JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2007, **40,** 45–71 NUMBER 1 (SPRING 2007)

ARBITRARILY APPLICABLE COMPARATIVE RELATIONS: EXPERIMENTAL EVIDENCE FOR A RELATIONAL OPERANT

NICHOLAS M. BERENS AND STEVEN C. HAYES

UNIVERSITY OF NEVADA, RENO

Arbitrarily applicable derived relational responding has been argued by relational frame theorists to be a form of operant behavior. The present study examined this idea with 4 female participants, ages 4 to 5 years old, who could not perform a series of problem-solving tasks involving arbitrary more than and less than relations. In a combined multiple baseline (across responses and participants) and multiple probe design (with trained and untrained stimuli), it was shown that reinforced multiple-exemplar training facilitated the development of arbitrary comparative relations, and that these skills generalized not just across stimuli but also across trial types. The sequence of training identified potential prerequisites in the development of comparative relations (e.g., nonarbitrary comparative relations). Taken as a whole, the present data, along with previous work by others in this area, suggest that relating arbitrary events comparatively is an operant. The implications of this conclusion for the analysis of complex behavior are discussed.

DESCRIPTORS: relational operants, relational frames, comparative relations, relational frame theory, verbal behavior, multiple-exemplar training

---

Understanding language ability is one of the greatest challenges in behavior analysis. Relational frame theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001a) provides a comprehensive approach to this challenge. Suppose, for example, that a typically developing child is told that "Jack is faster than Bob" and "Mike is faster than Jack." From these two simple statements the child is able to infer that (a) Bob is slower than Jack, (b) Jack is slower than Mike, (c) Mike is faster than Bob, and (d) Bob is slower than Mike. Furthermore, if this child is told that "Jack is too slow to catch the rooster," he or she may be able to tell us that Bob is also too slow to catch the rooster. What are the contingencies that select and shape this type of responding? Finding an answer to that question is at the core of RFT, and is the primary purpose of the current investigation.

These data were part of the first author's master's thesis conducted under the supervision of the second author. We thank Joe Spradlin and Chris Ninness for their thoughtful reviews of the manuscript.

Address correspondence to Nicholas Berens, Psychology Department, University of Nevada, Reno, Reno, Nevada 89557.

doi: 10.1901/jaba.2007.7-06

A broad body of evidence exists to support RFT concepts, but more needs to be done in two major areas. First, more direct experimental evidence on the operant nature of relating is necessary before relational operants will be fully admitted into the conceptual armamentarium of behavior analysis. Second, a vast amount of applied work needs to be done to test the pragmatic implications of RFT. These two needs come together in some areas of applied behavior analysis. For example, behavioral education focused on relational tasks can provide evidence both on their operant nature and on the applied relevance of such performances.

Although RFT is becoming better known and RFT studies have begun to appear in this journal (e.g., Murphy, Barnes-Holmes, & Barnes-Holmes, 2005; Ninness et al., 2005; Rehfeldt & Root, 2005), it has a technical vocabulary that is necessary for clarity about the operant unit being discussed. Applications of work on derived stimulus relations have appeared for many years (e.g., de Rose, de Souza, & Hanna, 1996; de Rose, de Souza, Rossito, & de Rose, 1992; Joyce & Wolking, 1989; Matos & d'Oliveira, 1992; Stromer &

MacKay, 1992; Stromer, MacKay, & Stoddard, 1992), but these have used the language of stimulus equivalence classes or exclusion, neither of which is adequate to cover nonsymmetrical forms of derived stimulus relations. Thus, we will briefly review the concept of a relational frame, and then describe the theoretical and applied importance of work on their acquisition through multiple-exemplar training.

Relational Frames

If selecting Stimulus B and Stimulus C in the presence of Stimulus A has been reinforced, most individuals will subsequently emit a range of derived responses that were not part of the specific training: selecting A or C in the presence of B and B in the presence of C and vice versa. These are standard characteristics of stimulus equivalence, the most commonly studied relational frame, and reflective of its defining features (i.e., reflexivity, symmetry, transivity— see Steele & Hayes, 1991). When relations other than equivalence are of interest, the situation is more complex, and the specific derived relational response will depend on the relational context provided during training.

For example, suppose a child who has learned to respond appropriately to the cues "more than" and "less than" is presented with this same network of stimuli, but selecting B given A is reinforced in the presence of the cue more than and selecting C given A is reinforced in the presence of the less than cue. A more complex set of derived relational responses may now be predicted. For example, although selecting B given A was reinforced in the presence of more than, A will likely be selected given B only in the presence of less than. This is not *symmetry*, and a more generic term is needed: RFT uses the term *mutual entailment*.

Similarly, selecting B given C will only be likely in the presence of more than as a result of a combination of a mutually entailed more than relation ($A > C$ resulting from the trained $C < A$ relation) and a trained more than relation ($B > A$). This combination is neither *symmetry* nor *transitivity*, and thus a more generic term is needed: RFT uses the term *combinatorial entailment*. Because none of these relations are based solely on formal properties, there must be cues (in this case, more than or less than) that specify the trained and derived relations among the stimuli in the case of arbitrarily applicable relational responses. In RFT these are denoted by the abbreviation C_{rel}, for relational contextual cues.

Finally, if A has a psychological function (e.g., suppose it was a conditioned reinforcer), in contexts that make that function relevant (RFT uses the abbreviation C_{func}), it is likely that B will function more as a reinforcer than C, and so on. This active and relative change is not *transfer*, and thus a more generic term is needed: RFT terms this phenomenon the *transformation of stimulus function*.

Behaviors with all of these features established by operant learning are forms of *arbitrarily applicable relational responding*, and specific types (e.g., relations of difference, opposition, comparison, etc.) are called *relational frames*.

The Applied and Basic Relevance of Relational Operants

Until recently, RFT researchers examined the idea that there are relational operants through indirect means. Derived stimulus relations were shown by RFT researchers to develop over time (Lipkens, Hayes, & Hayes, 1991), to come under contextual control (Dymond & Barnes, 1995; Steele & Hayes, 1991; Wulfert & Hayes, 1988), and to be controlled by consequences (Healy, Barnes, & Smeets, 1998; Healy, Barnes-Holmes, & Smeets, 2000; Wilson & Hayes, 1996), but none of this provided direct confirmation of the operant nature of derived stimulus relations.

Examining the impact of an experimentally manipulated history of reinforcement with derived stimulus relations is the proper direct test, but this was difficult with frames of coordination (i.e., equivalence relations) because these develop so early (Lipkens et al., 1991) and multiple-exemplar training with infants is technically challenging. Supportive data are emerging even here (Luciano, Becerra, & Valverde, unpublished manuscript), but a bigger change has been to focus on more advanced types of relational frames with older children (Y. Barnes-Holmes, Barnes-Holmes, Roche, & Smeets, 2001a, 2001b; Y. Barnes-Holmes, Barnes-Holmes, & Smeets, 2004; Y. Barnes-Holmes, Barnes-Holmes, Smeets, Strand, & Friman, 2004).

In this area the basic and applied questions raised by RFT come together. Until recently, RFT applications have been studied in the form of a clinical research program in acceptance and commitment therapy (ACT; Hayes, Strosahl, & Wilson, 1999; see Hayes, Luoma, Bond, Masuda, & Lillis, in press, and Hayes, Masuda, Bissett, Luoma, & Guerrero, 2004, for recent reviews of the ACT evidence). That is beginning to change as behavior analysts begin to apply RFT concepts to areas such as education (Ninness et al., 2005), and language learning (Murphy et al., 2005; Rehfeldt & Root, 2005).

Relational abilities have abundant applied significance, a prime example of which involves the cognitive abilities of children. A frame of comparison is a good example. Most complex organisms can readily learn comparisons based on relative physical properties such as size (e.g., Andrews & Halford, 1998; Lowenkron, 1989; Wright & Dowker, 2002; see Reese, 1968, for a book-length review). Such nonarbitrary relations may initially dominate over arbitrary forms in humans as well. For example, a young child who has learned directly to treat coins as a conditioned reinforcer may prefer a nickel over a dime because of its relative physical size. But as children develop they need to learn to evaluate one event relative to another simply by social attribution, not necessarily direct experience. For example, as an arbitrarily applicable comparative relation emerges, an older child will prefer a dime over a nickel because a dime is more than a nickel by social attribution.

Y. Barnes-Holmes, Barnes-Holmes, Smeets, Strand, and Friman (2004) published the first study showing that arbitrarily applicable comparative relations can be trained using multiple exemplars. Three children ages 4 to 5 years old were presented with two or three coins on a piece of paper (e.g., A-B-C), were told the relative values of each, and were asked which one they would use to buy candy. Baseline tested both mutual entailment and combinatorial entailment. During baseline, all participants responded below 50% accuracy. Following baseline, participants were exposed to a program of reinforced multiple-exemplar training of increasing complexity: (a) more than with 2 coins, (b) less than with 2 coins, (c) more than with 3 coins, and (d) less than with 3 coins. To clarify the training procedures, we will describe a more than trial involving three coins.

Three coins were presented horizontally in front of the child (A-B-C). The experimenter said, "This [pointing to Coin A] is more than this [pointing to Coin B], and this [pointing to Coin B] is more than this [pointing to Coin C]. Which would you use to buy more candy?" If the child pointed to Coin A, the experimenter provided reinforcement. Once a participant reached 90% or better on a particular relation (e.g., more than with two coins), he or she was exposed to training on the next level (e.g., less than with two coins). When all four trial configurations had been trained, each participant was reexposed to a baseline condition involving novel stimulus sets. Results indicate that after the multiple-exemplar training across the different relations described above, participants responded at above 90% accuracy during this baseline condition. (Y. Barnes-Holmes, Barnes-Holmes, Smeets, Strand, and Friman, 2004, included several subsequent procedures that evaluated generalizations and the sensitivity to contextual control of these trained responses. The current study was concerned only with this first portion of the study; therefore, the other details of the earlier study will not be elaborated.)

There are limitations to the Y. Barnes-Holmes, Barnes-Holmes, Smeets, Strand, and Friman (2004) study, however. In their study, multiple relational features (i.e., both mutual and combinatorial entailment) were simultaneously established, so the necessary and sufficient aspects of relational training required to establish the repertoire are not known. Because all trial types were trained, it is not known whether successful posttesting involved only generalization to new stimuli or also generalization to new trial types. In addition, because only linear trial types were employed (e.g., A > B > C), the tests of derived relations within comparative stimulus networks were somewhat limited. Finally, the baselines were relatively short and the impact of training relatively quick, which raises the possibility that the training methods merely established a context for the display of existing behavior rather than showing the acquisition of new behavior.

The purpose of the present study was to replicate and extend the findings of Y. Barnes-Holmes, Barnes-Holmes, Smeets, Strand, and Friman (2004). Specifically, this study evaluated the degree to which multiple-exemplar training can be used to establish derived relational responding in accordance with a comparative frame. To address the question, procedural and methodological variances from the previous study were needed. Specifically, the current study systematically tested the impact of each phase of training on the entire comparative frame, employed nonlinear trial types, and provided more elaborate and lengthier baseline trial blocks. These modifications to the original procedure were made to isolate more precisely the sources of control and to determine more clearly the degree to which multiple-exemplar training facilitates the development of arbitrarily applicable derived relational responding. A successful demonstration that arbitrary comparative relations can be trained as an operant would strengthen the central thesis of RFT and expand its basic and applied implications.

