0	0	2	8
		A2	N	B1	B3	C3	0	0	0	2	8
		A3	N	B2	C1	C2	0	0	0	2	8
	Reflexive tests						24	24	24	24	24
		A1	A1	A2	A3	N[b]	8	8	8	8	8
		A2	A2	A1	A3	N[b]	8	8	8	8	8
		A3	A3	A1	A2	N[b]	8	8	8	8	8
	Symmetrical tests						24	24	24	24	24
		B1	A1	A2	A3	N[b]	4	4	4	4	4
		B2	A2	A1	A3	N[b]	4	4	4	4	4
		B3	A3	A1	A2	N[b]	4	4	4	4	4
		C1	A1	A2	A3	N[b]	4	4	4	4	4
		C2	A2	A1	A3	N[b]	4	4	4	4	4
		C3	A3	A1	A2	N[b]	4	4	4	4	4
	Combined tests						48	48	48	48	48
		B1	C1	C2	C3	N[b]	8	8	8	8	8
		B2	C2	C1	C3	N[b]	8	8	8	8	8
		B3	C3	C1	C2	N[b]	8	8	8	8	8
		C1	B1	B2	B3	N[b]	8	8	8	8	8
		C2	B2	B1	B3	N[b]	8	8	8	8	8
		C3	B3	B1	B2	N[b]	8	8	8	8	8
6	Equivalence posttest										
	Same as Phase 5 for the control group.										

[a] For the control, no-training, and 0% groups, a blank comparison box substituted for the none option during the mixed training review.

[b] For the control group, a blank comparison box substituted for the none option during equivalence testing.

groups. The number of catch trials per session was defined as a proportion of the total number of trials involving previously trained relations: 0 (zero trials) for the control, no-training, and 0% groups; .20 (six trials) for the 25% group; and .50 (24 trials) for the 50% group.

If performance on any of the prerequisite relations dropped below 90% in a session, the mixed training review (Phase 4) was reinstated until the mastery criterion was met, and then testing resumed. As long as accuracy on each of the prerequisite relations remained at 90% or higher, emergent relations testing continued until one of two criteria was met. A subject who scored 90% or higher on all three emergent relations tests on each of three consecutive sessions was considered to have demonstrated stimulus equivalence. A subject who exhibited stable performance (based on visual inspection) on reflexivity, symmetry, and combined tests across three consecutive sessions, but scored consistently less than 90% on any of the three types of tests, was considered not to have demonstrated stimulus equivalence.

Phase 6: Control testing for failed subjects. Subjects in the "none" groups who failed to demonstrate stimulus equivalence completed an additional emergent relations test phase in which conditions were identical to those for the control group in Phase 5 (i.e., no "none" option available, and no catch trials). One control-group subject (171) who failed Phase 5 equivalence testing was not retested. At the start of this phase subjects were told not to select blank boxes. Completion criteria were the same as for Phase 5.

RESULTS

Data analyses evaluated (a) performance on prerequisite relations trained during the initial phases of the study; (b) maintenance of these prerequisite relations during equivalence testing; (c) performance on the initial emergent relations tests; and, if applicable, (d) performance on the control posttest. Because the mastery criterion for emergent relations during Phases 5 and 6 specified performance over three consecutive sessions, "terminal" data refers to the final three sessions per phase. Group differences were corroborated statistically using analyses of variance followed by post hoc comparisons between the control and other groups using the Dunnett test (one tailed, $p < .05$).

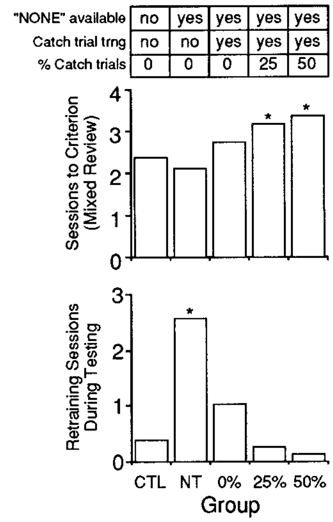

Fig. 2. Group differences in performance on the trained relations. Top panel: sessions required to meet the mastery criterion in the mixed training review (Phase 4). Bottom panel: number of retraining sessions required, during equivalence testing, after accuracy on trained relations dropped below 90%. Asterisks indicate significant differences from control-group performance.

Training Results

All subjects mastered the A-B and A-C relations prerequisite to equivalence class formation, and most did so readily. One-factor ANOVAs applied to the data in Appendix A found no statistically reliable group differences in the number of sessions needed by subjects to meet the mastery criteria during the A-B and A-C training of Phases 1 and 2 (combined) or the catch-trial training of Phase 3 (0%, 25%, and 50% groups). There was a significant group effect for sessions needed to complete the mixed training review of Phase 4, $F(4) = 4.14$, $p = .0029$. Post hoc comparisons confirmed that the 25% and 50% groups (for whom catch trials and the "none" response option were included in the mixed training review) took longer, on average, to complete this phase than did the control group (Figure 2, top panel).

