CONTROLLING RELATIONS IN CONDITIONAL DISCRIMINATION AND MATCHING BY EXCLUSION

WILLIAM J. McILVANE, JOANNE B. KLEDARAS, LEE C. MUNSON,
KATHERINE A. J. KING, JULIO C. DE ROSE, AND
LAWRENCE T. STODDARD

EUNICE KENNEDY SHRIVER CENTER AND NORTHEASTERN UNIVERSITY

Normally capable adults learned two-choice identity matching of three-digit numerals and arbitrary matching of physically dissimilar nonsense syllables. The stimuli were displayed on a computer terminal, and responses consisted of typing on the terminal's keyboard. In Experiment 1, every trial displayed a sample numeral, a comparison numeral, and three equal signs (===). The comparison stimulus was to be selected if it was identical with the sample; otherwise the equal sign was to be selected. This "single comparison" method was then used to show that arbitrary matching could be based upon either sample-S+ or sample-S- relations. In Experiment 2, a series of probe trials displayed a novel sample, a comparison stimulus from the arbitrary matching baseline, and ===. Subjects typically selected ===; they apparently were excluding the baseline comparison stimulus. Experiments 3 through 5 investigated which variables in training would lead to the selection of baseline comparison stimuli in response to novel samples. Behavior was usually unchanged when baseline training included relating comparison stimuli to as many as four different samples. Punishment contingencies were effective, but performance did not generalize unless those contingencies were applied in relation to more than one baseline comparison stimulus.

Key words: stimulus control, matching to sample, controlling relations, exclusion, punishment, typing, normally capable adults

A major concern of behavior analysis is to account for behavior that occurs without a history of direct reinforcement and cannot be explained by primary stimulus generalization. Research topics relevant to these issues have included generalized matching-to-sample (Cumming & Berryman, 1965), generalized imitation (Baer, Peterson, & Sherman, 1967), generalized instruction-following (Striefel, Bryan, & Aikins, 1974), grammatical behavior

(Guess & Baer, 1973), functional equivalence (Goldiamond, 1966), and stimulus equivalence (Sidman, 1971; Sidman & Tailby, 1982). Each of these topics concerns new behavior that emerges in response to novel stimuli or to novel arrangements of familiar stimuli.

Dixon (1977) conducted one of the earlier studies of behavior that emerges when novel and familiar stimuli are displayed together in certain arrangements. Eight mentally retarded subjects were trained to select an invariant printed letter only after its name was spoken. Two other letters alternated irregularly as the incorrect choice over trials, and neither of these two letters was ever correct during training. With one set of stimuli, for example, the name spoken was "pi" and the correct letter was the Greek letter *pi*; the incorrect letter was *upsilon* on some trials and *theta* on others. After subjects selected *pi* reliably on baseline trials, probe trials were introduced. Baseline and probe trial displays were identical, but the names spoken were novel. For the example displays, the novel names were "upsilon" or "theta," depending on which letter was present with *pi*. Immediately, at first presentation of the probe trial, each subject responded to these novel names

This research was supported in part by NICHD Grants HD 17211, HD 17445, and HD 10210. Julio de Rose received postdoctoral support from Fulbright Grant No. 06669. We also acknowledge the generous support of Ben Heffron, whose private donation provided some of the computing equipment. These experiments were part of a dissertation submitted by William J. McIlvane to the faculty of Northeastern University in partial fulfillment of the requirements for the PhD degree. Portions of these data were presented at the meeting of the Association for Behavior Analysis, Columbus, Ohio, May 1985. The authors thank Harry Mackay and Jack Carroll for their comments on an earlier version of the manuscript and William Dube and Faith Sheiber for help in manuscript preparation. Special thanks go to Murray Sidman for his many suggestions that greatly improved the final product. Reprint requests may be addressed to William J. McIlvane, Behavioral Neurology Department, E. K. Shriver Center, 200 Trapelo Rd., Waltham, Massachusetts 02254.

by selecting *upsilon* or *theta* and not *pi*. Moreover, all subjects continued to select *pi* reliably when "pi" was spoken on baseline trials. The remarkable finding was the immediate differential control by the novel spoken letter-names. To explain this performance, Dixon hypothesized that subjects had learned a specific relation between the S+ letter (e.g., *pi*) and the name spoken during training (e.g., "pi"), discriminated the words spoken on the probes (e.g., "upsilon" or "theta") as different from that spoken during training, and rejected or excluded the S+ letter, because the probe-names differed from the name related to that letter. She called the performance *exclusion*.

Notably, Dixon's training contingencies did not require subjects to discriminate the baseline training word from other words. They could have learned merely to select the trained letter whenever any word was spoken. Moreover, Dixon did not teach the subjects how to respond to novel words. Nonetheless, all subjects behaved in the same way—apparently excluding the S+ letter and selecting previous S- letters when novel words were spoken. Dixon concluded that the subjects' performances were "based on a previously established . . . abstraction of a relationship between spoken words and environmental events" (p. 441), suggesting that exclusion performance might depend upon a highly developed speech repertoire.

Recent studies, however, have demonstrated exclusion in profoundly retarded, nonverbal individuals (McIlvane, Bass, O'Brien, Gervac, & Stoddard, 1984; McIlvane & Stoddard, 1981). Exclusion has also been demonstrated when none of the task stimuli is explicitly verbal (Dixon, Dixon, & Spradlin, 1983; Stromer, 1986). Thus, exclusion appears not to be a speech-related phenomenon; it seems to occur in the context of nonverbal conditional-discrimination procedures.

Many questions remain about the stimulus control of exclusion. The present study addressed the following three: (a) Does the *selected* stimulus on the exclusion probe control the subject's performance, or is its selection a byproduct of a relation between the novel sample and the stimulus that is excluded? (b) Does exclusion occur only when each stimulus selected during training is conditionally related to only one sample stimulus? (c) What training procedures make exclusion performances likely

and what procedures make them unlikely? Five experiments were conducted, with normally capable adults as subjects.

GENERAL METHOD

This section describes features that were common to all five experiments. Procedures specific to a given experiment will be described when that experiment is presented.

Subjects

Normally capable adult subjects were recruited through advertisements and personal contacts. A total of 70 individuals, 49 women and 21 men, served in one or more experiments. Most subjects (58) were college students. All but 2 of the remaining subjects had college degrees and were currently employed in clerical, administrative, or technical positions.

Apparatus and Experimental Setting

The apparatus was a PDP-11[®] microcomputer (Digital Equipment Corp.) with a floppy disk drive and a CRT terminal that displayed green characters on a black background. Stimuli were presented on the screen of the terminal, which was located on a desk in a small office. Subjects sat at the desk and responded by typing on the terminal keyboard.

Pretraining

The experimenter first familiarized the subject with the keyboard, including the RETURN and DELETE keys. Some subjects who were unfamiliar with computer terminals were permitted to play a computer game to practice input procedures. Before the start of the first session, the subject received instructions describing the nature of the tasks that he or she would perform. (See Appendix for the text of the instructions.)

Matching to Sample

Conditional discriminations were examined with a variant of the matching-to-sample procedure. Stimuli were composed of printed characters, displayed at the bottom of the CRT screen, centered above the keyboard. In all trials one sample stimulus and, below it, two comparison stimuli were displayed. The sample was presented with enhanced brightness. Two types of matching-to-sample trials were pre-

sented. Illustrative trials are shown in Figure 1.

Arbitrary matching. Stimuli were three-letter nonsense syllables (e.g., CUG, VEK). The sample and the two comparison stimuli always differed in at least two letters. Reinforcement contingencies defined which comparison stimulus was to be selected (i.e., reproduced) in the presence of each sample; for convenience, those stimuli are designated here as "correct." The first row of Figure 1 shows illustrative trials on which the comparison stimuli are ZID and PAF. When the sample was CUG, reinforcers could follow selections of ZID but not PAF. When the sample was VEK, reinforcers could follow selections of PAF but not ZID.

Identity matching. Stimuli on these trials were three-digit numerals, and any numeral could appear as a sample or comparison stimulus (Figure 1, third row). Reinforcers could follow selections of the comparison numeral that was the same as the sample.

Trial control. Programs written in BASIC controlled all stimulus presentations and collected the data. Matching-to-sample trial types were presented in quasi-random order; all samples appeared approximately equally often. Position of the correct comparison stimulus varied unsystematically; the same position was correct on no more than four consecutive trials.

Responses. The subject selected a comparison stimulus by typing its constituent characters on the keyboard. As this was done, the characters appeared on the CRT screen, centered below the comparison stimuli. Like the sample, stimuli produced by the subject's response were displayed with enhanced brightness. The subject could use the terminal's DELETE key to erase any mistyped letters and could reenter the correct ones. The response was completed by striking the RETURN key on the keyboard.