METHOD

Procedure

Participants

Participants were 4 typically developing girls (Laura, Valerie, Emma, and Sally) whose parents responded to a flyer posted on the campus of the University of Nevada, Reno. During an initial preexperimental meeting, the participants' primary caregivers were given an informal questionnaire regarding their child's toy preference and the participants were administered the Vineland Adaptive Behavior Scale (VABS). Although a more direct test of verbal abilities may have been desirable, the VABS was used to estimate the participants' abilities without exhausting the participants prior to their participation in this potentially demanding study. One additional participant was excluded from the study because he responded with perfect accuracy on all baseline trial blocks. The ages at initiation and completion, sessions to completion, and VABS performances (expressed as receptive and expressive verbal age) of the participants are shown in Table 1.

Setting and Stimuli

During all sessions, the participant was seated at a small table to the right of the experimenter. When integrity data were collected, the participant was seated between the experimenter and secondary observer such that neither the child nor the secondary observer could see the experimenter's data sheet. Sessions for Laura and Valerie were conducted in a small therapy room on the campus of the University of Nevada, Reno. Sessions for Emma and Sally were conducted in rooms in their homes because of transportation difficulties. In these rooms, the tables were placed against a blank wall so as to minimize distractions. In addition,

Table 1
Demographic Information and Sessions to Completion

Participant	Age at initial session	Verbal age (receptive)	Verbal age (expressive)	Sessions to completion	Age at completion
Sally	4 years 7 months	5 years 2 months	5 years 3 months	12	4 years 9 months
Emma	3 years 11 months	4 years 6 months	4 years 9 months	23	4 years 2 months
Valerie	4 years 10 months	3 years 8 months	3 years 10 months	38	5 years 3 months
Laura	4 years	4 years	4 years	30	4 years 6 months

sessions were conducted at a time of day when the child's primary caregiver was the only other person at home.

Experimental materials included three sets of three paper pictures (see Figure 1). For clarification purposes, each picture within a set will be referred to as either A, B, or C, but the participants were not informed of these labels. To make the stimuli more interesting for the young children, each stimulus had a unique colored picture (Pilgrim, 1998, p. 25). Other materials included a sticker page, a table and chairs, stickers, a reinforcer bin, and reinforcers (reinforcers were either small toys or candies such as M&Ms®, lollipops, small chocolates, or Skittles®).

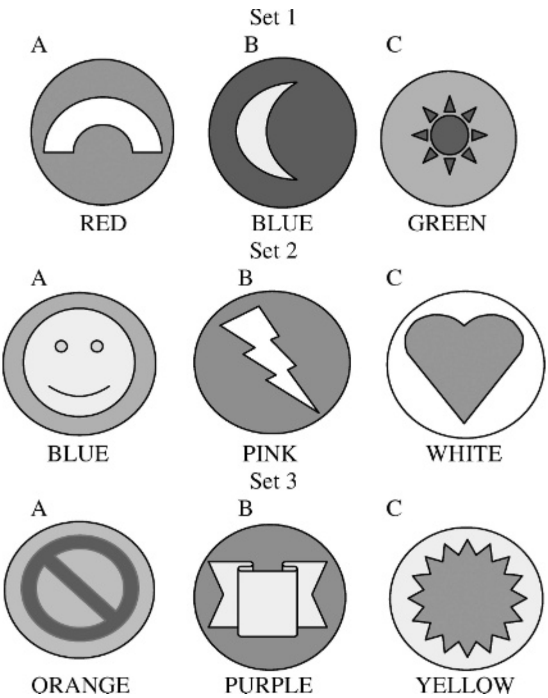

Figure 1. The pictures represent the different stimulus sets used. The color label below each stimulus represents its border color.

General Procedure

Every session began with the experimenter telling the child "We are going to play a game. Your job is to pick the picture that will buy you the most candy." Training and testing occurred in trial blocks. In each trial block, there were between 4 and 20 different trial types (see Table 2). Each trial type was designed such that each possible stimulus configuration and relation specification for each trial type was distributed equally. For example, when training more than using two stimuli (e.g., A and B), there are four possible configurations of stimulus presentation and specification of the more than relation (see Table 2). All of these possible trial types for each phase of training were presented two times per trial block. The total trials per block ranged from 8 to 40 based on the number of trial types.

The relational value of each stimulus (e.g., more than or less than) and thus value relative to other stimuli changed from trial to trial. The purpose of this procedure was to ensure that participants' responding reflected relational stimulus control exerted by the C_{rel} term rather than their history with the experimental stimuli.

During baseline, Sally and Valerie were exposed to three trial blocks of each of the three stimulus sets. Thus, they responded to nine trial blocks, or 360 trials, during baseline. Likewise, Emma and Laura were exposed to six

Table 2

	More than trials	Less than trials	Mixed nonlinear trials
Phase 1 (eight trials per block)	A(1) > B(2) B(1) > A(2) B(2) < A(1) A(2) < B(1)	none	none
Phase 2 (16 trials per block)	A(1) > B(2) B(1) > A(2) B(2) < A(1) A(2) < B(1)	A(1) < B(2) B(1) < A(2) B(2) > A(1) A(2) > B(1)	none
Phase 3 (eight trials per block)	A(1) > B(2) > C(3) C(1) > B(2) > A(3) C(3) < B(2) < A(1) A(3) < B(2) < C(1)	none	none
Phase 4 (16 trials per block)	A(1) > B(2) > C(3) C(1) > B(2) > A(3) C(3) < B(2) < A(1) A(3) < B(2) < C(1)	A(1) < B(2) < C(3) C(1) < B(2) < A(3) C(3) > B(2) > A(1) A(3) > B(2) > C(1)	none
Phase 5 (eight trials per block)	none	none	A(1) > B(2 and 4) > C(3) C(1) > B(2 and 4) > A(3) C(3) < B(2 and 4) < A(1) A(3)<B(2&4)<C(1)
Baseline and probes (40 trials per block)	A(1)>B(2) B(1)>A(2) B(2)<A(1) A(2)<B(2) A(1)>B(2)>C(3) C(1)>B(2)>A(3) C(3)<B(2)<A(1) A(3)<B(2)<C(1)	A(1)<B(2) B(1)<A(2) B(2)>A(1) A(2)>B(1) A(1)<B(2)<C(3) C(1)<B(2)<A(3) C(3)>B(2)>A(1) A(3)>B(2)>C(1)	A(1)>B(2&4)>C(3) C(1)>B(2&4)>A(3) C(3)<B(2&4)<A(1) A(3)<B(2&4)<C(1)

Note. This table details the trial types for each phase and the number of trials used for each block. The letters indicate the stimulus; its position for that trial is shown sequentially from left to right and the order in which the experimenter pointed to the stimuli is shown by the number in parentheses. For instance, the less than trial A(1) < B(2) < C(3) indicates that A was the left stimulus, B was the center stimulus, and C was the right stimulus, and that they were pointed to in that order. > and < indicate the relation specified between stimuli. Thus, in the less than example A(1) < B(2) < C(3), the experimenter said "This [pointing to A] is less than this [pointing to B] and this [pointing to B] is less than this [pointing to C]. Which one would you use to buy more candy?" The underlined stimulus indicates the correct choice for each trial type.

trial blocks of each of the three stimulus sets. Thus, Emma and Laura responded during 18 trial blocks, or 720 trials, during baseline (see Table 2). Similarly, test probes for each participant involved one trial block for each stimulus set, thus there were 120 trials per test probe.

For every trial block, the experimenter had a data sheet that also served as a script for every trial (see the Appendix). Data sheets were constructed by randomly selecting the order of the presentation of the trial types. As each trial was selected, it was then transcribed to the data sheet, which noted the arrangement of the stimuli, the relation among them, the order that the relation was to be specified, and the correct response. A secondary observer was presented with a duplicate data sheet.

During each trial, the experimenter arranged the stimuli according to the data sheet, and the child was told the relation between the stimuli (see Table 2). For example, on a trial in which A is more than B (A > B) the experimenter said "This [pointing to Picture A] is more than that [pointing to Picture B]." On a trial in which A is less than B (A < B), the experimenter said "This [pointing to Picture A] is less than that [pointing to Picture B]." On mixed nonlinear trials (Phase 5), in which A was more than B and B was more than C (A > B > C), the experimenter said "This [pointing to A] is more than that [pointing to B] and this [pointing to C] is less than that [pointing to B]." On all occasions, the child was then asked, "Which would you use to buy candy?" To clarify, Phases

1 through 4 were considered to be linear trials in that the relation that was specified proceeded from the right stimulus to the left stimulus or the left stimulus to the right stimulus. Phase 5 trials were considered mixed nonlinear trials. These trials were mixed because both a more than and a less than relation were specified by the experimenter during each trial, and they were nonlinear because the specification of these relations did not proceed in succession from the far left stimulus to the far right stimulus (see Table 2).

Sessions occurred one to three times per week and lasted between 40 and 60 min. Duration of the study varied for each participant due to differential learning and variables out of the experimenter's control (e.g., illness, holidays, or the participant fell asleep in the car ride to the session). The number of sessions and the length of time of the experiment for each participant are shown in Table 1. Following each trial block, participants were given an opportunity to take a 5- to 10-min break. At the end of the break, participants were asked if they wanted to continue. The number of trial blocks encountered for each session varied because of the breaks and the variable number of trials among the different phases (e.g., there were eight trials per block in Phase 1 and 40 trials per block in baseline and probes; see Table 2).

Response Definition and Reinforcement Procedure

Following the emission of a response, irrespective of accuracy, contingent feedback was provided and the next trial was arranged and presented. If the participant emitted a correct response (e.g., selected the picture that was more on any given trial by pointing to that picture), the experimenter provided verbal praise and presented the child with a token. If a participant emitted an incorrect response (e.g., selected a picture that was not more for a given trial, selected two pictures, or did not emit a response), the experimenter withheld the tokens and said in a gentle voice, "No, that is not it." These were the same contingencies used by Y. Barnes-Holmes, Barnes-Holmes, Smeets, Strand, and Friman (2004), and closely approximate vocal feedback statements in the natural environment. Error correction and prompting procedures were avoided to evaluate fully the effects of the experimental contingencies in establishing the targeted repertoires. No children showed external signs of distress over the contingent negative feedback following incorrect responses.

Programmed Consequences

Stickers and small candies were used as tokens. Participants kept their tokens following all trial blocks independent of meeting the goal. However, meeting the goal also resulted in an additional larger prize.

A goal was established for every trial block other than in baseline. The first goal for the first trial block for all training phases was 50% correct. The response requirement for each subsequent trial block was set at one more correct response than was achieved during the previous trial block. This was done to make the contingency more salient. If the participant responded at chance levels, she would still get her chosen tokens. However, to be able to receive the selected larger prize on a trial block, correct responding had to be better than the previous performance. This contingency was maintained until she reached 100% correct. The criterion for sessions following the attainment of 100% was maintenance of that level.

Prior to the start of a trial block, each participant was informed of her goal, was allowed to choose a prize, and was told that she would be given the prize if she met the goal. If she did not meet this goal, the prize was withheld. Because it did not seem feasible for such young children to respond without reinforcement for the duration of a long trial block (40 trials) during the baseline and probe conditions, each participant was given noncontingent reinforcers during those phases. One token was presented every eight trials for Laura and Sally; due to their greater distractibility, Valerie and Emma were given a token every five trials. All participants were allowed to choose a prize in these phases regardless of their performance.

