JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1991, 56, 489-504 NUMBER 3 (NOVEMBER)

PROTOCOL ANALYSIS OF THE CORRESPONDENCE OF VERBAL BEHAVIOR AND EQUIVALENCE CLASS FORMATION

EDELGARD WULFERT, MICHAEL J. DOUGHER, AND DAVID E. GREENWAY

STATE UNIVERSITY OF NEW YORK AT ALBANY AND
UNIVERSITY OF NEW MEXICO, ALBUQUERQUE

In two equivalence experiments, a "think aloud" procedure modeled after Ericsson and Simon's (1980) protocol analysis was implemented to examine subjects' covert verbal responses during matching to sample. The purpose was to identify variables that might explain individual differences in equivalence class formation. The results from Experiment 1 suggested that subjects who formed equivalence classes described the relations among stimuli, whereas those not showing equivalence described sample and comparison stimuli as unitary compounds. Because Experiment 1 only demonstrated a correlation between describing stimulus compounds and the absence of equivalence classes, a second study was conducted. In Experiment 2, equivalence class formation was brought under experimental control through pretraining manipulations that facilitated responding either to stimulus compounds or to relations among stimuli. The results demonstrated that a history of describing stimulus compounds, when compared with describing the relations among the stimuli, interfered with the emergence of stimulus equivalence. These findings clarify individual differences in stimulus equivalence. They also demonstrate the utility of analyzing verbal reports to identify possible variables that can be manipulated experimentally.

Key words: stimulus equivalence, matching to sample, conditional discrimination, stimulus compounds, verbal behavior, protocol analysis, adult humans

When humans are taught a number of conditional discriminations within a set of stimuli, the stimuli often evolve into a class of equivalent members, with untrained conditional relations emerging among them (Sidman & Tailby, 1982). Although this phenomenon, termed stimulus equivalence (Sidman, 1971), has been demonstrated in numerous studies, both with children (e.g., Devany, Hayes, & Nelson, 1986; Lazar, Davis-Lang, & Sanchez, 1984) and with adult humans (e.g., Sidman, 1971; Wulfert & Hayes, 1988), equivalence classes do not always emerge from conditional discrimination training. With nonhumans, for example, attempts to demonstrate stimulus equivalence have not yet been successful (e.g., Sidman et al., 1982) or, when seemingly successful (McIntire, Cleary, & Thompson, 1987; Vaughan, 1988), were open to alternative interpretations (Hayes, 1989; Saunders, 1989). Sometimes even humans fail to show equivalence despite extensive training in the underlying conditional discriminations (e.g., Devany et al., 1986; Lazar, 1977), and the reasons for this variability are not well understood. It is possible that the differences between human and nonhuman responding are phylogenic in origin (Sidman, 1990). The performance differences among people, however, are more likely the product of complex ontogenic histories whose nature is largely unknown.

One barrier to a better understanding of individual differences in equivalence class formation is that subjects' matching to sample does not directly reveal what stimulus characteristics control behavior. Typically, we infer equivalence relations from performance when subjects reliably match the "correct" stimuli (as defined by the experimenter), but without additional tests we do not know what aspects of the situation control behavior when subjects match the "wrong" stimuli. Additional experimentation would likely reveal the sources of control, but it could well be a laborious endeavor. This process might be shortened considerably by examining subjects' verbal behavior to see what it reveals about the variables controlling performance in the experimental task. Gaining access to private self-talk, descriptions of figural images, and other covert behavior correlated with matching to

We thank two anonymous reviewers for their helpful comments. Please direct corespondence to Edelgard Wulfert, Department of Psychology, State University of New York, Albany, New York 12203.

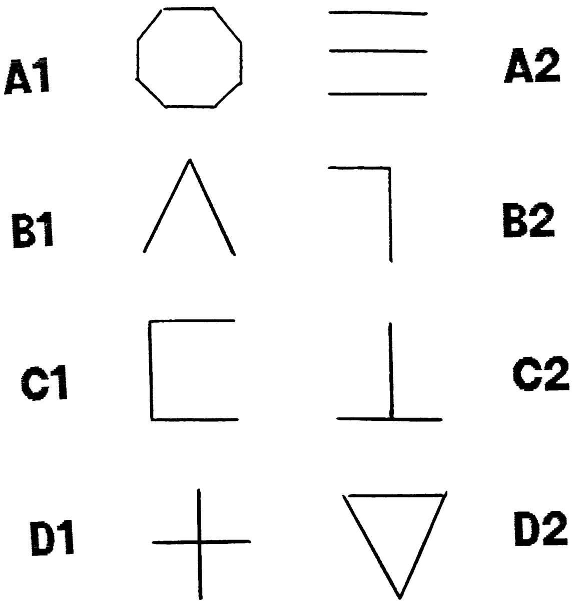

Fig. 1. Experiment 1: Visual symbols used for the conditional discriminations.

sample might be a convenient way to shed light on the sources of variability in equivalence class formation.

A common method of assessing what subjects say to themselves when they perform a task is by retrospective reports. Such reports may, however, be inaccurate because they are open to numerous distortions and misinterpretations (e.g., Nisbett & Wilson, 1977). To avoid these problems, cognitive psychologists have developed a more direct method of "observing thinking." This method, termed *protocol analysis* (Ericsson & Simon, 1980, 1984), requires subjects to "think aloud" while performing problem-solving tasks. Cognitive scientists assert that having subjects verbalize their thoughts as they enter consciousness, without trying to explain, analyze, or interpret them, will have no disruptive effect on performance. This claim is supported by a large number of studies comparing subjects thinking aloud with subjects performing the same task silently. None of these studies found differences in success rates or methods employed, except that in tasks in which subjects used visual images their verbal descriptions slightly slowed down their performance (see Ericsson & Simon, 1984, for a review). Whereas cognitive psychologists are interested in protocol analysis because of what it reveals about underlying cognitive processes, behavior analysts may find this method useful as a way of assessing the relationship between covert and overt behavior as well as examining what private behavior may reveal about important sources of subtle stimulus control (see Hayes, 1986, for a review of Ericsson and Simon's book and for a discussion of its implications for behavior analysis).

The two studies presented below used protocol analysis in an attempt to identify sources of variability in equivalence class formation. Experiment 1 was designed to teach subjects a think aloud procedure and to examine their verbal behavior for possible discrepancies correlated with performance differences in matching to sample. In Experiment 2 we then trained subjects in a task designed to increase the likelihood that they would emit specific types of self-talk we had found to be correlated with performance differences in Experiment 1. Through this manipulation we attempted to bring the emergence of equivalence classes under experimental control.

EXPERIMENT 1

This experiment was conducted as an exploratory study to examine what subjects say to themselves while performing conditional discriminations. Its specific purpose was to compare task-relevant verbal behavior of subjects who show stimulus equivalence with those who do not.

METHOD

Subjects

Twenty-nine college undergraduates of both sexes, enrolled in an introductory psychology course, participated in the study for course credit. However, the data from only 10 subjects were analyzed (see below).

