JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2000, **74,** 207–227 NUMBER 2 (SEPTEMBER)

DERIVED RELATIONAL RESPONDING AS GENERALIZED OPERANT BEHAVIOR

OLIVE HEALY, DERMOT BARNES-HOLMES, AND PAUL M. SMEETS

UNIVERSITY COLLEGE CORK, IRELAND/CABAS©,
NATIONAL UNIVERSITY OF IRELAND, MAYNOOTH, AND
LEIDEN UNIVERSITY, THE NETHERLANDS

The major aim of the present study was to demonstrate that derived relational responding may be viewed as a form of *generalized* operant behavior. In Experiment 1, 4 subjects were divided into two conditions (2 in each condition). Using a two-comparison matching-to-sample procedure, all subjects were trained and tested for the formation of two combinatorially entailed relations. Subjects were trained and tested across multiple stimulus sets. Each set was composed of novel stimuli. Both Conditions 1 and 2 involved explicit performance-contingent feedback presented at the end of each block of test trials (i.e., delayed feedback). In Condition 1, feedback was accurate (consistent with the experimenter-designated relations) following exposure to the initial stimulus sets. When subjects' responding reached a predefined mastery criterion, the feedback then switched to inaccurate (not consistent with the experimenter-designated relations) until responding once again reached a predefined criterion. Condition 2 was similar to Condition 1, except that exposure to the initial stimulus sets was followed by inaccurate feedback and once the criterion was reached feedback switched to accurate. Once relational responding emerged *and* stabilized, response patterns on novel stimulus sets were controlled by the feedback delivered for previous stimulus sets. Experiment 2 replicated Experiment 1, except that during Conditions 3 and 4 four comparison stimuli were employed during training and testing. Experiment 3 was similar to Condition 1 of Experiment 1, except that after the mastery criterion was reached for class-consistent responding, feedback alternated from accurate to inaccurate across each successive stimulus set. Experiment 4 involved two types of feedback, one type following tests for mutual entailment and the other type following tests for combinatorial entailment. Results from this experiment demonstrated that mutual and combinatorial entailment may be controlled independently by accurate and inaccurate feedback. Overall, the data support the suggestion, made by relational frame theory, that derived relational responding is a form of generalized operant behavior.

Key words: generalized operant class, relational frame theory, mutual entailment, combinatorial entailment, equivalence relation, matching to sample, humans

Derived relational responding has recently stimulated much interest and research activity among behavior analysts. One of the main reasons for focusing on such responding is the fact that it cannot be readily accounted for by the concept of conditional discrimination. In the typical equivalence experiment, for example, if subjects are trained to match Comparison Stimulus B1 to Sample Stimulus A1, and Comparison Stimulus C1 to Sample Stimulus B1, they will likely match Comparison Stimulus A1 to Sample Stimulus C1, demonstrating derived equivalence responding, without additional training. A conditional discrimination, as normally defined, does not predict the emergence of this untrained performance. Neither A1 nor C1 has a history of differential reinforcement as a conditional discriminative stimulus with regard to the other, and thus, neither stimulus would be expected to reliably control selection of the other.

Relational frame theory (RFT) offers one account of derived relational responding. The important feature of RFT for the current research is the fact that a relational frame, as an analytic unit, is conceptualized as a three-term contingency. For RFT, the contextual cue is the third term, the relational response (e.g., responding to Stimulus B in terms of A and responding to A in terms of B) is the second term, and a history of differential reinforcement correlated with the contextual cue is the first term in the contingency. From

This research was conducted as part of Olive Healy's doctoral research program under the supervision of Dermot Barnes-Holmes. Portions of these data were presented at the annual conference of the Experimental Analysis of Behavior Group, London, April 1998, and at the annual convention of the Association for Behavior Analysis, Orlando, Florida, May 1998.

Requests for reprints may be sent to Dermot Barnes-Holmes at the Department of Psychology, National University of Ireland, Maynooth, Maynooth, County Kildare, Ireland (E-mail: Dermot.Barnes-Holmes@may.ie).

this perspective, therefore, responding to B given A and to A given B may be considered as a single response unit controlled by a relevant contextual cue by virtue of its previous correlation with differential reinforcement. In effect, the RFT approach invokes a purely functional concept of an operant, and the term *generalized operant class* (e.g., Barnes, 1994, 1996; D. Barnes-Holmes & Barnes-Holmes, 2000; Hayes, 1992; Hayes & Barnes, 1997) is used to emphasize this fact.

The RFT view of equivalence responding as a form of generalized operant behavior may be difficult to appreciate if one typically conceptualizes operants in structuralist terms (see D. Barnes-Holmes & Barnes-Holmes, 2000, for a detailed discussion of this issue). The concept of a response class with an infinite range of topographies is a defining property of operant behavior. Nonetheless, topographical and functional classes of behavior–environment interactions quite often overlap, and thus the two may become confused. Lever pressing, for instance, may be defined by the effect of activity upon the lever, but almost all lever presses involve ''pressing'' movements. A sensitive lever may be activated by coughing, but for most purposes such instances can normally be ignored. Sometimes, however, the independence between topographical and functional classes is made very clear. The concept of generalized imitation (e.g., Baer, Peterson, & Sherman, 1967; Gewirtz & Stengle, 1968; Poulson, Kymissis, Reeve, Andreatos, & Reeve, 1991) provides one excellent example. After a generalized imitative repertoire is established, an almost infinite variety of response topographies may be substituted for the forms used in the earlier training. The behavior of imitating is generalized because it is not limited to any particular response topography. In a similar vein, some behavior analysts have argued that it is possible to reinforce ''generalized attending'' (McIlvane, Dube, & Callahan, 1995; McIlvane, Dube, Kledaras, Iennaco, & Stoddard, 1990), although *what* is being attended to will change. Broadly similar arguments have also been made with respect to many other phenomena, such as generalized identity matching and mismatching (e.g., Cumming & Berryman, 1965; Dube, McIlvane, & Green, 1992; Saunders & Sherman, 1986), exclusion (e.g., Lipkens, Hayes, & Hayes, 1993; McIlvane et al., 1987), arbitrary assignment (e.g., Saunders, Saunders, Kirby, & Spradlin, 1988), and one-trial learning (e.g., Catania, 1996; Dube et al., 1992).

Although these and yet other examples (see Neuringer, 1986; Pryor, Haag, & O'Reilly, 1969; Stokes & Baer, 1977) constitute a simple extension of the three-term contingency as an analytic unit, specific qualifiers are often included when operant classes are not readily defined topographically. Such classes are referred to as *generalized, higher order,* or *overarching*. These qualifiers are not used in this instance as technical terms, and they do not imply the existence of mediational processes leading to the formation of operants of this type. Instead, these qualifiers emphasize that a specific functional class cannot be defined by its response forms or stimulus forms, a fact that is true in principle for all functional classes (D. Barnes-Holmes & Barnes-Holmes, 2000). At the present time, very little is known about the determinants of operant class formation in those cases in which there is minimal overlap between function and topography (see Pilgrim & Galizio, in press). As a possible starting point for addressing this important issue, RFT argues that derived relational responding, including stimulus equivalence, can be explicitly interpreted as a type of discriminated operant behavior that is more usefully defined by functional rather than topographical properties (D. Barnes-Holmes & Barnes-Holmes, 2000).

Consistent with earlier literature on the topic, Hayes (1994) suggested that there are four particularly important properties of discriminated operant behavior: (a) operants develop, (b) operants are flexible and can be shaped, (c) operants can come under stimulus control, and (d) operants are controlled by their consequences. Obviously, if deriving stimulus relations is to be viewed as operant behavior, all four of these properties should apply. Supportive research has been provided on all four points (e.g., Barnes, Browne, Smeets, & Roche, 1995; Barnes & Hampson, 1993, 1997; Barnes, Hegarty, & Smeets, 1997; Dymond & Barnes, 1995, 1996; Lipkens et al., 1993; Roche & Barnes, 1996; 1997; Roche, Barnes, & Smeets, 1997; Roche, Barnes-Holmes, Smeets, Barnes-Holmes, & McGeady, 2000; Steele & Hayes, 1991; Wilson & Hayes, 1996). The least support, however, has been obtained for the final point.

In line with this operant approach to relational responding, we recently reported a series of experiments that examined the effects of differential consequences on derived relational responding (Healy, Barnes, & Smeets, 1998). Specifically, adult humans were trained on four matching-to-sample (MTS) trial types that established the following conditional discriminations: A1-B1, A2-B2, B1-C1, B2-C2 (the stimuli were three-letter nonsense syllables but are labeled, for the purposes of communication, with alphanumerics). Following training, the subjects were exposed to a derived relations test that probed for transitivity (A1-C1 and A2-C2) and combined symmetry and transitivity (C1-A1 and C2-A2); the generic term, *combinatorial entailment,* will be used here to label these relational responses (see Hayes, 1991). Using the same set of nonsense syllables, each subject completed a total of 11 cycles of training and testing. However, unlike a traditional emergent relations experiment, after each exposure to the test, a specific form of ''test-performance feedback'' was presented to the subject.

In Experiment 5 of Healy et al. (1998), for example, following each of the 11 cycles of training and testing, subjects were presented with a point system that indicated the exact number of class-consistent responses made during the test phase (e.g., after training A1-B1 and B1-C1, a class-consistent response would be C1-A1). For some subjects the feedback was accurate following the first five stimulus sets and was inaccurate following the next six (and vice versa for the remaining subjects). During accurate feedback, if a subject emitted five class-consistent responses, for example, five star shapes were presented in a vertical column on the screen; if the feedback was inaccurate then 15 stars were presented (i.e., 20 stars minus 5 for the class-consistent responses). Eight of the 10 subjects responded in accordance with the accurate and inaccurate feedback. Taken together, the five experiments reported by Healy et al. demonstrated that *derived* relational responding is often sensitive to differential consequences provided at the end of consecutive test phases. The authors argued that the data supported the operant interpretation of such responding.