Design

Multiple Probes Across Stimulus Sets

A multiple probe across two stimulus sets (Sets 2 and 3) was employed to evaluate the degree to which reinforced responding with the targeted stimulus set generalized to untrained stimulus sets. The initial baseline phase included all sets. Following intervention with Set 1, probes were conducted after mastery of each phase to evaluate generalization to Sets 2 and 3. Sets were tested in random order.

Multiple Baseline Across Participants

A multiple baseline across participants design was employed to control for maturation and extraexperimental contingencies. The multiple baselines were conducted in groups of 2 participants. It is more common to use groups of 3 in a multiple baseline, but the logic of the design does not require this (Hayes, Barlow, & Nelson-Gray, 1999); the present procedure led to rather extended baselines for the 2nd child. In addition, in previous research the degree to which multiple-exemplar training established a novel repertoire or served as a context for an already-existing repertoire has yet to be fully clarified (e.g., Y. Barnes-Holmes et al., 2001a, 2001b; Y. Barnes-Holmes, Barnes-Holmes, & Smeets, 2004; Y. Barnes-Holmes, Barnes-Holmes, Smeets, Strand, & Friman, 2004). The extended baselines for the 2nd participant in each dyad allows further clarification of this issue. Furthermore, having two linked dyads provides additional control. Finally, it is important to note that in the second dyad, the 2nd participant became the lead after Phase 1. This was due to quicker acquisition of the response being trained in Phase 1. Indeed, this occurrence makes the results of the multiple baseline less compelling, but the extended initial baseline provides evidence that exposure to the targeted trials was not sufficient to establish the repertoire.

Component Analysis

The component training sequence was used to evaluate the contribution of multiple-exemplar training for specific forms of comparative relations to participants' overall performance. After baseline, relational training occurred in five phases of gradually increasing complexity, which allowed an analysis of the impact of training on specific relational components. For this part of the design, a mastery criterion of 100% accuracy for two consecutive trial blocks was employed. When participants reached this mastery criterion, they were then exposed to a baseline probe across all three stimulus sets. Probes were conducted either on the same day that the mastery criterion was met or at the beginning of the next session. If responding during the baseline probe showed 80% or higher accuracy across each relational response and each stimulus set, participation in the study was ended. Otherwise, participants were exposed to the next phase in the training sequence. This ensured that deficits in relational responding on any of the trial types, or failure to generalize to untrained stimulus sets, led to additional training.

The reader is directed to Table 2 for a description of each phase including the number of trials and trial types. The phases were as follows: baseline; Phase 1: training A-B relations (more than); Phase 2: training A-B relations (more than and less than); Phase 3: training A-B-C relations with linear cues (more than); Phase 4: training A-B-C relations with linear cues (more than and less than); and Phase 5: training A-B-C relations with nonlinear cues (more than and less than).

Supplementary Nonarbitrary Training

RFT proposes that the ability to derive arbitrary relations is initially dependent on a rich history of reinforcement for responding with regard to nonarbitrary relations (e.g., Hayes, Fox, et al., 2001, p. 25). Thus, when the multiple-exemplar procedure in the arbitrary context was not successful, participants were exposed to nonarbitrary training with more than and less than. Such supplemental training employed the same consequences for correct and incorrect responses that were used in the typical training protocol. Only Valerie and Emma required additional nonarbitrary relation training.

Valerie: Phase 1.1. Trials were identical to typical training trials with the exception that during these trials, piles of pennies were placed on the picture cards. There were more pennies on the picture that was to be specified as more for that trial. For all nonarbitrary training sequences, the number of pennies used varied on each trial, but there was always a visually discernible difference in amount between the piles.

Phase 1.2. This phase involved the presentation of nonarbitrary pretrials. Valerie was presented with two piles of pennies (one large and one small) and was asked a series of five questions: (a) "Which is more?" (to help establish a more and less C_{rel}). (b) Picture cards from Set 1 were then placed under pennies and the experimenter asked, "Which one is more?" (c) "Which one would you use to buy candy?" (d) "If this [pointing to the large pile of pennies] is more than this [pointing to the smaller pile of pennies], which one would you use to buy candy?" (e) This was the same as (d) but the position of the piles and pictures was switched. If she answered each question correctly, she was then exposed to a traditional trial block.

Phase 2.1. This phase involved the use of nonarbitrary pretrials. These trials were presented as follows: (a) "Which pile of pennies has more?" (b) "Which pile of pennies has less?" (c) "Which one would you use to buy more candy?" After achieving 100% correct responding on these pretrials, Valerie was immediately exposed to a Phase 2 trial block. This pattern of nonarbitrary pretrials followed by a Phase 2 trial block continued until Valerie was exposed to Phase 2.2.

Phase 2.2. This phase was exactly like Phase 2.1 with the exception that nonarbitrary contextual cues (a big pile of pennies and a little pile of pennies) were placed on the picture cards during the Phase 2 trial blocks. Once Valerie reached 100% correct, the contextual cues were systematically faded (e.g., all but two trials would be presented with nonarbitrary contextual cues, then three trials, etc.).

Phase 2.3. Phase 2.3 involved a series of pretrial questions that were designed to promote more active responding to the stimuli such that the experimentally desired stimulus functions of more and less could be enhanced and captured. Using different-sized piles of pennies, Valerie was asked the following questions: (a) "Is this more or less?" (the experimenter pointed to one pile). (b) "Is this more or less?" (the experimenter pointed to the other pile). (c) "Which one has more?" (d) "Which one has less?" (e) "Which would you use to buy candy?" (f) "If this [pointing to a pile] is more [less] than this [pointing to the other pile], which would you use to buy candy?" (g) "If this [pointing to a pile] is more [less] than this [pointing to the other pile], which would you use to buy candy?" If she responded correctly to every question, she was then exposed to typical trials for Phase 2; if she did not answer all of the questions correctly, she was recycled through Phase 2.3.

Phase 3.1. This phase was exactly like Phase 2.3, with the exception that three piles of pennies were used and only more than trials were trained.

Phase 4.1. This phase was exactly like Phase 2.3, with the exception that three piles of pennies were used.

Phase 4.2. This phase was exactly like Phase 4.1, with the exception that when she was exposed to the standard trial blocks, the instructions were changed to include "If this one has more *pennies* than this one which would you use to buy more candy?"

Emma: Phase 1.1. This phase involved the incorporation of nonarbitrary pretrials similar to those used with Valerie in Phase 2.3 except that only more than was targeted.

Phase 1.2. This was exactly like Phase 1.1 for Valerie, with the exception that more was emphasized during the final question: "If this one [pointing to the larger pile of pennies] is more [said both longer and louder] than this one [pointing to the smaller pile of pennies] which one would you use to buy more candy?"

Phase 1.3. This was exactly like Phase 1.2, with the exception that during training only the picture that was more for a given trial was pointed to.

Interobserver Agreement

For 39% of all trials, a secondary data collector independently scored each trial as correct or incorrect based on the criteria described above. Secondary data collectors were required to reach 100% accuracy on three consecutive mock baseline sessions before they could score an experimental session. Agreement data were collected across all participants for all phase types. An agreement was scored when both observers scored a trial as being either correct or incorrect. A disagreement was scored if the observers recorded the trial differently. A percentage agreement score was calculated by dividing the total agreements by the total agreements plus total disagreements and multiplying by 100%. This resulted in a total agreement score of 99.8%.

Procedural Integrity

During trial blocks in which interobserver agreement data were collected, procedural integrity data were also collected. Three measures of integrity were scored for every trial: trial arrangement, trial presentation, and correct consequence provided. For each of these categories, either a yes or no was scored. If any item was scored as no, the entire trial was scored as incorrect. The total number of trials scored as correct were divided by the total number of trials scored. This resulted in an integrity score of 99.8%.

RESULTS

The results for all participants are described in Figures 2 through 8. The current study contained many design elements. To clarify these data and their coinciding controls, the data are displayed in multiple ways. Figures 2 and 3 depict the multiple baseline across participants design element and for the purpose of clarity show only responding on Set 1. Figures 4 through 7 present individual-participant data on each trial type. These data are depicted in this way to clarify the component analysis. Figure 8 presents data from all participants on only their baseline and probe trial blocks across all stimulus sets. These data reflect the multiple probe design element of the study.

An analysis of the baseline performances of all participants indicates that no participant showed the targeted arbitrary relational responses during baseline. Figures 2 and 3 show the data from the multiple baseline across participants on trial blocks with Set 1 for Dyads 1 and 2, respectively. These data indicate that no participants were able to respond proficiently to the relational tasks using stimuli from Set 1. Furthermore, Emma's data (Figure 2) and Laura's data (Figure 3) indicate that the detected deficits did not improve with repeated exposures to the baseline condition. Figure 8 shows baseline and baseline probe data for all participants across all sets of stimuli. The first clusters of data for each participant represent baseline performances. When responding across all sets is taken into consideration, it is clear that all participants performed

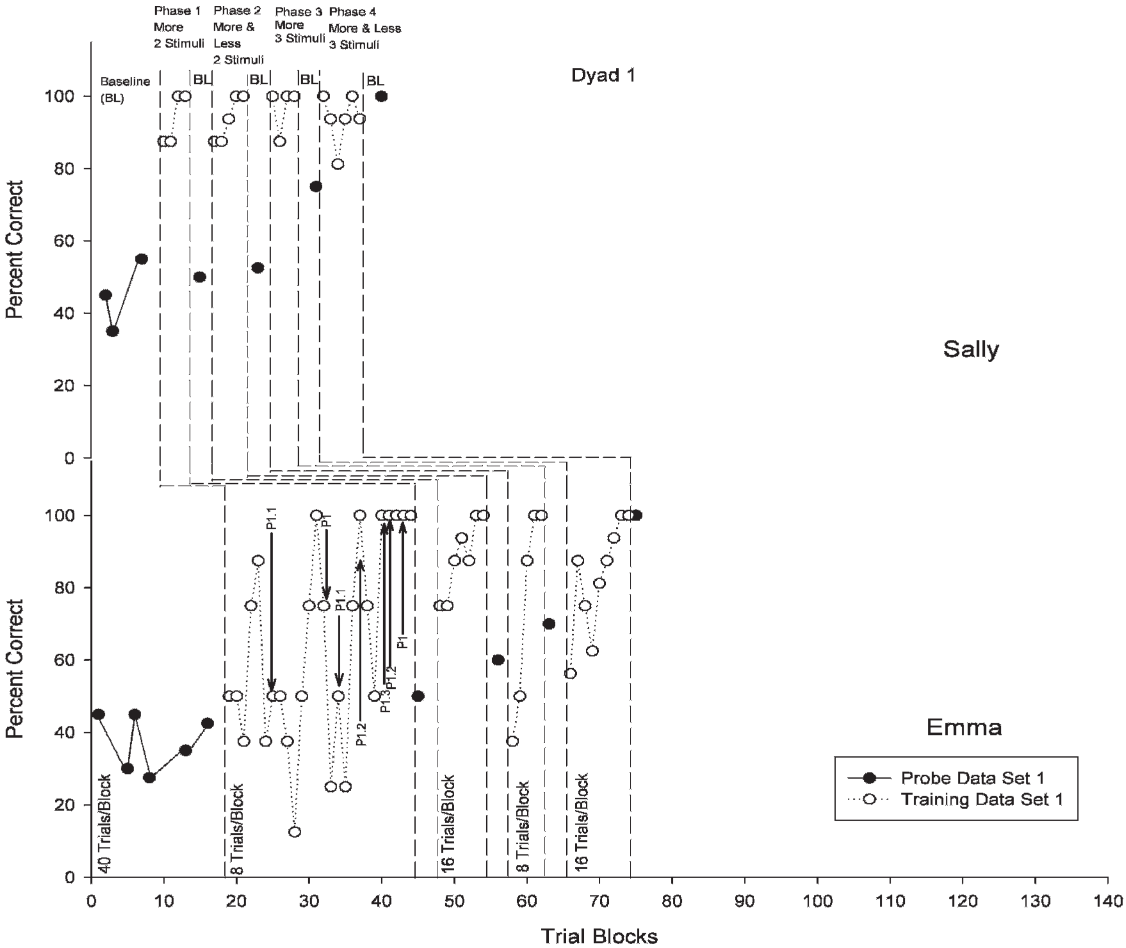

Figure 2. Dyad 1: Sally (top) and Emma (bottom). Data shown are for all stimulus sets. Along the bottom of the data series, the number of trials for each trial block of each phase is shown. The arrows indicate when particular interventions were implemented.

poorly on the relational tasks of baseline, and that there was no improvement for any participant during baseline.

