The Psychological Record, 2000, 50, 339-354

INSTRUCTIONS, STIMULUS EQUIVALENCE, AND STIMULUS SORTING: EFFECTS OF SEQUENTIAL TESTING ARRANGEMENTS AND A DEFAULT OPTION

PAUL M. SMEETS
Leiden University, The Netherlands

SIMON DYMOND
Anglia Polytechnic University, England

DERMOT BARNES-HOLMES
National University of Ireland, Maynooth

The present study was a modified replication of a paper-and-pencil format study by Eikeseth, Rosales-Ruiz, Duarte, and Baer (1997) on equivalence relations derived from instructionally induced conditional relations. The study consisted of three experiments, all with Dutch psychology students as subjects. After being instructed to memorize four printed examples of arbitrary A-B and B-C conditional discrimination tasks and completing A-B and B-C trials in the presence of these examples, the subjects received a series of probe trials (no access to the examples): baseline and symmetric transitivity (C-A) probes (Experiment 1), or baseline, symmetry (B-A, C-B), and symmetric transitivity probes followed by a sorting test (Experiments 2 and 3). Without the option to skip "impossible-to-solve" probe trials (Experiments 1 and 2), almost all subjects (99%) completed all training and all probe trials. Most subjects (87%) who responded correctly on the baseline training trials also responded correctly on the baseline probes. Most of these subjects responded correctly on the symmetry trials (87%), the symmetric transitivity probes (81%), and on the sorting test (76%). Symmetric transitivity was seen most often when tested after symmetry. The performances on the sorting test corresponded with the numbers of derived relations (symmetry *and* transitivity; symmetry *or* symmetric transitivity; no symmetry nor symmetric transitivity) rather than with equivalence per se. The introduction of the default option (Experiment 3) resulted in most subjects skipping and responding inaccurately on the symmetry and symmetric transitivity probes.

Stimulus equivalence research has shown that after being trained on multiple arbitrary match-to-sample tasks with common sets of stimuli

Correspondence may be sent to Paul M. Smeets, Behavior Analysis Unit, Department of Psychology, Leiden University, P.O. Box 9555, 2300 RB, Leiden, The Netherlands.

(e.g., A1-B1, A2-B2 and B1-C1, B2-C2), most verbal humans relate all directly and indirectly linked stimuli conditionally with one another without further training: A1-A1, A2-A2, B1-B1, B2-B2, C1-C1, C2-C2 (reflexivity); B1-A1, B2-A2, C1-B1, C2-B2 (symmetry), and C1-A1, C2-A2 (symmetric transitivity). Stimulus equivalence is assumed when all three properties are demonstrated (Barnes, 1994; Sidman, 1994; Sidman, Willson-Morris, & Kirk, 1986).

Although many stimulus equivalence studies deal with normal adults, the procedures for training the prerequisite (baseline) match-to-sample tasks are similar to those used with nonverbal organisms. The training trials are presented one at a time, instructions are kept to a minimum, and responses are followed by differential though nondescriptive feedback (e.g., "Correct"). This training continues until the subject responds without or almost without errors on a series of consecutive trials. These procedures are consistent with the operant tradition and are designed to study stimulus class formation as a function of programmed contingencies and to isolate stimulus equivalence from other forms of stimulus control. The procedures, however, are not consistent with the way normal adults learn conditional stimulus relations. Normally, such relations are established through verbal and printed instruction, frequently without feedback, and in settings where multiple tasks are presented simultaneously rather successively, thereby providing subjects the opportunity to check and correct written answers to previous problems.

Recently, Eikeseth, Rosales-Ruiz, Duarte, and Baer (1997) examined whether stimulus equivalence in adults can emerge also from printed instructions. The study was conducted in a regular classroom setting. Two groups of undergraduate psychology students participated, one of 58 students (training assembly condition) and one of 20 students (nontraining assembly condition). The training assembly received four pages with instructions and match-to-sample tasks with Roman letters as samples and as comparisons. Page 1 showed a description of an (identity) match-to-sample task and the requested response mode (marking the correct comparison). Pages 2 and 3 showed four examples of marked arbitrary match-to-sample tasks (A1-B1, A2-B2, B1-C1, B2-C2), and instructions to (a) memorize these examples, hereafter referred to as rules (Eikeseth et al., 1997) and (b) complete 40 unmarked (training) A-B and B-C matching trials with the rules present for visual inspection. Page 4 showed 16 baseline trials mixed with 16 symmetric transitivity (C-A) trials but without access to the rules. The nontraining assembly received only pages 1 and 4. Twenty-eight training assembly subjects (48%) responded correctly on 15/16 baseline probes. Of these 28 subjects, 14 (50%) responded correctly on 15/16 symmetric transitivity probes. Of the 20 nontraining assembly subjects, only 1 (5%) demonstrated the designated baseline relations (A1-B1, A2-B2, B1-C1, B2-C2) and none the designated symmetric transitivity relations (C1-A1, C2-A2).