Maintenance of Trained Relations During Testing

As they were taught to do during training, subjects exposed to catch trials during testing

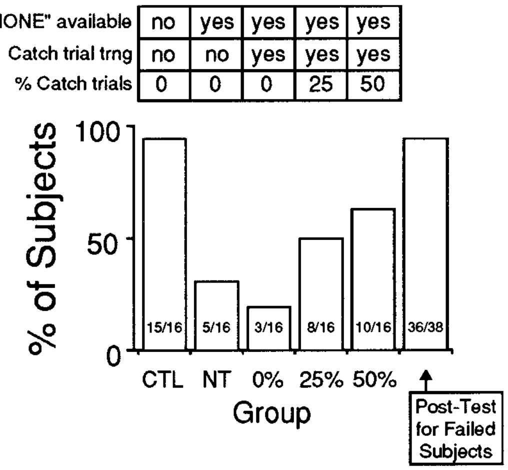

Fig. 3. Percentage of subjects demonstrating stimulus equivalence in the initial emergent relations tests of Phase 5. The rightmost bar shows the percentage of failed subjects (all experimental groups combined) who demonstrated stimulus equivalence in the control post-test of Phase 6.

Table 3

Patterns of nonequivalence outcomes (P = pass, F = fail) in the equivalence tests of Phase 5, and the number of subjects in each group showing these patterns. See text and Appendix B for details.

Group	Relation	Outcome				
	Reflexive	F	F	P	F	P
	Symmetrical	P	P	P	F	P
	Combined	P	F	F	F	P
Control		1	0	0	0	15
No training		7	4	0	0	5
0%		4	7	1	1	3
25%		3	5	0	0	8
50%		2	3	1	0	10
Total		17	19	2	1	41

(25% and 50% groups) reliably made the "none" response when it was appropriate to do so (mean correct on catch trials = 99%; Appendix B). Although the prerequisite relations (A-B and A-C) typically were maintained at high levels of accuracy during testing, 15 subjects required retraining when performance on these relations dropped below 90% during unreinforced test sessions (Appendix B). The bottom panel of Figure 2 shows that, based on a one-factor ANOVA, there was a statistically reliable group effect for number of retraining sessions necessary to complete the testing phase, $F(4) = 2.57$, $p = .045$. Subjects in the no-training group tended to require more retraining sessions than subjects in the control group (Figure 2, bottom panel). During the terminal sessions of testing, which provided the main data, performance on the trained relations was near 100% for all subjects (Appendix B).

Emergent Relations Test Performance

Figure 3 shows the percentage of subjects in each group who exhibited stimulus equivalence, that is, scored above 90% accuracy on reflexivity, symmetry, and combined tests for three consecutive sessions. In the control group, which was trained and tested conventionally, 94% of subjects (15 of 16) demonstrated the emergent relations indicative of stimulus equivalence. In the remaining groups, all of which had some exposure to the "none" response option, 41% of subjects (26 of 64) demonstrated equivalence. Few subjects showed equivalence in the no-training group, and among the remaining groups, the number of subjects exhibiting equivalence was positively related to amount of experience with catch trials (i.e., 50% > 25% > 0%). Appendix B shows the corresponding terminal data for individual subjects during the initial equivalence tests of Phase 5. Figure 3 also shows that 95% (36 of 38) of the subjects who failed the initial equivalence test of Phase 5 showed equivalence when retested in Phase 6 under conditions equivalent to those of the control group (no catch trials or "none" option).

Overall in Phase 5, 99% of subjects (79 of 80) passed the symmetry tests, 75% (58 of 80) passed the combined symmetry-transitivity tests, and 54% (43 of 80) passed the reflexivity tests. Table 3 shows that the most common outcomes among subjects who did not demonstrate equivalence were to fail reflexivity tests while passing the other two tests and to fail reflexivity and combined tests while passing the symmetry tests.

Figure 4 shows the percentage of subjects in each group who demonstrated each of the emergent relations in the initial equivalence test of Phase 5. A 5 (group) × 3 (type of emergent relation) ANOVA, conducted on these outcomes using percentage of correct (i.e., class consistent) responses as the depen-

