JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2000, **73,** 5–22 NUMBER 1 (JANUARY)

TESTING FOR SYMMETRY IN THE CONDITIONAL DISCRIMINATIONS OF LANGUAGE-TRAINED CHIMPANZEES

NEIL DUGDALE AND C. FERGUS LOWE

UNIVERSITY OF WALES, BANGOR

If subjects are taught to match Stimulus A to B and then, without further training, match B to A, they have passed a test of symmetry. It has been suggested that nonhumans' lack of success on symmetry tests might be overcome by giving them a history of symmetry exemplar training, that is, by directly teaching a large number of conditional relations (e.g., AB, CD, EF, . . .) and also directly training the "reverse" of these relations (e.g., BA, DC, FE, . . .). The chimpanzee subjects of the present study, Sherman, Austin, and Lana, had already received extensive symmetry exemplar training as a result of attempts to teach a selection-based language system of lexigrams. The present study systematically subjected 2 of these chimps (Sherman and Lana), for the first time, to standard symmetry tests in controlled conditions. Both chimps failed the tests, even when their correct responses on test trials were reinforced. The findings do not support the exemplar training hypothesis, and cast doubt upon whether the chimps can pass tests of stimulus equivalence.

Key words: symmetry, stimulus equivalence, verbal behavior, language, conditional discrimination, key press, chimpanzees

This paper reports a systematic attempt to assess the performances on standard (matching-to-sample; MTS) tests of symmetry of 3 chimpanzees, Sherman, Austin, and Lana, that have featured prominently in language learning programs (see Rumbaugh, 1977; Savage-Rumbaugh, 1986). Testing these chimps is especially valuable because, as we aim to show in this section, (a) unequivocal demonstrations of symmetry in nonhumans are nonexistent, and (b) if positive results were to be achieved they might well be expected to come from chimpanzees, and particularly chimps with learning histories like those of Sherman, Austin, and Lana. These histories are specified in some detail below, but to show why they are important it is first necessary to consider recent theoretical accounts of performance on symmetry tests.

In preparation for a symmetry test, subjects are explicitly taught a series of baseline conditional discriminations, typically via an arbitrary MTS procedure. For example, the subject may be presented with an array of visual comparison stimuli (B1, B2, B3, etc.), and reinforcers are contingent upon the subject selecting B1 given Stimulus A1, B2 given A2, and so on. Trials continue until the subject matches A samples to B comparisons with a high level of accuracy (i.e., learns the AB matching task). A test would then be given to assess whether the subject can, without requiring further training or differential reinforcement, perform BA matching correctly. Passing this symmetry test shows that new conditional discrimination performances have emerged from those previously trained as baseline relations.

Verbally able human subjects have frequently produced high and sustained levels of correct responding on symmetry tests, but several studies of pigeons, monkeys, and baboons have failed to find comparable levels of emergent performance (see reviews by Dugdale & Lowe, 1990, Hayes & Hayes, 1992, and Saunders, Williams, & Spradlin, 1996). In two studies (a study of macaques by McIntire, Cleary, & Thompson, 1987, and a study of a sea lion by Schusterman & Kastak, 1993) that have reported positive results, it has been argued that the tested relations were directly

This research was previously reported in Neil Dugdale's PhD thesis (Dugdale, 1988) submitted to the University of Wales, Bangor, and was supported by a studentship from the Economic and Social Research Council. We thank D. M. Rumbaugh and E. S. Savage-Rumbaugh for providing us with the opportunity to work with their very special subjects, and for the invaluable assistance that they and the staff of the Language Research Center, Atlanta, Georgia, gave throughout the study. We also thank Bill McIlvane and an anonymous reviewer for their helpful comments on the manuscript.

Requests for reprints should be addressed to Neil Dugdale, School of Psychology, University of Wales, Bangor, Gwynedd, LL57 2DG, United Kingdom (E-mail: n.dugdale@bangor.ac.uk).

trained rather than emergent (see Dugdale, 1988, Dugdale & Lowe, 1990, Hayes, 1989, and Saunders, 1989, for critiques of McIntire et al., and Horne & Lowe, 1996, 1997, and Lowe & Horne, 1996, for critiques of Schusterman & Kastak).