However, the authors also pointed out that the use of the same stimulus set across the entire 11 sessions for each subject limits the implications the results have for the interpretation of relational responding as a generalized or overarching operant class. Consistent with RFT, the term *generalized* is used here to denote an operant that cannot be defined topographically. Because only one stimulus set was used, the consistent changes in responding that occurred across stimulus sets could have been due to the direct effects of the feedback on responding to the same stimulus set. The effects of the feedback do not, therefore, require an interpretation in terms of emergence or in terms of generalized operant behavior. To constitute generalized operant behavior, there should be no history of differential consequences contingent on responding to a particular stimulus set. For example, if each cycle of training and testing involved a novel set of nonsense syllables, and accurate performance-contingent feedback on tests facilitated the emergence of relational responding *across sets,* this would indicate that generalized derived relational responding, as a response unit, was being shaped by the differential feedback. This approach is consistent with the following quotation from Estes (1971):

> In more mature human beings, much instrumental behavior and, more especially, a great part of verbal behavior is organized into higher-order routines and is, in many instances, better understood in terms of the operation of rules, principles, strategies and the like than in terms of successions of responses to particular stimuli. . . . In these situations it is the selection of strategies rather than the selection of particular reactions to stimuli which is modified by past experience with rewarding or punishing consequences. (p. 23)

Although Estes used loosely defined terms, such as *strategies,* RFT adopts a similar position but includes the term *relational operant,* so that the final sentence of the above quotation would read, ''In these situations it is the selection of relational operants rather than the selection of particular reactions to stimuli which are modified by past experience with rewarding and punishing consequences.''

Another important issue that arose from the Healy et al. (1998) study was the fact that some subjects showed relatively stable performances across multiple exposures, whereas others did not. One procedural reason for this variability across subjects may be that no mastery criterion was used before terminating the experiment or changing from one type of feedback to another (i.e., all subjects completed 11 stimulus sets irrespective of performance). To demonstrate clearly the effects of differential feedback, a mastery criterion should be employed that requires a predetermined level of consistent responding in accordance with the feedback before there is any change in the type of feedback delivered, or before the experiment is terminated.

In Experiment 1 of the present series, all subjects were first exposed to conditional discrimination training on four MTS trial types (A1-B1, B1-C1, A2-B2, B2-C2). The subjects then were tested for the formation of four combinatorially entailed derived relations (A1-C1, C1-A1, A2-C2, C2-A2). Following exposure to this first cycle of training and testing, accurate or inaccurate feedback (the conditions were counterbalanced) was delivered for the test performances. The next cycle of training and testing then began, but with a novel set of stimuli. This cycle of training and testing, using novel sets of stimuli for each cycle, continued until a subject responded in accordance with the feedback across three consecutive stimulus sets. Once this stability criterion was reached, the type of feedback switched (accurate to inaccurate or vice versa) and the training and testing continued, using novel stimulus sets, until the test performance again reached the stability criterion. In the interests of clarity, the other experiments will be introduced subsequently.

GENERAL METHOD

Subjects

Participants in the study were 13 students pursuing both undergraduate and postgraduate studies at University College Cork. Eight were female, and 5 were male. Their ages ranged from 18 to 25 years. Each subject was randomly assigned to one of the experimental conditions. Subjects 1 through 4 participated in Experiment 1, Subjects 5 through 8 in Experiment 2, Subjects 9 through 11 in Experiment 3, and Subjects 12 and 13 in Experiment 4. None of the subjects had any knowledge of stimulus equivalence or related phenomena. All but 1 subject (Subject 1) completed an entire experiment in one sitting.

Apparatus

The experiment was conducted in a small experimental room (4 square m) in which the subject was seated at a table in front of an Apple Macintosh LCIII® personal computer. The computer was programmed in BBC BASIC and it controlled stimulus presentations and recorded subjects' responses. During all conditions (except Conditions 3 and 4), the Z and M keys were designated as response keys and were marked with yellow paper squares. During Conditions 3 and 4, the G, R, U, C, and N keys were designated as response keys. The pool of stimuli consisted of 168 nonsense syllables. Each of the stimulus sets used for a particular subject was constructed by randomly selecting nonsense syllables, without replacement, from this pool, and then assigning them randomly to their roles as sample and comparison stimuli. Although the same pool of nonsense syllables was used for each subject, the random selection and random assignment of these syllables allowed us to construct a completely novel stimulus set for every cycle of training and testing across all of the subjects in the study (i.e., a particular set was only used with 1 subject).

General Procedure

All subjects were exposed to individual cycles of training and testing, using novel stimulus sets for every cycle. Each cycle of training and testing was as follows. All subjects were trained in four MTS trial types using a randomly generated set of six nonsense syllables (each set is designated using the alphanumerics A1, A2, B1, B2, C1, C2; subjects never saw these labels). Subjects were seated in the experimental room and the following instructions were presented on the computer screen:

> During this stage of the experiment you must look at the nonsense syllable at the top and then choose one of the two nonsense syllables at the bottom by pressing one of the marked keys on the keyboard. To choose the left nonsense syllable press the marked key on the left.

> To choose the right nonsense syllable press the marked key on the right. Press the spacebar twice to continue.

On each MTS trial, the sample and the two comparison stimuli always differed in at least two letters. The sample appeared centered in the top half of the monitor screen, followed 1.5 s later by the comparison stimuli, which were positioned to the left and right of the sample, 5.00 cm from the bottom of the screen. On each MTS trial, the position of the comparison stimuli was varied randomly (i.e., the correct nonsense syllable could appear on either the left or the right with equal probability). Subjects were trained on the two A-B and two B-C MTS trial types. The four trial types were presented in a quasirandom order, each trial type occurring twice every eight trials until the subject produced eight consecutive correct responses. For each stimulus set, when A1 was the sample, B1 was correct; when B1 was the sample, C1 was correct; when A2 was the sample, C2 was correct; when B2 was the sample, C2 was correct.

The correct completion of an MTS trial removed the stimulus display and produced ''correct'' in the center of the screen accompanied by a high-pitched beep for 1.5 s. The incorrect completion of an MTS trial removed the stimulus display and produced ''wrong'' in the center of the screen but without auditory feedback, again for 1.5 s. A 1-s intertrial interval followed both types of feedback, during which the computer screen remained blank. When eight consecutive correct responses had been emitted, the computer program progressed without interruption to an MTS test for combinatorial entailment. The test consisted of four MTS trial types presented in a quasirandom order, with each of the four trial types occurring five times across 20 trials (i.e., no intermixed baseline trials). No feedback was given during testing.

In summary, each training and testing cycle employed a novel set of stimuli and consisted of training to a mastery criterion (eight consecutive correct) followed by exposure to a single block of 20 test trials. In effect, subjects were required to learn a *novel* set of interrelated conditional discriminations, and were tested for *novel* examples of combinatorial entailment, across every cycle of training and testing.

Posttest Feedback

Feedback was delivered after each test for combinatorial entailment. This feedback provided differential consequences contingent upon the exact number of class-consistent responses emitted during the previous derived relations test. Once a test phase had been completed, the words ''Minimum Points Earned'' and ''Maximum Points Earned'' appeared at the very bottom and top of the screen, respectively. A computer-generated visual sequence presented a star shape for every relational-consistent response (accurate feedback) or nonrelational-consistent response (inaccurate feedback) emitted during the test. Each star appeared on the screen accompanied by a high-pitched beep. The first star appeared directly above the words ''Minimum Points Earned'' and each subsequent star appeared above the other in a vertical line. A broken line appeared midway on the screen as a half-way mark. Once the stars earned during a test exposure had been presented, they remained on the screen for 10 s before the screen cleared, and the following instructions appeared on the monitor screen: ''That is the end of this part of the experiment. Please report to the experimenter.'' When the subject reported to the experimenter, he or she was asked to ''Please wait here [i.e., outside the experimental room], until I ask you to continue with the experiment again.'' When the experimenter had examined and saved the data to a diskette and reset the computer for the next cycle of training and testing (with a novel stimulus set), she returned to the subject and asked him or her to ''Please go back to the computer and continue with the experiment.'' No other instructions were provided, and no other interactions occurred between the experimenter and subject during the course of the experiment.

EXPERIMENT 1

Procedure

Condition 1 (Subjects 1 and 2). In Condition 1, accurate feedback was delivered following each cycle of training and testing until responding reached a predefined mastery criterion. This criterion determined when a switch in feedback would occur. The criterion required that a subject produce a pattern of responding consistent with the feedback across three consecutive test exposures. Thus, for example, during accurate feedback a subject was required to emit the conditional responses A1-C1, A2-C2, C1-A1, and C2-A2 at least four times across each of the five exposures to each trial type, across three consecutive stimulus sets. Once a subject had achieved this mastery criterion, the experimenter altered the computer program and the feedback switched to inaccurate. During inaccurate feedback the subject was required to emit the conditional responses A1-C2, A2-C1, C1-A2, and C2-A1 at least four times across each of the five exposures to each trial type, across three consecutive stimulus sets, before the experiment was terminated. If a response pattern fell outside either of the above categories, it was labeled as intermediate responding.

Condition 2 (Subjects 3 and 4). Condition 2 was identical to Condition 1 except that inaccurate feedback was delivered following each of the initial cycles of training and testing, until the criterion was reached. Feedback then switched to accurate, and exposure to test conditions continued until performance once again reached the stability criterion.