Although this study clearly showed that instructions alone can be a cost-effective procedure for establishing derived conditional stimulus relations with some adults, some of the data, notably those of the training assembly, are puzzling. The poor performances on the baseline probes could be related to the limited number of practice trials. Perhaps, overtraining was required. Alternatively, given that no incentives were used, the students may have found it more expedient to complete the practice trials without memorizing the rules. The low outcome on the symmetric transitivity probes could be seen as a demonstration of delayed equivalence. Stimulus equivalence is not always immediately evident and frequently occurs only after repeated training and testing (Green, 1990; Leader, Barnes, & Smeets, 1996; Sidman et al., 1986; Spradlin, Cotter, & Baxley, 1973). Delayed equivalence, however, would not be expected, at least not to such a degree, because Roman letters were used and equivalence occurs more readily with familiar and easy-to-name stimuli rather than with abstract stimuli (Holth & Arntzen, 1998). Moreover, the correctly marked and interspersed A-B and B-C probes should have made it easy for these adults to derive the C-A relations. The low outcome of the nontraining assembly condition, however, should be expected because these subjects did not receive conditions (i.e., training on multiple arbitrary match-to-sample tasks) that would normally be required for generating generalized conditional discrimination responding (Saunders, Wachter, & Spradlin, 1988).

The present study was designed to replicate the Eikeseth et al. research (1997). The study assessed also whether equivalence relations derived from instructionally established conditional discriminations are different from those derived from contingency-shaped conditional discriminations (Hayes, 1989). The study consisted of three experiments, all with psychology students from Leiden University as subjects. Experiment 1 was a replication of the Eikeseth et al. study (1997) but with different stimuli. Experiment 2 examined (a) whether the emergence of the symmetric transitivity relations would be sensitive to the order in which they were tested: before, together with, or after the symmetry probes (Adams, Fields, & Verhave, 1993; Fields, Adams, Newman, & Verhave, 1992; MacDonald, Dixon, & LeBlanc, 1986; Smeets, Leader, & Barnes, 1997) and (b) the correspondence between the performances on equivalence measures (symmetry and symmetric transitivity) and on a sorting test (e.g., Green, 1990; Pilgrim & Galizio, 1996). Experiment 3 examined the degree to which the performances on the equivalence and sorting tests would be affected by the introduction of a default-response option (Duarte, Eikeseth, Rosales-Ruiz, & Baer, 1998; Innis, Lane, Miller, & Critchfield, 1998).

Experiment 1

This experiment was a modified replication of the Eikeseth et al. study (1997). Two conditions were used. Condition 1 was similar to training assembly (training A-B and B-C, testing C-A). Condition 2 was the same except the subjects were trained on A-B and D-C tasks. This condition permitted generalized conditional discrimination responding but not equivalence class formation.

Method

Setting, Subjects, and Experimenter

The experiment was conducted during a regular classroom session. The instructor (first author) served as experimenter. Twenty-three 3rd-year psychology students following an introductory course on behavior analysis served as subjects. The students were asked to participate in an exercise that might help them get a better idea of the types of tasks that are used in some behavior analytic research. No compensatory credits were given for their performance or participation. To secure anonymity, the students were instructed to mark all materials with a self-generated identification code (e.g., fictitious name or date). None of the students refused.

Materials and Procedures

Thirteen subjects received Condition 1 and 10 subjects Condition 2. All subjects received three pages with printed instructions and match-to-sample tasks. The materials were distributed such that each subject of Condition 1 was seated between two subjects of Condition 2. After finishing the third page, the students were permitted to leave the classroom and take a break. Time required for the experiment was 15 to 43 min.

Figure 1 shows the materials used for Condition 1. Pages 1 and 2 were presented simultaneously (page 2 was printed on the reverse side of page 1). Page 1 described and presented an example of an identity matching-to-sample task, and the response mode (i.e., encircling the matching comparison). Page 2 showed four marked arbitrary match-to-sample tasks (A1-B1, A2-B2, B1-C1, B2-C2). Six symbols served as experimental stimuli (A1: #, A2: %, B1: §, B2: £, C1: ≈, C2: @). The examples were followed by 40 practice (training) trials, 10 trials on each relation. The rules were present for visual inspection but with the instruction to memorize the rules while completing the practice trials. After handing in page 2, the subjects received page 3 with 32 probe trials (no access to the printed rules): 16 baseline trials mixed with 16 C-A trials.

The materials for Condition 2 were the same except that on pages 2 and 3, the B-C trials had been replaced by D-C trials (D1: *, D2: µ). In both conditions, the performance criteria were set at 38/40 (95%) training trials correct, 14/16 (88%) trials correct (positive criterion) or incorrect (negative criterion) on each type of probe (baseline, symmetric transitivity).

Results

All subjects in Conditions 1 and 2 demonstrated criterion performance on the baseline trials during training *and* testing (see Table 1). Ten subjects (77%) of Condition 1 matched the same class C and A stimuli with one another. Of the other 3 subjects, 1 mismatched (C1-A2, C2-A1), 1 skipped all C-A tasks, and 1 responded at chance level. In Condition 2, 5 subjects (45%) demonstrated performances congruous