Programmed consequences. Two types of reinforcement schedules were used at different points in training. When new performances were taught, the bell in the computer terminal sounded after every correct selection. Simultaneously, "Correct" appeared below and to the left of the stimuli produced by the subject's response. Additional consequences followed about one-third of the "Correct" presentations; the program printed a message indicating that a coin had been earned (e.g., "Ten Cents"). The first subjects earned dimes; later

ARBITRARY MATCHING			
UNPROMPTED			
CUG		VEK	
PAF -	ZID +	ZID -	PAF +
PROMPTED			
CUG		VEK	
ZID +	PAF -	PAF +	ZID -
IDENTITY MATCHING			
UNPROMPTED			
347		961	
961 -	347 +	961 +	347 -
PROMPTED			
347		961	
347 +	961 -	347 -	961 +

Fig. 1. Representative arbitrary- and identity-matching trials. Two unprompted and two prompted examples of each trial type are shown. In each trial, a sample was centered above two comparisons. The correct and incorrect comparison stimuli are indicated by + and -, respectively, but + and - did not appear on the displays. Boldfaced stimuli indicate those displayed with enhanced brightness.

subjects earned quarters. Subjects could earn about \$10 to \$20 in the course of their participation. The computer program kept track of money earned across trials and printed the total amount at the end of the session. If the subject's response was incorrect, "Wrong" was printed below and to the left of the response-produced stimuli and neither "Correct" nor a money amount was presented. The screen was then erased and an intertrial interval of approximately 1.5 s preceded the next trial.

After new performances were learned, reinforcement became intermittent; reinforcers followed about 50% of the subjects' correct responses (VR 2). On unreinforced trials, the program merely advanced to the intertrial in-

terval. "Wrong" no longer followed incorrect responses.

Typographical errors. Because subjects typed their comparison selections, typographical errors could occur (i.e., letters incorrect, repeated, or out of order). The computer program screened for these events. If any two of the three letters typed appeared in the correct comparison stimulus, the response was called a "typo," and the comparisons were presented again, under the original stimuli. The subject then could correct the typo, reentering his or her selection before proceeding to the next trial. If the reentry was correct, it was followed by a reinforcer, if appropriate to the prevailing schedule.

EXPERIMENT 1

The research questions of Experiment 1 derived from studies of nonhumans' (primarily pigeons') performance on matching-to-sample procedures (e.g., Carter & Werner, 1978; Cumming & Berryman, 1965; Skinner, 1950). We asked whether adults' matching-to-sample performances could be based on relations between (a) the sample and the correct comparison stimulus, (b) the sample and the incorrect comparison stimulus, or (c) both types of relations. Relations of the first type may be called S+ (or S^D) relations and those of the second, S- (or S^A) relations. An S+ relation involves the sample and the comparison stimulus whose selection is (at least occasionally) reinforced; the performance does not depend on any particular negative comparison. An S- relation involves the sample and the negative comparison stimulus, whose selection is never reinforced; the performance does not depend on any particular positive comparison stimulus. The bulk of experimental evidence suggests that S+ relations are readily established in infrahumans (Carter & Werner, 1978). Experiments testing for S- relations in nonhuman species have generally reported negative results, and the few reports of success (e.g., Zentall, Edwards, Moore, & Hogan, 1981) may be open to other interpretations (cf. D'Amato, Salmon, & Colombo, 1985).

Both mentally retarded and normally capable children, however, do seem to acquire both S+ and S- relations (Dixon & Dixon, 1978; McIlvane, Withstandley, & Stoddard, 1984; Saunders & Sherman, 1986; Sherman,

Saunders, & Brigham, 1970; Stromer & Osborne, 1982). In Experiment 1 we tested the prediction that normally capable adults would acquire them as well; it also examined a novel procedure, termed "single-comparison" matching to sample, that would be used in subsequent analyses of exclusion performance.

The single-comparison procedure was first reported by McIlvane, Withstandley, and Stoddard (1984). In that study, four severely retarded males were first taught two-comparison (a) arbitrary auditory-visual and (b) visual-visual identity matching-to-sample performances. On arbitrary-matching trials, samples were spoken food names and comparison stimuli were food items. On identity-matching trials, the subject was to select a comparison food item that was physically identical to a sample food item. Next, a stimulus-shaping program progressively covered one of the food comparisons in the auditory-visual task with an opaque screen. In the final performance, subjects could observe only one food item on each trial. If this food item matched the spoken sample, the subject selected it as before. If the food item did not match, the subject rejected it and selected the screen. Then, additional auditory-visual and visual-visual identity-matching trials were presented in the single-comparison format; none of the stimuli displayed on these trials had been involved in stimulus shaping. Subjects also displayed S+ and S- relations on these trials.

A critical aspect of the single-comparison procedure was that the screen was at some times the S+ and at other times the S- in the presence of each sample; the screen's function as S+ or as S- depended on the comparison stimulus that was displayed with it. In this regard, the screen functioned differently from the comparison food items. Specific food items served only as S+ or only as S- in the presence of each sample; no food item ever served as both S+ and S- in the presence of any sample.

Because the screen served as both S+ and as S- in the presence of every sample, subjects could not make consistently correct selections if they observed only the sample and the screen. Consistently correct behavior on single-comparison trials had to be based on relations between the food comparison and the sample. This aspect leads us to use the term "single-comparison" to describe this procedure, although the screen also can be viewed as a com-

parison stimulus. The screen can be seen as "uninformative," however, because its function as S+ or S- is determined solely by the combination of sample and comparison stimulus with which it is displayed. When the comparison stimulus matched the sample, the subject's selection of that stimulus revealed an S+ controlling relation between the sample and the correct comparison. When the comparison stimulus did not match the sample, the subject's selection of the screen revealed an S- controlling relation between the sample and the incorrect comparison.

A major advantage of the single-comparison method is that it provides a direct demonstration of S+ and S- controlling relations. Previous studies have inferred such relations from results of procedures that replaced familiar comparison stimuli with novel ones (e.g., Dixon & Dixon, 1978; Stromer & Osborne, 1982). It was assumed but not demonstrated that novelty per se did not influence performance on S+ and S- control tests. The potential methodological difficulties of procedures that use novel stimuli have been widely acknowledged (Dixon & Dixon, 1978; Dixon et al., 1983; Farthing & Opuda, 1974; McIlvane, Withstandley, & Stoddard, 1984; Zentall et al., 1981). Because the single-comparison method introduces no novel stimuli, its use simplifies the analysis of controlling relations.

In Experiment 1, the single-comparison procedures were systematically replicated. First, adults learned both visual arbitrary and identity matching, as described above under General Method. Then, a form of single-comparison identity matching was taught. Finally, arbitrary matching was examined in the single-comparison format to test for S+ and S- controlling relations.

METHOD

Procedure

Sixty-four subjects participated. For the first subjects (1 through 8), experimental procedures covered three sessions; later subjects served in a single session.

Initial teaching: Baseline arbitrary and identity matching. All baseline trial types are listed in Table 1. Left-right position of the comparison stimuli varied unsystematically across trials. For Subjects 1 through 8, arbitrary-matching trials displayed one of three samples (CUG, VEK, and DAX) and two of three

comparison stimuli (ZID, PAF, and BEH). For Subjects 9 through 64, a fourth sample (XER) and a fourth comparison stimulus (FUA) were added.

Teaching procedures. In initial teaching, correct selections were prompted by presenting each in enhanced brightness (Figure 1). Whenever the subject typed the brighter comparison stimulus, the response was followed by reinforcing consequences. If the subject typed the darker, incorrect comparison stimulus, the program printed "Wrong—The two bright items go together." Failure to respond correctly to the brightness prompt was rare, however. Only a few subjects experienced this contingency.

A teaching sequence consisted of 50 to 60 intermixed identity- and arbitrary-matching trials. The proportion of the former to the latter trials was approximately 1:9. After six to eight prompted arbitrary-matching trials with each of the comparison stimuli, or two to three prompted identity-matching trials, the prompts were withdrawn. For the remaining trials, differential reinforcement contingencies prevailed unless the subject made two or more errors in any three successive trials. If this occurred, the prompting procedure was reinstituted for two or three more trials. Successful performance resulted in return to the differential-reinforcement-alone condition.

The initial teaching sequence was repeated if a subject achieved less than 80% accuracy on the unprompted trials in that sequence. The criterion for moving to the next phase was 90% accuracy on a 20-trial test that reviewed each possible type of arbitrary-matching trial several times. If a subject achieved an accuracy score between 80% and 90% correct on the test, the test was repeated until criterion was met.

Intermittent reinforcement was designed to prepare subjects for later test trials, which would be conducted without differential consequences. The following instructions appeared on the CRT screen to inform subjects of the change in schedules:

In this part, you will not always be told whether or not your choice was correct. In other words, you will sometimes respond correctly, but the program will not print a "Correct," as it has before. This is part of the study. If there is no "Correct," do not assume that you have responded incorrectly.

Table 1

Numeral and nonsense-syllable sample and comparison stimuli displayed in training and probe trials in Experiments 1 through 3. In each trial, one comparison stimulus was correct and one was incorrect in the presence of each sample. Across baseline trials, all comparison stimuli were displayed with all others. In single-comparison trials, the === was sometimes correct and sometimes incorrect in the presence of each sample. Probe trials in Experiment 1 assessed S+ and S- controlling relations; the indicated probe trials were presented twice each for Subjects 1 through 8 and once each for Subjects 9 through 64. For each group of subjects in Experiments 2 and 3, the first three probe trials listed were single-comparison exclusion trials, and the fourth was the undefined-stimulus probe.