Chimpanzees are more closely related genetically to humans than any other nonhuman species, having DNA 99% the same as ours. Although chimps might thus be considered more likely than other nonhumans to pass symmetry tests, only two studies have formally assessed this hypothesis and neither has confirmed it. A chimp in the study by Yamamoto and Asano (1995) produced highly accurate performance on symmetry probes in only one of 34 test sessions (18 probes per session), an outcome that could clearly be produced by chance alone. Three chimps in a study by Tomonaga, Matsuzawa, Fujita, and Yamamoto (1991) each had three symmetry tests (eight probes per test). High accuracy occurred in only one of the nine sessions; one chimp was correct on her first eight trials but made 10 errors on the remaining 16, resulting in an overall score of 75% correct. This was above the chance level of 50% correct but still too low to demonstrate unequivocally the emergence of the experimenter-specified symmetrical conditional discriminations (see Sidman, 1987).

Hayes (1989) and Boelens (1994) have suggested that nonhumans' lack of success on MTS tests of symmetry and equivalence might be overcome by giving them an extensive history of symmetry exemplar training (also see Sidman et al., 1982). This would involve directly training a large number of conditional discriminations and their symmetrical counterparts (e.g., train A1B1, A2B2, A3B3, etc., *and* B1A1, B2A2, B3A3, etc.). Directly training B1A1 after A1B1 does not in itself constitute symmetry. However, the hypothesis is that, if one continues directly training many such symmetry exemplars, eventually the subject will perform accurately on symmetry trials without exposure to the reinforcement contingency (i.e., pass a BA test after training on AB with any set of novel stimuli, thus showing "generalized symmetrical responding"). Hayes and Hayes (1989) and Boelens (1994) assume that verbally able humans learn generalized symmetrical responding in childhood from an extensive history of reinforced reversals of conditional relations that occur incidentally in learning verbal repertoires of speaking and listening (see also Devany, Hayes, & Nelson, 1986; Steele & Hayes, 1991).

The exemplar-training account of symmetry remains to be confirmed with respect to humans and nonhumans. The chimps in the present study afforded an opportunity to test the account because they had received extensive symmetry exemplar training prior to testing, and, just like the training that is purported to enable human children to learn generalized symmetry, theirs occurred incidentally during their caregivers' attempts to teach them speaker and listener repertoires with lexigrams.

History of Conditional Discrimination Learning with Lexigrams

The lexigram system. The heart of the system is a keyboard with a large number of keys, each of which could be back-projected with a unique arbitrary shape or "lexigram." New lexigrams were added as training progressed, until Sherman's and Austin's keyboards contained 92 lexigrams and Lana's contained just over 100. Of these, some 30 or so were assigned to various kinds of food, 20 or so to inedible objects such as tools, 13 or so to actions, and the remainder to various items including locations, caregivers, and attributes and states. The location of each lexigram on the keyboard was regularly changed, and the number active at any time could be restricted. As lexigrams were touched, they appeared in sequence on a screen above the keyboard (for further details of the system, see Savage-Rumbaugh, 1986, pp. 45–50).

The chimps had over 10 years' experience with the lexigram system. What follows is a selective review of their training history that focuses on those conditional discrimination training procedures that seemed to be effective in promoting arbitrary matching of objects to lexigrams and vice versa.

Matching lexigram comparisons to object samples: "Speaker" behavior. The first conditional discriminations Sherman, Austin, and Lana learned were between object samples and lexigram comparisons. Initially, these object → lexigram arbitrary matching relations were designed to function as requests, or mands (see Rumbaugh, 1977; Savage-Rumbaugh, 1984, 1986). In Sherman's and Austin's case, the first objects manded were food items. One of several clear plastic food dispensers was filled with one particular type of food at a time. Sherman and Austin were required to select the corresponding food lexigram from an array of comparisons that included at least one other food lexigram. Correct lexigram selections triggered the filled dispenser to deliver a piece of the sample food, whereas incorrect lexigram selections produced no food (i.e., activated an empty dispenser). This, then, was a *differential outcome* procedure (Brodigan & Peterson, 1976) in which each comparison correctly chosen produced a specific and unique consequence (in this instance, the sample item itself). Once a dispenser was empty of food, it (or an adjacent dispenser) was reloaded with a different type of food, and trials continued. The chimps were deemed to have learned a sample–comparison relation when they responded correctly on at least 90% of Trial 1 presentations following a refilling of the dispenser with that sample food. Sherman and Austin learned their first 11 food → lexigram relations this way (see Savage-Rumbaugh, 1986, pp. 92–101; Savage-Rumbaugh & Rumbaugh, 1978, pp. 285–292).