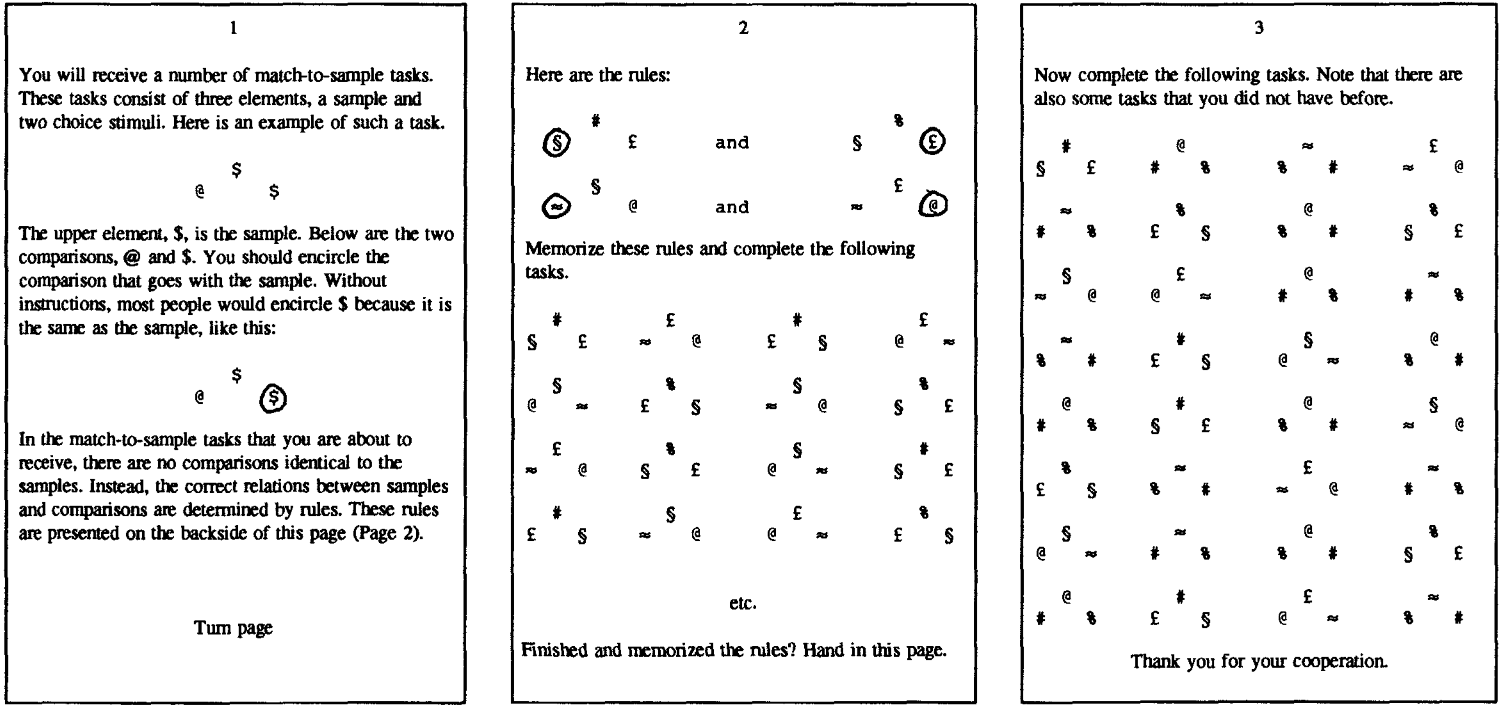

Figure 1. Pages 1 to 3 for Condition 1 in Experiment 1.

with and 4 subjects (36%) performances incongruous with the experimenter designated C-A relations. Of the other 2 subjects, 1 skipped all C-A tasks and 1 responded at chance level.

Table 1

Numbers and Percentages of Subjects Demonstrating Criterion Performance in Experiment 1

		Training		Testing		
Conditions	Number of Subjects	A-B, B-C	A-B, D-C	A-B, B-C	A-B, D-C	C-A
Condition 1	13	13/13 100%		13/13 100%		10/13 77%
Condition 2	11		11/11 100%		11/11 100%	5/11 45%

Discussion

The performances on the baseline probes (Conditions 1 and 2) and on the symmetric transitivity probes in Condition 1 were quite superior to those reported by Eikeseth et al. (1997). These discrepancies could be related to at least two variables. First, one might argue that the positive outcomes of the baseline and symmetric transitivity probes were biased because our criteria for mastery performance (correct responding on 14/16 trials) were less stringent than those (correct responding on 15/16 trials) used by Eikeseth et al. (1997). Implementation of the more stringent criterion, however, would have produced only one more “failure” in Condition 1, a subject who responded correctly on 14 baseline probe trials and on all 16 symmetric transitivity probe trials.

Second, our 3rd-year students may have been more advanced than those used by Eikeseth et al. (1997). Perhaps our students had become acquainted with issues related to stimulus equivalence (e.g., concept formation and reasoning) in other courses. Moreover, our students may have been particularly compliant because the instructor acted as experimenter. Although the students' selected codes prevented disclosure of identity, the instructor's presence may have encouraged the students to perform well.

Nine subjects (82%) in Condition 2 related the C stimuli conditionally to the A stimuli: C1-A1, C2-A2 (5 subjects) or C2-A1, C2-A1 (4 subjects). These performances can be seen as a case of generalized conditional discrimination responding and suggest that the C-A relations obtained in Condition 1 resulted largely from the trained baseline tasks rather than from extraneous sources of stimulus control.

Experiment 2

Previous research has shown that symmetric transitive relations derived from contingency-shaped conditional discriminations emerge more readily when these probes are presented after rather than before or together with symmetry probes (Adams et al., 1993; Fields et al., 1992; Green, 1990; Smeets et al., 1997). Experiment 2 examined whether this testing order effect is also evident with instructionally induced conditional discriminations.