Apparatus and Materials

A Tandy 1000 SX® microcomputer with a color monitor controlled the experiment (stimulus presentations, timing, data collection, and printing). The stimuli for the conditional discriminations consisted of eight arbitrary visual symbols (see Figure 1). Subjects were seated before the computer and responded on the numeric pad of the keyboard. Presses on Key 1 selected the left comparison stimulus on the monitor, and presses on Key 2 selected the right one. Subjects' verbal behavior was recorded with a sensitive clip-on mini-microphone (Realistic® Tie Pin Microphone 33-1063) connected to a cassette recorder (Realistic® VSC-2000).

General Procedure

All subjects were trained and tested individually in a session that lasted a maximum of 2 hr (including obtaining informed consent, providing instructions, training, testing, and debriefing).

Think aloud procedure. At the start of the experiment, subjects were trained to think aloud on a task unrelated to the conditional discrimination procedure. The experimenter read the following instructions:

> When the experiment begins you will see sets of symbols appear on the screen. Depending on which one you think is correct, you will choose either the left or the right symbol at the bottom of the screen by pressing the left key, 1, or the right key, 2 (the experimenter here pointed to the keys). At first the screen will tell you whether your choices are correct, but later no further feedback will be given.
>
> As we are interested in understanding how people solve problems, we want you to think out loud during the entire experiment. We will record what you say and later transcribe it. So that you understand what I mean by thinking out loud, let me give you an example. Assume I asked you, "How much is 127 plus 35?" Now think out loud so I can hear how you solve this problem.

If subjects simply stated the solution ("162"), the experimenter corrected them and modeled an example: "Assume the problem is 123 plus 66. To solve it, I will think, 123 plus 6 makes 129, plus 60 makes 189. Now here's another problem. Solve it thinking out loud." Similar arithmetic problems were presented until subjects verbalized the process of the solution on two consecutive trials. Then the computer program with the conditional discriminations was started. The experimenter remained in the subject booth for 3 to 5 minutes and prompted, "Don't forget to think out loud," if on any two consecutive trials no utterance was forthcoming.

Matching to sample. On each trial a sample stimulus (A1 or A2) appeared at the top of the screen and was followed 2 s later by two comparison stimuli at the bottom (B1-B2, C1-C2, or D1-D2). A response removed the stimulus display, produced brief written feedback during training ("correct" or "wrong"), and was followed by a 2-s intertrial interval. To facilitate comparing the verbal protocols with the computer-generated records of stimulus presentations and responses, every trial started with a single computer tone, and every 10th trial began with a double tone. These tones ensured that the computer printouts and transcripts of the audiotapes could be matched trial by trial.

Table 1

Training and test sequence for Experiments 1 and 2.

	Composition of presentation	Criterion (consecutive trials correct)
I. Training steps		
1. A-B relations	100%	
A1-B1		20
A2-B2		20
2. A-B relations	25%	
A-C relations	75%	
A1-C1		15
A2-C2		15
3. A-B relations	12.5%	
A-C relations	12.5%	
A-D relations	75%	
A1-D1		15
A2-D2		15
4. Random mix of A-B, A-C, and A-D relations	100%	40
II. Fading out feedback		
Over 10 consecutive trials		
III. Testing		
1. Symmetry test:		
Symmetry probes	75%	40
Trained relations	25%	
2. Equivalence test:		
Equivalence probes:	50%	
B1-C1, B2-C2		5, 5
B1-D1, B2-D2		5, 5
C1-D1, C2-D2		5, 5
C1-B1, C2-B2		5, 5
D1-B1, D2-B2		5, 5
D1-C1, D2-C2		5, 5
Symmetry probes	25%	
Trained relations	25%	

Training and test sequences and requirements are presented in Table 1. Briefly, the A-B, A-C, and A-D conditional discriminations were trained successively. Each conditional discrimination was trained to criterion (either 15 or 20 consecutive trials correct) before the next one was added. Then quasi-randomized trials of all trained relations were presented (with the restriction that the same trial never appeared more than three times in a row), until subjects correctly completed 40 consecutive trials. After reaching this criterion, the written feedback ("correct") was faded over the following 10 trials by replacing one letter at a time with dots on each successive trial (i.e., Correc.; Corre. .; Corr. . .; Cor. . . .; etc.) and finally having the dots disappear. No further feedback was provided for the remainder of the session.

During the 10 fading trials, an alternating S− procedure was introduced. The comparison stimuli no longer consisted of the B1-B2, C1-C2, or D1-D2 pairs used during training; instead, the positive comparison stimulus for a given sample (e.g., B1 for sample A1) was now paired randomly with any of the negative ones (i.e., A2, B2, C2, or D2). This method was introduced to determine whether responding was controlled by sample/S+ relations. If a subject instead had learned sample/S− relations (e.g., choosing "not B2" instead of "B1"), the change in stimulus presentations when fading feedback would have decreased responding from 100% accuracy to chance level and subjects would not have entered the test phase. The alternating S− procedure also served to rule out the possibility that equivalence relations (if they emerged) were context dependent; that is, that they would only occur in the context of specific sets of positive and negative comparisons used in training (Fields, Verhave, & Fath, 1984).

Unpublished work from our laboratory suggested that the formation of equivalence classes is facilitated when symmetry probes are presented before introducing equivalence probes; so the first test consisted of 75% symmetry probes interspersed with 25% unreinforced baseline trials. If subjects met the symmetry test criterion (40 consecutive trials correct), equivalence probes were added such that the ensuing test consisted of a quasi-random mixture of approximately 50% equivalence probes, 25% symmetry probes, and 25% baseline trials. To complete the test, subjects had to solve a minimum of five consecutive probes of each of the 12 equivalence relations presented in Table 1. Meeting this criterion was taken as evidence that two four-member equivalence classes had emerged: A1/B1/C1/D1 and A2/B2/C2/D2. If subjects did not meet this criterion within the 2-hr time limit, the test was discontinued and subjects were debriefed.

Verbal protocols. Subjects' utterances were transcribed verbatim from the audiotapes. The transcribers in the first study were undergraduate assistants familiar with the experiment; those in the second study were blind to the purpose of the experiment and unfamiliar with equivalence research in general. To evaluate the accuracy of the transcripts, one of the authors and a graduate student transcribed four tape sections and compared them trial by trial to those of the undergraduate assistants. The transcripts matched on all but two of 500 trials.

To categorize subjects' verbalizations, the authors inspected the verbal protocols, derived four formal categories of verbal responses, and trained three assistants to assign the utterances to these classes on a trial-by-trial basis. The four categories were termed:

1. "Relational responding," including references to the relationship between two stimuli ("Circle goes with the open triangle"), to the nodal stimulus ("Both symbols went with the three bars"), or to a superordinate class name ("Both belong to 'complexity' ").
2. "Common physical features,"which referred to relating pairs of stimuli by specific, nonarbitrary features ("They look alike" or "Both have 90-degree angles").
3. "Stimulus compounds," suggesting that subjects visually integrated sample and comparison stimuli ("Together they look like a house").
4. "Other," including responses not previously classified ("I'm choosing the left") or remaining silent.