Results and Discussion

The number of training trials required to meet the mastery criterion, per stimulus set, by each subject is presented in Table 1. The minimum number of trials required was 10 (Subject 1) and the maximum was 308 (Subject 3). The number of training trials per exposure varied unsystematically for each subject. In general, however, subjects' earlier exposures to the training involved larger numbers of trials than many of their later exposures, but there were some notable exceptions (e.g., Subjects 1 and 2, Exposures 14 and 6, respectively).

Performances on each of the tests for combinatorial entailment are shown in Figure 1. Subject 1 required a total of 15 exposures before achieving the stability criterion during accurate feedback. This was followed by a rapid shift from high levels of combinatorial entailment to three consecutive performances in accordance with low levels with the introduction of inaccurate feedback. For ease of communication, the descriptor *low level of combinatorial entailment* is used to label the following relational responses: A1-C2, A2-C1, C1-A2, C2-A1. We recognize, however, that this pattern may itself constitute combinatorial entailment rather than the absence of entailed relations. For instance, the inaccurate feedback might establish the relational frame of opposition (e.g., A1 opposite to B1 and B1 opposite to C1 derives A1 *not* opposite to C1) (e.g., Steele & Hayes, 1991). Alternatively, the feedback might establish a frame of coordination but with negative stimulus control emerging during the test phases (see Experiment 2). In either case, the A and C stimuli participate in a type of combinatorially entailed relation. In the current article, therefore, the distinction between high and low levels of entailment is used simply to reflect the position of the data points on the graphs. Subject 2 reached the criterion on Exposure 5, and switched responding promptly during inaccurate feedback. In Condition 2, Subject 3 produced intermediate responding followed by a variable pattern during inaccurate feedback before achieving the criterion, and when the feedback switched so did the responding to high levels of combinatorial entailment. Subject 4 reached the criterion rapidly during both types of feedback, requiring a total of eight exposures.

These data suggest that the feedback for test performances, combined with the use of a stability criterion, produced consistent patterns of combinatorial entailment across all 4 subjects. Furthermore, these consistent patterns were established across novel stimulus sets. These data thus provide support for the argument that derived relational responding may be conceptualized as generalized operant behavior.

EXPERIMENT 2: INCREASING THE NUMBER OF CLASSES

Although the data from Experiment 1 showed that feedback affected derived relational responding across stimulus sets, an important interpretive problem arises. The use of completely reversed feedback contingencies, combined with only two comparisons on each trial, may have facilitated control by

Table 1

Number of training trials across successive stimulus sets for each subject in Experiment 1.

Stimulus set (exposure)	Training trials			
	S1	S2	S3	S4
1	148	132	308	232
2	68	176	272	143
3	104	158	196	96
4	116	86	84	96
5	112	96	114	124
6	56	194	167	86
7	70	48	84	72
8	96	68	86	36
9	10	42	102	
10	128		72	
11	92		42	
12	72			
13	148			
14	186			
15	104			
16	92			
17	114			
18	56			
19	82			

what are called Type R and Type S relations (see Carrigan & Sidman, 1992). Accordingly, subjects may either select a particular comparison when presented with a sample (a Type S or select relation) or they may reject a particular comparison stimulus (a Type R or reject relation). Both types of relation (select and reject) may generate consistent patterns of combinatorial entailment. One way to reduce the likelihood of Type R control in MTS is to employ three or more comparisons, because a greater number of reject, relative to select, relations are then required (see Sidman, 1987). All of our research on the effects of feedback on derived relational responding has employed only two comparisons, and thus by implication Type R relations may have played an important role in generating the observed effects. The possibility remains, therefore, that similar results would not be obtained if the number of comparisons were increased. Experiment 2 was conducted to address this possibility.

Procedure: Conditions 3 and 4

The procedures for Conditions 3 and 4 in this experiment were similar to Conditions 1 and 2 in Experiment 1, respectively, except that subjects were trained in eight rather than four MTS trial types using sets of 12 rather than six randomly selected nonsense syllables (each set is designated using the alphanumerics A1, A2, A3, A4, B1, B2, B3, B4, C1, C2, C3, C4; subjects never saw these labels). Subjects 5 and 6 participated in Condition 3 (accurate-inaccurate feedback), and Subjects 7 and 8 participated in Condition 4 (inaccurate-accurate feedback). On each MTS trial, the sample appeared centered on the monitor screen. Once the subject pressed the U key on the keyboard (marked with colored paper), the four comparison stimuli each appeared on the four corners of the monitor screen. On each MTS trial, the position of the comparison stimuli was varied randomly (i.e., the correct nonsense syllable could appear on any of the four corners with equal probability). In these two conditions, a one-to-many training and testing design was employed (i.e., train A-B, A-C, and test B-C, C-B). Furthermore, because of the increased number of MTS trial types, training was divided into two blocks.

Subjects were trained on the four A-B MTS trial types. The four trial types were presented in a quasirandom order, each trial type occurring twice across each block of eight trials, until the subject produced eight correct consecutive responses. Subjects were then trained on the four A-C MTS trial types. These were also presented in a quasirandom order, until the subject produced eight consecutive correct responses. The four A-B and four A-C trial types were then presented quasirandomly in blocks of eight trials (each trial type occurring once in each block), until the subject produced eight consecutive correct responses. The A-B trial types always presented the four B stimuli as comparisons, and the A-C trial types always presented the four C stimuli as comparisons. For each stimulus set, when A1 was the sample, B1 was correct; when A2 was the sample, B2 was correct; when A3 was the sample, B3 was correct; when A4 was the sample, B4 was correct; when A1 was the sample, C1 was correct; when A2 was the sample, C2 was correct; when A3 was the sample, C3 was correct; when A4 was the sample, C4 was correct.

When subjects had successfully completed the mixed A-B and A-C training, the computer program progressed without interruption to an MTS test. The test consisted of eight MTS trial types (B1-C1, B2-C2, B3-C3, B4-C4,

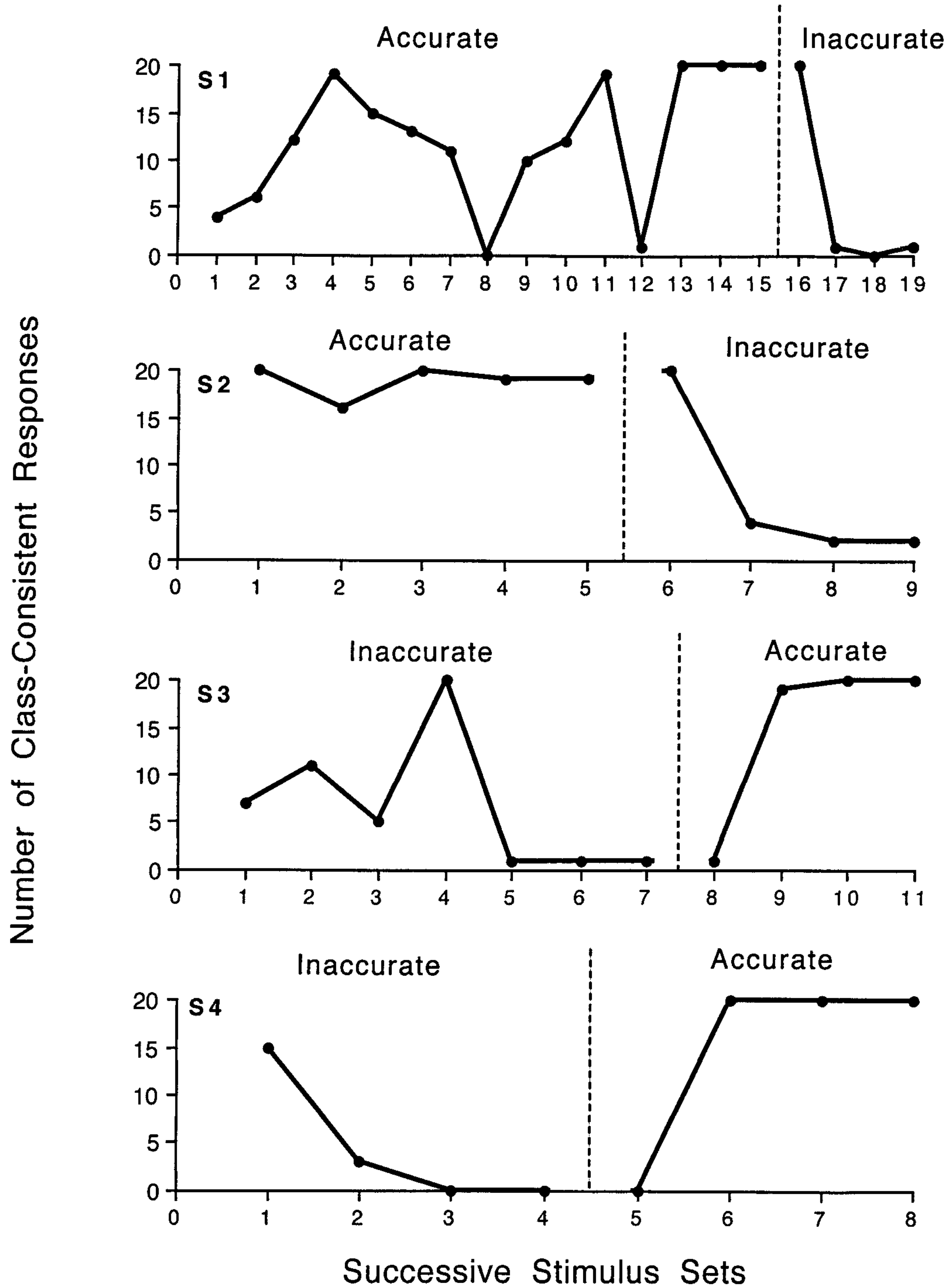

Fig. 1. Class-consistent responding across successive test exposures for each subject in Experiment 1. Each data point represents one exposure to a 20-trial test block involving a novel stimulus set. Subjects 1 and 2 were exposed to accurate test-performance feedback followed by inaccurate feedback. Subjects 3 and 4 were exposed to inaccurate feedback followed by accurate feedback.