It is clear that no participant demonstrated strong responding on their overall performance during baseline, but it is possible that responding was stronger with certain trial types. Figures 4 through 7 show a breakdown of each participant's responding to each trial type. The baseline data indicate that correct and incorrect responses were equally distributed across all trial types. Thus, these data indicate that these participants did not have the targeted arbitrary comparative relational responses in their repertoire. We now turn our attention to individual training and probe data.

Sally

Sally took four trial blocks to reach mastery criteria for Phase 1 (Figure 2). Her relatively rapid acquisition of the targeted relational response raises the possibility that the multiple-exemplar training served as a contextual cue for previously learned responding. Sally's responding during her second exposure to the baseline trial blocks showed no improvements across all three of the stimuli sets (Figure 2). In addition, there was no noted improvement

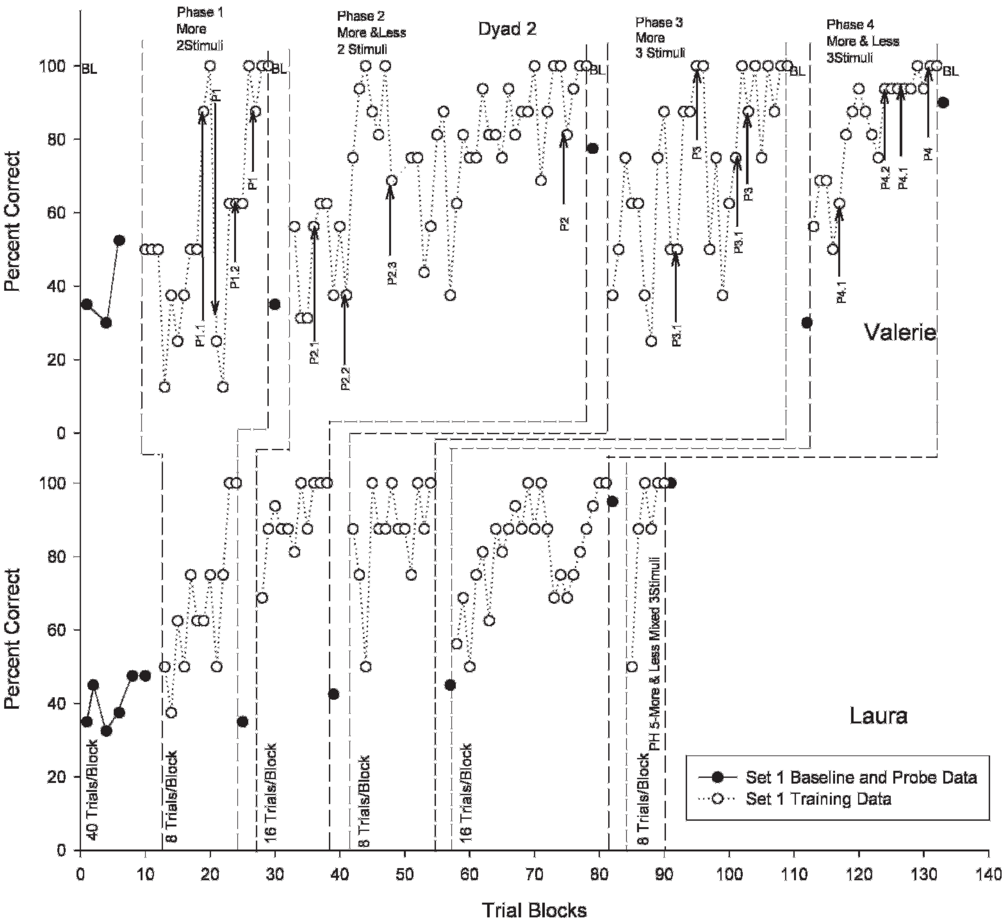

Figure 3. Dyad 2: Valerie (top) and Laura (bottom). Arrows indicate additional interventions.

across the different trial types (Figure 4). Thus, even though Sally rapidly acquired the mutually entailed relational response involving two stimuli, her responding was not maintained when reinforcement was withheld. Furthermore, it did not generalize when she was tested using other sets of stimuli. This pattern undermines the possibility that the rapid acquisition seen in Phase 1 did not represent real acquisition.

Sally required five trial blocks to reach our mastery criteria for Phase 2, again showing rapid acquisition (Figure 2). When Sally was exposed to the third baseline condition her correct responding showed increases over the previous two baseline conditions on her responding to Sets 2 and 3 (53%, 65%, and 63% correct for Sets 1, 2, and 3, respectively; see Figure 8).

Sally required four total trial blocks to reach our mastery criteria in Phase 3 (Figure 2). During her fourth exposure to baseline, she showed improvements in her correct responding across all three stimulus sets when compared to her previous exposures to baseline conditions (75%, 78%, and 73% correct for Sets 1, 2, and 3, respectively, see Figure 8). In addition to generalization to new stimulus sets, the trial types that had been targeted thus far in the experiment during training particularly improved (Figure 4).

Sally required six trial blocks to reach our mastery criteria during the fourth phase

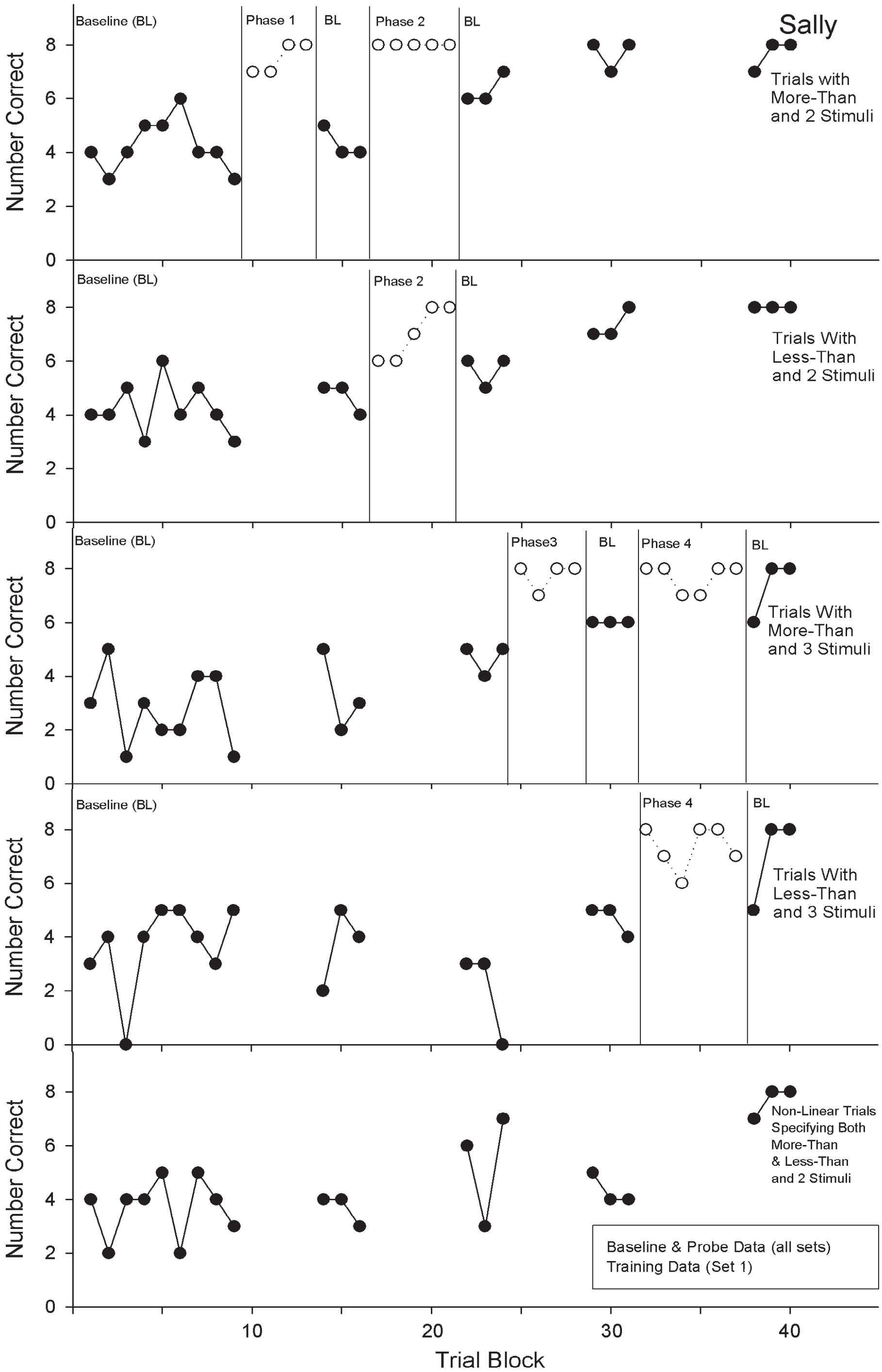

Figure 4. Results for Sally. The top graph presents data from more than linear trials with two pictures. The second graph presents data from less than linear trials with two pictures. The third graph presents data from more than linear trials with three pictures. The fourth graph contains data from less than linear trials with three pictures. The fifth graph presents data from more than and less than nonlinear trials with three pictures.