RESULTS

Given that the transcription and analysis of verbal protocols are extremely labor intensive and time consuming, a protocol analysis was performed only for those 5 subjects who failed to reach the test criterion (here arbitrarily designated as Subjects 1 through 5) and for another 5 subjects (designated as 6 through 10) who were selected for comparison from those who successfully completed the experiment. Table 2 presents the number of training trials, test probes, and percentage of errors subjects made on the tests. The experimenter accidentally exceeded the session limit for Subject 3, which accounts for the large number of test

Table 2

Training and test performances for subjects failing (1 through 5) or passing (6 through 10) the equivalence test (Experiment 1).

Subject	Training to criterion (number of trials)	Test probes: Trained Trials	Trained Errors (%)	Symmetry Trials	Symmetry Errors (%)	Equivalence Trials	Equivalence Errors (%)
1	246	117	17	153	8	214	49
2	205	113	34	346	48	—	
3	222	134	7	220	10	188	29
4	355	127	7	145	16	231	47
5	226	56	0	78	0	95	35
6	209	62	0	100	1	93	2
7	265	69	0	93	0	121	2
8	198	41	0	62	0	67	0
9	226	41	0	65	0	71	1
10	208	37	3	64	2	67	0

trials presented. The session for Subject 5 was prematurely terminated upon the subject's request.

Among those subjects who did not show equivalence, Subject 2 did not complete the symmetry test, and Subjects 1, 3, 4, and 5 failed up to one half of the probes on the equivalence test (Figure 2). In contrast, Subjects 6 through 10 completed the symmetry and equivalence tests with few, if any, errors (Figure 3).

A comparison of the verbal protocols revealed that the performance differences on the conditional discriminations were correlated with different types of utterances. The results of the protocol analysis are summarized in Table 3. Of those subjects whose performance did not demonstrate equivalence, Subjects 1, 4, and 5 referred to common physical features of the stimuli on up to 28% of the trials, and on up to 50% of the trials they remained silent or made ("other") utterances that revealed little about the problem-solving process. Most interestingly, however, all of these subjects named compound stimuli to varying degrees (e.g., A1-B1: "Together they look like a person with hat"; A1-D1: "Cross goes inside the circle"). Subject 2 did so on virtually all test trials, and Subjects 1, 4, and 5 did so on at least one third of the trials. Only Subject 3 did not describe compounds (except on a few initial A-B trials). During training she emitted progressively fewer utterances, only occasionally naming the position of the symbols. From Trial 100 on she remained silent, which made it impossible to determine whether she responded to compound stimuli during testing.

The utterances of Subjects 6 through 10 (whose performance showed equivalence) differed markedly from those of Subjects 1 through 5. On at least 77% of the training trials and 93% of the test trials, these subjects emitted relational responses; that is, they named and linked the stimuli with relational phrases, and on equivalence probes they explicitly referred to the nodal stimulus. Over the course of the experiment, Subjects 6 and 9 stopped naming the symbols and began to relate all stimuli pertaining to the same class by category names. All 5 subjects completed the equivalence test virtually without errors.

Incidentally, all subjects changed their utterances over the course of the experiment. At the beginning, their verbalizations tended to be lengthy (e.g., "I guess these lines could go with the box, but then, I don't know, they could also go with the other one"). However, over multiple repetitions of the same stimulus configurations, all subjects essentially ended up using telegram style, simply naming sample and comparison stimuli ("cross-triangle") or occasionally only referring to the comparison stimulus ("triangle").

Discussion

In Experiment 1, subjects who formed equivalence classes tended to name the symbols and match them with relational phrases. In contrast, subjects whose performance did not demonstrate equivalence mainly described stimulus compounds, which suggested that responding to compound stimuli might be one possible source of variability in equivalence

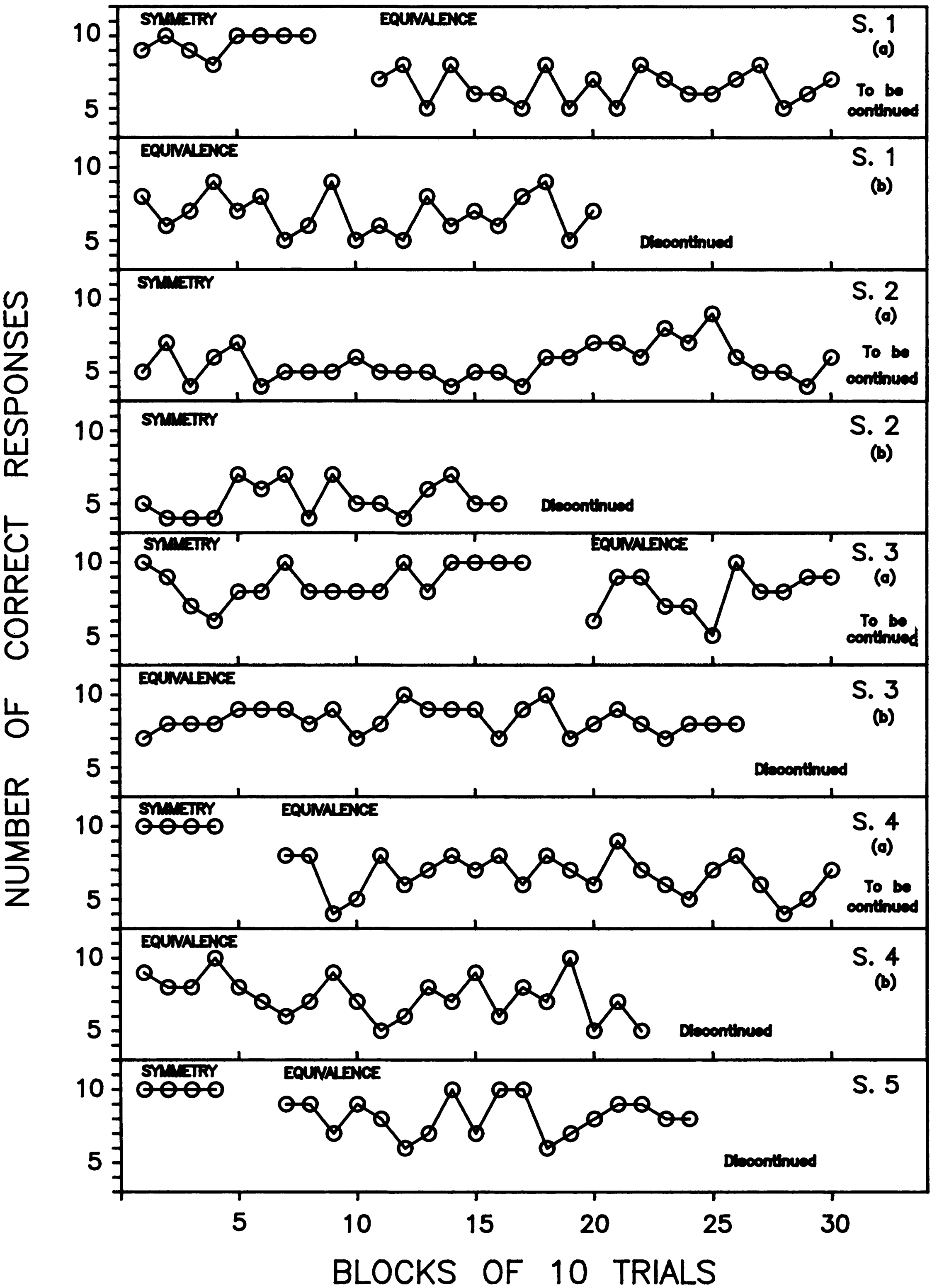

Fig. 2. Experiment 1: Performance of Subjects 1 through 5 on unreinforced symmetry and equivalence tests. Each data point reflects the number of correct responses per blocks of 10 trials. (The last block in each test may reflect fewer or more than 10 trials.)