Lana learned her first object → lexigram conditional discriminations through similar differential outcome contingencies, with some procedural differences. Lana had six dispensers available simultaneously, each one filled with one of nine different sample items (seven food types and two inedible objects known to be reinforcing in her case; see Rumbaugh, 1977, p. 134). Reinforcement for choosing any of the nine lexigram comparisons was therefore conditional upon the presence of the corresponding sample. If a given sample item was not present, then selections of the corresponding lexigram comparison would not be reinforced. Tests given after some 7 months of exposure to this contingency confirmed its effectiveness in teaching the conditional discriminations (see Gill & Rumbaugh, 1974). On each test trial, Lana's trainer randomly selected one of the items, showed Lana this sample by holding it up and pointing to it, and then reinforced her selection of the corresponding lexigram comparison (as before, by giving Lana the sample in question).

Sherman and Austin, too, proved capable of correctly selecting food lexigrams when corresponding food samples were held up and pointed to by their trainers, rather than placed in dispensers (see Savage-Rumbaugh, 1986, p. 103). They also learned to select the corresponding lexigram comparison for each of several sample food items presented together on a tabletop (Lana did not have this training). Each correct lexigram selection resulted in the trainer giving the corresponding food to the chimp (see Savage-Rumbaugh, 1984, p. 230; Savage-Rumbaugh, Pate, Lawson, Smith, & Rosenbaum, 1983, pp. 470–472).

Once a food → lexigram relation had been established via the differential outcome/manding procedure, attempts were made to make it also function as a tact of sorts by presenting reinforcers other than the sample item for correct lexigram choices (i.e., nondifferential outcomes; see Savage-Rumbaugh, 1984). In Sherman's and Austin's case (but apparently not Lana's), initial attempts to switch from differential to nondifferential outcomes made performance fall to chance levels of accuracy. This disruption was eventually overcome by presenting differential and nondifferential outcomes together as reinforcers on each trial and then gradually reducing the amount of the differential outcomes with each successive correct response (see Savage-Rumbaugh, 1984). This fading program was no longer required after it had been successfully applied to three food → lexigram relations, because Sherman and Austin maintained performance on the remaining relations in their repertoires even when the reinforcers were switched immediately from differential to nondifferential outcomes.

Sherman and Austin eventually learned some 33 food → lexigram matching relations. Lana learned just as many, and in addition learned to select lexigram comparisons conditional upon five more objects (e.g., bowl, box, can) and each of six colors with which the objects had been painted (see Rumbaugh & Gill, 1977, pp. 176–177). These relations, too, were initially learned via differential outcomes, because when Lana chose a correct comparison she was given the sample object. Once the relations were learned, the differential outcomes were dropped and Lana simply got a food reinforcer and not the sample item for each correct response (see Essock, Gill, & Rumbaugh, 1977, pp. 194–198).

All 3 chimps were also taught to select lexigram comparisons conditional upon nine different tools. Sherman and Austin first learned their tool → lexigram relations with differential outcomes (if correct, they were given the tool to extract food from various contraptions that otherwise rendered the food inaccessible), and then did so with nondifferential outcomes (see Savage-Rumbaugh, 1986, pp. 180–189; Savage-Rumbaugh, Rumbaugh, & Boysen, 1978). Lana also learned the relations with a differential outcome procedure (see Savage-Rumbaugh, 1986, pp. 246–249). Finally, the chimps could do object → lexigram matching when the sample objects were depicted two-dimensionally, on television (Savage-Rumbaugh, 1986, pp. 306–307) or in photographs (e.g., see Rumbaugh & Gill, 1976, p. 111).