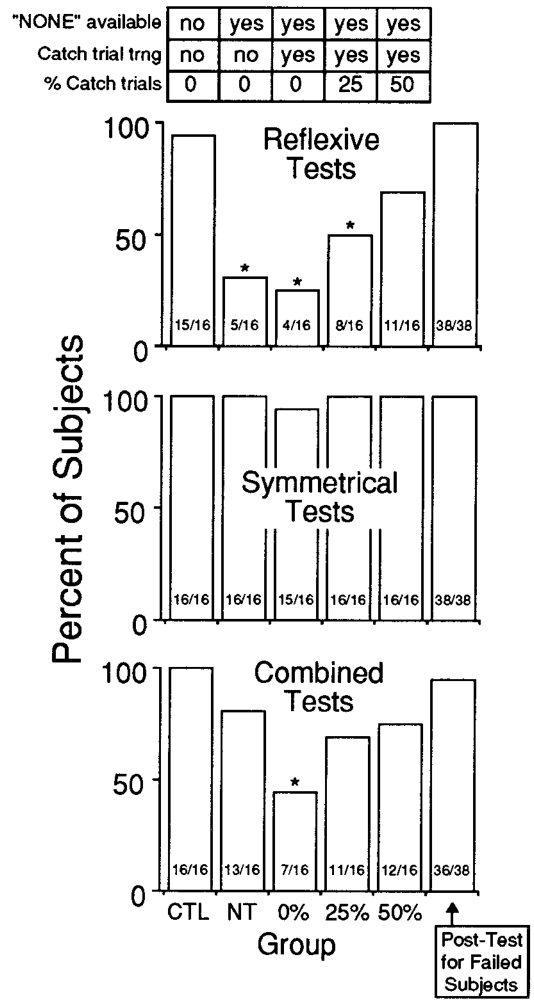

Fig. 4. Percentage of subjects demonstrating reflexive, symmetrical, and combined relations. Asterisks indicate significant differences from control-group performance, according to paired comparisons of match-to-sample accuracy. See text for details. Performances of failed subjects in the control posttest of Phase 6 (rightmost bar in each panel) were not included in the statistical analyses.

dent measure,[1] largely corroborated the visually evident patterns in Figure 4. Following a Groups × Relation Type interaction, $F(4, 2) = 3.66$, $p = .0006$, one-factor ANOVAs showed statistically reliable group differences for the reflexivity relation, $F(4) = 6.13$, $p = .0003$, and the combined relation, $F(4) = 3.89$, $p = .0067$, but not the symmetry relation. Post hoc comparisons following the significant ANOVAs confirmed that the no-training, 0%, and 25% groups performed especially poorly on reflexivity tests, and that the 0% group performed especially poorly on combined tests.

Relation Between Accuracy and "None" Selections

During equivalence test sessions in Phase 5, when the "none" option was available, subjects rarely made incorrect responses by selecting a comparison stimulus from the wrong class. Instead, they tended to select the "none" option even though a class-consistent selection was available. The top left panels of Figure 5 show the relation between accuracy and percentage of "none" selections in pooled group data from equivalence test sessions (catch trials excluded). For emergent relations, the figure shows three points per subject (reflexivity, symmetry, and combined relations) for each test session. For previously trained prerequisite relations, the figure shows two points per subject (A-B and A-C relations) for each test session. At the group level, there was a strong negative correlation ($r = -.995$ for all relations combined) between percentage correct and percentage of "none" responses. Because many subjects passed or failed the equivalence tests in close to the minimum number of sessions (Appendix B), and therefore showed little variability in performance, analogous individual analyses were possible only in selected cases. The individual-subject scatter plots in Figure 5 represent (a) 6 subjects who showed gradual emergence of an untrained relation across

[1] This analysis serves in lieu of statistically corroborating categorical outcomes (here, each subject = pass or fail) shown in Figures 3 and 4. The most commonly used tool for evaluating categorical outcomes, the chi square test, is logically inconsistent with the present experimental design because it compares actual group outcomes with an expected frequency of outcomes based on chance (here, eight pass and eight fail). In the present study the comparisons of primary interest are between the control group and other groups. An alternative approach, of defining expected frequencies in terms of control-group outcomes, is also invalid because chi square is regarded as inappropriate for expected frequencies of less than five (e.g., Camilli & Hopkins, 1978). Because percentage correct is a continuous measure, its use in the analyses may be seen as inconsistent with the definition of equivalence as a categorical outcome (e.g., Sidman, 1994). But it is important to note that although operational definitions of success or failure on equivalence tests may be categorical, the possible outcomes in individual performances are not.