The experiment examined also the correspondence between the performances on the symmetry and symmetric transitivity probes and on a sorting test. Sorting tests are commonly used in studies on categorization and concept formation (e.g., Ludvigson & Caul, 1964; Rosch & Mervis, 1977) but seldom in studies on stimulus equivalence (Green, 1990; Pilgrim & Galizio, 1996). The evidence available shows that, after demonstrating equivalence, most subjects also sort the stimuli in a class-consistent fashion. Would this form of generalized partitioning also be evident with instructionally based equivalence relations?

Finally, the implications of the students' educational history and the status of the experimenter were assessed. First-year psychology students served as subjects and 3rd-year students as experimenters. Would the probe performances of these 1st-year students be comparable with those reported by Eikeseth et al. (1997)?

Method

Subjects, Experimenters, and Setting

One hundred-and-twenty 1st-year (1st semester) students served as subjects and three 3rd-year (female) students as experimenters. The subjects were divided into three classes, one experimenter per class. The class sessions were conducted concurrently. The subjects received compensatory credits for their participation in partial fulfilment of a required practical in experimental psychology. The experimenters were not familiar with the concept of equivalence and were paid for their participation.

Materials and Conditions

Three conditions were used, each involving five pages. All three conditions used the same stimuli and started with pages 1 and 2 that were also used in Condition 1 of Experiment 1 (see Figure 1). Pages 3

Table 2

Training and Test Trials for Each Condition in Experiment 2

Conditions	Pages 1&2*	Page 3**	Page 4**	Page 5**
BT-S	Instructions 40 B trials	16 B trials 16 T trials	16 S trials	Sorting test
BS-T	Instructions 40 B trials	16 B trials 16 S trials	16 T trials	Sorting test
ST-B	Instructions 40 B trials	16 S trials 16 T trials	16 B trials	Sorting test

Note. * = training, ** = testing. B = baseline, S = symmetry, T = symmetric transitivity.

and 4 tested the trained and derived stimulus relations (baseline, symmetry, symmetric transitivity). These pages were similar to page 3 used in Condition 1 of Experiment 1 but with the instruction to complete all trials. The conditions differed with regard to the order in which the stimulus relations were tested and are referred to as BT-S, BS-T, and ST-B (see Table 2). The letters refer to specific probes, B to baseline, S to

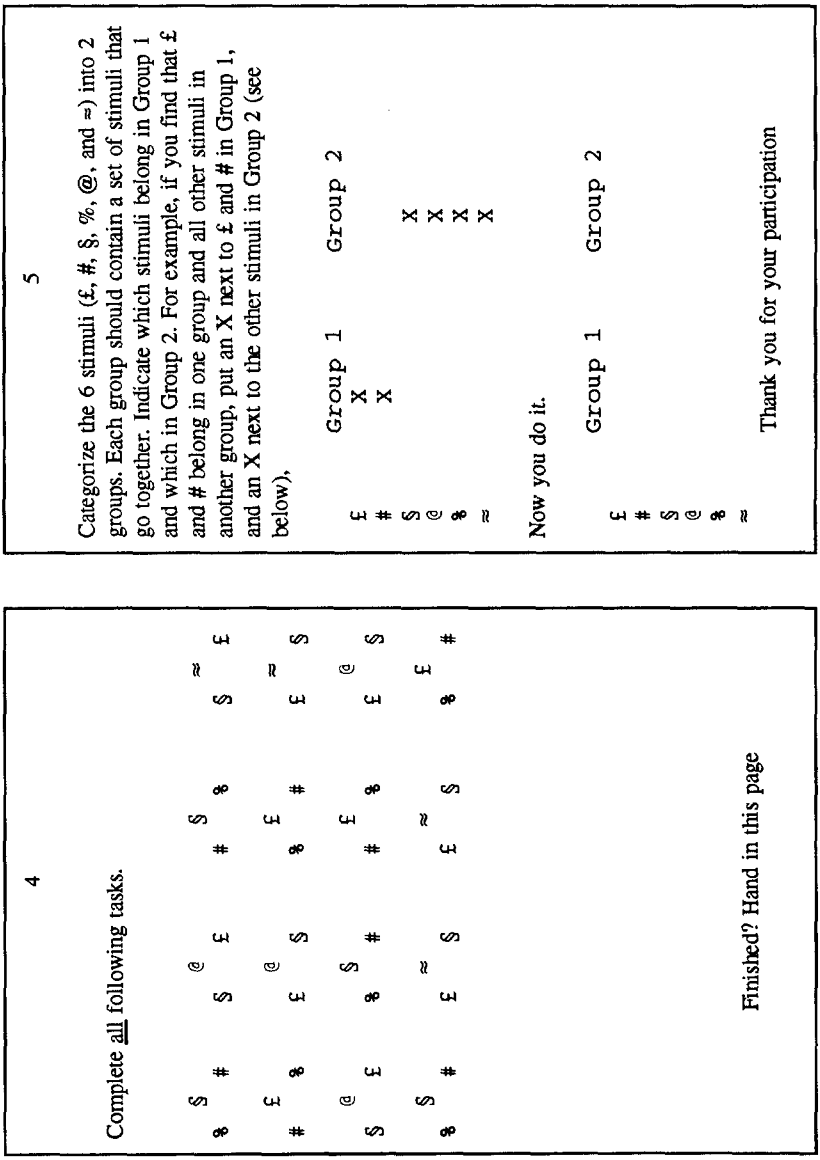

Figure 2. Pages 4 and 5 for condition BS-T in Experiment 2.

symmetry, and T to symmetric transitivity. The subjects received page 3 when handing in page 2, and page 4 when handing in page 3.