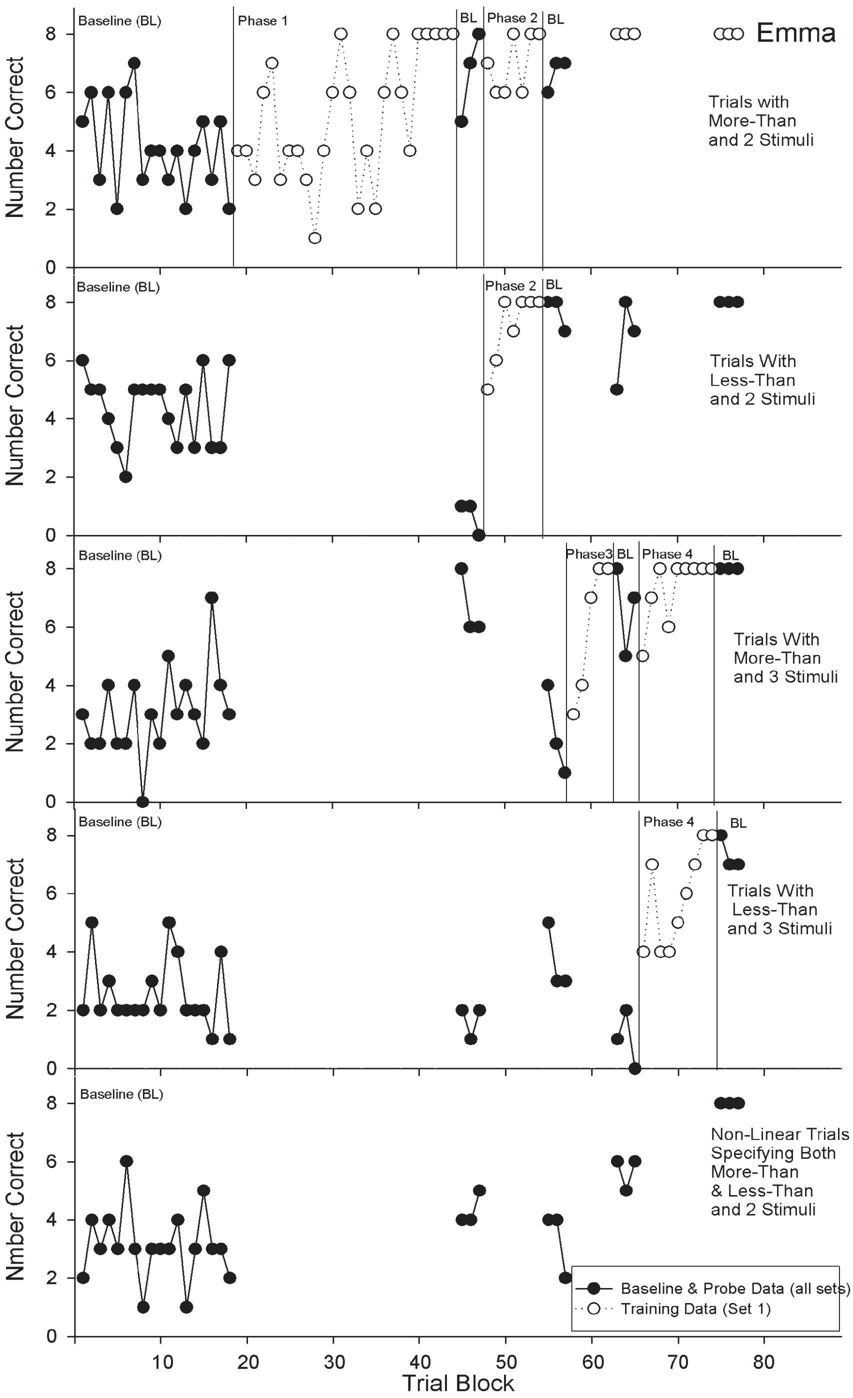

Figure 5. Results for Emma. See Figure 4 for details.

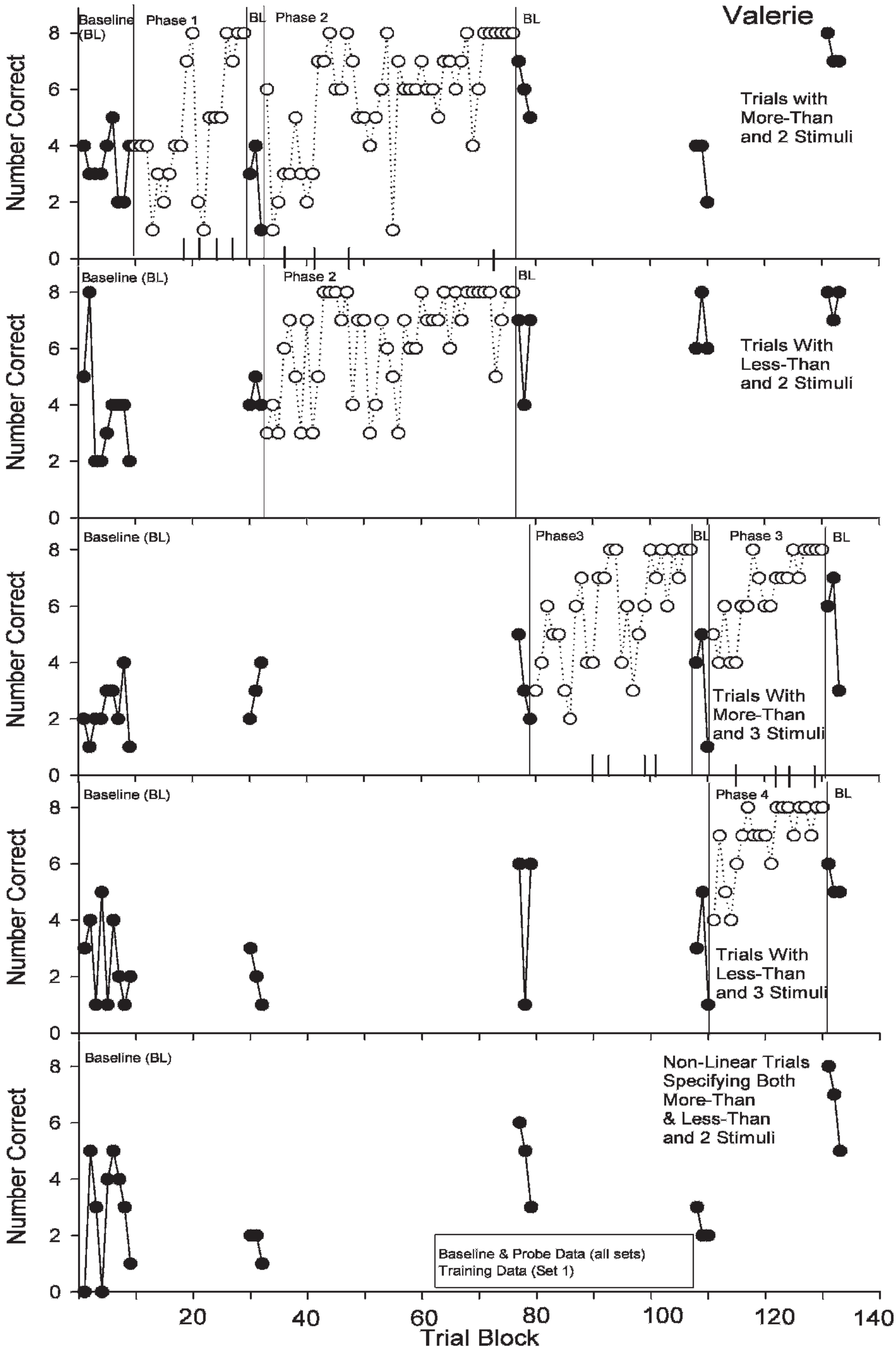

Figure 6. Results for Valerie. See Figure 4 for details. Hash marks on the abscissa indicate when nonarbitrary training occurred and on what trials it occurred. Refer to Figure 3 for specific information on the level of nonarbitrary training.

(Figure 2). Her terminal performance was 94% correct but was considered mastered because she had made only two errors across the last three trial blocks (or 48 trials; Figure 2). Her fifth and final exposure to baseline showed large improvements in responding over the previous baseline conditions (Figure 8). She answered 83% of the questions correctly with Set 2 and

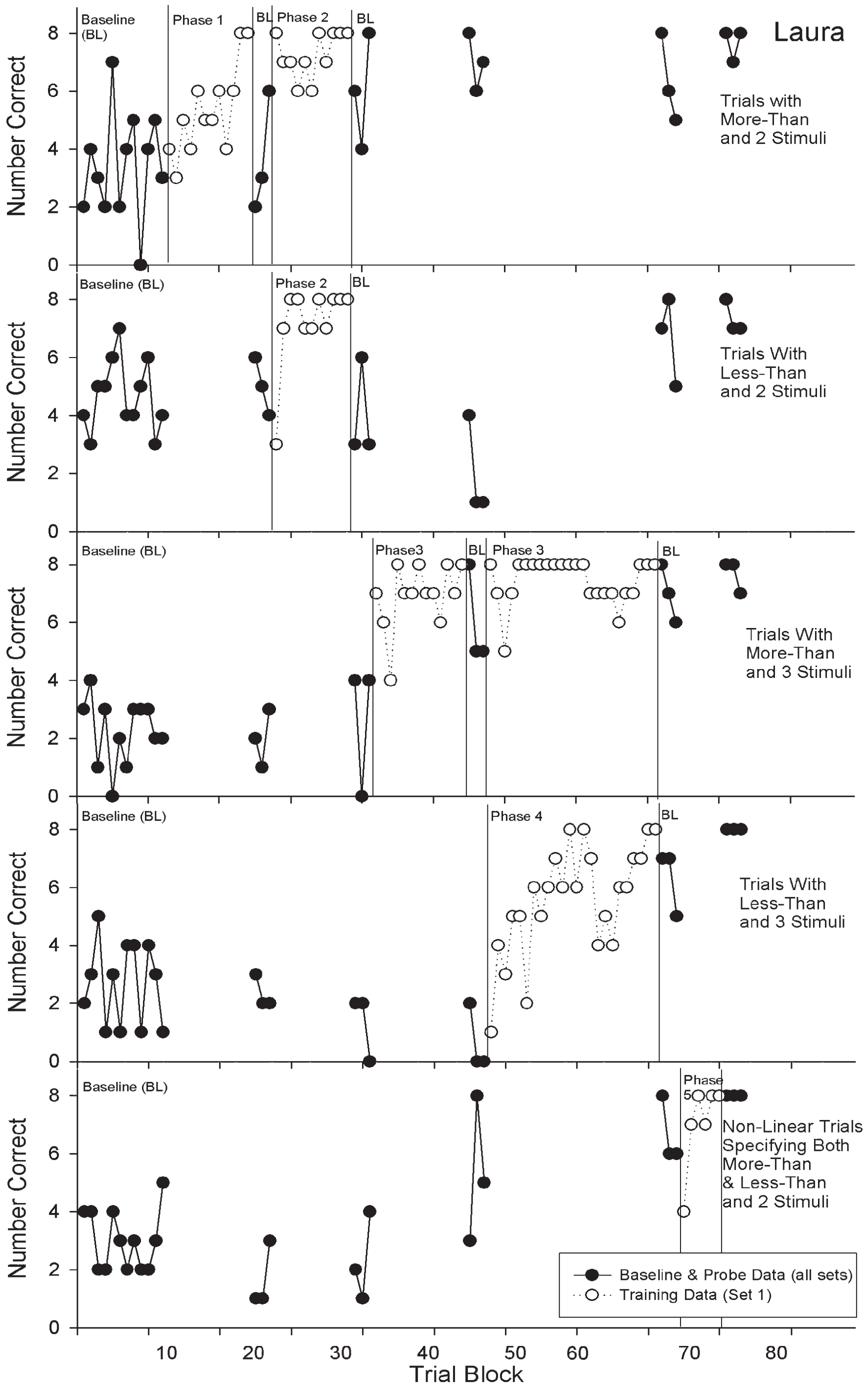

Figure 7. Results for Laura. See Figure 4 for details.