Experi- ment	Sub- jects	Baseline trials			Probe trials	
		Sample	Comparison stimuli		Sample	Comparison stimuli
			Correct	Incorrect		
1	1-64	347	347,===	961,528,804,===		
		961	961,===	347,528,804,===		
		528	528,===	347,961,804,===		
		804	804,===	347,961,528,===		
	1-8	CUG	ZID	PAF,BEH	CUG	ZID,===; PAF,===
		VEK	PAF	ZID,BEH	VEK	PAF,===; BEH,===
		DAX	BEH	ZID,PAF	DAX	BEH,===; ZID,===
	9-64	CUG	ZID	PAF,BEH,FUA	CUG	ZID,===; BEH,===
		VEK	PAF	ZID,BEH,FUA	VEK	PAF,===; ZID,===
		DAX	BEH	ZID,PAF,FUA	DAX	BEH,===; FUA,===
		XER	FUA	ZID,PAF,BEH	XER	FUA,===; PAF,===
2	1-8	CUG	ZID,===	PAF,BEH,===	GAW	ZID,===
		VEK	PAF,===	ZID,BEH,===	NEY	PAF,===
		DAX	BEH,===	ZID,PAF,===	YOM	BEH,===
					FUA	XER,===
	9-14	CUG	ZID,===	PAF,BEH,FUA,===	TUP	ZID,===
		VEK	PAF,===	ZID,BEH,FUA,===	EIS	PAF,===
		DAX	BEH,===	ZID,PAF,FUA,===	FAJ	BEH,===
		XER	FUA,===	ZID,PAF,BEH,===	CLO	RAS,===
3	1-8	CUG	ZID,===	PAF,BEH,===	TAH	ZID,===
		VEK	PAF,===	ZID,BEH,===	HUW	PAF,===
		JOM	PAF,===	ZID,BEH,===	PIV	BEH,===
		DAX	BEH,===	ZID,PAF,===	ROG	KEZ,===
		WEF	BEH,===	ZID,PAF,===		
		SUY	BEH,===	ZID,PAF,===		
		YIM	BEH,===	ZID,PAF,===		

The criterion for going on to the next phase was 90% accuracy on a series of 20 intermit-
tently reinforced arbitrary-matching trials.

Single-comparison identity matching. A sys-
tematic replication of the experiment of
McIlvane, Withstandley, and Stoddard (1984)
was conducted by first establishing single-com-
parison matching performances with the stim-
uli from the subjects' identity-matching base-
lines. In the final performance, the sample was
presented along with one comparison and
another stimulus that was analogous to the
blank screen used in the original study. This
stimulus, composed of three equal signs (===),
appeared in the position formerly occupied by
the other comparison stimulus. If the number
sample and the number comparison were iden-

tical, the subject was to type that number; if
they were not identical, the subject was to type
===.

The single-comparison performance was
taught with a stimulus-shaping program in
which individual numerals of one comparison
(the correct one on some trials and the incorrect
one on others) were replaced in a series of steps
(e.g., 347, 34=, 3==, ===). If a subject did
not type the equal signs where appropriate
(i.e., if they reproduced the sample exactly),
the computer program printed the message
"Next time, type the =." All stimuli from the
subject's identity-matching baseline were used
in this shaping. The upper portion of Figure
2 shows representative stimulus-shaping and
final performance single-comparison trials.

For Subjects 1 through 8, stimulus-shaping trials were interspersed among an equal number of two-comparison arbitrary-matching trials with intermittent reinforcement. These subjects received six shaping trials at each step of the program. For Subjects 9 through 64, shaping trials were programmed in a continuous series, with only four trials at each step. The criterion for proceeding to the next phase was 95% accuracy on a 20-trial test at the final performance.

Single-comparison arbitrary-matching test. Without further training, arbitrary-matching performances were tested in the single-comparison format. Representative test trials are shown in the lower portion of Figure 2. Each nonsense syllable was displayed as a comparison stimulus along with ==. On half of the test trials (S+ probes), the comparison syllable had previously been correct in the presence of the sample; on the other half (S- probes), the comparison had previously been correct in the presence of a different sample and incorrect in the presence of the current sample.

Subjects 1 through 8 and Subjects 9 through 64 received 12 and eight probe trials, respectively (Table 1). The probes were interspersed among an equal or greater number of single-comparison identity-matching trials. Responses on single-comparison identity-matching trials were reinforced intermittently. All responses on probe trials were followed only by the intertrial interval.

RESULTS

Acquisition of identity- and arbitrary-matching baselines. Only 7 subjects required more than one exposure to prompting conditions. Overall accuracy on the last 20 unprompted trials exceeded 99%, and no subject made more than one error on these trials. When matching performances were reinforced intermittently, all subjects continued to perform with comparable accuracy.

Acquisition of single-comparison identity matching. During shaping, the 64 subjects made only five errors, and no subject made more than one error. On the 20-trial test of the final performance, overall accuracy exceeded 99%.

Single-comparison arbitrary-matching test. Figure 3 shows the number of subjects who selected the nonsense-syllable comparison on probes of S+ and S- relations (white and dark bars, respectively). For Subjects 9 through 64,

STIMULUS SHAPING			
347		961	
347 +	961 -	347 -	961 +
961		528	
528 -	96= +	528 +	96= -
528		804	
804 -	5== +	5== -	804 +
347		961	
== +	961 -	961 +	== -
PROBE			
CUG		CUG	
== (S-)	PAF	== (S+)	ZID
VEK		VEK	
ZID	== (S-)	PAF (S+)	==

Fig. 2. Representative stimulus-shaping and probe trials presented in Experiment 1. Correct and incorrect comparison stimuli in shaping trials are indicated by + and -, respectively, but the + and - did not appear on the trial displays. S+ and S- probe trials are indicated by (S+) or (S-) under the comparison stimulus whose selection would be consistent with that type of relation. Neither (S+) nor (S-) appeared on the trial displays.

all probe data are shown. For Subjects 1 through 8, data from only the first eight of the 12 probe trials are shown. (For the latter subjects, the seventh and eighth probe trials were repeats of earlier probe trials [see Table 1]; they are included simply for convenience in plotting the data.) On S+ probes, subjects almost always selected the syllable, thereby demonstrating S+ relations. On S- probes, however, subjects rarely selected the syllable, typically typing == instead; they apparently rejected the syllable because it was not related

Fig. 3. The number of subjects who selected the nonsense-syllable (rather than the ===) comparison stimulus on each of eight successive probe trials in Experiment 1. White bars, marked +, indicate performance on S+ probes. Dark bars, marked -, indicate performance on S- probes.

to the probe sample, thereby demonstrating S- relations. The only notable exceptions occurred on the first S- test trial. There, about 10% of the subjects (6 of 64) selected the syllable instead of ===. Further deviations from consistent S+ and S- relations were rare.

DISCUSSION

Our data are consistent with previous observations that human subjects can learn both S+ and S- relations in the course of two-comparison conditional-discrimination training. Because the single-comparison tests introduced no novel comparisons, the subjects' behavior could not have been influenced by stimulus novelty per se. A striking finding was the high consistency of behavior in the test trials. Across subjects, 496 probe trials tested specific S+ and S- relations for the first time; only 13 probe responses were not consistent with performance predicted by acquisition of such relations.

The finding of highly reliable S- relations in matching performances is pertinent to the potential for exclusion performances. Conditional-discrimination training does apparently establish the potential for subjects to select one comparison stimulus (the === in the present

case) by rejecting the other comparison stimulus when that comparison is not conditionally related to the sample. An exclusion test differs from an S- test in that the sample is novel. Control by novel samples, reflecting generalized rather than specific controlling relations, was investigated next.

EXPERIMENT 2

Previous studies have used two-comparison procedures to analyze exclusion performances (e.g., Dixon, 1977; McIlvane & Stoddard, 1981; Stromer, 1986). When novel samples were presented on exclusion tests, subjects virtually always selected comparison stimuli that had never before been conditionally related to a sample; we will henceforth term such comparisons *undefined*. Samples that are novel in the experimental context will also be termed undefined. Samples and comparison stimuli will be termed *defined* when they have appeared in conditional-discrimination baseline trials. The terms *defined* and *undefined* are used (a) to emphasize the functional nature (i.e., defined by prevailing reinforcement contingencies) of relations between sample and comparison stimuli and (b) to avoid the surplus meaning implied by terms such as "known," "associated with," and so on.

When subjects selected undefined comparison stimuli in studies of exclusion, that performance appeared to have been based upon explicit rejection of defined comparisons. Further, the rejection appeared to have been based on the stimulus difference between a given undefined sample and the sample previously related to the comparison stimulus to be excluded. These two interpretations could be evaluated with the single-comparison method.