Matching object comparisons to lexigram samples: "Listener" behavior. Having learned to match lexigram comparisons to object samples, the chimps were not able, without further training, to match the same objects now presented as comparisons to the corresponding lexigrams, now presented as samples (i.e., they did not show symmetry; see Savage-Rumbaugh, 1984). In Sherman's and Austin's case, the object comparisons in initial tests were food items placed on a table in front of them. On each trial, a lexigram for one of the items was randomly presented as a sample on the screen above the keyboard. If the chimp gave the corresponding food to the trainer, then the trainer divided it and returned half to the chimp as reinforcement (a differential outcome procedure; see Savage-Rumbaugh, 1984, p. 238). At first, Sherman's and Austin's food selections had no clear correspondence to the food their trainers requested. They had to learn a number of component skills before progress could be made. These included (a) giving the food they had selected to their trainer (Savage-Rumbaugh, 1984, p. 238), (b) inhibiting the selection of preferred food when designated as incorrect comparisons (Savage-Rumbaugh et al., 1983, p. 476), and (c) learning that the location of the food in the comparison array was irrelevant (Savage-Rumbaugh et al., 1983, p. 476). In the case of Lana, Savage-Rumbaugh (1981) reported that she "did not want to hand the experimenter portions of food displayed in front of her, and, therefore, we used only photographs" (p. 39). There appear to be no published details of the reinforcement contingencies used to teach Lana to match food lexigrams to food photographs.

Single-case demonstrations of symmetry with reinforced probe trials require correct performance beginning with Trial 1 of the test. The chimps fell well short of this criterion, taking many trials (although exactly how many is not clear) to learn each lexigram → food relation after having learned the reverse. It also took them several reinforced test trials to learn to select tools conditional upon lexigram samples, the symmetrical counterparts of their previously acquired tool sample → lexigram conditional relations (see Savage-Rumbaugh, 1986, p. 249; Savage-Rumbaugh et al., 1978, pp. 542–543). Lana was apparently never required to demonstrate symmetry with the other objects (e.g., cup, box, can, etc., colored or otherwise) by giving them conditional upon their lexigram samples.

The lexigram training history: A summary. Once learned, Sherman's, Austin's, and Lana's object → lexigram and lexigram → object repertoires were maintained in daily sessions over several years. A sense of just how much this amounted to can be gained by considering an example. Savage-Rumbaugh (1986, p. 212) reported that, in the space of just 4 days, Sherman and Austin participated in 255 "communicative bouts," each successive bout requiring them to act alternately as "speaker" (match lexigrams to objects) and "listener" (match objects to lexigrams). Clearly, both chimps had therefore experienced thousands of trials of symmetry exemplar training with a large number of lexigrams and their corresponding objects. There is, however, no evidence that this reinforcement history had established generalized symmetrical responding, because the chimps had never undergone systematic symmetry testing under standardized, controlled conditions. Nevertheless, Cerutti and Rumbaugh (1993) claim that Sherman and Austin demonstrated symmetry in a study conducted by Savage-Rumbaugh, Rumbaugh, Smith, and Lawson (1980). The study was conducted after the chimps had learned most of their object → lexigram and lexigram → object relations, so its apparent success might be taken as evidence in favor of the exemplar-training hypothesis. The study employed three sets of stimuli; food items and tools (Set A), lexigrams corresponding to each specific food and tool (Set B), and two other lexigrams, one intended to function as a general category label for *food* and the other for *tool* (Set C). In summarizing the study, Cerutti and Rumbaugh report that (a) the chimps had already learned to select specific food and tool lexigrams conditional upon corresponding food items and tools (AB); (b) Sherman and Austin were then trained to select the appropriate category lexigrams conditional upon the food items and tools (AC); (c) Sherman and Austin were then able, without further training, to select the category lexigrams conditional upon specific lexigrams for food items and tools (BC), scoring correctly on 15 of 16 and 17 of 17 Trial 1 presentations, respectively; and (d) this outcome required symmetry, because for BC to emerge, BA would itself have to emerge from the AB training via symmetry before combining transitively with the trained AC relations to yield BC. However, the BA food or tool lexigram → object relations did not have to emerge in this way because they had already been trained. This is confirmed in Table 1 of the source article, and in figures showing that the BA relations were maintained by Sherman and Austin throughout the duration of the study (see, e.g., Savage-Rumbaugh, 1981, Figure 1, p. 41; BA is labeled as ''receptive''). Because BA relations were trained, the chimps' success with the BC relations at best demonstrates only transitivity (BA and AC yielding BC), a form of unidirectional transfer that has none of the bidirectional properties of symmetry. We say ''at best'' because Epstein (1982) has given an alternative account that suggests that the BC performance was itself directly trained by Pavlovian contingencies present during the study. That being so, the BC relations would not be truly emergent as transitivity demands. As far as we are aware, Epstein's account has never been contradicted by the chimps' teachers.