C1-B1, C2-B2, C3-B3, C4-B4) presented in a quasirandom order, with each of the eight trial types occurring twice across 16 trials. Similar to the previous experiment, the mastery criterion required that subjects respond in accordance with the feedback across three consecutive stimulus sets. However, because there were only two exposures to each of the eight trial types, the mastery criterion required that all 16 responses be in accordance with the feedback. During accurate feedback, one star for each of the following relational responses was delivered: B1-C1, B2-C2, B3-C3, B4-C4, C1-B1, C2-B2, C3-B3, C4-B4. When the feedback was inaccurate, a star was delivered for any three of the class-inconsistent responses (i.e., B1-C2/C3/C4, B2-C1/C3/C4, B3-C1/C2/C4, B4-C1/C2/C3, C1-B2/B3/B4, C2-B1/B3/B4, C3-B1/B2/B4, C4-B1/B2/B3).

At the beginning of each exposure to a cycle of training and testing (with a novel stimulus set), each subject was seated in the experimental room and the following instructions were presented on the computer screen (i.e., the same procedure as Experiment 1, but with modified instructions to accommodate the four-comparison MTS trial types):

> You must look at the nonsense syllable in the center of the screen, press the marked center key, and then choose one of the 4 nonsense syllables that appear at each corner of the screen by pressing one of the four marked keys on the keyboard. To choose the top-left syllable press the marked key on the top-left. To choose the top-right syllable press the marked key on the top-right. To choose the bottom-left syllable press the marked key on the bottom-left. To choose the bottom-right syllable press the marked key on the bottom-right. Sometimes the computer will give you feedback and sometimes it will not. Press the space-bar twice to continue.

Results and Discussion

The number of training trials required by each subject to meet the mastery criterion during each stimulus set is presented in Table 2. The minimum number of trials required was 42 (Subjects 5 and 6) and the maximum was 542 (Subject 8). As in Experiment 1, the number of training trials per exposure tended to vary unsystematically for each subject, but again, in general, early training exposures tended to involve larger numbers of trials than many of the later exposures (see Subject 8, Exposure 7, for a notable exception).

Table 2

Number of training trials across successive stimulus sets for each subject in Experiment 2.

Stimulus set (exposure)	Training trials			
	S5	S6	S7	S8
1	231	286	214	236
2	135	246	186	111
3	148	106	128	168
4	121	81	321	112
5	66	140	124	118
6	81	148	121	121
7	56	96	216	542
8	146	42	138	84
9	102	46	114	97
10	163	64	85	86
11	42	84	92	
12		36	72	
13			85	
14			84	

Performances on each of the tests for combinatorial entailment are shown in Figure 2. In Condition 3, Subject 5 reached the stability criterion on Exposure 5 and maintained high levels of combinatorial entailment following a switch in feedback during Exposure 6. Intermediate responding was shown during Exposures 7 and 8, before low levels of combinatorial entailment on Exposures 9 to 11. Subject 6 reached the stability criterion on Exposure 7, and maintained a high performance on the test for combinatorial entailment across Exposures 8 and 9, before rapidly shifting to low levels during Exposures 10, 11, and 12. In Condition 4, Subject 7 reached the criterion on Exposure 7 and produced low levels of combinatorial entailment during Exposures 8, 9, and 10, despite the switch to accurate feedback. On Exposure 11, this subject produced intermediate responding, and then on Exposure 12 responding shifted to high levels of entailment in accordance with accurate feedback. Subject 8 reached the mastery criterion during Exposures 3, 4, and 5. After accurate feedback was introduced, this subject switched to high levels of entailment during Exposures 8 to 10.

These data show that the effects of posttest feedback are not restricted to two-choice MTS trial types. Although these data extend the basic feedback effect beyond the two-choice MTS format, they do not exclude Type

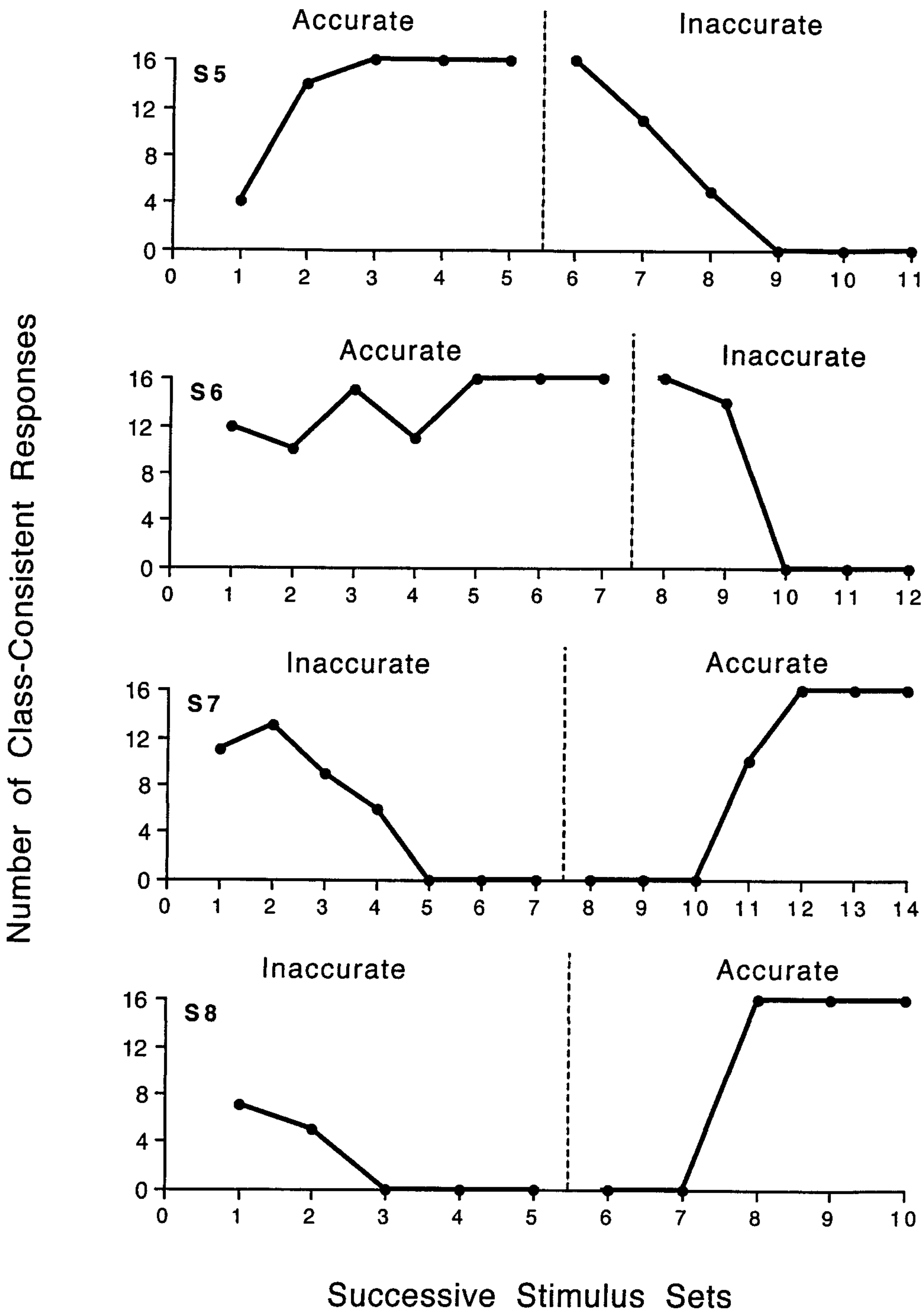

Fig. 2. Class-consistent responding across successive test exposures for subjects in Experiment 2. Each data point represents one exposure to a 16-trial test block involving a novel stimulus set.

R relations from involvement in test performance. The use of four comparisons may have reduced the likelihood of Type R control during the conditional discrimination *training*, but the introduction of inaccurate feedback during the *tests* may have facilitated Type R control. In fact, visual inspection of the raw data (not presented) showed that all 4 subjects tended to distribute their responses across the three class-inconsistent comparisons when they were responding in accordance with the inaccurate feedback. Such a pattern of responding suggests that subjects were not choosing a particular comparison, but were simply responding away from the "correct" choice (i.e., Type R control). Future analysis might reveal this effect in greater detail.

Table 3

Number of training trials across successive stimulus sets for each subject in Experiment 3.

Stimulus set (exposure)	Training trials		
	S9	S10	S11
1	127	142	162
2	141	116	112
3	135	178	108
4	101	86	61
5	88	52	82
6	104	111	101
7	92	54	81
8	58	88	89
9	123	45	122
10	127	109	113
11		66	98
12		32	62
13			88
14			106
15			85
16			94

EXPERIMENT 3: REPEATED FEEDBACK REVERSALS

In the previous two experiments, and in our previously published research (Healy et al., 1998), we employed separate AB and BA designs, in which A constitutes accurate feedback and B is inaccurate feedback. These designs produced relatively systematic outcomes. When feedback was accurate, the subjects eventually produced stable, combinatorially entailed relational responses, and when feedback was inaccurate subjects eventually produced stable responding that was not in accordance with combinatorial entailment. A traditional single-subject experimental design, however, would involve not only AB or BA but also a return to baseline (e.g., ABA or BAB, respectively). In fact, a highly robust design would involve repeated reversals (e.g., ABABAB) (e.g., Sidman, 1960). In Experiment 3, patterns of responding were examined when the type of feedback was reversed repeatedly across cycles of training and testing (again using a novel stimulus set for each cycle).