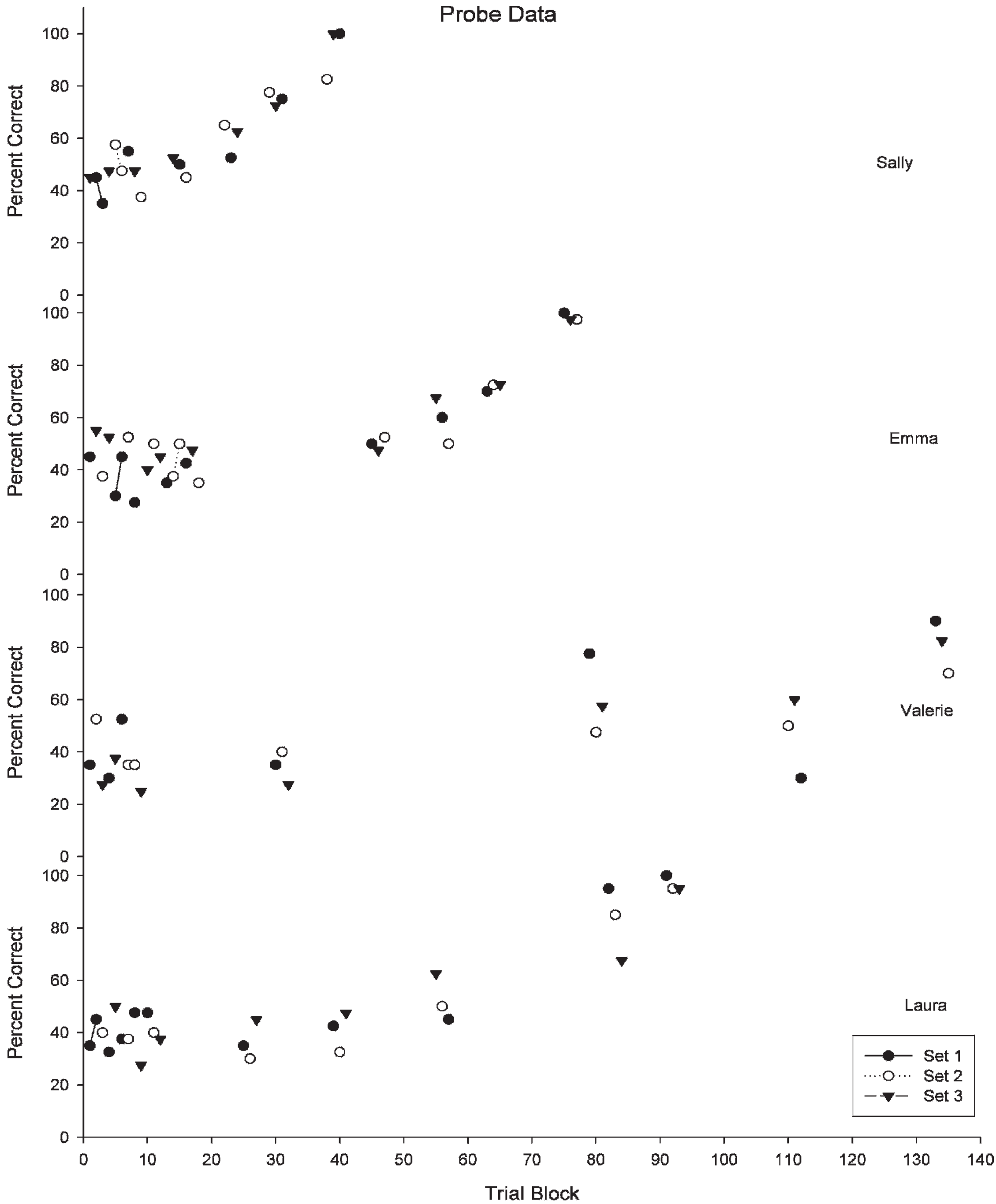

Figure 8. Probe data for all stimulus sets and all participants.

100% of the questions with Sets 1 and 3. Errors on Set 2 were equally distributed across the different trial types (e.g., errors were not made with just one trial type; Figure 4). Also, the improvements in this baseline performance occurred with regard to trial types in which explicit training had never been given (e.g., mixed nonlinear trials with three stimuli;

Figure 4). Because of the high degree of accuracy during this baseline probe, Sally's participation in the study was completed.

Emma

Emma required 26 trial blocks to reach our mastery criteria for Phase 1 (Figure 2). Because of her initial difficulties in acquiring the targeted relational response, on the seventh trial block she was exposed to the nonarbitrary training procedures described above. When exposed to the second baseline, her collective responding showed no improvement over the original baseline for Sets 1, 2, and 3 (see Figure 2 for data on Set 1 and Figure 8 for data on all three sets). However, her responding to trials that involved more than with two and three stimuli showed improvement (Figure 5).

Emma showed rapid acquisition of the less than relation with two stimuli when combined with more than trials in Phase 2 (Figure 2). During the third baseline her performance showed a slight improvement over her previous exposures to baseline when responding to stimuli in Sets 1 and 3 (Figure 8). Figure 8 suggests that her responding may not have generalized to Set 2, in that responding on this set was near the previous baseline level. Figure 5 indicates although her overall performance did not show much improvement, her responding to the relations that had been trained was stronger than baseline.

Emma required five trial blocks to meet the mastery criteria in Phase 3 (Figure 2). Emma's responding to Set 1 stimuli during the fourth baseline showed further improvements over the previous baseline conditions (Figure 2). Increases in baseline were related to improvements on only those responses that had been exposed to the training procedures (Figure 5).

Emma required nine trial blocks to meet the mastery criteria in Phase 4 (Figure 2). Her final exposure to the baseline condition showed complete acquisition of all trial types, including mixed nonlinear trial types, even though she had yet to be exposed to training on these responses (Figure 5). This acquisition occurred across all three sets of stimuli, including Sets 2 and 3 in which no direct reinforcement for responding had been provided (Figure 8). This included the mixed nonlinear trial types, even though she had yet to be exposed to training on these responses (Figure 5).

Valerie

Valerie required 20 trial blocks to reach the mastery criteria for Phase 1 (Figure 3). Given her difficulty in acquiring the targeted response, we modified the procedure and on the 10th trial block Valerie began the nonarbitrary training sequence described above. There was no improvement over her original baseline performance across the three stimulus sets (Figure 8) or for any specific trial type for this second exposure to baseline (Figure 6).

Valerie's initial responding for Phase 2 was variable and inaccurate across the two trial types (Figure 6). Following four trial blocks with poor performance in Phase 2, Valerie was exposed to the nonarbitrary training procedures. Valerie's responding during her third exposure to baseline showed overall improvements across Sets 1 and 3 (Figure 8). These increases were related to improvements in responding in both more than with two stimuli and less than with two stimuli (Figure 6), indicating that multiple-exemplar training in combination with nonarbitrary training facilitated the development of mutual entailment with arbitrary comparative relations that generalized to new stimulus sets.

Valerie required 23 trial blocks to reach the mastery criteria in Phase 3. On the seventh trial block she was exposed to the nonarbitrary training procedures. Responding during her fourth exposure to the baseline condition showed a marked degradation in Set 1 when compared to her previous baseline performances (Figure 3) and a slight improvement in responding to Sets 2 and 3 (Figure 8). Furthermore, no one trial type was more accurate than any other trial type during this baseline (Figure 6). Thus, although multiple-exemplar training in combination with non-arbitrary training improved Valerie's responding, her accuracy was not maintained when reinforcement was withheld and trials were mixed.

Valerie required 21 trial blocks to reach our mastery criteria in Phase 4 (Figure 3). We intervened with Phase 4.1 (nonarbitrary pretrials) after three trial blocks. This intervention increased the accuracy of her responding; however, she was still consistently making errors (Figure 3). It was noticed that the word "pennies" was frequently used in her previous nonarbitrary training. We presumed that the spoken stimulus "pennies" may have acquired some relational functions during this training. Thus, we intervened with Phase 4.2, which incorporated the word "pennies" in the arbitrary trials. This intervention immediately improved and stabilized Valerie's responding (Figure 3). After three trial blocks of Phase 4.2, we removed the C_{rel} "pennies" (Phase 4.1) and she reached 100% accuracy on the third trial block. She was then reexposed to Phase 4 and responded at 100% correct for two trial blocks; thus, she was exposed to a fifth baseline.

On Valerie's fifth exposure to the baseline condition, she showed marked improvements in responding across all three sets of stimuli when compared to her responding on the previous four baseline conditions (Figure 8). Improvement was shown in trial types that had been directly trained as well as the one that had not been trained (e.g., mixed nonlinear trials; Figure 6). It is interesting to note that her responding degraded across the three baseline trial blocks, indicating sensitivity to the lack of contingent reinforcement in that phase. Valerie was then withdrawn from the study because her primary caregiver was no longer able to transport her to sessions.

Laura

Laura required 11 trial blocks to reach mastery criteria in Phase 1 (Figure 3). Her second exposure to the baseline condition showed no improvements over the original baseline performance (Figure 8). She required 11 trial blocks to reach the mastery criteria for Phase 2 (Figure 3). She showed no improvements during her third exposure to baseline over her previous exposures (Figure 8).

It took Laura 13 trial blocks to meet the mastery criteria in Phase 3 (Figure 3). She showed improvements only during her third exposure to baseline with Set 3 (Figure 8). These improvements were related to improvements in trials that specified more than relations between two and three stimuli (Figure 7).

Laura required 24 trial blocks to reach the mastery criteria in Phase 4 (Figure 3). Her fifth exposure to the baseline condition showed strong improvements over her previous exposures (Figure 8). These increases were related to increases in all of the trial types used in baseline, including the mixed nonlinear trials that had not yet been targeted in training (Figure 7). She required six trial blocks to meet the mastery criteria in Phase 5 (Figure 3). Her final exposure to baseline showed near-perfect responding on all trial types (Figure 7) and across all sets (Figure 8).

DISCUSSION

The core hypothesis of RFT is twofold: (a) There are relational operants, and (b) they constitute the essential behavioral core of human language and cognition. The present study is focused on the first of these two claims. The primary empirical support for the concept of relational operants has been a substantial and growing body of indirect data showing that derived stimulus relations develop, come under antecedent and consequential control, and can be modified into multiple forms, all features of instrumental behavior (Hayes et al., 2001b). More recently, a small number of studies have directly provided an operant history focused on specific types of relational responding, which is a more direct test of the concept (e.g., Y. Barnes-Holmes et al., 2001a, 2001b; Y. Barnes-Holmes, Barnes-Holmes, & Smeets, 2004; Y. Barnes-Holmes, Barnes-Holmes, Smeets, Strand, & Friman, 2004; Luciano et al., unpublished manuscript). The present study builds on these previous studies and provides controlled evidence that relational frames are learned. Furthermore, the data support the idea that nonarbitrary relational responding, when abstracted and brought under contextual control, fosters the development of arbitrarily applicable derived relational responding.

The baseline condition in this study was critical to demonstrating that comparative relational framing is operant. All participants were apparently deficient in the targeted relational responses. When responding to each trial type was individually analyzed, these deficits were shown across the range of specific relational tasks tested. The extended baselines for Emma and Laura showed that these deficits were not merely artifacts of the novel testing situation.