Claims that the chimps have demonstrated symmetry with lexigrams should also be viewed skeptically, because the differential outcome procedures used in much of their lexigram training may render any such demonstration invalid. Take, for example, object → lexigram matching reinforced via differential outcomes. The sequence of events would be as follows: See Food or Tool A (sample) → see (and select) Lexigram B (comparison) → see (and select) same Food or Tool A (now presented as a reinforcer). Note that both AB and BA are embedded in the same training sequence; the experimenter may have only intended to train AB, but doing so via differential outcomes may also train BA (albeit incidentally). Cases in which AB and BA have been trained via a reinforcement contingency are not valid demonstrations of symmetry, because the latter require that BA emerges untrained after training only AB. Studies that train AB and BA, CD and DC, EF and FE, and so on, either with or without differential outcomes, provide trained exemplars of the kinds of performances that would have constituted symmetry if the second conditional discrimination of each pair had been produced under test (i.e., unreinforced) conditions after training only the first of each pair. It remains an open question whether or not the provision of such exemplars would eventually bring about symmetry.

This is why it is necessary to determine whether these chimps' extensive history of many hundreds of such reinforced exemplars established generalized symmetrical responding. To find out, one would have to test for the emergence of BA in standardized controlled conditions after training a new AB relation without differential outcomes (to avoid inadvertently training BA). That was the aim of our study.

GENERAL METHOD

Subjects

The subjects were 3 adult chimpanzees (*Pan troglodytes*), Sherman, Austin, and Lana, who were 13, 12, and 16 years old, respectively, at the start of the study. Consistent with their earlier lexigram training, no chimp was food deprived in the general sense. However, highly preferred food (e.g., candies, exotic fruits, etc.) were reserved as reinforcers, and the chimps rarely had access to these at any time other than during experimental sessions. These foods had been effective reinforcers in the chimps' previous training (see Rumbaugh, 1977; Savage-Rumbaugh, 1986).

Throughout our study the chimps consumed these food items immediately upon delivery, and there was no evidence of satiation despite the continuous availability of other less preferred food outside the experimental sessions. It should also be noted that the chimps did not have access to a lexigram keyboard at any time during the year in which the present study was conducted, so ''food manding'' with lexigrams was not possible.

Apparatus

The experimental chamber was the middle room of the chimpanzees' living quarters. During each experimental session, disturbances were minimized as much as possible by keeping the subject alone in the experimental chamber with the chamber doors locked. On a wall at one end of the chamber was a five-key stimulus–response panel. Each key was 5 cm square and made of transparent acrylic plastic. Four of the keys were located at the corners of a rectangle (21 cm by 12 cm), with the fifth key in the rectangle's center. The panel was mounted so that the center of the middle key was 76 cm above the floor. When a chimp was seated in front of the panel, its eyes were approximately level with the middle key.

The stimulus–response panel was positioned in front of a color monitor screen so that the stimuli displayed on the screen appeared in the center of the keys. The monitor was connected to an Apple® microcomputer that programmed the sequencing and display of the stimuli and recorded key presses. The stimuli were two colors (red and green, each drawn as a 4 cm square) and four shapes (a Y, a zig-zag, a triangle, and a cross, each drawn white on a black background to occupy a 4 cm square area).

A food chute was placed directly beneath the panel, approximately 30 cm from the floor. A variety of food items were dispensed down the chute via an automatic dispenser controlled by the microcomputer.

Procedure

Sessions of 48 trials (unless otherwise stated) were given at least once a day 5 days per week. Each trial of conditional discrimination training began with the presentation of one of a number of sample stimuli on the center key of the display panel. When the sample came on, pressing the center key produced two comparison stimuli, one on each of two of the outer keys (the other two outer keys remained dark). The appearance of the comparisons was accompanied by an audible beep from the computer. Pressing the sample key or either of the dark outer keys had no effect once the comparisons had appeared. On reinforced trials, pressing the key on which the correct comparison was shown produced a high-pitched tone from the computer and the delivery of food from the food dispenser. Pressing the key on which the incorrect comparison was shown produced a low-pitched tone and no food. After a correct or incorrect choice all stimuli were removed (i.e., the screen went blank) and a 5-s intertrial interval followed, at the end of which the next sample appeared. On unreinforced trials, correct and incorrect responses produced the intertrial interval only, and neither food nor tones were delivered. The procedure was noncorrection throughout; that is, errors did not cause trials to be repeated.