In *Condition BT-S* (40 subjects), the baseline probes (8 A-B and 8 B-C trials) and symmetric transitivity probes (16 C-A trials) were presented on page 3 (same as in Condition 1 of Experiment 1), and the symmetry probes (8 B-A and 8 C-B trials) on page 4. In *Condition BS-T* (36 subjects), the baseline and symmetry probes were presented on page 3, and the symmetric transitivity probes on page 4. *In Condition ST-B* (44 subjects), the symmetry probes were presented with the symmetric transitivity probes on page 3, and the baseline probes on page 4.

After handing in page 4, all subjects received page 5, the sorting test. The test was preceded by the instruction to categorize the stimuli into two groups and an example of how the stimuli could be classified. Figure 2 shows pages 4 (symmetry probe) and 5 (sorting test) that were used for the BT-S condition.

In all three conditions, criterion performance was set at 38/40 training trials correct, 14/16 trials correct on each stimulus relations probe (baseline, symmetry, symmetric transitivity), and sorting all stimuli correctly (e.g., Group 1: #, §, ≈; Group 2: %, £, @).

Results and Discussion

All subjects marked all training and probe trials. Table 3 shows the numbers and percentages of subjects in each condition who demonstrated criterion performance on the baseline, symmetry, and symmetric transitivity probes, and on the sorting task. Most subjects (88-95%) responded accurately on the baseline training trials. Most of these subjects (83-88%) responded accurately on the baseline probes, most of whom also responded correctly on the symmetry trials (79-93%) and on the symmetric transitivity trials (72-93%). These proportions are similar to those in Experiment 1. Thus, the much lower proportions reported in the study by Eikeseth et al. (1997) were probably not related to the students' educational history or the status of the experimenter.

Table 3

Numbers and Percentages of Subjects Demonstrating Criterion Performance in Experiment 2

	Conditions							
Tasks	BT-S	(N=40)	BS-T	(N=36)	ST-B	(N=44)	Overall	(N=120)
Baseline Training	35/40	88%	32/36	89%	42/44	95%	109/120	84%
Tests								
Baseline	29/35	83%	28/32	88%	35/42	83%	92/109	84%
Symmetry	23/29	79%	26/28	93%	31/35	88%	80/92	87%
Sym.Trans.	21/29	72%	26/28	93%	28/35	80%	75/92	82%
Sorting	23/29	79%	23/28	82%	24/35	69%	70/92	76%

The BT-S, BS-T, and ST-B conditions yielded small but systematic differences on the symmetry and symmetric transitivity probes (see Table 3). Criterion performance on these probes was seen least often in BT-S (symmetry: 79%; symmetric transitivity: 72%), slightly more often in ST-B (symmetry: 88%; symmetric transitivity: 80%), and most often in BS-T (symmetry: 93%; symmetric transitivity: 93%). The difference between the symmetric transitivity outcomes in BT-S (72%) and BS-T (93%) was statistically significant [$\chi^2(1) = 5.52$, $p < .02$]. This finding is consistent with previous research on the efficacy of simple-to-complex versus complex-to-simple protocols (Adams et al., 1993; Fields et al., 1992; Smeets et al., 1997).

Most subjects (69-82%) who responded correctly on the baseline probes also sorted the stimuli class consistently. The sorting performances were closely related to those on the derived relations tests. Class-consistent sorting was shown by 64/73 of the subjects (88%) who responded correctly on the symmetry *and* symmetric transitivity tests, by 5/9 subjects (56%) who responded correctly on only one test (symmetry *or* symmetric transitivity), and by only 1/10 subjects (10%) who failed on both these tests [$\chi^2(2) = 31.47$, $p < .01$]. These findings tentatively suggest that the sorting performances were related to the numbers of derived stimulus relations (symmetry *and* symmetric transitivity, symmetry *or* symmetric transitivity, no symmetry nor symmetric transitivity) rather than to equivalence per se.

Experiment 3

Experiments 1 and 2 demonstrated that most subjects who responded correctly on the baseline probes also responded to the symmetry, symmetric transitivity, and sorting probes in a class-consistent fashion. These findings could be related to the fact that the subjects felt obligated (Experiment 1) or were instructed (Experiment 2) to respond on all trials. These conditions mimic laboratory situations in which a following trial is presented only after responding to the previous trial has occurred. Everyday acts of categorization, however, typically occur in rather unconstrained conditions that permit no responding (Duarte et al., 1998; Innis et al., 1998). Experiment 3 replicated the BT-S and BS-T conditions of Experiment 2 except that the subjects were given the option to skip any test trial they found "impossible to solve."

Method

Seventy-seven new 1st-year students served as subjects and two 3rd-year students (same as in Experiment 2) as experimenters. Two conditions were used, BT-S (40 subjects) and BS-T (37 subjects). The recruiting procedure, setting, and materials were the same as in Experiment 2 except that on pages 3, 4, and 5 (baseline, symmetry, symmetric transitivity tests, and sorting tests), the instruction to mark all trials was replaced by the instruction, "You may skip problems that you find impossible to solve."