The first interpretation would be supported if subjects reliably excluded in the single-comparison format, showing that exclusion did not depend upon unspecified features of two-comparison exclusion-test displays. The second interpretation could be evaluated by presenting a new trial type that was made possible by the single-comparison method. This trial type also displayed an undefined sample. However, the comparison syllable displayed with === was undefined rather than defined. Consistent rejection of the undefined comparison stimuli on these *undefined-stimulus* probes could not be based on the exclusion principle, as formulated above, because the comparison syllable had not

Table 2

Performances of subjects on probe trials in Experiment 2. Selection of the comparison syllable or === are indicated by + and -, respectively. For exclusion probes, the sample previously defined as correct for a given comparison is shown in parentheses. Undefined-stimulus probes displayed undefined syllables as both sample and comparison stimuli; therefore, no defined sample is shown.

Displayed with ===	Sample	Subjects							
		1	2	3	4	5	6	7	8
ZID	GAW (CUG)	--	--	--	--	+	--	--	--
PAF	NEY (VEK)	--	--	--	--	+	--	--	--
BEH	YOM (DAX)	--	--	--	--	--	--	--	--
FUA	XER	--	+-	+	--	+-	--	--	++
		Subjects							
		9	10	11	12	13	14		
ZID	TUP (CUG)	---	---	---	---	---	---		
PAF	EIS (VEK)	---	---	---	---	---	---		
BEH	FAJ (DAX)	---	---	---	---	---	---		
CLO	RAS	---	---	++	---	---	---		

been conditionally related to a sample. Inconsistent selections would be consistent with this principle.

METHOD

Subjects

Subjects were the first 8 subjects from Experiment 1 and 6 subjects selected unsystematically from the remainder.

Procedure

This experiment was conducted after the single-comparison arbitrary-matching tests in Experiment 1. Subjects 1 through 8 were tested in a separate session, usually conducted the next day. Tests for Subjects 9 through 14 occurred in the same session. For all subjects, the tests were conducted after a 20 to 30 trial review of the single-comparison arbitrary-matching baseline with intermittent reinforcement.

Exclusion and undefined-stimulus probes. On exclusion probes, the samples were undefined and the comparison syllables were defined (see Table 1); selections of === would be consistent with exclusion. On undefined-stimulus probes, both the sample and the comparison syllable were undefined (FUA-XER for Subjects 1 through 8; CLO-RAS for Subjects 9 through 14). All probes were interspersed among single-comparison arbitrary-matching baseline trials with intermittent reinforcement. No reinforcers followed any probe selections. For Subjects 1 through 8, each probe was pre-

sented twice; for Subjects 9 through 14, it was presented three times.

RESULTS AND DISCUSSION

All subjects' baseline performances were almost errorless both before and during testing. Table 2 shows the test results. On exclusion probes, 13 of the 14 subjects selected === on every trial, apparently excluding the defined comparison stimulus. Subject 5 responded inconsistently. She selected === on the first three exclusion probes and the undefined nonsense syllable on the first defined-stimulus probe; after the latter trial, she selected the defined comparison syllable on two of the three remaining exclusion probes and === on the other undefined-stimulus probe. At the end of her test session, this subject spontaneously told the experimenter that she had intentionally made the "wrong" responses on some trials "just to see what would happen."

On undefined-stimulus probes, subjects' performances differed. Nine of the 14 subjects selected === on all trials. The undefined comparison nonsense syllable was selected less frequently. Subject 8 selected it on both trials and 4 others selected it on one trial each. Overall, therefore, there was a marked preference for selecting === on the undefined-stimulus probes. Most deviations from this preference occurred in Subjects 1 through 8 who had (a) one fewer defined sample-comparison relation in their baselines and (b) at least 1 day between initial training and the probe sessions. It is not

clear whether these differences in experimental history influenced probe performance.

In summary, these data suggested that rejection of a defined comparison stimulus was a sufficient basis for exclusion performance. Almost all subjects rejected such stimuli in the presence of undefined samples. Although performance on the undefined-stimulus probes was not consistent across subjects, the fact that most subjects rejected undefined comparison stimuli in the presence of undefined samples suggested the need for further analysis of the stimulus control basis for exclusion performances.

EXPERIMENT 3

Dixon (1977) suggested that exclusion performances might be based on a generalized controlling relation that might be described as "any [comparison] stimulus that is correct for [Sample A] is incorrect for [any other sample]" (p. 441). Dixon *et al.* (1983) have proposed a more general form, termed the S^X rule, that predicts exclusion if

"the [subject learns] that in the presence of a standard stimulus *a*, it [is] appropriate to select the comparison stimulus *A*, and if some other standard stimulus besides *a* occurs, it [is] not appropriate to select *A*." (p. 103)

To establish such a relation, it would seem that training contingencies must maintain consistent "one-to-one" sample-comparison relations; no comparison can be related to more than one sample. Stromer (1986) has developed this point explicitly, offering the example of a subject who learns to select a given comparison stimulus, B, conditionally upon three different samples—A, C, and D. After such training, Stromer suggested that the subject could learn that B was correct in a variety of contexts, and might therefore tend to select B when the sample was novel. Thus, he suggests that exclusion "may be partially attributable to a subject's learning one-to-one conditional relations between stimuli" (p. 70).

Experiment 3 assessed whether "one-to-one" relations were in fact prerequisites for exclusion. Subjects learned to select (a) one comparison stimulus in the presence of two samples (a "two-to-one" relation), and (b) another comparison stimulus in the presence of four samples (a "four-to-one" relation). Selection of a third comparison stimulus continued to be reinforced in the presence of only one sample (a "one-to-one" relation).

METHOD

Subjects

Subjects 1 through 4 and 6 through 8 participated in this experiment.

Procedure

Teaching "two-to-one" and "four-to-one" relations. These relations were established by elaborating matching performances that had been learned previously (Table 1, lower portion). In the "two-to-one" relation, subjects learned to select PAF conditionally upon not only VEK but also JOM as samples. In the "four-to-one" relation, subjects learned to select BEH conditionally upon WEF, SUY, and YIM in addition to DAX. In the "one-to-one" relation, ZID continued to be selected conditionally upon only CUG as sample.

Subjects were returned to the two-comparison format and reinforcement of every correct selection during a 70-trial training sequence. Brightness cues were again used to prompt new performances. The introduction of new sample-comparison relations was staggered; subjects received three to five prompted trials for each new relation before the next new relation was introduced. The criterion for moving on was 95% accuracy on a 20-trial test that included at least two arbitrary-matching performances with each sample. Next, a 20-trial test verified 95% accurate single-comparison arbitrary matching with these elaborated baseline performances under intermittent reinforcement conditions.

Exclusion and undefined-stimulus probes. Finally, exclusion performances were tested in the context of the elaborated single-comparison arbitrary-matching baseline. To minimize potential influence of the subjects' Experiment 2 probe experience on the current probe series, undefined stimuli used on exclusion and undefined-stimulus probes differed from those presented in Experiment 2 (Table 1, lower right portion). Again, probes were interspersed among baseline trials with intermittent reinforcement, and the ratio of probes to baseline trials was approximately 1:3. Each probe trial was presented twice, and no reinforcer followed any response on those trials.

RESULTS

Learning "two-to-one" and "four-to-one" relations. All subjects learned the baseline performances in one teaching sequence, and per-

Table 3

Performances of subjects on probe trials in Experiments 3 and 4. Selections of the comparison syllable or === are indicated by + and -, respectively. On exclusion probes, the samples for which a given comparison stimulus had previously been defined as correct are shown in parentheses. The undefined-stimulus probe displayed undefined syllables as both sample (ROG) and comparison stimuli (KEZ). Prior to the probes, Subjects 67 through 70 had also been exposed to two undefined comparison stimuli in the presence of the VEK sample (see text for details).

Dis- played with ===	Sample	Experiment 3							Experiment 4						
		Subjects							Subjects						
		1	2	3	4	6	7	8	65	66	67	68	69	70	
ZID	TAH (CUG)	--	--	--	--	--	--	-+	--	--	--	--	--	--	
PAF	HUW (VEK) (JOM)	--	--	--	--	--	--	+ -	--	--	--	--	--	--	
BEH	PIV (DAX) (WEF) (SUY) (YIM)	--	--	--	--	--	--	+ -	--	--	--	--	--	--	
KEZ	ROG	--	--	--	--	--	--	++	--	--	+ -	--	--	--	

formed with 99% accuracy when tested in the single-comparison format. Baseline performances remained highly accurate thereafter.

Exclusion and undefined-stimulus probes. Table 3 (left portion) shows the test results. Typically, the training of two-to-one and four-to-one relations did not change the subjects' performances on the exclusion probes; 6 of the 7 subjects always selected ===, as they had done in Experiment 2. Subject 8, however, responded inconsistently on the probes. On several trials, she chose the comparison syllable instead of ===. The previously observed preference for selecting === on undefined-stimulus probes was even more pronounced in this experiment. Only Subject 8 selected the comparison syllable.