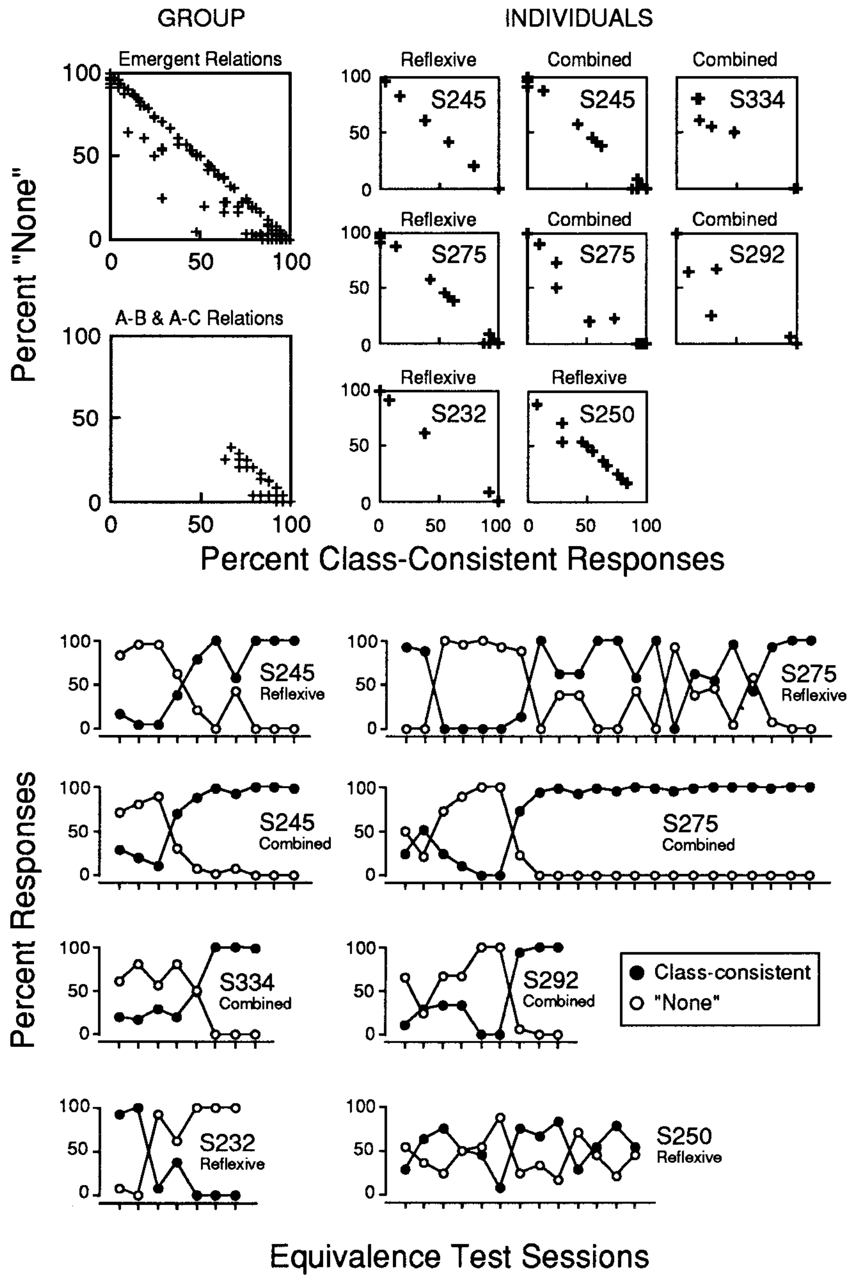

Fig. 5. Top panels: relation between accuracy and selection of the "none" option during equivalence test sessions (catch trials excluded). Top left two panels show pooled data for all subjects. Top right panels show data for selected individuals. Bottom panels: session-by-session plots of data shown in the individual-subject scatter plots.

eight or more test sessions, producing a range of accuracy values; (b) 1 subject (S232) who produced essentially the opposite pattern (after early signs of mastering the reflexivity relation, shifting, over several sessions, to exclusive use of the "none" option); and (c) 1 subject (S250) who showed unsystematic variability across 13 test sessions before being judged as failing the reflexivity tests. All eight individual scatter plots are similar to the group pattern ($r = -.92$ to -1.00). The lower panels of Figure 5 show the session-by-session covariation on which these scatter plots are based.

DISCUSSION

Across groups, a ''none'' response option controlled much behavior during equivalence tests, and with this default-response option present, almost all class-inconsistent responses were ''none'' responses. Both within-subject and between-groups contrasts indicated that the ''none'' response option tended to prevent subjects from demonstrating stimulus equivalence. In illustrating constraints on class formation, these results are broadly consistent with those of Fields et al. (1997), who included a default-response option during tests for expansion of previously demonstrated equivalence classes based on primary generalization. In that study, a default-response option reduced the likelihood that a stimulus, novel but physically similar to an equivalence class member, would be treated as consistent with previously formed classes. To our knowledge, however, the present study is the first to examine the effects of a default-response option on initial equivalence class formation and is the first to compare the effects of different default-response procedures with control conditions in which subjects were tested traditionally.

That a default-response option can disrupt equivalence class formation should come as no surprise based on precedents from other research areas. For example, discrete-choice response patterns, similar to those in conditional discrimination procedures, can be altered by including a ''don't know'' option in surveys (Sudman et al., 1996), an ''other'' option in human drug-discrimination procedures (Bickel et al., 1993), and an ''uncertain'' option in psychophysical judgments (Shields et al., 1997; Smith et al., 1995; Thomson, 1920). Disruption of equivalence also might be predicted from Sidman's observation that, for emergent relations to appear,

> the subject must have learned, first, that each trial . . . has a correct choice . . . ; and second, that each trial has only one correct choice. Without such a history, there is no reason why the subject's choices in a test without reinforcement should show any consistency at all. (Sidman, 1992, p. 24)

> A well designed test will arrange test trials so that the experimentally established equivalence class provides the only basis for classification that remains possible—that *works* on every trial. (Sidman, 1994, p. 512)

Poorly designed equivalence tests allow alternative factors, such as control by negative stimuli, to interfere with emergence of untrained relations (Carrigan & Sidman, 1992). Default responses, like the ''none'' option of the present study, provide another type of alternative to class-consistent responses—one that, from a subject's perspective, might be appropriate on any trial involving untrained relations. It is interesting, therefore, that many subjects in the present study demonstrated equivalence despite the presence of a ''none'' alternative, and most subjects, regardless of whether they passed or failed the equivalence tests, performed consistently throughout testing (note the systematic control of class-inconsistent responses by the ''none'' option in Figure 3, and, in Appendix B, the typically low number of sessions to criterion during Phases 5 and 6). Whether these findings run counter to Sidman's (1992, 1994) predictions depends on whether default responses are regarded as a generic alternative (i.e., part of a generalized response class) or as a response under more specific stimulus control, as we discuss below.