Results and Discussion

Table 4 shows the major results of Experiment 3. All subjects responded to all baseline training and test trials. Sixty-nine subjects responded correctly on the baseline training trials, 37 in BT-S (93%) and 32 (87%) in BS-T. Sixty-three of these subjects also responded correctly on the baseline probes, 33 (89%) in BT-S and 30 (94%) in BS-T.

Table 4

Numbers and Percentages of Subjects Demonstrating Criterion Performance in Experiment 3

	Conditions					
Tasks	BT-S	(N=40)	BS-T	(N=37)	Overall	(N=77)
Baseline						
Training	37/40	93%	32/37	87%	69/77	90%
Testing	33/37	89%	30/32	94%	63/69	91%
S and T Complete	7/33	21%	7/30	23%	14/63	22%
Symmetry	5/7	71%	6/7	86%	11/14	79%
Sym.Trans.	4/7	57%	6/7	86%	10/14	71%
Sorting	4/7	57%	6/7	86%	10/14	71%
S and T Incomplete	26/33	79%	23/30	77%	49/63	78%
Sorting	7/26	27%	9/23	39%	16/49	33%

Note. B = baseline, S = symmetry, T = symmetric transitivity.

Only 14 (22%) of these 63 subjects responded to all symmetry and all symmetric transitivity trials, 7 in each condition. BS-T and BT-S produced only minor differences on the symmetry and symmetric transitivity tests. Of the 7 BS-T subjects, 6 (86%) responded accurately on the symmetry tests and on the symmetric transitivity tests. Of the 7 BT-S subjects, 5 (71%) responded accurately on the symmetry test, and 4 (57%) on the symmetric transitivity test. Although these differences were not statistically significant, they showed the same trend as those obtained in Experiment 2.

Again, the sorting performances corresponded with the numbers of derived stimulus relations rather than with equivalence per se. Eight subjects evidenced symmetry and symmetric transitivity. Seven of these subjects (88%) sorted the stimuli correctly. Four other subjects responded accurately on only the symmetry test or only the symmetric transitivity test. Three of these subjects (75%) sorted correctly. Two subjects failed on both stimulus relations tests. Both these subjects also failed on the sorting test [$\chi^2(2) = 6.04$, $p < .05$].

The other 49 subjects, 26 in BT-S and 23 in BS-T, failed to complete the symmetry and symmetric transitivity trials. Forty of these subjects, 23 in BT-S and 17 in BS-T, skipped all symmetry and all symmetric transitivity trials. The other 9 subjects marked only a few symmetry and symmetric transitivity trials. Except for 1 subject, the accuracy of these responses was always low (below 80%). All 49 subjects, however, completed the sorting test, 16 of them correctly, 7 (27%) in BT-S and 9 (39%) in BS-T.

General Discussion

Present findings corroborate those reported by Eikeseth et al. (1997) and in other related studies (see below), some of which came to our attention only after the data collection had been completed. However, potentially important differences were evident.

The performances on the trained and untrained match-to-sample tasks were quite superior to those in similar paper-and-pencil format studies (Duarte et al., 1998; Eikeseth & Baer, 1997; Eikeseth et al., 1997). Most of our subjects who completed the training successfully also responded correctly on the *baseline probes*: 84% in Experiment 2, 91% in Experiment 3, and 100% in Experiment 1. These percentages are substantially higher than the 48% reported in the study by Eikeseth et al. (1997) and similar or even higher than in studies [72% in Eikeseth & Baer, (1997); 88% in Duarte et al., (1998)] in which nonarbitrary relations were trained.[1]

Except when given the option to skip "impossible-to-solve" trials (Experiment 3), almost all subjects (99%) completed the *derived stimulus relations probes*, most of them correctly: 87% symmetry and 84% symmetric transitivity (Experiments 1 and 2). In the training assembly condition of the Eikeseth et al. study (1997) and in the no-default-option (NCA) condition of the Duarte et al. study (1998), 75% to 86% of the subjects marked all derived stimulus relations trials, only 50 to 60% of them correctly.

These discrepancies are difficult to explain because, except for the stimuli, the procedures of the present study and those of Eikeseth and his colleagues were basically the same. Of course, the stimulus dimensions (abstract vs. familiar) could have affected the performances on the derived relations probes (Eikeseth et al., 1997; Holth & Arntzen, 1998), but that would not explain the different performances on the baseline probes. These discrepancies notwithstanding, present findings and those reported by Eikeseth clearly indicate that instructionally induced paper-and-pencil format conditional discrimination tasks can be a very efficient and cost-effective procedure for generating equivalence relations in adults.

The performances on the symmetric transitivity tests were sensitive to the order of presentation: before symmetry (BT-S) or after symmetry (BS-T). Experiments 2 and 3 comprised 36 BT-S and 35 BS-T subjects who completed all tasks and responded accurately on the baseline probes. Criterion performance on the symmetric transitivity probe was shown by 25 BT-S subjects (69%) and by 32 BS-T subjects (91%) [$\chi^2(1) = 5.52$, $p < .02$]. This finding indicates that, also when subjects use instructionally induced conditional discrimination tasks, symmetry is a precursor for successful performance on symmetric transitivity tests.

The introduction of the default option led to a drastic increase of unmarked symmetry and symmetric transitivity trials. These findings are consistent with those reported by Duarte et al. (1998) and Innis et al.

[1]The subjects were trained to relate conditionally letters and/or numerals with preestablished order relations: A-B, D-E and B-C, D-E (Eikeseth & Baer, 1997), or A-1, B-2 and 1-X, 2-Y (Duarte et al., 1998).