DISCUSSION

Despite the development of two-to-one and four-to-one relations, 6 of 7 subjects continued to exclude defined comparison stimuli in the presence of undefined samples. Subject 8, however, showed disruption of exclusion performance, apparently as a result of the more complex sample-comparison relations. Why she differed from the others was not immediately clear, but data collected after the probes suggested a possible explanation. These data, collected in an exploratory investigation, were limited but possibly meaningful. We summarize them briefly here because of their implications for future research.

Stimulus equivalence probes. Several studies

have shown that training contingencies like those in Experiment 3 may establish relations of equivalence among conditionally related stimuli (cf. Sidman, 1986). The subjects' two-to-one and four-to-one relations permitted tests for equivalence by examining potentially emergent relations among the training samples. An illustrative test trial was as follows: DAX was the sample, and YIM and JOM were comparison stimuli. Both DAX and YIM had previously controlled selection of the comparison stimulus BEH. This trial constituted a combined test for symmetry and transitivity of the DAX:BEH and YIM:BEH relations. This and other similar trial types would evaluate two of the three logical requirements necessary to document equivalence relations (Sidman & Tailby, 1982); the third, reflexivity, was verified earlier when all subjects demonstrated generalized identity matching in initial training.

There were 14 potentially emergent relations among the training samples. We evaluated six of them, presenting each of the six samples used in two-to-one and four-to-one training as the sample on an unreinforced probe trial of the type described above. All symmetrical relations were also tested, again once each; on these trials, the former comparison stimuli served as samples and the former samples as comparison stimuli (e.g., BEH as sample with DAX and JOM as comparison stimuli; see Table 1, bottom). Of the 6 subjects who had demonstrated perfect exclusion, 5 passed all

tests for equivalence. Data from Subject 6 were inconclusive but interesting. She passed all symmetry tests. When relations among the former samples were tested, however, she always typed ==, not displayed on those trials, rather than one of the available comparison syllables. Subject 8 was the only subject who conclusively failed the equivalence tests and also was the only one whose exclusion performance was affected. Although these results are far from conclusive, they do suggest a relationship between acquisition of stimulus equivalence and performance on exclusion trials, particularly after subjects learn multiple samples with a single comparison. We will return to this point after additional data are presented.

EXPERIMENT 4

All subjects in Experiment 3 had been previously exposed to the exclusion probes in Experiment 2, after learning one-to-one baseline matching relations and prior to learning two-to-one and four-to-one relations. The probe history may have influenced performance in Experiment 3 as follows: If exclusion depended on one-to-one relations between samples and comparison stimuli, then development of two-to-one and four-to-one relations would eliminate the basis for exclusion. These subjects, however, may have had another basis for selecting == on the exclusion probes; they had recently emitted that behavior on similar probes in Experiment 2. Therefore, Experiment 4 examined exclusion performances after developing two-to-one and four-to-one relations in subjects who had no prior probe history.

Experiment 4 also evaluated another variable that might affect exclusion performance. Previously, new conditional relations were always taught directly. We never before required subjects to relate an undefined and defined stimulus without first providing a brightness cue to prompt the relation. Confronted with a defined comparison stimulus and an undefined sample on an exclusion probe, subjects in Experiment 3 may have rejected a potential relation between these two stimuli simply because no prompt stimulus was provided. The present experiment used a procedure that might reduce the likelihood of this form of control as follows: On two-comparison teaching trials, we displayed a defined comparison stimulus

with an undefined one; the sample was also defined, but not in relation to the defined comparison stimulus that was available for selection. On this type of trial, the subject could display a form of S- control by rejecting the nonmatching defined comparison stimulus and selecting the undefined one. Analogous behavior has been observed in several studies (Dixon & Dixon, 1978; McIlvane & Stoddard, 1985; Stromer & Osborne, 1982), although the phenomenon is not always evident in all subjects (Dixon *et al.*, 1983; McIlvane, Munson, & Stoddard, 1987). If the subject did select the undefined stimulus, however, it would show that he or she related a defined and undefined stimulus even though the relation had never been prompted with a brightness cue.

METHOD

Subjects

Six experimentally naive subjects served. They did not participate in Experiment 1 and are therefore numbered 65 through 70.

Procedure

The baseline training was the same as in Experiment 3 (Table 1, lower left). After training, 4 subjects (67 through 70) also received the following trials:

VEK:YUN,ZID
VEK:KIB,BEH

In the diagrams, samples are shown to the left of the colon, and the correct (underscored) and incorrect comparison stimuli to the right. The sample VEK had previously controlled selection only of PAF. ZID and BEH were defined comparison stimuli selected conditionally upon one sample and four samples, respectively (see Table 1, lower left). YUN and KIB were undefined stimuli. Four trials of each type were presented, interspersed among baseline trials with continuously reinforced arbitrary matching. Selecting the undefined comparison stimulus on the first opportunity to do so would suggest an S- relation between VEK and the nonmatching defined comparison stimuli. Selection of the undefined comparison stimulus was followed by a reinforcer. If the defined comparison stimulus was selected, "Wrong" was printed on the screen and there was no reinforcer display. This combination of a presumably punishing consequence and no reinforcer was expected to increase the likelihood

that on the next such trial the defined (VEK) and undefined (YUN or KIB) syllables would become related. This history might make it more likely that subjects would later relate undefined and defined syllables when they were given an opportunity to do so on exclusion probes.

Afterward, subjects learned single-comparison matching via the stimulus-shaping program. In contrast to the training in Experiment 1, single-comparison arbitrary-matching performances were reinforced intermittently from the outset. Finally, exclusion and undefined-stimulus probes were conducted. Procedures were identical to those in Experiment 3 (Table 1, lower right).

RESULTS

Subjects made few or no errors acquiring the baseline performance of arbitrary matching and maintained their performances under intermittent reinforcement and on single-comparison arbitrary-matching trials. On trials that displayed VEK as the sample, ZID or BEH as the defined comparison stimulus, and YUN or KIB as the undefined comparison stimulus, 2 of 4 subjects selected the undefined stimulus on all eight trials. The other 2 subjects selected ZID on the first such trial. After this response was followed by "Wrong," both subjects then selected the undefined comparison stimulus on each of the remaining seven trials. Thus, all 4 subjects related a defined stimulus to an undefined one, even without a brightness prompt to cue the performance.

The right portion of Table 3 shows performance on the exclusion and undefined-stimulus probes. All 6 subjects always selected === in response to the undefined sample on the exclusion probes, as well as on all but one of the undefined-stimulus probes. Thus, variations in training had no apparent impact on probe performance; the results were virtually the same as those in Experiment 3.

DISCUSSION

Together, data from Experiments 3 and 4 show that teaching subjects to select individual comparison stimuli conditionally upon two or more samples will not prevent exclusion performances. These findings are not consistent with previous accounts of exclusion, which explicitly or implicitly stressed the importance of a previously established baseline of one-to-one

sample-comparison relations (Dixon, 1977; Dixon et al., 1983; McIlvane & Stoddard, 1981; Stromer, 1986). Moreover, no previous account of exclusion predicts consistent rejection of an undefined comparison stimulus in the presence of an undefined sample—the predominant outcome on the undefined-stimulus probes. Perhaps the most parsimonious account of these data is to presume differing bases for selection on exclusion and undefined-stimulus probes. The latter trial type displayed an undefined comparison, which had no reinforcement history, with ===, which did have a reinforcement history (albeit only with defined samples). Conceivably, subjects might have selected === merely on this basis. On the exclusion probes, however, both the defined comparison and === had reinforcement histories; selections of === must have been based on conditional control by the sample.

Another interpretation of these data, substantially different from previous accounts of exclusion, is that a form of generalized conditional control occurred that one might verbally describe as "select a given comparison stimulus if and only if that stimulus is functionally related to the sample." Such a generalized controlling relation predicts selection of === on exclusion probes, regardless of the number of samples related to each comparison stimulus in the subjects' conditional-discrimination baselines. It also predicts rejection of undefined comparison stimuli in the presence of undefined samples, because a functional relation involving these two stimuli had not been established. Such generalized control was also suggested by some subjects' comments in informal debriefing interviews. If such a relation was responsible for our results, further attempts to prevent exclusion by teaching ever more complex arbitrary matching relations might prove futile. Hence, in the final experiment we took a different approach.

EXPERIMENT 5

In Experiment 5 we examined the effects of punishing exclusion performances and reinforcing selections of defined comparison stimuli conditionally upon undefined samples. The primary goal was to examine further the relationship between performance on exclusion probes and that on undefined-stimulus probes. Because most subjects selected === on both,

Fig. 4. Schematic of procedures used with Subjects 15 through 30 in Experiment 5. Each major training or test phase is represented by a rectangle. Critical trial types presented in the first, third, and fifth phases are enclosed in the rectangles. Trial types presented in the second and fourth phases are omitted for clarity. Stimuli indicated by + and - were S+ and S- on training trials, respectively, but + and - did not appear on the displays.

it was possible that the two performances shared a common stimulus control basis. For example, subjects may have selected === whenever the sample was novel, perhaps because such samples were not functionally related to comparison stimuli. Procedures used so far had not demonstrated that subjects discriminated exclusion probes from undefined-stimulus probes. Experiment 5 sought to give subjects a training history that might foster such a discrimination.