Procedure

The procedure of Experiment 3 was similar to Condition 1 in Experiment 1 (e.g., two-choice MTS, with a novel stimulus set employed in each cycle of training and testing) with two important differences. First, Subjects 9 through 11 were exposed to training and testing cycles until they responded in accordance with accurate feedback across two, rather than three, consecutive cycles. Second, once this two-cycle mastery criterion had been reached, the feedback then alternated from inaccurate to accurate across six (Subjects 9 and 10) or 10 (Subject 11) cycles of training and testing.

Results and Discussion

The number of training trials required per stimulus set by each subject is presented in Table 3. The minimum number of trials required was 32 (Subject 10) and the maximum was 178 (Subject 10). As in Experiments 1 and 2, the number of training trials per exposure varied unsystematically for each subject, but again, in general, early training exposures tended to involve larger numbers of trials than many of the later exposures. Subject 9, however, appeared to depart a little from this pattern, with relatively high numbers of training trials required across the final two stimulus sets.

Performance on each of the tests for combinatorial entailment are shown in Figure 3. Subject 9 reached the stability criterion on Exposure 4 and with the introduction of inaccurate feedback switched responding from high levels of combinatorial entailment to intermediate responding. On remaining exposures as the feedback alternated from accurate to inaccurate, responding switched

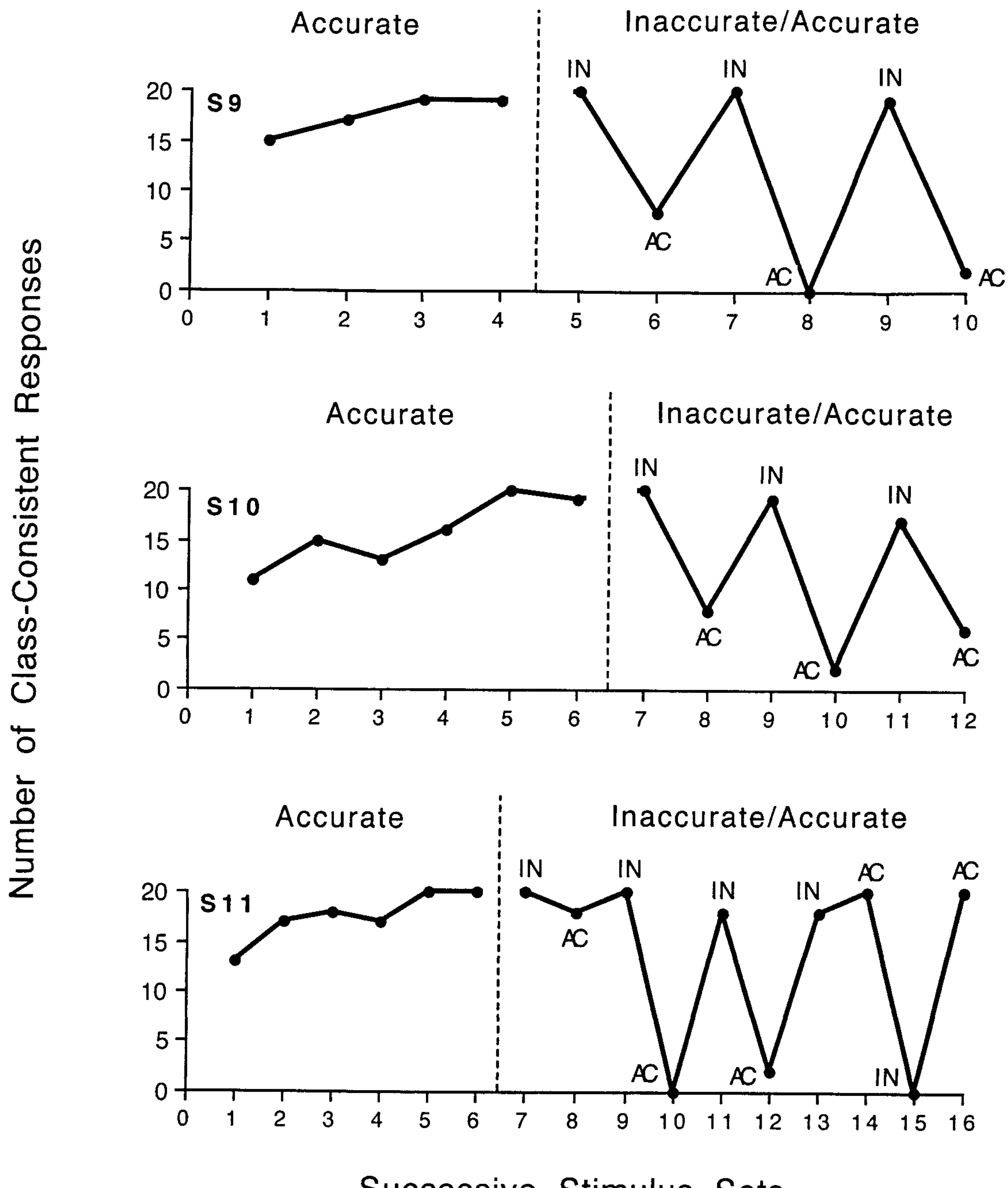

Fig. 3. Class-consistent responding across successive test exposures for each subject in Experiment 3. Each data point represents one exposure to a 20-trial test block involving a novel stimulus set. The presentation of inaccurate or accurate test-performance feedback after each test is indicated by the letters IN and AC, respectively (see text for details).

abruptly from high levels of entailment to low levels in accordance with the type of feedback delivered. Subject 10 reached the criterion on Exposure 6 and also produced a variable pattern of responding across Exposures 7 to 12 when the feedback alternated from accurate to inaccurate. For both of these subjects, the alternating feedback clearly produced dramatic changes in the patterns of responding. However, in both cases, neither subject successfully adapted responding to the feedback contingencies. For example, when the feedback changed from accurate to inaccurate at the end of Exposure 5 for Subject 9, a shift from a high to an intermediate level of combinatorial entailment occurred. However, at the end of Exposure 6, accurate feedback was again presented, and the subject shifted back to a high level of combinatorial entailment. Thus after both Exposures 5 and 6, Subject 9 received relatively few points for his performances during those exposures, and this pattern continued for the remaining stimulus sets. That is, he switched his pattern of responding back and forth between high and low levels of entailment across each stimulus set, thereby losing almost all of the feedback points available. A similar pattern was also observed for Subject 10. To determine whether this pattern of consistent point loss would eventually break down, Subject 11 was exposed to the same procedure, but additional exposures to the alternating feedback were provided. With sufficient exposure to this form of feedback, would a pattern of responding eventually emerge that allows the subject to obtain most of the feedback points available across stimulus sets?

Subject 11 reached the mastery criterion during Exposures 5 and 6, but maintained high levels of combinatorial entailment when inaccurate feedback was introduced following Exposure 7. This pattern of responding was maintained until inaccurate feedback was delivered following Exposure 9, and responding switched from high to low levels of entailment. Responding across Exposures 10, 11, 12, and 13 alternated from high levels to low levels of entailment, and thus, this subject lost most of the feedback points available during these exposures. During Exposure 14, however, a high level of entailment was maintained despite the delivery of inaccurate feedback. Responding then switched to low levels (Exposure 15) and reverted to high levels on the final exposure. Therefore, across Exposures 14, 15, and 16, Subject 11 achieved the maximum number of points obtainable.

These data indicate that not only does differential feedback establish, maintain, and produce rapid shifts in test performances, but it also produces responding that comes increasingly under the control of repeated reversals of the feedback contingencies. This finding may be considered evidence of yet another example of generalized operant behavior. We return to this issue in the General Discussion.

EXPERIMENT 4: FEEDBACK ON THE COMPONENT OPERANTS OF DERIVED RELATIONAL RESPONDING

The experiments conducted thus far have shown that the use of differential consequences can bring specific patterns of derived relational responding under relatively reliable control. In each of the two-choice MTS experiments (including those reported by Healy et al., 1998) the feedback contingencies were designed to reverse the pattern of derived relational responding. For example, if the A1-C1, A2-C2, C1-A1, and C2-A2 relational responses were class consistent, the inaccurate feedback contingencies aimed to generate a complete reversal in these relational responses (i.e., A1-C2, A2-C1, C1-A2, and C2-A1). This constitutes one way to test the operant nature of derived relational responding, but additional tests are possible. For example, the current work focused only on combinatorial entailment, which is only one of the defining properties of relational framing. Another property is mutual entailment, which we have not yet examined. Some relevant work has been conducted in this area (e.g., Pilgrim & Galizio, 1990, 1995; Saunders et al., 1988; Spradlin, Saunders, & Saunders, 1992) showing that the component properties of the relational frame of equivalence may dissociate under specific environmental control.