class formation. Although this interpretation is necessarily speculative (after all, one might argue that our "think aloud" instructions gave rise to verbalizations that would not have occurred otherwise), it is plausible, as the following example shows. Let us assume that some subjects during conditional discrimination training respond to sample and comparison stimuli as unitary compounds rather than as independent stimuli. Responding to compound stimuli would not interfere with the training phase of a standard equivalence experiment, which can be completed by solving conditional (if A1, choose B1) or simple discriminations (choose A1 plus B1). However, control by compound stimuli would interfere with the emergence of equivalence relations because the individual stimuli can no longer function independently in derived relations, thus violating one of the defining characteristics of equivalent stimuli (see Sidman, 1986). The results from Experiment 1 were consistent with this interpretation; however, the correlational nature of the design did not permit causal inferences.

To study the effects of responding to compound stimuli on equivalence class formation under more controlled circumstances, we designed a second experiment in which we manipulated our subjects' histories through pretraining to engender responding either to

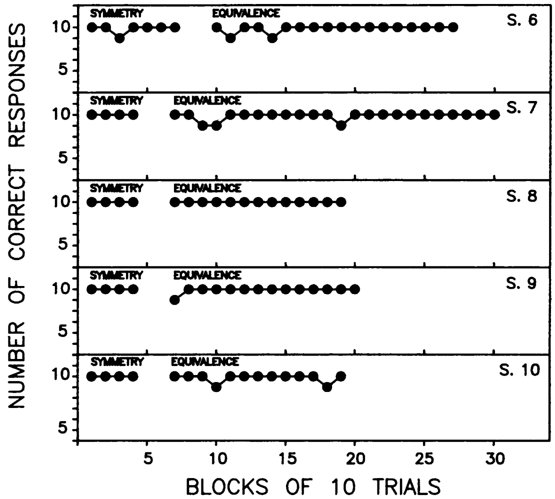

Fig. 3. Experiment 1: Performance of Subjects 6 through 10 on unreinforced symmetry and equivalence tests. Each data point reflects the number of correct responses per blocks of 10 trials. (The last block in each test may reflect fewer or more than 10 trials.)

Table 3

Verbal responses grouped by categories for subjects failing (1 through 5) or passing (6 through 10) the equivalence test (Experiment 1).

Subject		Relational responses		Naming stimulus compounds		Common physical features		Other responses/ no answer	
		n	%	*n*	%	*n*	%	*n*	%
1	Training	45	18	88	36	34	14	79	32
	Test	62	13	154	32	97	20	171	35
2	Training	—		167	81	—		38	19
	Test	—		452	98	—		7	2
3	Training	—		4	2	—		218	98
	Test	—		—		—		542	100
4	Training	58	16	162	46	101	28	34	10
	Test	121	24	206	41	123	24	53	11
5	Training	—		69	31	42	19	115	50
	Test	—		76	33	63	28	90	39
6	Training	168	81	—		36	17	15	7
	Test	243	95	—		—		12	5
7	Training	237	89	—		—		28	11
	Test	180	99	—		—		3	1
8	Training	164	83	—		12	6	4	2
	Test	167	98	—		—		3	2
9	Training	174	77	—		19	8	33	15
	Test	170	96	—		—		7	4
10	Training	166	80	—		8	4	34	16
	Test	156	93	—		—		12	7

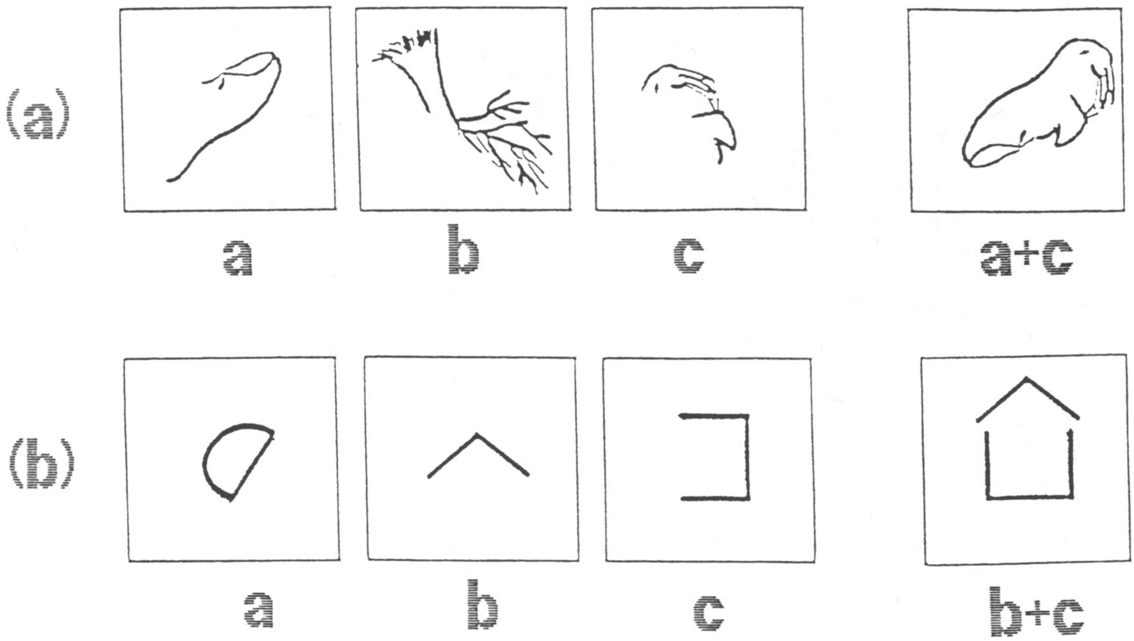

Fig. 4. Experiment 2: Sample stimuli from the pretraining procedure for the compounding group. Subjects received sets of three partial stimuli and were to mark "those two that together form a picture." From the sample stimuli in the first row (a), Stimuli a and c together form a walrus, as shown in the picture on the right. From the stimuli in the second row (b), Stimuli b and c together form the outline of a house, as shown on the right.

compound stimuli or to the relation between stimuli. We predicted that subjects pretrained in responding to compound stimuli would fail equivalence tests and that those pretrained in relational responding would demonstrate equivalence.