Figure 4 illustrates the procedures. In initial training, subjects acquired baseline performances of two-comparison and single-comparison arbitrary matching. (Single-comparison baseline trials are not shown in Figure 4; rectangular boxes represent phases in which these trials appeared.) Then, a new two-comparison trial type was presented (shown in the middle portion of Figure 4). On these trials, the sample and one comparison stimulus were undefined, and the other stimulus was defined. These were two-comparison exclusion trials. We expected highly reliable selections of the undefined comparison stimulus on the first of these trials, because subjects could exclude the defined comparison stimulus in the presence of the undefined sample (cf. McIlvane & Stoddard, 1981; Stromer, 1986). When the subjects did so, no reinforcer followed; instead, "Wrong" was printed on the screen. Together, these consequences were expected to decrease the likelihood of exclusion performances and to increase the likelihood that subjects would select defined comparison stimuli conditionally upon undefined samples; the latter behavior was then reinforced. Pilot data suggested that the presentation of "Wrong" would lead to faster suppression of undefined comparison selections than would simple extinction alone. For this reason, and for convenience in exposition, we will call these trials *punishment* trials, recognizing that extinction and reinforcement could occur on those trials.

Only two defined comparison stimuli from the arbitrary-matching baseline, PAF and BEH, were displayed on punishment trials. On these trials, subjects could learn, specifically, to select PAF and BEH conditionally upon undefined samples, and, more generally, that undefined samples and undefined comparison stimuli were not to be related. They could also learn a generalized performance involving defined comparison stimuli—to select

any defined comparison stimulus in the presence of an undefined sample. To examine these possibilities, we presented the single-comparison probe trials shown in the lower part of Figure 4.

On single-comparison exclusion probes, three defined comparison stimuli, ZID, PAF, and BEH, were displayed along with === in the presence of new undefined samples, TUP, EIS, and FAJ, respectively. Only PAF and BEH had been displayed as comparison stimuli on the punishment trials; ZID had never been involved in punishment contingencies. Undefined-stimulus probes displayed the undefined comparison stimulus, RAS, along with ===, in the presence of the undefined sample, CLO. These probe trials were designed to answer three questions: First, would the two-comparison punishment history lead subjects to select the defined stimuli PAF or BEH in the presence of an undefined sample, even when the alternative was === rather than an undefined comparison stimulus? Second, would they select ===, rejecting RAS, an undefined comparison stimulus (with no punishment history), in the presence of an undefined sample? Third, would they select the defined stimulus ZID (with no punishment history) or === in the presence of an undefined sample? The answer to the last question provided the critical data, as follows. Because ZID had not appeared on punishment trials, subjects could not have learned explicitly to select it in the presence of an undefined sample. Therefore, its selection in the presence of such a sample would suggest the acquisition of a generalized controlling relation. The relation would specify selection rather than exclusion of defined comparison stimuli in the presence of undefined samples. Further, if subjects reliably selected === on the undefined-stimulus probes, this would demonstrate that a comparison stimulus had to be defined in order for it to be related to an undefined sample. Discrimination of single-comparison exclusion probes from undefined-stimulus probes would be shown.

Another question asked in Experiment 5 (and related to the third question above) was whether exclusion performances in general could be prevented by a restricted punishment procedure that involved only one of the defined comparisons. If exclusion did reflect a generalized controlling relation, then punishment of any performance controlled by that relation

Table 4

Subjects' trial-by-trial performances when undefined samples and comparison stimuli were displayed with the defined comparison(s) PAF and BEH (Subjects 15 through 30) or PAF alone (Subjects 31 through 38). Shown below the trial number is the defined comparison stimulus that was displayed on that trial; P and B stand for PAF and BEH, respectively. Selections of the defined and undefined comparison stimuli are shown by + and U, respectively.

Subject	Trial number											
	1	2	3	4	5	6	7	8	9	10	11	12
	B	B	P	P	B	P	B	B	P	B	P	P
15	U	U	+	+	+	+	+	+	+	+	+	+
16	U	U	U	+	+	+	+	+	+	+	+	+
17	U	U	+	+	+	+	+	+	+	+	+	+
18	U	U	+	+	+	+	+	+	+	+	+	+
19	U	+	U	U	+	+	+	+	+	+	+	+
20	U	+	+	+	+	+	+	+	+	+	+	+
21	U	U	+	+	+	+	+	+	+	+	+	+
22	U	U	+	+	+	+	+	+	+	+	+	+
23	U	U	+	+	+	+	+	+	+	+	+	+
24	U	U	+	+	+	+	+	+	+	+	+	+
25	+	+	U	U	U	U	U	+	U	U	+	+
26	U	+	+	+	+	+	+	+	+	+	+	+
27	U	U	+	+	+	+	+	+	U	+	U	+
28	U	+	+	+	+	+	+	+	+	+	+	+
29	U	+	+	+	+	+	+	+	+	+	+	+
30	U	+	+	+	+	+	+	+	+	+	+	+
	P	P	P	P	P	P	P	P				
31	U	+	+	+	+	+	+	+				
32	U	+	+	+	+	+	+	+				
33	U	U	+	+	+	+	+	+				
34	U	U	+	+	+	+	+	+				
35	U	U	+	+	+	+	+	+				
36	U	U	U	U	U	+	+	+				
37	U	+	+	+	+	+	+	+				
38	U	+	+	+	+	+	+	+				

might affect all others. On the other hand, suppression of exclusion performance might be specific to the comparison stimulus present at the time of punishment. Punishment in more than one context might be required to alter the generalized relation (cf. Holland, Solomon, Doran, & Frezza, 1976; Hovland, 1952).

METHOD

Subjects

Twenty-four subjects served. All had previously participated only in Experiment 1.

Procedure

Preliminary training was described in Experiment 1. All learned the arbitrary-matching baseline performances diagrammed in the upper portion of Figure 4, ultimately displaying accurate single-comparison arbitrary-matching performances under conditions of intermittent reinforcement.

Punishment of exclusion performances. Next, the two-comparison arbitrary-matching format and continuous reinforcement conditions were conducted for 36 trials. Subjects 15 through 30 received the 12 punishment trials shown in Figure 4, once each; the defined comparison stimulus was PAF on half of these trials and BEH on the remainder. Subjects 31 through 38 received the following eight punishment trials, also once each; the defined comparison stimulus was always PAF:

NEY:PAF,ZAL; YOT:PAF,HIW;
JOM:PAF,LUR; KUK:PAF,DEP
BIP:PAF,REZ; ARZ:PAF,KOH;
MUV:PAF,LYD; GAW:PAF,TAZ

Samples are shown to the left of the colon, with the correct (underscored) and incorrect comparison stimuli to the right. All punishment trials were interspersed among arbitrary-matching baseline trials.

Single-comparison exclusion and undefined-stimulus probes. Finally, the single-comparison format with intermittent reinforcement was presented. After verifying 95% or greater accuracy on baseline performance, the probe trials shown in the lower portion of Figure 4 were interspersed among 20 baseline trials. Each type of trial was presented three times.

RESULTS

Punishment of exclusion. The top portion of Table 4 shows the trial-by-trial performance of Subjects 15 through 30 during the punishment trials. Overall, the procedure proved highly effective in suppressing exclusion performance. Fifteen of the 16 subjects selected the undefined comparison stimulus on the first trial (apparently excluding it); these selections were punished. By the fifth trial, most subjects selected defined comparison stimuli, either BEH or PAF, on every trial. Two subjects, however, differed from the others. Subject 25 selected the defined comparison on the first two trials. Although reinforcers followed each selection, she then selected the undefined comparison on the next eight trials before returning to select defined comparisons once again. Subject 27's performance was typical of other subjects until Trial 9, when her behavior became inconsistent.

The lower portion of Table 4 shows the performances of Subjects 31 through 38; all selected the undefined comparison stimulus on the first trial, and these responses were punished. Over trials, selection of the undefined comparison stimulus occurred with decreasing frequency. By the sixth trial, all subjects selected the defined comparison stimulus, PAF, in the presence of undefined samples. Performances of both groups were virtually errorless on two-comparison and single-comparison arbitrary-matching baseline trials.

Single-comparison exclusion and undefined-stimulus probes: Subjects 15 through 30. The left portion of Table 5 shows the outcome of these tests for 14 of the 16 subjects. Data are not shown for Subjects 25 and 27, who did not respond consistently on the last few trials of the punishment procedure. Subjects rarely selected === on the single-comparison exclusion probes. Seven subjects selected the defined comparison stimulus in all exclusion trials; they never excluded. Six others selected the defined comparison stimulus in most but not all trials.

Subject 18 selected BEH in the presence of the undefined sample FAJ, but selected === on the remaining two types of exclusion trials.