The arrangement of training by types of relational tasks and the probe data advance the methodology used by Y. Barnes-Holmes, Barnes-Holmes, Smeets, Strand, and Friman (2004). In this study, generalization tests were conducted after all relational responses had been trained. The present data permit a considerably more precise evaluation of the concept of a comparative relational frame, because specific forms of generalization are central to that concept.

When participants were exposed to reinforcement across multiple examples of comparative relational responding, subsequent probes also improved on both the training stimulus set and the probe sets. As the term *frame* suggests, this kind of stimulus generalization is critical to the concept of a relational frame. Relational framing is arbitrarily applicable in the sense that C_{rel} cues (in this case, words like "is more than") can produce coherent patterns of relational responding with virtually any stimuli, regardless of their formal properties.

More important, when responding on individual trial types was analyzed, improvements were largest on the relation tasks that had been trained, but all participants showed improvement in performances on untrained trial types as well. For example, when Emma learned the targeted more than relations with two pictures, she immediately improved on the untargeted more than relations with three pictures (see Figure 5, first and third graphs). Similarly, all participants showed marked improvements on the mixed nonlinear trial types before being exposed to specific training on that trial type; only 1 participant required such training.

It is important that these forms of generalization occurred only after the putatively critical behavioral features of a comparative relational frame were reinforced. Relational frames are psychological, not logical, units. More than and less than are logically mutually related, for example, but the present data suggest that this logical relation is not the source of mutual relational responding. The causal influence is in the opposite direction: The history of reinforcement for a relational response pattern led to the kind of overall comparative relation that we call logical. Once established, however, it generalized to new networks of stimulus relations and to new forms of stimuli.

This effect is what justifies considering a relational frame to be a unit. It is not a unit in the sense of being primitive, but it is a unit in the sense that when its elements are assembled these keystone features can be flexibly extended to novel and much more elaborate networks, as is done in natural language in novel sentences. In other words, a relational frame seems to be the smallest verbal unit capable of capturing processes of meaning and understanding as they occur in natural language (Hayes, Fox, et al., 2001, p. 34).

In summary, we draw the conclusion that the training contingencies were necessary for acquisition of the generalized comparative relational performance from five consistent patterns in the data: (a) There were no improvements in responding in the extended baselines, (b) generalization across stimuli and trial types emerged gradually, (c) improvement in responding in the baseline probes was greatest for targeted trial types, (d) generalization to new trial types occurred as key relational elements had been trained in other trial types, and (e) the apparent difficulty of the training tasks for 3 participants showed that a new form of responding was being acquired. If, as was the case with Sally, all participants rapidly acquired the targeted responses, then alternative interpretations would be warranted.

One participant in the Y. Barnes-Holmes, Barnes-Homes, and Smeets study (2004) required nonarbitrary relational training before responding could be established in arbitrary contexts. This finding was replicated in the current study with Valerie and Emma. Taken together, their data suggest that nonarbitrary relational responding may be an important component of the acquisition of arbitrarily applicable derived relational responding, as has been suggested from the beginning of RFT research (e.g., Steele & Hayes, 1991). Participants in this study were not preexperimentally assessed for such responding, but given Valerie's difficulty during Phases 1.1, 2.1, and 2.2, which involved the use of nonarbitrary cues during training, it seems likely that her ability to respond relationally in a nonarbitrary context was weak. In contrast, Emma required nonarbitrary relational training only for Phase 1, and the rest of the relational responses were readily acquired, suggesting that she had a repertoire of nonarbitrary relational responding. The interventions used in Phase 1 brought an abstracted and arbitrarily contextually controlled version of the repertoire to bear in the current context.

There are limitations to this investigation. Participation in the study lasted from 2 to 7 months. Given the developmental nature of the study, the longer participants remained in the study the greater the probability that extraexperimental variables influenced their responding. It was noted that for each participant the number of trial blocks required to pass each phase got shorter. The possibility that participants' experiences outside the study influenced their responding cannot be ruled out, although the multiple baseline design does provide broad protection against extraexperimental history as the source of the specific effects seen. There were also inconsistencies in when baseline probes occurred (e.g., immediately following a training session if time allowed or on subsequent days; at the end of the week or the beginning). Performance on the baseline trial blocks may have been influenced by this inconsistency.

Implications

The implications of this study are both basic and applied. In the basic area, RFT claims that relational operants suggest a new behavioral principle (Hayes, Fox, et al., 2001, pp. 45–46). As the present study shows, this principle is not invoked to explain relational operants. The contingencies that gave rise to a comparative relational frame in the current study were entirely typical. Rather, a new behavioral principle is argued to be an implication of relational frames (Hayes & Barnes-Holmes, 2004; Hayes, Barnes-Holmes, & Roche, 2003).

Consider the arbitrary network $A < B < C$ among three coins that are said to be able to buy candy. It is the relative value functions of these three stimuli that demand an alternative account. To see this more clearly, suppose B was given a discriminative stimulus function through normal means, perhaps by reinforcing a particular rate of behavior in its presence. Given the $A < B < C$ relational network, if A and C were then unexpectedly presented, one might expect the rate of responding to decrease in the presence of A and increase in the presence of C. Similarly, suppose B was given a conditional stimulus function through normal means, perhaps by pairing B with food. If A and C were unexpectedly presented, one might expect lower levels of salivation to A but higher levels to C, perhaps even higher than the response to B, which had been directly paired with food. The present study shows that a comparative relational frame can be learned, but there is no behavioral principle that describes a situation in which a learned operant then alters other behavioral processes, such as discriminative control or classical conditioning. The discrimination and classical conditioning transformation of stimulus function experiment just described is not a thought experiment. It has recently been conducted, and the results are exactly as described, with the exception that shock was used as the unconditioned stimulus (Dougher, Hamilton, Fink, & Harrington, in press). Several other studies have shown such transformational effects, both with classical and operant functions (e.g., Dymond & Barnes, 1995; Roche & Barnes, 1997). When these data are considered in total, the applied implications are unlimited. For example, it may be possible to program training such that otherwise neutral stimuli become powerful reinforcers for individuals with limited sets of reinforcing stimuli or who satiate quickly.

No existing behavioral term fits such situations. Hayes and Hayes (1989) suggest the use of the terms *relational* or *verbal* in these conditions. For example, stimuli that acquire discriminative-like functions through relational frames (e.g., Kohlenberg, Hayes, & Hayes, 1991) might be usefully called relational discriminative stimuli. They are not conventional discriminative stimuli because they have neither the history that fits that term nor the similar formal properties that would provide such functions via stimulus generalization; rather, their functions are discriminative-like but are established via a transformation of stimulus functions through a relational frame.

Perhaps the best place to test the progressivity of the basic RFT account is in applied work, because it is there that verbal and cognitive phenomena are most central to prediction and influence in important domains (Hayes & Berens, 2004). An example is the kind of educational situation examined in the present study. Virtually all educational tasks arguably involve relational frames, and a growing body of literature shows that derived relational responding is correlated with intellectual tasks (e.g., O'Hora, Palaez, & Barnes-Holmes, 2005) and can be used to foster educational and language performances (e.g., D. Barnes-Holmes, Barnes-Holmes, & Cullinan, 2000; Murphy et al., 2005; Rehfeldt & Root, 2005).

As an applied matter, a desirable characteristic of RFT is that it specifies a precise unit to target that appears to be central to human language. Knowing the unit one is trying ultimately to train is critical in applying behavior-analytic training procedures. For example, in any operant training procedure it is important to vary the irrelevant features of the task and its context so that functional control is not captured by irrelevant invariant features. It is also important to arrange proper contrasts with similar but functionally distinct contexts and actions, and to ensure that terminal responding incorporates the range of response and stimulus control topographies intended. As these principles are applied to relational frames, they suggest areas in which care should be taken in training verbal and cognitive skills. For example, although relational responding seems often to emerge from nonarbitrary relational training, if the ultimate goal of this responding is to become arbitrarily applicable, nonarbitrary features should be varied and should be faded into arbitrary features. Such procedures were used with Valerie and Emma with good results, but more research will be needed to work out how best to produce transfer from nonarbitrary to arbitrary stimuli.

It is also important to train the relevant aspects of the relational frame being established and to bring it under flexible contextual control. In the absence of guidance regarding the key unit being trained, it might be easy to overtrain in certain areas or undertrain in others. For instance, suppose, as seems likely from an RFT point of view, that frames of coordination are sometimes being established as a side effect of tact training. If establishing such frames were part of the purpose of tact training, it could be critical to include multiple aspects of the relational frame and to distinguish it from other relational forms during training. This could be done in several ways, such as including both productive and receptive examples in the training, including transformations of stimulus functions in the tasks (as was done in this study), or including training of frames of distinction along with training in frames of coordination (e.g., "which one is *not* the ball?"). Indeed, there was some indication in the present data that training in the multiple aspects of a relational frame was helpful. For example, Valerie did not show improvements on more than and less than until both were trained together (Figure 6).

When applying relational frames to nonarbitrary relations, as occurs in most natural language situations, RFT suggests that it is important to establish flexible forms of contextual control so that the arbitrary nature of the underlying relation is made even more evident (e.g., when approaching a stoplight, it might be useful to go beyond asking "What should I do now?" to asking "If red were green what should I do now?"). This is precisely the kind of work that is creating advancement in the establishment of perspective-taking skills in the RFT laboratory (Rehfeldt, Dillen, Ziomek, & Kowalchuk, in press). Excessively narrow training curricula in all of these areas that bear on relational frames could prevent the kind of behavioral flexibility needed for good verbal and intellectual development. Whether such ideas are helpful is an empirical matter, but they are logical extensions of RFT.

There are other reasons that the applied laboratory is well suited to the analysis of relational operants: A purely functional analytic approach is more common in applied behavior analysis than in basic behavior analysis, which is often populated by those interested in associative forms of learning theory (e.g., Burgos, 2003; Tonneau, 2004). As with the classic research in an operant analysis of imitation, applied behavior analysts did much of the analytic work on this relatively basic question (e.g., Baer, Peterson, & Sherman, 1967; Baer & Sherman, 1964; Gewirtz & Stengle, 1968; Peterson, 1968; Peterson & Whitehurst, 1971). Imitation is now an integral part of the applied armamentarium of the field (Young, Krantz, McClannahan, & Poulson, 1994).

The original RFT volume stated that "An important empirical question, therefore, is whether we can design effective RFT-based interventions that establish or facilitate new repertoires of derived relational responding in young children. Positive evidence in this regard would provide firm support for RFT's approach to derived relational responding" (Hayes, Fox, et al., 2001, p. 28). The present study is one of several recent findings (e.g., Y. Barnes-Holmes, Barnes-Holmes, & Smeets, 2004; Y. Barnes-Holmes, Barnes-Holmes, Smeets, Strand, & Friman, 2004) that support this possibility. Given the growing body of data on the link between such relational behavior and language and cognitive abilities (e.g., Hayes & Bissett, 1998; O'Hora et al., 2005), this result opens operant approaches to the experimental analysis of a much wider range of verbal and cognitive phenomena than was previously the case.