EXPERIMENT 2

METHOD

Subjects

Fourteen college undergraduates of both sexes participated in this study for credit as part of an introductory psychology course. In addition, two $20.00 prizes were offered to those subjects who completed the equivalence test with the smallest number of errors. Subjects were randomly assigned to either the compounding group (n = 7) or the relational group (n = 7).

Apparatus and Materials

Pretraining. The material for the pretraining phase consisted of two types of stimuli: (a) drawings of animals and plants and (b) simple line drawings (e.g., the outline of a house, a glass, etc.). The latter more closely resembled the arbitrary visual symbols later used for matching to sample. Both groups of subjects received 100 sets of three stimuli each, 70 containing drawings of Type a and 30 of Type b. For the compounding condition, all drawings were cut in half. In each set of three partial drawings, two of the parts fit together to make a whole (e.g., a walrus), which was intended to convey the idea of compounding stimuli. Sample sets are presented in Figure 4.

For the relational condition, each set consisted of three intact pictures or line drawings, with two drawings "going together" (e.g., bird and elephant vs. flower) because they belonged to the same category (see sample sets in Figure 5). This condition was intended to convey the idea of matching physically different symbols that were members of a given class.

Matching to sample. The second part of the experiment was computerized and set up identically to Experiment 1, except that different visual symbols (Figure 6) were used for the conditional discriminations.

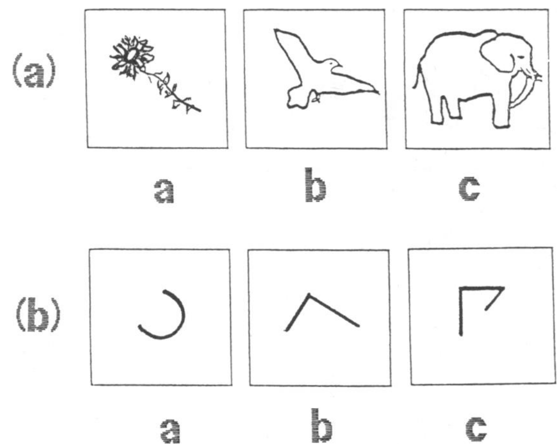

Fig. 5. Experiment 2: Sample stimuli from the pretraining procedure for the relational group. Subjects received sets of three stimuli and were to mark "those two that go together." From the sample stimuli in the first row (a), Stimuli b and c go together (class: animals); from the stimuli in the second row (b), Stimuli a and b go together (class: letters).

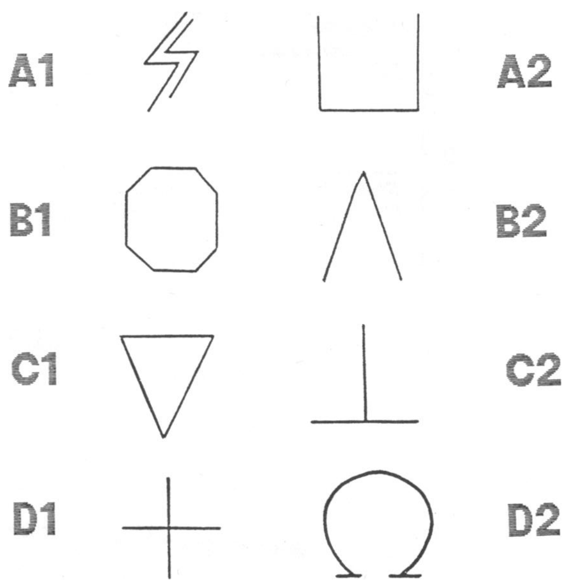

Fig. 6. Experiment 2: Visual symbols used for the conditional discriminations.

Procedure

All subjects were trained and tested individually.

Pretraining. For the pretraining manipulations, subjects first received a packet of 70 sets of either divided (Figure 4(a)) or intact (Figure 5(a)) drawings and the written instructions: "For each row, please mark those two pictures that go together." The experimenter observed the first 10 trials and gave corrective feedback, if necessary, until all subjects in the compounding condition marked partial stimuli that together formed a whole and subjects in the relational condition marked stimuli that belonged to the same class. Then subjects received a second packet with 30 sets of divided (Figure 4(b)) or intact (Figure 5(b)) line drawings and instructions to "Do with these symbols the same as with the pictures." Because the stimuli for this task were considerably more abstract, for the first 10 trials the experimenter had subjects state why they matched a given pair. If necessary, he or she verbally prompted or modeled responding to the relation between symbols or to compound stimuli, for example, by saying "These two go together, they look like letters," or "If you fit both pieces together it'll look like a glass." In no case was prompting necessary beyond the 10th trial.

Equivalence experiment. After pretraining, all subjects were trained to think aloud and were exposed to the equivalence experiment. All procedures and instructions were identical to those described in Experiment 1 with the following additions:

1. Motivational manipulation. To increase subjects' motivation for error-free responding on the equivalence tests and thus to provide a strong basis against which to evaluate the effects of pretraining, one sentence was added to the written instructions: "Those two subjects who make the fewest errors during that part of the experiment where no feedback is given will win $20.00."

2. Verbal protocols. For the protocol analysis, the four categories used in Experiment 1 were extended to seven. Specifically, relational responding was refined and divided into three subcategories: (a) naming two stimuli (e.g., "Octagon goes with cross"), (b) naming the nodal stimulus (e.g., "Cross and triangle both went with the octagon"), and (c) relating stimuli by superordinate class (e.g., designating all Class 1 symbols as simple and Class 2 symbols as complex figures). The latter category was added because we observed in Experiment 1 that 2 subjects initially named individual stimuli, but during testing generated class names and matched the stimuli accordingly. In ad-