Pilgrim and Galizio (1990), for example, trained adult subjects on a series of conditional discriminations (i.e., A1-B1, A2-B2, A1-C1, A2-C2) that led to the emergence of two three-member equivalence classes during testing (i.e., A1-B1-C1, A2-B2-C2). Following equivalence testing, subjects received further training in which the original A-C relations were reversed (i.e., A1-B1, A2-B2, A1-C2, A2-C1). This altered the symmetry responding of 3 of 4 subjects, but none of them responded in accordance with the transitive relations that would be expected to follow from the new (reversed) conditional discriminations. In effect, equivalence test performances were not controlled by the modified conditional discrimination contingencies that were in effect, despite the fact that performances on symmetry probes *were* sensitive to the novel reinforcement contingencies. This finding is consistent with both earlier and more recent research with adult subjects showing the resistance of equivalence relations to modification via the manipulation of baseline conditional discriminations (Pilgrim & Galizio, 1995; Saunders et al., 1988; but see Pilgrim, Chambers, & Galizio, 1995; Spradlin et al., 1992, for evidence that equivalence responding is less resistant to change when children, rather than adults, are subjects).

These data may indicate that symmetry and transitivity (as examples of mutual and combinatorial entailment) do not always appear together but may instead represent independent stimulus-control relations. In the words of Pilgrim and Galizio (1996), "The dissociation between symmetry and transitivity patterns raises questions about the integrity of the equivalence phenomenon, as defined by cohesiveness among the properties of reflexivity, symmetry, and transitivity" (p. 177). This view provides a contrast to previous definitions of equivalence classes, that of congruent patterns of responding on tests of symmetry and transitivity (Sidman & Tailby, 1982). For Pilgrim and Galizio, therefore, the formation of equivalence classes may be due to a correspondence between multiple, smaller behavioral units. Similar arguments have been made by relational frame theorists (e.g., Roche et al., 1997; Wilson & Hayes, 1996). The feedback procedures employed in the current study may prove to be useful in analyzing the component properties of relational frames. That is, if relational frames are composed of component operant classes, then it should be possible to gain control over the component operants by employing differential consequences. That was the purpose of Experiment 4.

Procedure

The MTS training and testing procedures for Experiment 4 were broadly similar to Experiments 1 and 3 (i.e., two-choice MTS), except that Subjects 12 and 13 were trained and tested for both mutual *and* combinatorial entailment. Training was identical to Experiments 1 and 3 (e.g., four two-choice MTS trial types were used and eight consecutively correct responses were required to complete the training). During the initial cycles of training and testing, training was followed directly by a test for mutual entailment, and then by a test for combinatorial entailment. The tests for mutual entailment consisted of four MTS trial types involving B-A and C-B relations (i.e., B1-A1/not A2; B2-A2/not A1; C1-B1/not B2; C2-B2/not B1). These trial types were presented in a quasirandom order, with each of the four trial types occurring five times across 20 trials. The tests for combinatorial entailment were similar to those employed in Experiments 1 and 3 (i.e., four trial types involving the A-C and C-A relations, presented five times across 20 trials). The computer screen remained blank for 30 s between the last trial of the mutual entailment test and the first trial of the combinatorial entailment test. No test-performance feedback was delivered at this point in the experiment. This cycle of training and testing, using a novel stimulus set for each cycle, continued until a subject produced both mutual and combinatorial entailment across three successive stimulus sets. In effect, subjects had to produce at least four out of five class-consistent responses on each of the eight trial types, across three consecutive exposures to the tests for mutual and combinatorial entailment (i.e., the mastery criterion employed in Experiment 1 was applied to both mutual and combinatorial entailment tests). This provided a baseline of derived relational responding (Condition A) from which to determine the effects of inaccurate test feedback on mutual and combinatorial entailment.

Once a baseline had been established, both subjects continued to receive cycles of training and testing, with new stimulus sets employed for each cycle, but different types of test-performance feedback were introduced across the 2 subjects. Subject 12 received accurate feedback following tests for mutual entailment (see below for details) and inaccurate feedback following tests for combinatorial entailment (Condition B). Subject 13 received inaccurate feedback following tests for mutual entailment (see below for details) and accurate feedback following tests for combinatorial entailment (Condition C). When a subject produced at least four of five feedback-consistent responses on each of the eight trial types across three consecutive exposures to each test, the alternative condition was introduced, again using novel stimulus sets with each cycle (i.e., Subject 12 switched to Condition C, and Subject 13 switched to Condition B). When a subject once again produced at least four of five feedback-consistent responses on each of the eight trial types across three consecutive exposures to each test, the baseline condition (A) was reintroduced. That is, the test-performance feedback was omitted after each exposure to the tests for mutual and combinatorial entailment. Once again, novel stimulus sets were used with each cycle of training and testing. Because we were uncertain what response patterns would emerge during this return to baseline, we simply exposed each subject to six cycles and then examined the data. If a stable performance had not emerged across the three final cycles, additional exposures were planned, but this proved to be unnecessary.

The feedback delivered after each test for mutual entailment and each test for combinatorial entailment provided differential consequences contingent upon the exact number of class-consistent responses emitted during the two tests for derived relational responding. The format for both types of feedback was identical to the format used in the previous experiments (i.e., stars presented vertically on the screen). Similar feedback contingencies operated for mutual entailment as had operated for combinatorial entailment in the previous experiments. For example, emitting the relational responses B1-A1, B2-A2, C1-B1, and C2-A2 each produced a star under accurate feedback contingencies, but produced no stars under inaccurate feedback contingencies. Similarly, emitting the relational responses B1-A2, B2-A1, C1-B2, and C2-B1 each produced a star under inaccurate feedback contingencies, but produced no stars under accurate feedback contingencies. Mutual entailment feedback always followed the test for mutual entailment, and combinatorial entailment feedback always followed the test for combinatorial entailment. Each cycle of training and testing, including the two types of feedback, was presented without interruption. In effect, the computer asked the subject to report to the experimenter only after the following general sequence had been completed: conditional discrimination training → mutual entailment test → mutual entailment feedback → combinatorial entailment test → combinatorial entailment feedback. The computer screen remained blank for 10 s between the end of the mutual entailment feedback and the presentation of the first trial of the combinatorial entailment test.

In summary, Subject 12 was exposed to an ABCA design, and Subject 13 was exposed to an ACBA design. Condition A constituted no test-performance feedback, B constituted accurate feedback for mutual entailment and inaccurate feedback for combinatorial entailment, and C constituted inaccurate feedback for mutual entailment and accurate feedback for combinatorial entailment.

Results and Discussion

The number of training trials required per stimulus set by each subject is presented in Table 4. The minimum number of trials required was 26 (Subject 12) and the maximum was 162 (Subject 13). As in the previous experiments, the number of training trials per exposure tended to vary unsystematically for each subject, but once again early training exposures tended to involve larger numbers of trials than many of the later exposures.

Performances on each of the tests for mutual and combinatorial entailment are shown in Figure 4. Derived relational responding occurred with Subject 12 in accordance with the stability criterion across Exposures 3, 4, and 5 (no feedback). With the introduction of accurate feedback following the test for mutual entailment and inaccurate feedback following the test for combinatorial entailment, intermediate responding occurred on the test for combinatorial entailment along with high levels of mutual entailment. The stability criterion was reached across Stimulus Sets 9, 10,

Table 4

Number of training trials across successive stimulus sets for each subject in Experiment 4.

Stimulus set (exposure)	Training trials	
	S12	S13
1	126	162
2	64	116
3	54	88
4	121	76
5	98	96
6	32	54
7	28	32
8	46	28
9	72	44
10	38	104
11	26	122
12	66	53
13	30	45
14	54	32
15	34	52
16	138	48
17	44	66
18	36	88
19	72	62
20	40	42
21	36	50
22	54	
23	76	
24	42	

and 11 with high levels of mutual entailment and low levels of combinatorial entailment. Following this, feedback switched and the test performances were affected across Exposures 13, 14, and 15. The mastery criterion was reached across Exposures 16, 17, and 18 with high levels of combinatorial entailment and low levels of mutual entailment. With a return to baseline (no feedback), performance on the test for mutual entailment gradually shifted back to class-consistent responding, and high levels of combinatorial entailment were maintained across the final three exposures (i.e., meeting the mastery criterion).

Derived relational responding occurred with Subject 13 in accordance with the mastery criterion across Exposures 2, 3, and 4 (no feedback). With the introduction of inaccurate feedback following the test for mutual entailment and accurate feedback following the test for combinatorial entailment, responding on the test for mutual entailment switched rapidly to low levels, with high levels of combinatorial entailment maintained. The mastery criterion was reached across Stimulus Sets 7, 8, and 9 with high levels of combinatorial entailment and low levels of mutual entailment. Following this, feedback switched and the mastery criterion was reached once again across Exposures 13, 14, and 15 with high levels of mutual entailment and low levels of combinatorial entailment. With a return to baseline (no feedback), performance on the test for combinatorial entailment shifted gradually towards class-consistent responding. Intermediate levels of mutual entailment were produced during Exposure 17, but responding returned to high levels during Exposures 19, 20, and 21 (i.e., meeting the stability criterion).

In summary, both subjects produced derived relational responding across stimulus sets without feedback. With the introduction of differential feedback following both tests, responding was either maintained or changed by the type of feedback delivered. When the test-performance feedback was removed, response patterns similar to those observed during the first baseline occurred.

GENERAL DISCUSSION

The present study supports previous research by Healy et al. (1998) demonstrating that patterns of combinatorial entailment can be manipulated using differential feedback, at least in adult human subjects. The use of multiple stimulus sets in the current study, however, has shown that derived relational responding may be interpreted as a form of *generalized* operant behavior. The present data also indicate that derived relational responding, in certain contexts, may "fracture" into independent stimulus-control relations, thus supporting the findings of a number of previous studies (e.g., Pilgrim & Galizio, 1990, 1995; Roche et al., 1997; Wilson & Hayes, 1996).