(1998). In both these studies, however, subjects who used the default ("none" or "cannot answer") option on some trials responded accurately on other trials. In the present study, subjects who skipped trials frequently responded *in*accurately on other trials. This discrepancy may be related to the way in which the default options were used. In both aforementioned studies, the default option was presented at each trial (i.e., each match-to-sample probe included two or more comparisons and the default option). Thus, these studies (a) may have implied (Duarte et al., 1998) or required (Innis et al., 1998) that the subjects respond to each trial, while (b) allowing them to respond away from the comparisons. The default option used in the present study may have conveyed that the untrained tasks were less important, thereby setting the occasion for skipping or sporadic and careless responding to untrained trials, more so with the many and unfamiliar format match-to-sample probes than with the one-trial and more traditional sorting test.

The interference by the default option does not come as a surprise. Stimulus equivalence requires that the subjects' responses are exclusively based on the experimentally induced sample-comparison relations. If not (e.g., the sample-comparison responses are controlled by preestablished relations between stimuli), equivalence relations do emerge in a delayed fashion or not at all. The default option, as it was used here and in several other studies (e.g., Duarte et al., 1998), permitted interference by undefined competing variables (i.e., the subjects' preexperimental history). In retrospect, this problem could have been avoided by training the function of the default option before the introduction of the probe trials (Innis et al., 1998). Most everyday situations, however, permit subjects not to respond. Thus, equivalence seems to be a very robust phenomenon but only in a very narrow band of highly controlled conditions.

For the subjects who completed all untrained match-to-sample trials, the sorting performances covaried with the numbers of derived stimulus relations. In Experiments 2 and 3 (see Table 5), class-consistent sorting was demonstrated by (a) 71/81 subjects (88%) who responded accurately on the symmetry *and* symmetric transitivity probes, 8/13 subjects (62%) who responded accurately on only the symmetry *or* symmetric transitivity probe, and by (c) 1/12 subjects (8%) who respond incorrectly on both these probes [$\chi_2(2) = 37.07$, $p < .01$].

Table 5

Numbers and Percentages of Subjects Who Responded Class-Consistently During Sorting Task

	Symmetry and Sym-Trans Probes Complete						Symmetry and Sym-Trans Probes Incomplete	
	Symmetry and Sym-Trans		Symmetry or Sym-Trans		No Symmetry and No Sym-Trans			
Exp 2	64/73	88%	5/9	56%	1/10	10%		
Exp 3	7/8	88%	3/4	75%	0/2	0%	16/49	33%
Total	71/81	88%	8/13	62%	1/12	8%	16/49	33%

These findings may indicate that, at least for the test that was used here, class-consistent sorting does not provide convergent validity for stimulus equivalence (Pilgrim & Galizio, 1996). Clearly, subjects who responded correctly on only the symmetry or symmetric transitivity tests did not meet the criteria for stimulus equivalence. Yet, many of these subjects (61%) sorted the stimuli in a class-contingent fashion. This proportion was much higher than that of the correctly sorting no-symmetry/no-symmetric-transitivity subjects (8%) [$\chi^2(1) = 7.66$, $p < .01$]. Of course, this finding would be easily understood if the correct symmetry responses by the symmetric-transitivity-only subjects, and the correct symmetric transitivity responses by the symmetry-only subjects, approximated criterion level, or were at least substantially higher than those by the no-symmetry/no-symmetric-transitivity subjects. This, however, was not the case. The numbers of correct symmetric transitivity responses by the symmetry-only subjects ($M = 9.2$, Range 5-12) and of correct symmetric responses by the symmetric-transitivity-only subjects ($M = 10.0$, Range 4-13) were very similar to those by the no-symmetry/no-symmetric-transitivity subjects (symmetry: $M = 10.3$, Range 4-13; symmetric transitivity: $M = 8.7$, Range 2-12). Of course, the outcome of the sorting test, particularly by the symmetry-only and symmetric-transitivity-only subjects might have been quite different without the nodal B stimuli.

Researchers concerned with stimulus equivalence have shown little interest for sorting tests relative to functional equivalence. Functional equivalence assessment typically starts with the training of a novel function to one member of each class (e.g., A1-CLAP, A2-WAVE). Following this training, the experimenter presents the untrained stimuli in a quasirandom fashion to assess stimulus control transfer (B1&C1-CLAP, B2&C2-WAVE). Sorting tests could be seen as a functional equivalence test without training but with an instruction to classify, and in which the order in which the stimuli are grouped is determined by the subject. For example, when given all stimuli, a subject might first sort all class-1 stimuli in one group and then all class-2 stimuli in another group. It has been shown that stimulus equivalence is frequently, but not always, associated with functional equivalence, and vice versa (Barnes & Keenan, 1993; de Rose, McIlvane, Dube, Galpin, & Stoddard, 1988; Dube, McDonald, & McIlvane, 1990; Dymond & Barnes, 1994; Sidman, Wynne, Maguire, & Barnes, 1989; Smeets, Barnes, & Roche, 1997). How does sorting compare with functional equivalence? To our knowledge, there have been no studies showing a relationship between functional equivalence and the numbers of derived conditional relations (symmetry and symmetric transitivity, symmetry or symmetric transitivity only, no symmetry nor symmetric transitivity). Analyzing this relationship, or the absence thereof (no functional equivalence without stimulus equivalence), may contribute to an understanding of the effects of different types of stimulus classification procedures on derived relational responding (Cullinan, Barnes, & Smeets, 1997).