Working out how to study, train, and apply relational frames in basic and applied behavior analysis will take considerable effort, but behavior analysts have a notable track record of success with difficult methodological and empirical issues within their domain. The applied successes of technologies based on RFT in the clinical area (e.g., Hayes et al., 1999) suggest that it may be worth the effort in the applied areas that are more commonly associated with applied behavior analysis.

What seems most important about the present study it that it provides evidence that relational operants exist as an empirical phenomenon. If operants of this kind exist and if they affect other behavioral processes (Dougher et al., in press), an analysis of their impact is necessary. Whether or not RFT is helpful in dealing with these phenomena is a separate question; the present data suggest that relational operants are there.

REFERENCES

Andrews, G., & Halford, G. S. (1998). Children's ability to make transitive inferences: The importance of premise integration and structural complexity. *Cognitive Development*, *13*, 479–513.

Baer, D. M., Peterson, R. F., & Sherman, J. A. (1967). The development of imitation by reinforcing behavioral similarity to a model. *Journal of Applied Behavior Analysis*, *10*, 405–416.

Baer, D. M., & Sherman, J. A. (1964). Reinforcement control of generalized imitation in young children. *Journal of Experimental Child Psychology*, *1*, 37–49.

Barnes-Holmes, D., Barnes-Holmes, Y., & Cullinan, V. (2000). Relational frame theory and Skinner's *Verbal Behavior*: A possible synthesis. *The Behavior Analyst*, *23*, 69–84.

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. (2001a). Exemplar training and a derived transformation of function in accordance with symmetry. *The Psychological Record*, *51*, 287–308.

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. (2001b). Exemplar training and a derived transformation of function in accordance with symmetry: II. *The Psychological Record*, *51*, 589–604.

Barnes-Holmes, Y., Barnes-Holmes, D., & Smeets, P. M. (2004). Establishing relational responding in accordance with opposite as generalized operant behavior in young children. *International Journal of Psychology and Psychological Therapy*, *4*, 559–586.

Barnes-Holmes, Y., Barnes-Holmes, D., Smeets, P. M., Strand, P., & Friman, P. (2004). Establishing relational responding in accordance with more-than and less-than as generalized operant behavior in young children. *International Journal of Psychology and Psychological Therapy*, *4*, 531–558.

Burgos, J. E. (2003). Laudable goals, interesting experiments, unintelligible theorizing: A critical review of relational frame theory. *Behavior and Philosophy*, *31*, 19–45.

de Rose, J. C., de Souza, D. G., & Hanna, E. S. (1996). Teaching reading and spelling: Exclusion and stimulus equivalence. *Journal of Applied Behavior Analysis*, *29*, 451–469.

de Rose, J. C., de Souza, D. G., Rossito, A. L., & de Rose, T. M. S. (1992). Equivalence relations and reading. In S. C. Hayes, & L. J. Hayes (Eds.), *Understanding verbal relations* (pp.69–82). Reno, NV: Context Press.

Dougher, M. J., Hamilton, D., Fink, B., & Harrington, J. (in press). Transformation of the discriminative and eliciting functions of generalized relational stimuli. *Journal of the Experimental Analysis of Behavior*.

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.

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

Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist-practitioner: Research and accountability in the age of managed care*. Boston: Allyn & Bacon.

Hayes, S. C., & Barnes-Holmes, D. (2004). Relational operants: Processes and implications. A response to Palmer's review of *Relational Frame Theory*. *Journal of the Experimental Analysis of Behavior*, *82*, 213–224.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.) (2001a). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York: Kluwer Academic/Plenum.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001b). Relational frame theory: A précis. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp.141–154).New York: Kluwer Academic/Plenum.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2003). Behavior analysis, relational frame theory, and the challenge of human language and cognition: A reply to the commentaries on *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition*. *The Analysis of Verbal Behavior*, *19*, 39–54.

Hayes, S. C., & Berens, N. M. (2004). Why relational frame theory alters the relationship between basic and applied behavioral psychology. *International Journal of Psychology and Psychological Psychotherapy*, *4*, 341–353.

Hayes, S. C., & Bissett, R. (1998). Derived stimulus relations produce mediated and episodic priming. *The Psychological Record*, *48*, 617–630.

Hayes, S. C., Fox, E., Gifford, E. V., Wilson, K. G., Barnes-Holmes, D., & Healy, O. (2001). Derived relational responding as learned behavior. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 21–49). New York: Kluwer Academic/Plenum.

Hayes, S. C., & Hayes, L. J. (1989). The verbal action of the listener as a basis for rule-governance. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 153–190). New York: Plenum.

Hayes, S. C., Luoma, J., Bond, F., Masuda, A., & Lillis, J. (in press). Acceptance and commitment therapy: Model, processes, and outcomes. *Behaviour Research and Therapy*.

Hayes, S. C., Masuda, A., Bissett, R., Luoma, J., & Guerrero, L. F. (2004). DBT, FAP, and ACT: How empirically oriented are the new behavior therapy technologies? *Behavior Therapy*, *35*, 35–54.

Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. New York: Guilford.

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.

Healy, O., Barnes-Holmes, D., & Smeets, P. M. (2000). Derived relational responding as generalized operant behavior. *Journal of the Experimental Analysis of Behavior*, *74*, 207–227.

Joyce, B. G., & Wolking, W. D. (1989). Stimulus equivalence: An approach for teaching beginning reading skills to young children. *Education and Treatment of Children*, *12*, 109–122.

Kohlenberg, B. S., Hayes, S. C., & Hayes, L. J. (1991). The transfer of contextual control over equivalence classes through equivalence classes: A possible model of social stereotyping. *Journal of the Experimental Analysis of Behavior*, *56*, 505–518.

Lipkens, G., Hayes, S. C., & Hayes, L. J. (1991). Longitudinal study of derived stimulus relations in an infant. *Journal of Experimental Child Psychology*, *56*, 201–239.

Lowenkron, B. (1989). Instructional control of generalized relational matching to sample in children. *Journal of the Experimental Analysis of Behavior*, *52*, 293–309.

Luciano, C., Becerra, I. G., & Valverde, M. R. (unpublished). *The formation of generalized listener behavior and equivalence relations in an infant, from 15 months 24 days to 24 months*.

Matos, M. A., & d'Oliviera, M. M. H. (1992). Equivalence and reading. In S. C. Hayes, & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 83–96). Reno, NV: Context Press.

Murphy, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2005). Derived manding in children with autism: Synthesizing Skinner's *Verbal Behavior* with relational frame theory. *Journal of Applied Behavior Analysis*, *38*, 445–462.

Ninness, C., Rumph, R., McCuller, G., Harrison, C., Ford, A. M., & Ninness, S. K. (2005). A functional analytic approach to computer-interactive mathematics. *Journal of Applied Behavior Analysis*, *38*, 1–22.

O'Hora, D., Palaez, M., & Barnes-Holmes, D. (2005). Derived relational responding and performance on verbal subtests of the WAIS-III. *The Psychological Record*, *55*, 155–175.

Peterson, R. F. (1968). Some experiments on the organization of a class of imitative behaviors. *Journal of Applied Behavior Analysis*, *1*, 225–235.

Peterson, R. F., & Whitehurst, G. J. (1971). A variable influencing the performance of generalized imitative behaviors. *Journal of Applied Behavior Analysis*, *4*, 1–9.

Pilgrim, C. (1998). The human subject. In K. A. Lattal, & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 15–44). New York: Plenum.

Reese, H. W. (1968). *The perception of stimulus relations: Discrimination learning and transposition*. New York: Academic Press.

Rehfeldt, R. A., Dillen, J. E., Ziomek, M. M., & Kowalchuk, R. K. (in press). Assessing relational learning deficits in perspective-taking in children with high-functioning autism spectrum disorder. *The Psychological Record*.

Rehfeldt, R. A., & Root, S. L. (2005). Establishing derived requesting skills in adults with severe developmental disabilities. *Journal of Applied Behavior Analysis*, *38*, 101–105.

Roche, B., & Barnes, D. (1997). A transformation of respondently conditioned functions in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior*, *67*, 275–301.

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

Stromer, R., & MacKay, H. A. (1992). Spelling and emergent picture-printed word relations established with delayed identity matching to complex samples. *Journal of Applied Behavior Analysis*, *25*, 893–904.

Stromer, R., MacKay, H. A., & Stoddard, L. T. (1992). Classroom applications of stimulus equivalence. *Journal of Behavioral Education*, *2*, 225–256.

Tonneau, F. (2004). Who can understand relational frame theory? A reply to Barnes-Holmes and Hayes. *European Journal of Behavior Analysis*, *3*, 95–102.

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

Wright, B. C., & Dowker, A. D. (2002). The role of cues to differential absolute size in children's transitive inferences. *Journal of Experimental Child Psychology*, *81*, 249–275.

Wulfert, E., & Hayes, S. C. (1988). The transfer of conditional sequencing through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior*, *50*, 125–144.

Young, J. M., Krantz, P. J., McClannahan, L. E., & Poulson, C. L. (1994). Generalized imitation and response-class formation in children with autism. *Journal of Applied Behavior Analysis*, *27*, 685–697.

Received January 25, 2006
Final acceptance April 5, 2006
Action Editor, Ruth Anne Rehfeldt

APPENDIX

A sample data sheet for the secondary observer. The first five trials represent sample trials for Phases 1 through 5 with Set 1. R = red stimulus, B = blue stimulus, and G = green stimulus. The position of the stimuli on the data sheet corresponds to their position during the trial. The number under each stimulus indicates the order in which the stimuli were to be pointed to by the experimenter. The symbols (> and <) indicate the relation specified among the stimuli. The upper left corner of each box indicates which stimulus was selected. After a session, the secondary experimenter took the experimenter's data sheet and scored his or her agreement on which stimulus was selected for each trial. During trial blocks the secondary experimenter scored a yes or a no for trial arrangement, trial presentation, and provision of the correct consequence for each trial. The first two trials on this data sheet are samples of what may have been scored on a given trial by the secondary observer. The experimenter's data sheets were identical to the secondary observer's data sheets, except that they did not have the columns for interobserver agreement and procedural integrity.

Condition	**Baseline-DAY 1**		**SET: __3__**		**PARTICIPANT:Valerie_____**							
Name of Primary Observer:________________________												
Date:______________			Start Time_____________		End Time ________________							
IOA Taken YES / NO												
1)	X R	>	B						Arrange stimuli	present trial	correct	consequence
ORDER	1		2			Agree		YES	X			X
CORRECT						Disagree	X	NO		X		
2)	X B	<	R						Arrange stimuli	present trial	correct	consequence
ORDER	1		2			Agree	X		X	X		X
CORRECT						Disagree						
3)	R	>	B	>	G				Arrange stimuli	present trial	correct	consequence
ORDER	1		2		3	Agree						
CORRECT						Disagree						
4)	R	>	B	>	G				Arrange stimuli	present trial	correct	consequence
ORDER	3		2		1	Agree						
CORRECT						Disagree						
5)	R	>	B	>	G				Arrange stimuli	present trial	correct	consequence
ORDER	1		2,4		3	Agree						
CORRECT						Disagree						
6)									Arrange stimuli	present trial	correct	consequence
ORDER						Agree						
CORRECT						Disagree						
7)									Arrange stimuli	present trial	correct	consequence
ORDER						Agree						
CORRECT						Disagree						
8)									Arrange stimuli	present trial	correct	consequence
ORDER						Agree						
CORRECT						Disagree						