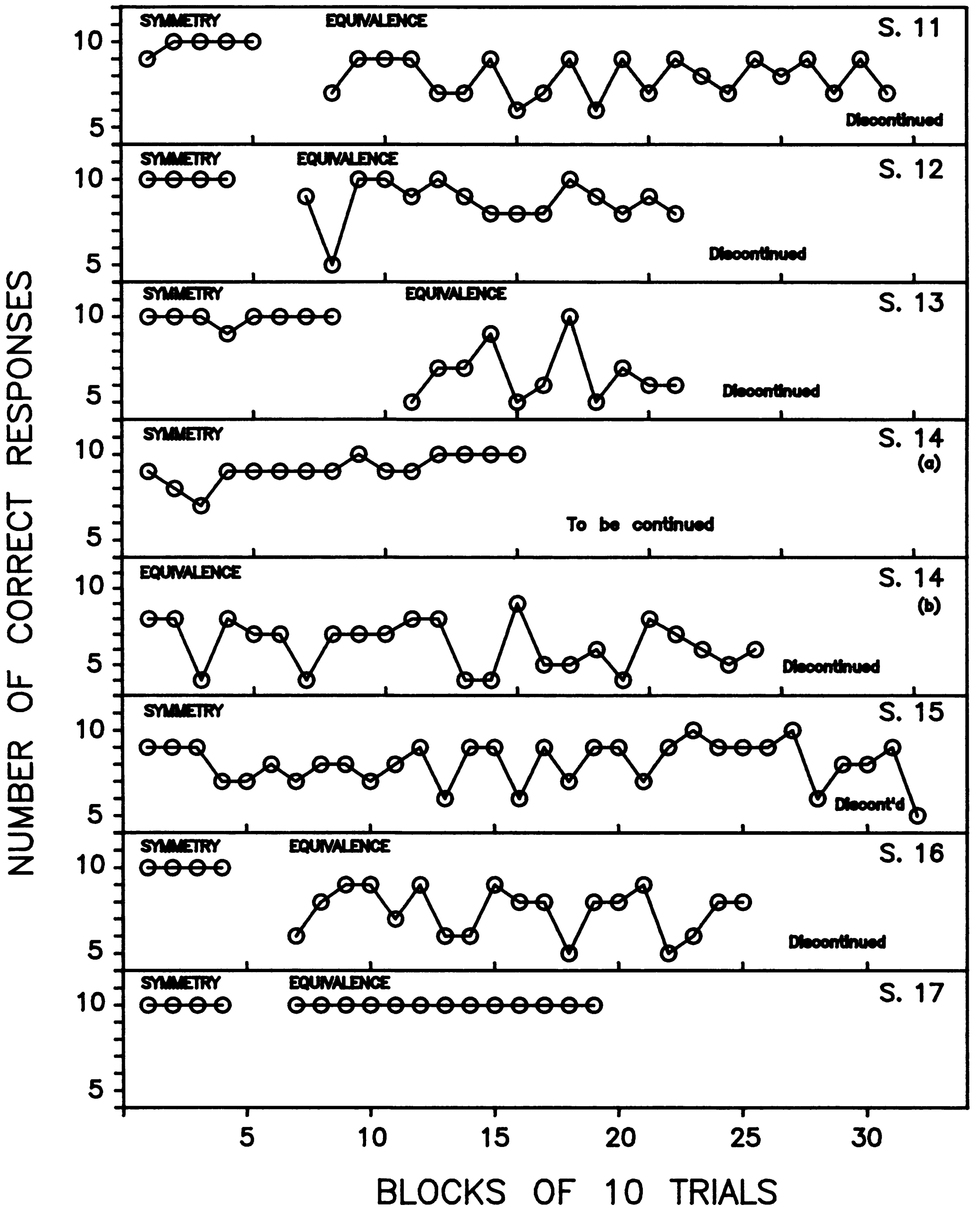

Fig. 7. Experiment 2: Performance of Subjects 11 through 17 (compounding group) on unreinforced symmetry and equivalence tests. Each data point reflects the number of correct responses per blocks of 10 trials. (The last block in each test may reflect fewer or more than 10 trials.)

Table 4

Training and test performances for subjects in the compounding (11 through 17) and the relational (18 through 24) groups (Experiment 2).

Subject	Training to criterion (number of trials)	Test probes: Trained		Symmetry		Equivalence	
		Trials	Errors (%)	Trials	Errors (%)	Trials	Errors (%)
11	175	65	2	92	4	108	39
12	177	46	0	66	0	71	28
13	176	43	0	85	6	55	58
14	169	102	1	180	9	118	70
15	199	79	4	242	24	—	
16	176	53	2	78	1	93	49
17	187	39	0	64	0	68	0
18	211	55	0	80	0	95	4
19	184	45	0	60	0	66	0
20	178	43	0	65	2	65	0
21	191	71	0	107	1	93	4
22	173	45	0	64	0	72	6
23	170	45	0	66	0	70	0
24	179	91	0	224	6	58	14

dition, a separate category for naming one stimulus was added because some subjects occasionally referred only to the comparison stimulus.

To examine whether the verbal responses were categorized reliably, one of the authors classified 50 training and 50 test trials each from four protocols and compared them to the assistants' categorizations. Only seven disagreements were found in 400 trials, which suggests that the response categories were well defined.

3. Postsession reports. Before being debriefed, subjects completed a written postsession questionnaire requiring them to draw the stimuli they remembered, specifying their names (if any), and noting what strategy (if any) they had used to relate them. These reports were compared to the utterances subjects had emitted during matching to sample.

RESULTS

Subjects' performances on the conditional discriminations and their concurrent verbalizations were analyzed and compared. Table 4 presents the number of training and test trials and percentage of errors for subjects in the compounding group (arbitrarily designated as Subjects 11 through 17) and the relational group (Subjects 18 through 24).

Subjects in both groups performed comparably during training, but significant differences emerged on the symmetry and equivalence tests. As shown in Figure 7, 6 of the 7 subjects in the compounding group failed the symmetry or equivalence tests, whereas an equal number of those in the relational group (described below) completed these tests with few, if any, errors.

The protocol analysis (Table 5) revealed that the 7 subjects in the compounding group described compound stimuli on 61% to 95% of the baseline conditional discriminations and on 36% to 100% of the test probes (e.g., A1-B1: "Lightning striking bull's eye"; A2-B2: "house"). On up to 34% of the trials, these subjects also emitted responses (classified as "other") that did not reflect the problem-solving process.

The verbal protocol of Subject 15, the only subject who did not complete the symmetry test, showed that on some trials she integrated a given sample with the positive and on other trials with the negative comparison stimulus, depending on which S− was presented. For example, with sample C2 and comparisons A1-A2 she integrated C2-A2 and named the compound the same as during training ("Stemmed object"); however, with the same sample and comparisons A2-C1 she related C2-C1 ("Yield sign").

Only 1 subject in the compounding group (Subject 17) did not show the predicted effect of the pretraining manipulation. During training, she named four sample/S+ compounds

Table 5

Verbal responses grouped by categories for subjects in the compounding (11 through 17) and the relational (18 through 24) groups (Experiment 2).

		Relational responses													
		Superordin. class		Naming nodal stimulus		Naming two stimuli		Naming one stimulus		Naming stimulus compounds		Common physical features		Other responses/ no answer	
Subject		*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
11	Training	—		—		—		—		128	73	—		47	27
	Test	—		—		—		—		241	91	3	1	21	8
12	Training	—		—		—		10		162	92	1	1	4	2
	Test	48	26	—		—		—	6	84	46	50	27	1	1
13	Training	—		—		—		—		119	68	—		57	32
	Test	—		—		—		—		163	89	2	1	18	10
14	Training	—		—		—		—		141	83	6	4	22	13
	Test	—		—		—		—		349	87	6	2	45	11
15	Training	—		—		—		—		190	95	—		9	5
	Test	—		—		—		—		321	100	—		—	
16	Training	—		—		—		—		165	94	—		11	6
	Test	—		—		—		—		207	92	—		17	8
17	Training	—		—		—		—		114	61	—		63	34
	Test	—		68	40	—		—		61	36	—		42	25
18	Training	—		—		4	2	4	2	—		5	2	198	94
	Test	—		94	41	32	14	—		—		—		104	45
19	Training	—		—		159	86	6	3	—		—		19	10
	Test	92	54	—		77	45	—		—		—		2	1
20	Training	—		—		168	94	2	1	—		2	1	6	4
	Test	—		36	21	137	79	—		—		—		—	
21	Training	—		—		138	72	—		—		13	7	40	21
	Test	133	49	9	3	125	46	—		—		—		4	1
22	Training	102	59	—		45	26	4	2	—		16	9	6	3
	Test	109	60	—		66	37	—		—		6	3	—	
23	Training	—		—		166	98	1	.5	1	.5	1	.5	1	.5
	Test	—		26	14	155	86	—		—				—	
24	Training	—		—		—		42	23	106	59	15	8	16	9
	Test	—		49	13	—		90	24	212	57	19	5	3	1

and referred to them on 61% of the baseline trials. On the very first equivalence probe, however, she said "Hmmm, I haven't seen this one before. It's the left (C1) because it goes with the lightning bolt (A1) which goes with the one on the top (D1)." Although she continued to describe compounds on baseline and symmetry probes, on equivalence probes she related the symbols by their nodal stimulus and completed the test error free.