Stimulus Control of Default Responses

Several group differences in the present study suggest that the effects of a default-response option during equivalence testing depend on a history of making, and producing reinforcement through, default responses. In particular, recall that subjects in three groups (0%, 25%, and 50%) completed preliminary training (Phase 3) during which only catch trials were presented and the use of the ''none'' option was reinforced. Thereafter, these groups differed in terms of the percentage of trained relations in each session that were comprised of catch trials (during the mixed training review of Phase 4, and during equivalence tests of Phase 5). Three effects of this differential experience were evident. First, the 25% and 50% groups took especially long, on average, to complete the mixed training review of Phase 4. In this sense, intermingling catch trials with other trained relations tended to slow training progress. Second, only in a group without catch-trial history (no-training group) did subjects exposed to the ''none'' option have difficulty maintaining the trained relations during equivalence testing. This suggests that, unless its function is well defined prior to the introduction of unreinforced probe trials, a default-response option can disrupt even directly established stimulus control.[2] Third, the ''none'' groups of the present study differed in terms of their overall success on the emergent relations tests. Among groups that completed catch-trial training, there was a positive correlation between amount of subsequent experience with catch trials and the likelihood of demonstrating equivalence.

Only further investigation can determine whether the effects just described resulted from mere exposure to catch trials, as provided during the equivalence tests; from feedback about catch trials, as provided during the mixed training review; or from a combination of the two factors. The performance of subjects in the no-training group, however, suggests that the answer is likely to be complex. The no-training group completed the mixed training review (Phase 4) and equivalence tests (Phase 5) without catch trials (like the 0% group), and encountered the default-response option for the first time during equivalence testing (unlike any of the other ''none'' groups). As in the 0% group, members of this group were especially likely to fail reflexivity tests, but unlike the 0% group, tended to pass tests for the combined symmetry-transitivity relations (B-C). Catch-trial training (Phases 3 and 4), therefore, did not necessarily promote emergent relations, and it seems likely that the relatively good outcomes of the 25% and 50% groups are attributable to an interaction of catch-trial training and experience with catch trials in the testing context.

[2] Experience with catch trials may not always prevent disruption of performance by a default-response option. The design of the present study originally included a group in which 75% of trained relations (during the mixed review and equivalence testing phases) were catch trials, but the group was discontinued after performance on trained relations was above 90% on only 18 of 49 total equivalence test sessions completed by 7 pilot subjects.

Methodological Issues Relevant to the Present Study

No independent assessment was conducted to determine whether the subjects assigned to the various groups had similar capabilities, but two group outcomes from the study suggest that they did. First, there were no systematic group differences in the number of sessions needed to complete training on the A-B and A-C relations that formed the basis for equivalence. Second, many subjects in the ''none'' groups failed initial equivalence tests (Phase 5), but nearly all demonstrated equivalence when retested under conditions like those used for the control group (i.e., without catch trials or the ''none'' response option). It seems likely, therefore, that group differences resulted from the experimental manipulations and not from quirks of subject assignment.

Although efforts were made to establish the operating characteristics of default responses during a preliminary training phase, the verbal label (''none'') used to designate the default-response option complicates matters by potentially bringing to bear complex, and undefined, preexperimental histories (R. Saunders, 1996). This shortcoming applies as well to default-response options used in two previous equivalence studies (''?'' in Roche & Barnes, 1996; ''neither'' in Fields et al., 1997). One useful research strategy would be to replicate these studies using an arbitrarily designated default-response option whose function is established entirely in the context of the experiment, as has been approximated in psychophysical studies with humans (Shields et al., 1997) and accomplished with nonverbal subjects such as bottlenose dolphins and rhesus monkeys (Shields et al., 1997; Smith et al., 1995). In such cases, results should reflect an interaction between the training history used to establish default responses initially and the charateristics of new situations in which they can occur subsequently (e.g., R. Saunders, 1996).

The pervasive disruption of reflexivity in the present study bears closer inspection than the present experimental design permits. Unfortunately, the equivalence tests of the present study included tests only for the A-A reflexivity relation. In taking this step to promote economy in testing, we endorsed a common belief that reflexivity relations are especially robust for normally functioning adults (e.g., some equivalence studies have omitted reflexivity tests altogether). But the omission of B-B and C-C tests creates ambiguity regarding three interesting possibilities. The first is that, because testing experience alone can sometimes help to promote equivalence (e.g., Harrison & Green, 1990), reflexivity might have emerged—despite the default-response option—had tests also been included for all reflexive relations. The second possibility is that the catch trials of the present study, by always employing the A stimuli as samples (see Table 2), created A-''none'' relations that could be expected to interfere *selectively* with the A-A reflexivity tests. If so, then subjects with catch-trial experience might often fail A-A tests, as in the present study, but pass B-B and C-C tests.

A third possibility is that, contrary to our own tacit assumptions at the start of the study, reflexivity may be especially fragile in some circumstances. Reflexive relations also were found to be unusual in a recent study of self-reports about emergent relations, in which discrepancies between reports and actual performance were most likely to occur for reflexive relations (Lane & Critchfield, 1996). These two studies share a reliance on training protocols in which each stimulus serves solely as a sample or a comparison (one-to-many, or sample-as-node; see Fields, Verhave, & Fath, 1984; K. Saunders, Saunders, Williams, & Spradlin, 1993; R. Saunders & Green, 1992). By contrast, in some other types of training protocols, stimuli may serve as a sample in one prerequisite relation and as a comparison in another. A default-response option might have different effects given different training protocols.