The performance of Subjects 15 and 19 was highly unusual. During the probes, both subjects developed a pronounced response bias that led to near-chance accuracy scores on single-comparison arbitrary-matching baseline trials, contrasting sharply with their previous errorless performances on such trials. Following the first and third exclusion probes, respectively, these subjects selected the comparison syllable on every trial, baseline or probe, except for the undefined-stimulus probes in which they consistently selected ===. The same type of bias was also observed during the last few trials of Subject 30's probe sequence, after most of the probe data had been collected. All other subjects maintained their baselines during the probes. On the undefined-stimulus probes, all subjects selected ===.

Single-comparison exclusion and undefined-stimulus probes: Subjects 31 through 38. The right portion of Table 5 shows that for 6 of 8 subjects the effects of the punishment procedure were restricted to the comparison stimulus, PAF, that had been displayed on punishment trials. Although these subjects did not exclude PAF, they apparently excluded ZID and BEH; the effects of punishment did not generalize to the latter stimuli. Subject 32, however, showed complete generalization. Only with this subject did punishment of exclusion in one context suffice to eliminate it in others. Again, all subjects selected === on each undefined-stimulus probe. All maintained their baselines during the probes.

DISCUSSION

Exclusion performances were markedly affected by the punishment procedure. Punishment in more than one context, however, was necessary to consistently produce generalized "nonexclusion." When exclusion of only one of the comparisons was punished, most subjects stopped excluding that comparison stimulus but did exclude the others. When exclusion of two comparisons was punished, however, subjects tended not to exclude any defined comparison stimuli. These results contribute to the growing body of data showing that generalization is highly dependent on the training procedures that are used to establish discrimination (see Stoddard & Sidman, 1971b;

Table 5

Performances of subjects on probe trials in Experiment 5. Selections of the comparison syllable or === are indicated by + and -, respectively. For each group, the first three data rows show behavior on exclusion probes; the fourth row shows behavior on the undefined-stimulus probes.

Displayed		Subjects									
Comparison	Sample	15	16	17	18	19	20	21	22	23	24
ZID	TUP	+++	+++	-++	---	+++	+++	+++	++-	+++	+++
PAF*	EIS	+++	+++	-++	---	+++	+++	+++	+++	+++	+++
BEH*	FAJ	+++	+++	+++	+++	+++	+++	+++	+++	++-	++-
CLO	RAS	---	---	---	---	---	---	---	---	---	---

Stokes & Baer, 1977). Further, it was clear that exclusion and nonexclusion—different and even seemingly conflicting forms of stimulus control—can coexist in the same discriminative baseline (cf. Ray & Sidman, 1970; Stoddard & Sidman, 1971a).

For subjects who showed generalized non-exclusion, our procedures were successful in demonstrating discrimination of single-comparison exclusion probes and undefined-stimulus probes. Establishing this discrimination, however, may require the explicit steps taken here or similar ones. Subjects may otherwise tend to reject any comparison stimulus that is not functionally related to the sample, as our earlier data suggested.

Two aspects of the present data are puzzling. The first is the somewhat inconsistent performance of some subjects on exclusion probes. On one hand, the punishment procedure seemed to influence most subjects' behavior. However, occasional selections of === continued to occur. Perhaps the inconsistencies resulted because probe responses were neither reinforced nor punished, unlike responses during the punishment procedure. One feasible option would have been to conduct probes during extinction (Sidman, Kirk, & Willson-Morris, 1985). Under such conditions, there would be no disparity in reinforcers obtained on baseline and probe trials, and probe behavior might have been more consistent.

The second puzzling finding was the response bias displayed by Subjects 15 and 19, and, to a lesser extent, by Subject 30. During the probes, these subjects selected the comparison syllable on every trial, unless that stimulus was undefined. Thus, it appeared that the probes substantially altered or obliterated control by the samples, despite the fact that control had been reliable throughout training. The loss of sample control was specific to the single-

comparison trials, however. When the subjects were returned to intermittently reinforced two-comparison trials, all performed without error. For these subjects, some feature(s) of the cumulative history apparently led to a highly specific breakdown of conditional control.

GENERAL DISCUSSION

In the present study we examined the performance of normally capable adults on several conditional-discrimination procedures. Experiment 1 showed that conditional discriminations could be based on both (a) S+ relations between the sample and the comparison stimulus that was selected and (b) S- relations between the sample and the comparison stimulus that was rejected. The single-comparison test procedure eliminated the potential influence of stimulus novelty per se in determining performance. Later experiments used the single-comparison procedure to examine exclusion performances. Experiment 2 showed that exclusion need not depend on the simultaneous availability of defined and undefined comparison stimuli; rejection of a defined comparison stimulus in the presence of an undefined sample was sufficient. Experiments 3 and 4 showed that development of two-to-one and four-to-one sample-comparison relations typically did not prevent or disrupt exclusion. Exclusion, therefore, did not appear to depend on one-to-one relations between sample and comparison stimuli.

A question for future research is whether exclusion performances of less capable subjects (i.e., young children or mentally retarded individuals) would be affected if individual comparison stimuli were related to more than one sample. So far, only one study has addressed this issue. McIlvane and Stoddard (1985) recently demonstrated apparent exclusion per-

Table 5 (Continued)

Comparisons involved in the punishment procedure are shown by asterisks (only PAF for Subjects 31 through 38). Arbitrary-matching baselines of Subjects 15 and 19 deteriorated to near chance accuracy in the course of probing (see text).

Subjects											
26	28	29	30	31	32	33	34	35	36	37	38
+++	+++	+++	---	---	+++	---	---	---	---	---	---
+++	+++	+++	+++	+++	+++	---	+++	+++	+++	+++	+++
+++	+++	+++	+++	---	+++	---	---	---	---	---	---
---	---	---	---	---	---	---	---	---	---	---	---

formances in a profoundly retarded subject who had learned to select individual comparison stimuli conditionally upon two different samples. This subject, however, had extensive prior experience with exclusion procedures, so generality of the finding needs to be assessed.

The single-comparison test procedure used in the present study permitted us to ask another basic question about exclusion, specifically whether a comparison stimulus had to be defined (that is, to have been previously established in relation to some sample stimulus) to be excluded. Results of undefined-stimulus probes showed that subjects were likely to reject the nonsense-syllable comparison stimuli, even though those stimuli had never been defined in relation to any sample. This finding was not predictable from accounts of exclusion that stress the importance of a previously established relation between the comparison stimulus to be excluded and a sample different from that presented on the exclusion probe trial (Dixon, 1977; Dixon et al., 1983; McIlvane & Stoddard, 1981).

Data from the present study suggest the possibility of a more general controlling relation: Subjects may exclude a given comparison stimulus whenever there is no previously established sample-S+ relation between that stimulus and the sample currently available. Here, sample-S+ relations must be defined to accommodate not only those relations taught directly but also those that emerge without explicit training through stimulus-class formation, as suggested by the results of Experiment 3. How might a subject learn, on a general basis, to reject any comparison stimulus that is not in the same class as the sample? Conceivably, such learning may have occurred outside the experiment. All subjects in this study had extensive histories with structured learning situations, the contingencies of which

were at least superficially similar to those used here. A relevant extraexperimental history, therefore, is one possibility. Performances learned within the experimental context, however, may provide another route to acquiring the more general relation. During conditional-discrimination training, subjects learn sample-S- controlling relations—to reject several specific defined comparison stimuli conditionally upon several specific nonmatching defined samples. Might they additionally learn to reject *any* comparison stimulus if it does not “match” the sample currently presented, an abstracted relation based on experience with several exemplars (cf. Holland et al., 1976)? According to this line of reasoning, teaching subjects to select individual comparison stimuli conditionally upon several different samples would, if anything, encourage exclusion; such teaching would necessarily increase the number of specific sample-S- relations in the baseline, thereby providing further examples on which to base a more general relation.

In the present experiments, the subjects' behavior in exclusion and undefined-stimulus probes suggested that they had learned to reject any comparison stimulus that was not related to the sample. How do these findings relate to performances observed on the more typical two-comparison exclusion trials? Such trials display a defined comparison stimulus with an undefined one in the presence of an undefined sample. Because the sample is undefined, neither comparison stimulus is related to it; according to the above analysis, both should be rejected. Why then do subjects almost always select the undefined comparison stimulus on their first opportunity to do so?

One might approach this question by modifying the two-comparison exclusion trial to add a third alternative, analogous to the === used in the present experiments. For example,

the exclusion probe trial might display a defined comparison stimulus, an undefined comparison stimulus, and === in the presence of an undefined sample. If subjects tend to reject both the defined and the undefined comparison stimuli in this three-choice context, the results would tend to support the analysis presented above. Further, these results would show that selection of an undefined comparison stimulus in a two-comparison exclusion trial may depend critically on the forced choice between it and the defined comparison stimulus. In the two-comparison context, selecting an undefined comparison stimulus conditionally upon an undefined sample may be due, in some cases, to their common property of novelty in the experimental context (Dixon et al., 1983; McIlvane, Bass et al., 1984). Moreover, as previously noted, subjects sophisticated in learning conditional discriminations may have previously abstracted rules that determine initial selections when none of the stimuli available has been explicitly defined as "correct" for a given display (cf. Dixon, 1977). Subjects might have selected undefined comparison stimuli conditionally upon undefined samples, for example, because those two stimuli were the only ones whose class membership had not yet been established.