One of the interesting findings from the current research is that subjects showed levels of responding in accordance with the feedback similar to those reported by Healy et al. (1998). One might expect that more subjects should have failed to respond in accordance with the feedback when a new stimulus set was introduced for each exposure than when the same set was used across all exposures. Perhaps, however, the use of novel stimuli for each successive stimulus set eliminated nonarbitrary forms of stimulus control that may

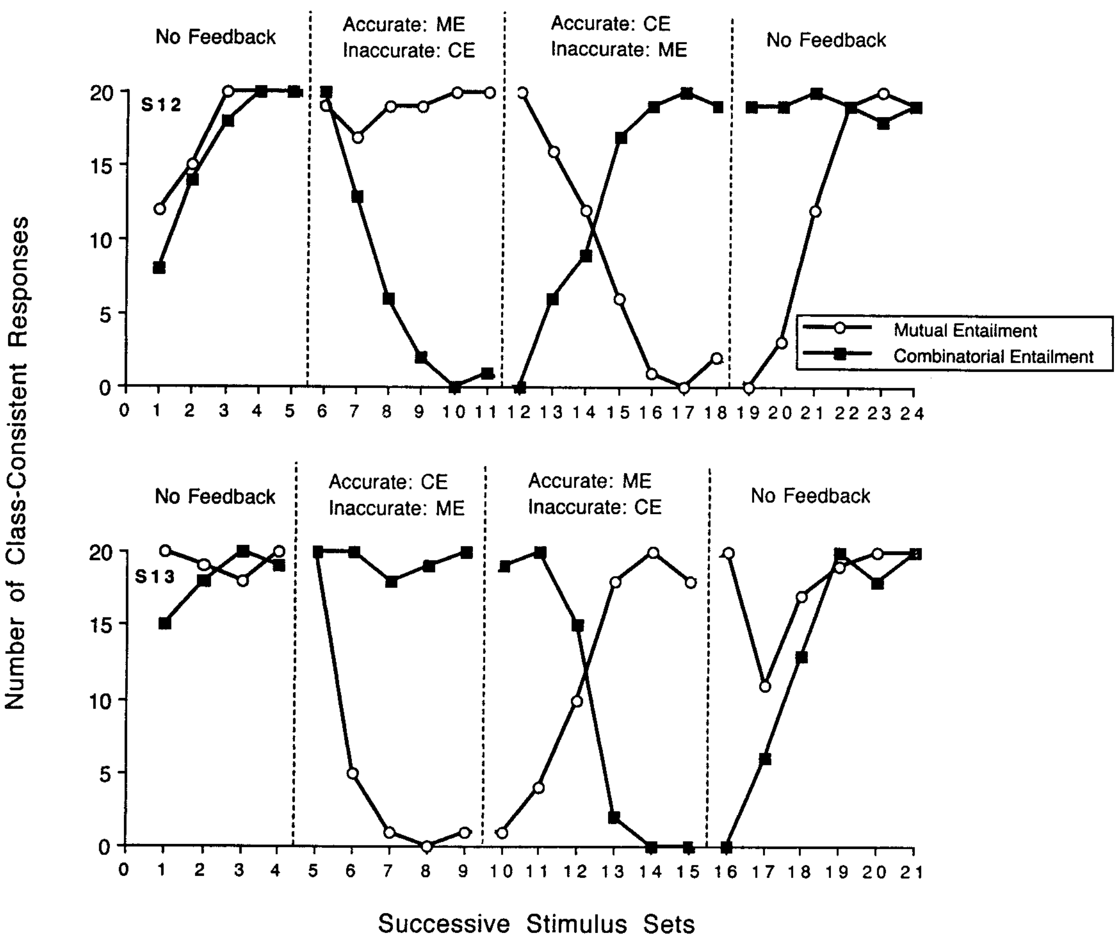

Fig. 4. Class-consistent responding across successive test exposures for each subject in Experiment 4. Each data point represents one exposure to a 20-trial test block. For each stimulus set two data points are shown, one for the mutual entailment test and one for the combinatorial entailment test. During Condition A no test-performance feedback was delivered; during Condition B accurate feedback was delivered for mutual entailment and inaccurate feedback was delivered for combinatorial entailment; during Condition C inaccurate feedback was delivered for mutual entailment and accurate feedback was delivered for combinatorial entailment.

have persisted when only one set was used. For example, during the first test (with Stimulus Set 1) a subject might have related the stimuli in terms of alphabetical order (e.g., ZID with VEK because Z and V are both towards the end of the alphabet). Using the same stimulus set again might continue this pattern, but if a new set were used this response pattern might be punished (e.g., with the second set, choosing DAX with CUG is punished). Thus, nonarbitrary forms of stimulus control would likely be filtered out across multiple sets, in a way that would not be possible with only a single set (see Hayes, Gifford, & Wilson, 1996, p. 287).

Healy et al. (1998) did not bring derived relational responding under the control of repeated feedback reversals, as has been shown with other types of operant response classes (e.g., Vaughan, 1988). In Experiment 3, however, we repeatedly reversed the type of feedback across successive exposures, and all 3 subjects rapidly shifted their levels of entailment. With sufficient exposure to the procedure, the behavior of 1 subject (11) adapted perfectly to the repeated reversals, in that he obtained the maximum number of points across the final three consecutive test exposures. In effect, this subject appeared to adopt a win-shift/lose-stay pattern of derived relational responding across successive stimulus sets. As indicated earlier, this type of performance could be interpreted as generalized operant behavior. Insofar as this was the case, it appears that the generalized operant of derived relational responding became a component operant of a win-shift/lose-stay generalized operant class. This finding is consistent with the RFT suggestion that many examples of complex human behavior may be usefully treated as generalized operants that are either composed of other generalized operants (see below) or function as the components of yet other generalized operants. In this regard, Barnes (1996) used the phrase *fractal-like* to characterize the way in which the generalized operant of relational framing could give rise to increasing orders of behavioral complexity. The data from Experiment 3 (and Experiment 4; see below) appear to provide an empirical example of this fractal-like effect (see D. Barnes-Holmes & Barnes-Holmes, 2000).

The current study provides evidence that derived relational responding is sensitive to differential consequences, and could thus be considered a form of generalized operant behavior. Insofar as this interpretation is correct, a number of possibly important implications arise for the experimental and conceptual analyses of derived relational responding. Consider, for example, the data from Experiment 4 showing that response patterns in accordance with mutual and combinatorial entailment could be fractured using the appropriate feedback contingencies. These data indicate that relational frames, as behavioral units, may be broken down into component operants by the appropriate reinforcement contingencies. These findings are consistent with previous relational-frame research that showed that mutual entailment developed before combinatorial entailment in the behavioral repertoire of a young child (Lipkens et al., 1993). In other words, it appears that these relational operants (i.e., relational frames) do not develop in whole cloth, but are established as component operants that are then combined through interaction with the contingencies operating in the verbal community. The current findings indicate that even when the component operants have combined into relational frames, they may be broken down again in an experimental context. This finding supports the RFT account of derived relational responding as a form of generalized operant behavior (D. Barnes-Holmes & Barnes-Holmes, 2000).

Other aspects of the data from Experiment 4 are also consistent with RFT. Consider the fact that both of the subjects in Experiment 4 showed resurgence of their earlier class-consistent responding when feedback was withdrawn during the return to baseline. According to RFT, the generalized operant of equivalence responding probably has the longest and richest history of reinforcement of all the relational operants (Barnes & Roche, 1996), and thus, in the absence of alternative controlling stimuli, equivalence responding becomes the most likely outcome (see Roche & Barnes, 1996). From the RFT perspective, this is particularly likely when using an MTS procedure because it may readily function as a cue for equivalence, based, for example, on its use in educational settings to establish word–symbol equivalences (D. Barnes-Holmes & Barnes-Holmes, 2000). The resurgence of equivalence responding during the return to baseline, when consequential control for nonequivalence responding was removed, is therefore consistent with the RFT interpretation of equivalence as a generalized operant class with a rich and protracted history of reinforcement. The present findings are also consistent with earlier RFT-based research reported by Wilson and Hayes (1996) that showed systematic behavioral resurgence of previously observed equivalence classes in a laboratory setting. Of course, the present data were obtained with only 2 subjects in a single experiment, and thus further research will be required to determine how general the resurgence effect might be.

Although the findings of the present study are consistent with RFT, three important issues should be addressed. First, although the present research showed consequential control over derived relational responding, such responding had almost certainly been established in the behavioral repertoires of the adult subjects during their preexperimental histories (consistent with this suggestion, Subjects 12 and 13 demonstrated class-consistent responding *before* test-performance feedback was introduced). From this perspective, therefore, the feedback influenced preexisting repertoires of generalized operant behavior, and did not establish those repertoires ab initio. Consequently, the current data do not provide strong evidence for the RFT view that derived relational responding is established, in the first instance, as generalized operant behavior. Nevertheless, given the paucity of research on the effects of consequences on equivalence class formation, demonstration studies such as this one constitute a first step in exploring the utility of the RFT interpretation of derived relational responding. Furthermore, preliminary research from the Maynooth laboratory has recently established, ab initio, patterns of derived relational responding in young children using operant techniques similar to those reported here (Y. Barnes-Holmes, Barnes-Holmes, & Roche, in press).

Second, the postsession feedback used as consequences for generalized relational responding was unusual in this area of research. Specifically, the operations necessary to establish the postsession stars as reinforcers remain unknown, but it seems likely that some type of instructional or verbal control was involved. Although some might argue that this possibility in some way diminishes the importance of the feedback effects demonstrated in the current study, we would argue that instructional control is difficult to avoid in any experiment using verbally able participants (including standard equivalence experiments). Due to their biological and personal histories, such subjects engage in vast amounts of verbal behavior, and any psychology experiment is therefore teeming with stimuli that may function as instructional cues. In fact, one of the major purposes of the current study, as an example of RFT research, was to contribute to developing a functional analysis of instructional or verbal control itself (see the final paragraph of this discussion). As such, we do not see the "intrusion" of instructional or verbal control as a "contaminant" to be removed from our experiments, but as the sine qua non of our research agenda (see Hayes & Barnes-Holmes, in press).