Conclusions and Possible Extensions

Discussions about default responses, dating back to the earliest days of psychophysics (e.g., Jastrow, 1888), frequently have echoed two themes. First, default-response options, when available, tend to be used by subjects. Second, the precise factors controlling their use are a matter of debate. Both themes are illustrated in the present study.

That default-response options control behavior means that they will compete with the control of behavior by other variables. Thus, default-response procedures may be useful in evaluating the robustness of emergent relations. At a general level, questions of robustness bear on the plausibility of claims that stimulus equivalence is a component of complex, naturally occurring human capabilities such as language (Hayes, Hayes, Sato, & Ono, 1994) and category formation (e.g., Fields, Reeve, Adams, & Verhave, 1991). Everyday acts of communication and conceptualization emerge under relatively unconstrained conditions in which the range of possible responses is far greater than the two or three provided in conditional discrimination trials. Among the options available are to not respond at all or to do something unrelated to the task at hand (e.g., see Smith et al., 1995). By providing competing sources of control, default-response options appear to mimic an important characteristic of the natural environments and thus improve the face validity of equivalence investigations (e.g., see Fields et al., 1997).

At a more specific level, default-response procedures could be useful when research questions focus on the differential ''associative strength'' of stimulus relations (e.g., Fields, Adams, Verhave, & Newman, 1993; Fields & Verhave, 1987). Examples might include different types of emergent relations (e.g., Lane & Critchfield, 1996), relations reflecting different nodal distances (e.g., Fields, Adams, Verhave, & Newman, 1990), relations derived from different amounts or types of training with prerequisite relations (e.g., K. Saunders et al., 1993), and relations that have been challenged by factors like class-reversal training (e.g., Pilgrim & Galizio, 1995), drug manipulations, or fatigue.

Inevitably, interpreting the effects of default options on other performances will depend on understanding what controls them. Default responses sometimes have been described as representing broad, situation-nonspecific response classes—for example, a generic expression of uncertainty (Roche & Barnes, 1996; Smith et al., 1995), or a generic means of escaping aversive situations in which the probability of reinforcement is low or the probability of punishment is high (e.g., Critchfield, 1996; Durand, Guffey, & Planchon, 1983; Johanson et al., 1993; see also Smith et al., 1995). But R. Saunders (1996) has argued that it is risky to assume generic properties of default responses, in part because generic repertoires, if they exist, likely result from different histories in different subjects. Although generic properties of default responses cannot be ruled out in the present study, between-groups differences in modal outcomes suggest specific forms of stimulus control over default responses, most plausibly acquired within the experiment. It therefore seems appropriate to extend Saunders' position by proposing that the operating characteristics of default responses will depend on each study's conditional discrimination training protocol, stimulus class structure, contingencies, and means of arranging, labeling, and training default responses. What remains to be determined is whether this complexity compromises the measurement opportunities provided by default-response procedures or merely introduces a fascinating new set of stimulus relations into equivalence research.

REFERENCES

Bentall, R., Dickins, D., & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 46B,* 187–214.

Bickel, W. K., Oliveto, A. H., Kamien, J. B., Higgins, S. T., & Hughes, J. R. (1993). A novel-response procedure enhances the selectivity and sensitivity of triazolam discrimination in humans. *Journal of Pharmacology and Experimental Therapeutics, 264,* 360–367.

Camilli, G., & Hopkins, K. D. (1978). Applicability of chi-square to 2 × 2 contingency tables with small expected frequencies. *Psychological Bulletin, 85,* 163–167.

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

Critchfield, T. S. (1996). Differential latency and selective nondisclosure in verbal self-reports. *The Analysis of Verbal Behavior, 13,* 49–63.

Dougher, M. J. (1994). Stimulus equivalence, functional equivalence, and the transfer of function. In S. C. Hayes, L. J. Hayes, M. Sato, & K. Ono (Eds.), *Behavior analysis of language and cognition* (pp. 71–90). Reno, NV: Context Press.

Dube, W. V., & McIlvane, W. J. (1996). Implications of a stimulus control topography analysis for emergent behavior and stimulus classes. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 197–218). Amsterdam: Elsevier.

Durand, R. M., Guffey, H. J., & Planchon, J. M. (1983). An investigation of the random versus nonrandom nature of item omission. *Journal of Marketing Research, 20,* 305–315.

Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior, 53,* 345–358.

Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1993). Are stimuli in equivalence classes equally related to each other? *The Psychological Record, 43,* 85–105.

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

Fields, L., Reeve, K. F., Adams, B. J., & Verhave, T. (1991). Stimulus generalization and equivalence classes: A model for natural categories. *Journal of the Experimental Analysis of Behavior, 55,* 305–312.

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48,* 317–332.

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.

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

Hayes, S. C., Hayes, L. J., Sato, M., & Ono, K. (Eds.). (1994). *Behavior analysis of language and cognition.* Reno, NV: Context Press.

Jastrow, J. (1888). A critique of psycho-physic methods. *American Journal of Psychology, 1,* 271–309.