There remains the question of whether the instructions used prior to the original conditional-discrimination training influenced the findings. The instructions were a practical solution to the problems of working with subjects who were not available for extended training and testing. With normally capable adult subjects, however, stating that particular samples "go with" particular comparison stimuli may be especially likely to produce equivalence class formation. If results obtained in exclusion procedures are influenced by whether or not such classes form, as the limited data from Experiment 3 suggest, such instructions should be avoided in the future. By hastening class development, instructions may make it especially difficult to disrupt exclusion by developing two-to-one and four-to-one matching relations.

Disrupting or preventing exclusion performance—other than by teaching explicit instances of nonexclusion, as with the punishment and differential reinforcement procedures used in the present study—may require provision of an intricate or unusual environmental history. Exclusion phenomena are among the

most reliable findings in the study of human operant behavior, considering both the variety of populations studied and the consistency of results across individuals. Behavioral prerequisites for such phenomena may be quite primitive. In research on child development, for example, performances resembling exclusion have been reported to emerge in the second year (cf. Kagan, 1981; Vincent-Smith, Bricker, & Bricker, 1974), coinciding with initial acquisition of verbal behavior. What variables account for the first instances of such behavior? To answer this question will require laboratory studies of the developing child and of more limited retarded individuals than studied heretofore. If these studies are conducted, we speculate that the relevant variables may be closely intertwined with and possibly overlap those that control equivalence class formation. Research with nonhumans is also necessary to define the limiting case.

REFERENCES

- Baer, D. M., Peterson, R. F., & Sherman, J. A. (1967). The development of imitation by reinforcing behavioral similarity to a model. *Journal of the Experimental Analysis of Behavior*, *10*, 405-416.
- Carter, D. E., & Werner, T. J. (1978). Complex learning and information processing by pigeons: A critical analysis. *Journal of the Experimental Analysis of Behavior*, *29*, 565-601.
- Cumming, W. W., & Berryman, R. (1965). The complex discriminated operant: Studies of matching-to-sample and related problems. In D. I. Mostofsky (Ed.), *Stimulus generalization* (pp. 284-330). Stanford, CA: Stanford University Press.
- D'Amato, M. R., Salmon, D. P., & Colombo, M. (1985). Extent and limits of the matching concept in monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes*, *11*, 35-51.
- Dixon, L. S. (1977). The nature of control by spoken words over visual stimulus selection. *Journal of the Experimental Analysis of Behavior*, *7*, 433-442.
- Dixon, M., & Dixon, L. S. (1978). The nature of standard control in children's matching-to-sample. *Journal of the Experimental Analysis of Behavior*, *30*, 205-212.
- Dixon, M., Dixon, L. S., & Spradlin, J. E. (1983). Analyses of individual differences of stimulus control among developmentally delayed children. In K. D. Gadow & I. Bialer (Eds.), *Advances in learning and behavioral disabilities* (Vol. 2, pp. 85-110). Greenwich, CT: JAI Press.
- Farthing, G. W., & Opuda, M. J. (1974). Transfer of matching-to-sample in pigeons. *Journal of the Experimental Analysis of Behavior*, *21*, 199-213.
- Goldiamond, I. (1966). Perception, language, and conceptualization rules. In B. Kleinmuntz (Ed.), *Problem solving: Research, method, and theory* (pp. 183-224). New York: Wiley.

- Guess, D., & Baer, D. M. (1973). An analysis of individual differences in generalization between receptive and productive language in retarded children. *Journal of Applied Behavior Analysis*, *6*, 311-329.
- Holland, J. G., Solomon, C., Doran, J., & Frezza, D. A. (1976). *The analysis of behavior in planning instruction*. Reading, MA: Addison-Wesley.
- Hovland, C. I. (1952). A "communication analysis" of concept learning. *Psychological Review*, *59*, 461-472.
- Kagan, J. (1981). *The second year: The emergence of self-awareness*. Cambridge, MA: Harvard University Press.
- McIlvane, W. J., Bass, R. W., O'Brien, J. M., Gerovac, B. J., & Stoddard, L. T. (1984). Spoken and signed naming of foods after receptive exclusion training in severe retardation. *Applied Research in Mental Retardation*, *5*, 1-27.
- McIlvane, W. J., Munson, L. C., & Stoddard, L. T. (1987). Some observations on controlling relations in children's conditional discrimination and matching by exclusion. Manuscript submitted for publication.
- McIlvane, W. J., & Stoddard, L. T. (1981). Acquisition of matching-to-sample performances in severe retardation: Learning by exclusion. *Journal of Mental Deficiency Research*, *25*, 33-48.
- McIlvane, W. J., & Stoddard, L. T. (1985). Complex stimulus relations and exclusion in severe mental retardation. *Analysis and Intervention in Developmental Disabilities*, *5*, 307-321.
- McIlvane, W. J., Withstandley, J. K., & Stoddard, L. T. (1984). Positive and negative stimulus relations in severely retarded individuals' conditional discrimination. *Analysis and Intervention in Developmental Disabilities*, *4*, 235-251.
- Ray, B. A., & Sidman, M. (1970). Reinforcement schedules and stimulus control. In W. N. Schoenfeld (Ed.), *The theory of reinforcement schedules* (pp. 187-214). New York: Appleton-Century-Crofts.
- Saunders, R. R., & Sherman, J. A. (1986). Analysis of the "discrimination-failure hypothesis" in generalized matching and mismatching behavior. *Analysis and Intervention in Developmental Disabilities*, *6*, 89-107.
- Sherman, J. A., Saunders, R. R., & Brigham, T. A. (1970). Transfer of matching and mismatching behavior in preschool children. *Journal of Experimental Child Psychology*, *9*, 489-498.
- 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. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213-245). Hillsdale, NJ: Erlbaum.
- Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior*, *43*, 21-42.
- Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, *37*, 5-22.
- Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, *57*, 193-216.
- Stoddard, L. T., & Sidman, M. (1971a). The removal and restoration of stimulus control. *Journal of the Experimental Analysis of Behavior*, *16*, 143-154.
- Stoddard, L. T., & Sidman, M. (1971b). Stimulus control after intradimensional discrimination training. *Psychological Reports*, *28*, 147-157.
- Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, *10*, 349-367.
- Striefel, S., Bryan, K. S., & Aikins, D. A. (1974). Transfer of stimulus control from motor to verbal stimuli. *Journal of Applied Behavior Analysis*, *7*, 123-135.
- Stromer, R. (1986). Control by exclusion in arbitrary matching-to-sample. *Analysis and Intervention in Developmental Disabilities*, *6*, 59-72.
- Stromer, R., & Osborne, J. G. (1982). Control of adolescents' arbitrary matching-to-sample by positive and negative stimulus relations. *Journal of the Experimental Analysis of Behavior*, *37*, 329-348.
- Vincent-Smith, L., Bricker, D., & Bricker, W. (1974). Acquisition of receptive vocabulary in the toddler-age child. *Child Development*, *45*, 189-193.
- Zentall, T. R., Edwards, C. A., Moore, B. S., & Hogan, D. E. (1981). Identity: The basis for both matching and oddity learning in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *7*, 70-86.

Received February 21, 1986
Final acceptance June 10, 1987

APPENDIX

Text of Initial Subject Instructions

Now that you are familiar with BASIC input operations, you will begin participation in the study. Initially, you will learn performances involving either nonsense syllables ("ZOM," "KAQ," etc.) or numbers. There are two different types of tasks. They will be described for you here before we ask you to do them.

On each trial, you will see two choice items (displayed side-by-side as shown below; the choice items are indicated here with the arrowheads (^). (There will be no arrowheads later. The choices will always be the bottom two items, however.) Soon, we will teach you how to make choices between the two items, based on a third item that is displayed above them—the centered one with no arrowhead beneath it.

ZOM
KAQ QAS
 ^ ^

For example, here your choices are either “KAQ” or “QAS.” You choose one that goes with “ZOM.” (You do not know which choice is correct yet, of course.) In other words, when “ZOM” is displayed, you choose one item from the two below that goes with “ZOM.” You choose by typing the correct item on the terminal keyboard.

 345
782 345
 ^ ^

In this example, you are given a choice between 782 and 345. The number on the top is 345 also. Here, you are asked simply to type the number that is the same as the one at the top.

In order to go on to participate fully in the study (and earn the full amount of money), you must master these basic performances. Initially, the program will give prompts to help you learn which items go together. Specifically, both the top item and the correct choice item below it will be very bright. The two bright ones go together. In addition to the brightness cues, the program will tell you whenever you have typed in the correct choice and if you have selected the wrong one.

Later on in the study, you will only sometimes be told whether your choice was correct. This is part of the study. It does not necessarily mean that you have responded incorrectly.

The program checks for typographical errors. If you make a response that is recognized as a typo, the choices will be represented, and you must enter the characters correctly. There is no other penalty for a typo. If you have any questions, ask the experimenter.