Third, although the findings of the current study are consistent with the RFT view that derived relational responding may be usefully approached as generalized operant behavior, the data do not directly contradict alternative theoretical positions, such as those developed by Sidman (1994) and Horne and Lowe (1996). It was not our purpose, however, to conduct a definitive study that would render one or more theoretical accounts invalid. In fact, we do not believe that such a study will ever be forthcoming (see Barnes, 1994; Barnes & Roche, 1996). Nevertheless, in due course one of the currently available theoretical accounts may be found to pertain to a broader array of data or to suggest a larger number of new and useful empirical investigations. In this regard, we simply note that RFT set the occasion for the present study and others like it (e.g., Y. Barnes-Holmes et al., in press; Healy et al., 1998). Perhaps more important, however, are the broader theoretical issues arising from the current work, and we will briefly consider these in closing.

At a conceptual level, approaching relational responding as generalized operant behavior may provide new and useful ways of conceptualizing complex human behavior typically referred to as language and cognition. From the perspective of RFT, relational activities are considered to be the functional-analytic bedrock of these complex behavioral repertoires. Relational frame theory avoids the typical approach to language and cognition taken by cognitive psychology, which has tended to emphasize "content" by the training of specific words or the acquisition of specific concepts applicable in the real world. For RFT the key focus should be on the relational activities per se, rather than on particular words or concepts. In the present study, for example, large numbers of nonsense syllables, with no clear reference to real-world events, were used for training and testing derived relational responding. Perhaps a similar approach could be taken in educational settings in which learners are trained in both real-world concepts and in various types of relational responding. Consider a classroom setting in which games could be designed to improve the flexibility of a child's relational responding. Questions could be asked such as: "If X is the same as Y, and Y is the same as Z, do I like Z if I like X?" Although there is currently little evidence to support this type of approach, the data reported in the present study suggest that the concept of the generalized operant may be central to the functional analysis of derived relational responding and human language and cognition.

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.

Barnes, D. (1994). Stimulus equivalence and relational frame theory. *The Psychological Record, 44,* 91–124.

Barnes, D. (1996). Naming as a technical term: Sacrificing behavior analysis at the altar of popularity. *Journal of the Experimental Analysis of Behavior, 65,* 264–267.

Barnes, D., Browne, M., Smeets, P. M., & Roche, B. (1995). A transfer of functions and conditional transfer of functions through equivalence relations in three- to six-year-old children. *The Psychological Record, 45,* 405–430.

Barnes, D., & Hampson, P. (1993). Stimulus equivalence and connectionism: The implications for behavior analysis and cognitive science. *The Psychological Record, 43,* 617–638.

Barnes, D., & Hampson, P. (1997). Connectionist models of arbitrarily applicable relational responding: A possible role for the hippocampal system. In J. Donahoe & V. P. Dorsel (Eds.), *Neural-network models of cognition* (pp. 496–521). Amsterdam: Elsevier.

Barnes, D., Hegarty, N., & Smeets, P. M. (1997). Relating equivalence relations to equivalence relations: A relational framing model of complex human functioning. *The Analysis of Verbal Behavior, 14,* 1–27.

Barnes, D., & Roche, B. (1996). Relational frame theory and stimulus equivalence are fundamentally different: A reply to Saunders' commentary. *The Psychological Record, 46,* 489–507.

Barnes-Holmes, D., & Barnes-Holmes, Y. (2000). Explaining complex behavior: Two perspectives on the concept of generalized operant classes. *The Psychological Record, 50,* 251–265.

Barnes-Holmes, Y., Barnes-Holmes, D., & Roche, B. (in press). Multiple-exemplar training and the development of relational framing in young children. In S. C. Hayes & D. Barnes-Holmes (Eds.), *Relational frame theory: Creating an alternative behavioral agenda in language and cognition.* Reno, NV: Context Press.

Carrigan, P. F., Jr., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 58,* 183–204.

Catania, A. C. (1996). On the origins of behavior structure. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 3–12). Amsterdam: Elsevier Science.

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.

Dube, W. V., McIlvane, W. J., & Green, G. (1992). An analysis of generalized identity matching-to-sample test procedures. *The Psychological Record, 42,* 17–28.

Dymond, S., & Barnes, D. (1995). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness, more than, and less than. *Journal of the Experimental Analysis of Behavior, 64,* 163–184.

Dymond, S., & Barnes, D. (1996). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness and opposition. *The Psychological Record, 46,* 271–300.

Estes, W. K. (1971). Reward in human learning: Theoretical issues and strategic choice points. In R. Glazer (Ed.), *The nature of reinforcement* (pp. 16–36). New York: Academic Press.

Gewirtz, J. L., & Stengle, K. G. (1968). Learning of generalized imitation as the basis for identification. *Psychological Review, 5,* 374–397.

Hayes, S. C. (1991). A relational control theory of stimulus equivalence. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 19–40). Reno, NV: Context Press.

Hayes, S. C. (1992). Verbal relations, time and suicide. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 109–118). Reno, NV: Context Press.

Hayes, S. C. (1994). Relational frame theory: A functional approach to verbal events. In S. C. Hayes, L. J. Hayes, M. Sato, & K. Ono (Eds.), *Behavior analysis of language and cognition* (pp. 9–30). Reno, NV: Context Press.

Hayes, S. C., & Barnes, D. (1997). Analyzing derived stimulus relations requires more than the concept of stimulus class. *Journal of the Experimental Analysis of Behavior, 68,* 235–233.

Hayes, S. C., & Barnes-Holmes, D. (Eds.). (in press). *Relational frame theory: Creating an alternative behavioral agenda in language and cognition.* Reno, NV: Context Press.

Hayes, S. C., Gifford, E. V., & Wilson, K. G. (1996). Stimulus classes and stimulus relations: Arbitrarily applicable relational responding as an operant. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 279–299). Amsterdam: Elsevier.

Healy, O., Barnes, D., & Smeets, P. M. (1998). Derived relational responding as an operant: The effects of between-session feedback. *The Psychological Record, 48,* 511–536.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65,* 185–241.

Lipkens, G., Hayes, S. C., & Hayes, L. J. (1993). Longitidunal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology, 56,* 201–239.

McIlvane, W. J., Dube, W. V., & Callahan, T. D. (1995). Attention: A behavior analytic perspective. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory and executive function* (pp. 97–117). Baltimore: Paul H. Brookes.

McIlvane, W. J., Dube, W. V., Kledaras, J. B., Iennaco, F. M., & Stoddard, L. T. (1990). Teaching relational discriminations to individuals with mental retardation: Some problems and possible solutions. *American Journal on Mental Retardation, 95,* 283–296.

McIlvane, W. J., Kledaras, J. B., Munson, L. C., King., K. A., de Rose, J. C., & Stoddard, L. T. (1987). Controlling relations in conditional discrimination and matching by exclusion. *Journal of the Experimental Analysis of Behavior, 48,* 187–208.

Neuringer, A. (1986). Can people behave "randomly"?: The role of feedback. *Journal of Experimental Psychology: General, 115,* 62–75.

Pilgrim, C., Chambers, L., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: II. Children. *Journal of the Experimental Analysis of Behavior, 63,* 239–254.

Pilgrim, C., & Galizio, M. (1990). Relations between baseline contingencies and equivalence probe performances. *Journal of the Experimental Analysis of Behavior, 54,* 213–224.

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

Pilgrim, C., & Galizio, M. (1996). Stimulus equivalence: A class of correlations or a correlation of classes. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 173–197). Amsterdam: Elsevier.

Pilgrim, C., & Galizio, M. (in press). Stimulus equivalence and units of analysis. In J. C. Leslie & D. E. Blackman (Eds.), *Experimental and applied analyses of human behavior.* Reno, NV: Context Press.

Poulson, C. L., Kymissis, E., Reeve, K. F., Andreatos, M., & Reeve, L. (1991). Generalized vocal imitation in infants. *Journal of Experimental Child Psychology, 51,* 267–279.

Pryor, K. W., Haag, R., & O'Reilly, J. (1969). The creative porpoise: Training for novel behavior. *Journal of the Experimental Analysis of Behavior, 12,* 653–661.

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

Roche, B., & Barnes, D. (1997). A transformation of respondently conditioned stimulus function in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior, 67,* 275–300.

Roche, B., Barnes, D., & Smeets, P. M. (1997). Incongruous stimulus pairing and conditional discrimination training: Effects on relational responding. *Journal of the Experimental Analysis of Behavior, 68,* 143–160.

Roche, B., Barnes-Holmes, D., Smeets, P. M., Barnes-Holmes, Y., & McGeady, S. (2000). Contextual control over the derived transformation of discriminative and sexual arousal functions. *The Psychological Record, 50,* 267–291.

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

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.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology.* New York: Basic Books.

Sidman, M. (1987). Two choices are not enough. *Behavior Analysis, 22,* 11–18.

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

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.

Spradlin, J. E., Saunders, K. J., & Saunders, R. R. (1992). The stability of equivalence classes. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 29–42). Reno, NV: Context Press.

Steele, D. L., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior, 56,* 519–555.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349–367.

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

Wilson, K. G., & Hayes, S. C. (1996). Resurgence of derived stimulus relations. *Journal of the Experimental Analysis of Behavior, 66,* 267–281.

Received April 28, 1999
Final acceptance June 6, 2000