The performance of subjects in the relational group stood in sharp contrast to those of the compounding group. As shown in Figure 8, 6 of them formed equivalence classes virtually error free; only Subject 24 failed the test. The protocol analysis also revealed considerable differences between the two groups and indicated that subjects in the relational group most frequently verbalized relational responses (see Table 5). Five of them (Subjects 19 through 23) referred to the relations among stimuli on 72% to 100% of the training trials and on 87% to 100% of the test probes by naming two symbols or using stimulus class names, and on equivalence probes by referring to the nodal stimulus.

During training, Subject 18 referred almost exclusively to the position of the stimuli. During testing, he continued to do so on 45% of the probes, but on 14% of the test trials he also named the symbols and referred to the nodal stimulus on virtually all the equivalence probes. For Subject 24, pretraining in relational responding did not lead to the intended effect.

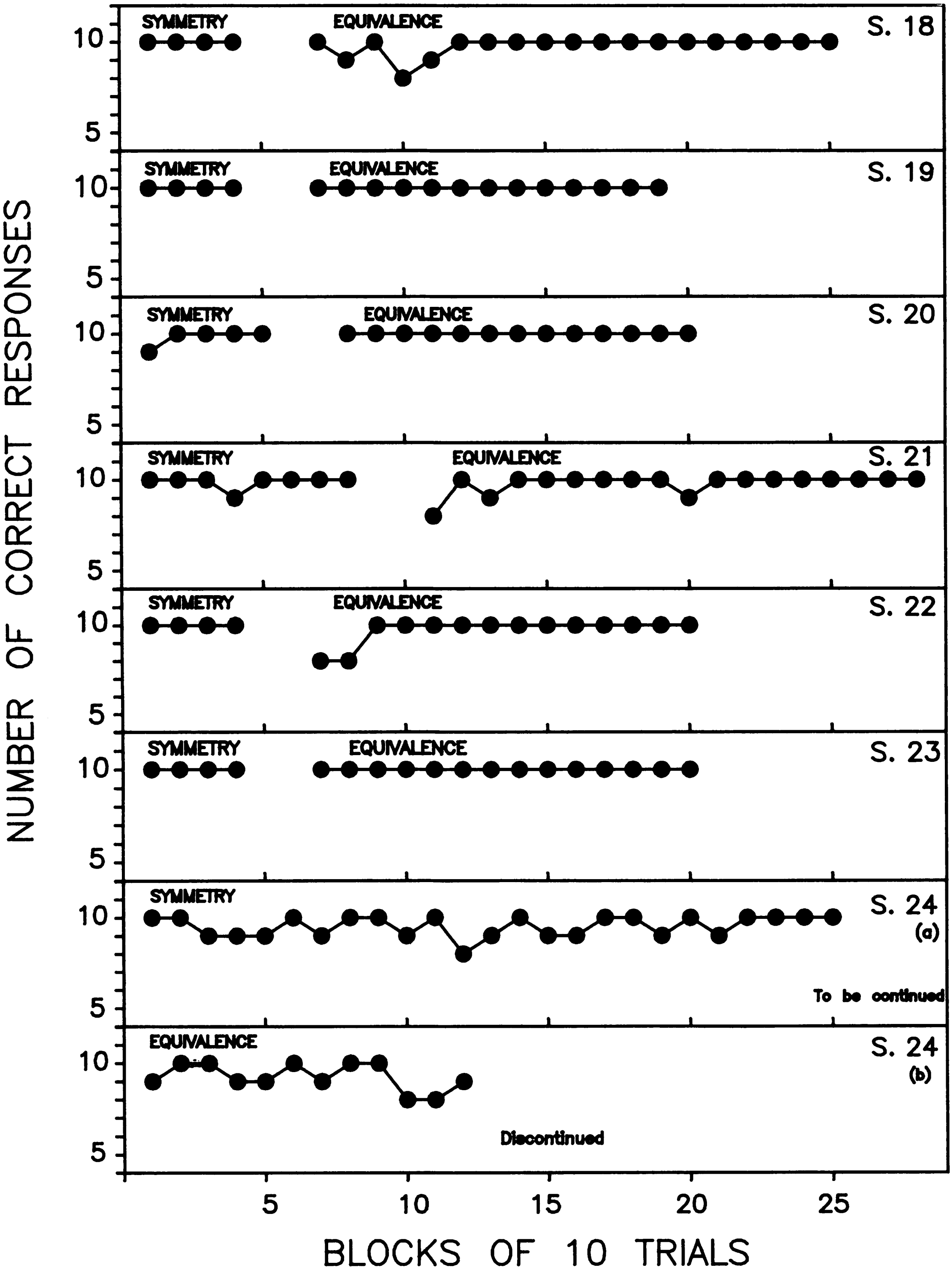

Fig. 8. Experiment 2: Performance of Subjects 18 through 24 (relational group) on unreinforced symmetry and equivalence tests. Each data point reflects the number of correct responses per blocks of 10 trials. (The last block in each test may reflect fewer or more than 10 trials.)

During training, he named one stimulus on only 23% of the trials, and on 59% of the trials he described stimulus compounds involving four of the six conditional discriminations. During testing, he related some symbols by their nodal stimulus, but on 57% of the test probes he continued to describe compounds. He was the only subject in the relational group to fail the equivalence test.

Discussion

The findings from Experiment 2 showed that particular manipulations of subjects' histories determined whether conditional discrimination training resulted in equivalence class formation. The results supported the hypothesis derived from Experiment 1 that subjects instructed to respond to sample and comparison stimuli as compounds would be less likely to demonstrate equivalence class formation than those who responded to relations.

These findings are significant not just because they help clarify individual differences in equivalence formation but also because they show the usefulness of analyzing subjects' verbal reports to identify possible sources of control over behavior. In contrast to cognitive research involving protocol analysis, however, the present research did not stop the analysis at the verbal reports obtained in Experiment 1, but experimentally manipulated the variables suggested by these reports. In Experiment 2 we were able to demonstrate that the verbal reports had identified a variable that influenced the experimental task. More importantly, however, it was also shown that it is unnecessary to assign causal status to self-talk or underlying hypothetical "cognitive processes" when it can be shown that both subjects' verbalizations and their task performance were the result of contingencies manipulated in the pretraining phase.