Johanson, G. A., Gips, C. J., & Rich, C. E. (1993). "If you can't say something nice." A variation on the social desirability response set. *Evaluation Quarterly, 17,* 116–122.

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

Pilgrim, C., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I. Adults. *Journal of the Experimental Analysis of Behavior, 63,* 225–238.

Roche, B., & Barnes, D. (1996). Arbitrarily applicable relational responding and sexual categorization: A critical test of the derived difference relation. *The Psychological Record, 46,* 451–475.

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

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

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

Shields, W. E., Smith, J. D., & Washburn, D. A. (1997). Uncertain responses by humans and rhesus monkeys (*Macaca mulatta*) in a psychophysical same-different task. *Journal of Experimental Psychology: General, 126,* 147–164.

Sidman, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 15–27). Reno, NV: Context Press.

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

Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor, R., & Erb, L. (1995). The uncertain response in the bottlenose dolphin (*Tursiops truncatus*). *Journal of Experimental Psychology: General, 124,* 391–408.

Spencer, T. J., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65,* 643–659.

Sudman, S., Bradburn, M., & Schwartz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Francisco: Jossey-Bass.

Thomson, G. H. (1920). A new point of view in the interpretation of threshold measurements in psychophysics. *Psychological Review, 27,* 300–307.

Wulfert, E., Dougher, M. J., & Greenway, D. E. (1991). Protocol analysis of the correspondence of verbal rules and equivalence class formation. *Journal of the Experimental Analysis of Behavior, 56,* 489–504.

Received September 18, 1996
Final acceptance April 1, 1998

APPENDIX A

Number of sessions required by each subject to achieve mastery (90% correct for two consecutive sessions) in each training phase.

Phase	Subjects															
Control group	166	168	169	171	172	173	174	175	176	177	180	181	183	184	185	263	*M*
1. A-B training	3	9	4	4	3	3	3	4	4	4	3	4	4	3	5	3	3.94
2. A-C training	3	3	3	2	3	3	2	2	2	2	2	3	2	5	4	4	2.81
3. None training																	
4. Mixed review	2	3	3	3	3	2	2	2	2	3	2	2	2	2	2	3	2.38
No-training group	279	280	281	282	284	285	288	289	292	293	300	301	310	325	334	341	*M*
1. A-B training	6	3	3	2	3	2	3	3	3	3	6	6	3	5	3	2	3.50
2. A-C training	3	3	2	3	4	2	3	3	3	2	3	5	3	3	3	3	3.00
3. None training																	
4. Mixed review	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	2	2.06
0% group	227	229	230	231	232	233	234	235	236	240	242	244	255	256	258	275	*M*
1. A-B training	3	3	3	3	3	4	6	3	4	4	4	4	3	3	3	7	3.75
2. A-C training	3	3	3	3	3	3	4	3	3	3	3	3	3	2	3	6	3.19
3. None training	3	2	3	2	3	2	3	2	2	2	2	2	3	2	2	3	2.38
4. Mixed review	2	2	2	2	3	4	4	2	3	2	4	2	2	2	4	5	2.81
25% group	200	202	203	206	208	211	213	245	246	247	248	250	257	259	277	278	*M*
1. A-B training	3	3	2	3	2	5	3	3	2	4	3	4	2	4	5	3	3.19
2. A-C training	3	2	3	3	3	3	3	3	2	3	5	4	3	3	2	3	3.00
3. None training	2	2	2	3	2	3	2	2	2	2	3	3	2	2	3	3	2.38
4. Mixed review	3	2	4	2	3	3	3	2	5	2	3	4	3	3	3	7	3.25
50% group	192	193	194	196	198	199	215	216	219	223	249	260	261	262	276	283	*M*
1. A-B training	4	4	3	3	11	4	5	3	3	6	4	3	3	3	3	6	4.25
2. A-C training	3	2	2	3	2	3	3	2	2	2	2	3	3	2	3	3	2.50
3. None training	2	2	3	4	4	3	3	2	2	2	2	2	3	3	3	2	2.63
D. Mixed review	3	2	2	4	6	6	3	3	2	3	2	6	3	3	3	3	3.38

APPENDIX B

Equivalence test performances in Phases 5 and 6. Sessions to demonstration of equivalence of failure; retraining necessary because of poor performance on trained relations during test sessions; and percentage correct on each relation type during the final three sessions per phase.