References

ADAMS, B. J., FIELDS, L., & VERHAVE, T. (1993). Effects of test order on intersubject variability during equivalence class formation. *The Psychological Record*, 43, 133-152.

BARNES, D. (1994). Stimulus equivalence and the relational frame theory. *The Psychological Record*, 44, 91-124.

BARNES, D., & KEENAN, M. (1993). The transfer of functions through arbitrary and nonarbitrary stimulus relations. *Journal of the Experimental Analysis of Behavior*, 59, 61-81.

CULLINAN, V. A., BARNES, D., & SMEETS, P. M. (1997). A precursor to the relational evaluation procedure: Analyzing stimulus equivalence. *The Psychological Record*, 48, 121-145.

DE ROSE, J. C., MCILVANE, W. J., DUBE, W. V., GALPIN, V. C., & STODDARD, L. T. (1988). Emergent simple discrimination establsihed by indirect relation to differential consequences. *Journal of the Experimental Analysis of Behavior*, 50, 1-20.

DUARTE, A. M., EIKESETH, S., ROSALES-RUIZ, J., & BAER, D. M. (1998). The effects of a can't-answer response option and instructions on stimulus equivalence. *The Psychological Record*, 48, 631-646.

DUBE, W. V., MCDONALD, S. J., & MCILVANE, W. J. (1990). A note on the relationship between equivalence classes and functional stimulus classes. *Experimental Analysis of Human Behavior Bulletin*, 9, 7-11.

DYMOND, S., & BARNES, D. (1994). A transfer of self-discrimination response functions through equivalence relations. *Journal of the Experimental Analysis of Behavior*, 62, 251-267.

EIKESETH, S., & BAER, D. M. (1997). Use of a preexisting verbal relation to prevent the properties of stimulus equivalence from emerging in new relations. In D. M. Baer & E. M. Pinkston (Eds.), *Environment and behavior* (pp. 138-143). Boulder, CO: Westview Press.

EIKESETH, S., ROSALES-RUIZ, J., DUARTE, A., & BAER, D. M. (1997). The quick development of equivalence classes in a paper-and-pencil format through written instructions. *The Psychological Record*, 47, 275-284.

FIELDS, L., ADAMS, B. J., NEWMAN, S., & VERHAVE, T. (1992). Interactions among emergent relations during equivalence class formation. *The Quarterly Journal of Experimental Psychology*, 45B, 125-138.

GREEN, G. (1990). Differences in development of visual and auditory-visual equivalence relations. *American Journal on Mental Retardation*, 95, 260-270.

HAYES, S. C. (1989). *Rule-governed behavior, cognition, contingencies, and instructional control.* New York: Plenum.

HOLTH, P., & ARTNZEN, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record*, 48, 81-110.

INNIS, A., LANE, S. D., MILLER, E. R., & CRITCHFIELD, T. (1998). Stimulus equivalence: Effects of a default-response option on emergence of untrained stimulus relations. *Journal of the Experimental Analysis of Behavior*, 70, 87-102.

LEADER, G., BARNES, D., & SMEETS, P. M. (1996). Establishing equivalence relations using a respondent-type training procedure. *The Psychological Record*, 46, 685-706.

LUDVIGSON, H. W., & CAUL, W. F. (1964). Relative effect of overlearning on reversal and nonreversal shifts with two and four categories. *Journal of Experimental Psychology*, 68, 301-306.

MACDONALD, R. F., DIXON, L. S., & LEBLANC, J. M. (1986). Stimulus class formation following observational learning. *Analysis and Intervention in Developmental Disabilities*, 6, 73-87.
PILGRIM, C., & GALIZIO, M. (1996). Stimulus equivalence: A class of correlations or a correlation of classes? In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 173-218). Amsterdam: Elsevier.
ROSCH, E., & MERVIS, C. B. (1977). Children's sorting: A reinterpretation based on the nature of abtstraction in natural categories. In R. C. Smart & M. S. Smart (Eds.), *Reading in child development and relationships* (pp. 140-148). New York: McMillan.
SAUNDERS, R. R., WACHTER, J., & SPRADLIN, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior*, 49, 95-115.
SIDMAN, M. (1994). *Equivalence relations and behavior: A research story.* Boston: Authors Cooperative.
SIDMAN, M., WILLSON-MORRIS, M., & KIRK, B. (1986). Matching-to-sample procedures and the development of equivalence relations: The role of naming. *Analysis and Intervention in Developmental Disabilities*, 6, 1-19.
SIDMAN, M., WYNNE, C. K., MAGUIRE, R. W., & BARNES, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior*, 52, 261-274.
SMEETS, P. M., BARNES, D., & ROCHE, B. (1997). Functional equivalence in children: Derived stimulus-response and stimulus-stimulus relations. *Journal of Experimental Child Psychology*, 66, 1-17.
SMEETS, P. M., LEADER, G., & BARNES, D. (1997). Establishing stimulus classes in adults and children using a respondent-type training procedure: A follow-up study. *The Psychological Record*, 47, 285-308.
SPRADLIN, J. E., COTTER, V. W., & BAXLEY, N. (1973). Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. *American Journal of Mental Deficiency*, 77, 556-566.