GENERAL DISCUSSION

In the present experiments, Ericsson and Simon's (1980, 1984) protocol analysis was used to examine the verbal responses subjects were instructed to emit concurrently with performing conditional discriminations. This think aloud procedure was implemented in an attempt to identify variables that might account for the variability in performance on equivalence tests that we have sometimes observed in human subjects of normal intelligence. The results of Experiment 1 identified responding to compound stimuli as one possible source of variability because virtually all subjects who described compound stimuli failed the equivalence test. To examine this possibility under more controlled circumstances we designed a second experiment, in which we randomly assigned subjects to one of two conditions and instructed them to describe either compound stimuli or the relation between stimuli. The results supported our hypothesis that subjects who were trained to verbalize stimulus compounds were unlikely to demonstrate equivalence.

The present studies suggest that verbal protocol analysis may be a useful methodological tool for behavior analysts. Asking subjects to think aloud while they perform a task may be a convenient shortcut to identify possible sources of control over behavior. Without a protocol analysis in Experiment 1, it would have been difficult to speculate that describing stimulus compounds was correlated with the failure to show equivalence.

A possible criticism of the present studies is that requiring subjects to think aloud might in itself have influenced their performance. Some subjects in the first experiment perhaps described compound stimuli not because this is what they would normally have done but because they were required to say something. Because the present studies did not include a comparison group without the talk aloud requirement, it is impossible to prove that verbalizing did not differentially affect performance. Future research should therefore examine this issue and compare subjects thinking aloud with subjects performing the same task silently.

One might also take issue with the way we tested equivalence relations, which differed from the standard procedure used by other researchers interested in equivalence. During testing we not only assessed all possible sample/positive comparison configurations in the context of the specific positive and negative comparison pairs used in training, but paired a given positive comparison with every possible negative comparison. One reason for this was to assess whether subjects' behavior during training had come under the control of specific sample/positive comparison combinations or of sample/negative comparison combinations (because responding to "not negative comparison" was also reinforced). A second reason was to demonstrate that the equivalence relations that did emerge were not context dependent and occurred in the presence of new comparison configurations not used in training. However, alternating the S− within trial types has recently been shown to engender responding to the S+ in the absence of reinforcement (Harrison & Green, 1990). One might therefore argue that our method of testing introduced alternative sources of control and that the performance of our subjects did not reflect equivalence. Although this argument is certainly reasonable, it appears implausible in the present case because subjects responding correctly on equivalence probes generally did so immediately. Immediate high accuracy would likely preclude control by the alternating S− because it would take more than one exposure to each trial type for the alternating S− to gain control.

A final limitation of the present studies concerns the generality of the findings. Because equivalence experiments rely almost exclusively on arbitrary visual symbols, it is possible that the tendency to compound stimuli or to associate them by nonarbitrary features is a method artifact. If the formation of equivalence classes is truly relevant to linguistic phenomena, neither compounding nor responding to arbitrary stimulus dimensions should play a role in naturalistic situations. After all, it would be difficult to think of a real-life example in which a person would compound a spoken or written word with the object it stands for, or associate words and objects according to similar physical features. Although this does not invalidate the present findings, it should alert researchers to the limited implications that can be drawn from the use of any given stimuli in equivalence experiments.

Although the present studies used protocol analysis to shed light on the variables controlling differences in equivalence class formation, the procedures have more general relevance for behavior analysts. Protocol analysis appears to be especially useful for assessing covert behavior and the extent to which it correlates with overt responding. For example, Hayes (1986) and Hayes, Zettle, and Rosenfarb (1989) have described the potential of protocol analysis as a method of investigating rule-governed behavior in general and self-rule governance in particular. In the present study, no attempt was made to identify specific functions of the subjects' covert behavior or their determinants. That is, the subjects may have been tacting the stimuli controlling their behavior, their covert behavior, or the relation between their covert and overt behavior (see Reese, 1989, for a discussion of rule governance and the problems inherent in its investigation). Recent evidence (Bentall & Lowe, 1987; Bentall, Lowe, & Beasty, 1985; Catania, Matthews, & Shimoff, 1982; Farber, 1963; Hayes et al., 1989) suggests that the relation between covert and overt responding is an important issue for behavior analysts, and protocol analysis appears to be a useful method for identifying variables that can be manipulated experimentally. As our two experiments demonstrate, an analysis of verbal protocols yielded a powerful independent variable that influenced the formation of equivalence classes.

REFERENCES

Bentall, R. P., & Lowe, C. F. (1987). The role of verbal behavior in human learning: III: Instructional effects in children. *Journal of the Experimental Analysis of Behavior,* **47,** 177–190.

Bentall, R. P., Lowe, C. F., & Beasty, A. (1985). The role of verbal behavior in human learning: II. Developmental differences. *Journal of the Experimental Analysis of Behavior,* **43,** 165–181.

Catania, A. C., Matthews, B. A., & Shimoff, E. (1982). Instructed versus shaped human verbal behavior: Interactions with nonverbal responding. *Journal of the Experimental Analysis of Behavior,* **38,** 233–248.

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

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review,* **87,** 215–249.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Farber, I. E. (1963). The things people say to themselves. *American Psychologist,* **18,** 185–197.

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. (1986). The case of the silent dog—verbal reports and the analysis of rules: A review of Ericsson and Simon's *Protocol Analysis: Verbal Reports as Data. Journal of the Experimental Analysis of Behavior,* **45,** 351–363.

Hayes, S. C. (1989). Nonhumans have not yet shown stimulus equivalence. *Journal of the Experimental Analysis of Behavior,* **51,** 385–392.

Hayes, S. C., Zettle, R. D., & Rosenfarb, I. (1989). Rule following. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 191–220). New York: Plenum Press.

Lazar, R. M. (1977). Extending sequence-class membership with matching to sample. *Journal of the Experimental Analysis of Behavior,* **27,** 381–392.

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

McIntire, K. D., Cleary, J., & Thompson, T. (1987). Conditional relations by monkeys: Reflexivity, symmetry, and transitivity. *Journal of the Experimental Analysis of Behavior,* **47,** 279–285.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review,* **84,** 231–259.

Reese, H. W. (1989). Rules and rule governance: Cognitive and behavioristic views. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 3–84). New York: Plenum Press.

Saunders, K. J. (1989). Naming in conditional discrimination and stimulus equivalence. *Journal of the Experimental Analysis of Behavior,* **51,** 379–384.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research,* **14,** 5–13.

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Erlbaum.

Sidman, M. (1990). Equivalence relations: Where do they come from? In D. E. Blackman & H. LeJeune (Eds.), *Behaviour analysis in theory and practice: Contributions and controversies* (pp. 93–114). Hillsdale, NJ: Erlbaum.

Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior,* **37,** 23–44.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior,* **37,** 5–22.

Vaughan, W., Jr. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes,* **14,** 36–42.

Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior,* **50,** 125–144.

Received April 3, 1990
Final acceptance June 5, 1991