		Phase 5: Initial equivalence tests								Phase 6: Control equivalence tests for failed subjects						
		Sessions to criterion	MTS retrainings		Emergent relations					Sessions to criterion	MTS retrainings		Emergent relations			
Group	Subject		Number	Sessions	Trained	Catch	Reflexivity	Symmetry	Combined		Number	Sessions	Trained	Reflexivity	Symmetry	Combined
Control	166	3	0	0	99		100	100	99							
	168	3	0	0	99		97	99	99							
	169	3	0	0	96		97	96	99							
	171	5	0	0	97		1	99	99							
	172	3	0	0	100		100	100	99							
	173	3	0	0	100		100	100	99							
	174	3	0	0	100		100	100	99							
	175	3	0	0	100		100	100	97							
	176	3	0	0	100		100	100	100							
	177	3	0	0	100		100	100	99							
	180	3	0	0	99		100	97	98							
	181	4	0	0	96		100	99	99							
	183	4	0	0	99		99	100	100							
	184	3	0	0	100		97	96	98							
	185	3	0	0	100		100	100	99							
	263	7	2	6	93		100	99	97							
	M	3.50	0.13	0.38	99		92	99	98							
No training	279	3	0	0	100		0	100	100	3	0	0	100	100	100	100
	280	7	1	2	97		0	99	66	3	0	0	99	95	100	97
	281	3	0	0	100		99	100	97							
	282	5	0	0	97		100	99	100							
	284	11	7	15	95		0	97	0	3	0	0	100	100	100	98
	285	3	0	0	96		0	97	1	4	0	0	100	97	100	100
	288	5	0	0	100		66	100	99	3	0	0	100	100	100	99
	289	7	1	2	99		0	100	99	4	0	0	100	99	100	100
	292	9	3	6	100		0	100	100	4	0	0	100	100	100	100
	293	3	0	0	100		4	96	97	3	0	0	100	99	99	100
	300	4	0	0	97		100	99	98							
	301	4	1	2	100		100	100	99							
	310	4	0	0	99		100	100	100							
	325	3	0	0	99		0	100	100	3	0	0	99	100	100	100
	334	8	4	10	100		0	96	99	4	0	0	100	100	100	99
	341	3	2	4	97		1	95	0	6	1	2	100	100	100	1
	M	5.13	1.19	2.56	99		36	99	40							
0%	227	3	0	0	97		0	100	1	3	0	0	100	100	100	99
	229	3	0	0	95		0	15	2	3	0	0	100	100	100	99
	230	3	0	0	97		0	96	0	5	0	0	100	97	100	100
	231	7	0	0	100		0	100	100	3	0	0	99	99	100	100
	232	7	0	0	100		0	100	100	3	0	0	100	100	100	98
	233	3	0	0	96		0	97	0	5	0	0	97	100	100	96
	234	3	0	0	96		100	97	1	7	1	2	100	100	97	97
	235	3	0	0	100		0	100	0	3	0	0	100	100	100	97
	236	6	1	2	100		100	100	99							
	240	3	0	0	100		99	100	100							
	242	3	0	0	99		0	97	98	3	0	0	99	100	100	100
	244	6	1	2	97		0	97	0	3	0	0	97	99	97	95
	255	3	0	0	100		0	100	97	3	0	0	100	99	100	100
	256	3	0	0	97		0	100	0	3	0	0	100	100	100	100
	258	7	2	4	99		0	100	0	3	0	0	100	97	100	98
	275	22	4	8	97		97	100	99							
	M	5.31	0.63	1.31	98		25	94	44							

APPENDIX B

(*Continued*)

		Phase 5: Initial equivalence tests								Phase 6: Control equivalence tests for failed subjects						
		Sessions to criterion	MTS retrainings		Emergent relations					Sessions to criterion	MTS retrainings		Emergent relations			
Group	Subject		Number	Sessions	Trained	Catch	Reflexivity	Symmetry	Combined		Number	Sessions	Trained	Reflexivity	Symmetry	Combined
25%	200	5	0	0	97	94	99	99	98							
	202	4	0	0	96	100	0	99	1	3	0	0	100	100	99	99
	203	5	1	2	99	100	0	97	0	3	0	0	100	99	100	98
	206	6	0	0	99	100	100	100	93							
	208	3	0	0	100	94	1	99	100	3	0	0	100	97	100	99
	211	3	0	0	100	100	100	97	99							
	213	3	0	0	99	100	0	100	0	3	0	0	97	99	99	99
	245	10	0	0	100	100	100	100	99							
	246	3	0	0	99	94	1	99	98	3	0	0	99	100	100	99
	247	3	0	0	100	100	99	99	97							
	248	5	1	2	97	100	1	99	0	8	0	0	100	100	100	100
	250	10	0	0	100	100	64	99	100	3	0	0	97	100	100	100
	257	3	0	0	100	100	100	100	99							
	259	3	0	0	99	100	100	100	99							
	277	3	0	0	100	100	0	99	1	7	0	0	99	100	99	99
	278	7	0	0	99	100	100	99	98							
	M	3.75	0.13	0.25	99	99	54	99	68							
50%	192	3	0	0	99	100	0	100	0	5	0	0	100	100	100	98
	193	3	0	0	99	97	100	100	95							
	194	3	0	0	99	100	100	99	98							
	196	3	0	0	99	100	100	100	99							
	198	5	1	2	99	100	100	97	1	3	0	0	99	99	100	43
	199	3	0	0	100	100	99	98	99							
	215	4	0	0	99	100	99	100	97							
	216	5	0	0	97	100	99	99	98							
	219	3	0	0	97	97	100	100	95							
	223	4	0	0	100	100	100	100	99							
	249	5	0	0	100	100	0	100	99	3	0	0	99	100	100	97
	260	3	0	0	99	100	0	97	0	4	0	0	100	100	100	98
	261	4	0	0	100	100	99	96	98							
	262	4	0	0	97	100	0	100	100	3	0	0	99	100	100	100
	276	4	0	0	97	100	0	100	0	6	0	0	100	100	100	100
	283	4	0	0	97	100	99	100	99							
	M	3.75	0.06	0.13	99	100	68	99	74							