All trial types (sample–correct comparison combinations) appeared equally often in a session and were presented quasirandomly, with the restrictions that a given trial type could appear no more than three times in a row and all four comparison keys had to be scheduled as correct before any could be correct again.

PHASE 1: IDENTITY MATCHING

We began by ensuring that the chimps could perform identity matching (i.e., select the comparison that is identical to the sample) with the shapes and colors that were to be used in subsequent phases of arbitrary matching and symmetry testing. After teaching the chimps shape matching, we tested their color matching, first with unreinforced probes and then with reinforced test trials. This preliminary phase acquainted the chimps with our MTS apparatus and general conditional discrimination procedures. It also allowed them to see each stimulus appear as sample and comparison prior to symmetry testing. This was important for the following reason. In a symmetry test, stimuli that formerly appeared as samples on the center key now appear as comparisons on the outer keys, and vice versa. If it is the very first time the subject has seen those stimuli in those locations, then that novelty itself, rather than an absence of symmetry, might be responsible for any subsequent failure. Our preliminary phase therefore mitigated against this potential source of artifactual responding. It also allowed us to determine whether stimulus relations already in the chimps' repertoires would be displayed under unreinforced test conditions. The chimps had already learned identity matching with these and other hues several years prior to our study (see Essock, 1977; Savage-Rumbaugh, 1986, p. 407). If the chimps gave no evidence of color matching on these unreinforced tests, then we would have reason to suspect any negative outcome on subsequent unreinforced symmetry tests. Moreover, if they persisted in failing our color-matching test even after introducing reinforcement for correct responding, then that would suggest that our test format differed from those used in their earlier training, not only in structural and procedural respects but also functionally. If we could not get transfer of a simple color-matching relation across these contexts, then we could hardly expect to show transfer of generalized symmety across those same contexts.

Method and Results

Sherman, Austin, and Lana were taught identity matching with the triangle and cross stimuli (Set 1) first, reaching the criterion of 90% correct in two consecutive sessions in six, three, and nine sessions, respectively. They were then taught identity matching with the Y and zig-zag stimuli (Set 2; stimuli in a set always appeared together as comparisons), reaching criterion in 11, three, and nine sessions, respectively. Then the feedback (i.e., the programmed consequence for a correct or incorrect response) on shape-matching trials was gradually reduced from a probability of 1.0 (feedback on every trial) to .2 (feedback, on average, every fifth trial), while maintaining criterion levels of accuracy. This took five sessions for Lana and Austin and three for Sherman. In Sherman's and Austin's case, these reduction sessions consisted of equal numbers of identity trials from both stimulus sets, whereas in Lana's case they comprised identity trials from Set 1 only.

In the test sessions that followed, 16 color-matching trials with the red and green stimulus set were interspersed among 32 baseline shape-matching trials. Sherman's 11 test sessions and Austin's eight consisted of eight trials of each of the four shape-matching baseline trial types and eight trials of each of the two color-matching probe trial types. In the first five of her 11 test sessions, Lana received 32 baseline shape-matching trials with the triangle and cross (16 trials each) and 16 color-matching probes (eight trials each). Thereafter, Lana's test sessions were the same as Sherman's and Austin's. For all but the final session, feedback was delivered on shape-matching trials according to the .2 probability schedule, whereas the color-matching probes were unreinforced. In the final test session, feedback was presented on every trial (i.e., all correct instances of identity matching were reinforced).

In each test session, each chimp's baseline shape-matching performance was at or above 90% correct per trial type. Figure 1 shows the chimps' performance on color-matching probes. Performance was poor on most unreinforced test sessions, the exceptions being Sherman's seventh and ninth sessions and Lana's seventh session, in each of which only one error was made in the 16 unreinforced color-matching trials. In the 11th session, when reinforcers were given for every correct response, Sherman and Austin made no errors and Lana made only one on the color-matching trials.

Discussion

All 3 chimps had learned color matching prior to our tests, yet only 2 showed any sign of this preexisting relational repertoire when tested on our apparatus with unreinforced probes. Prolonged testing in extinction, however, did not subsequently prevent the chimps from responding at criterion levels on color-matching probes once reinforcement for correct responding was introduced. (Note that although the chimps had only one reinforced color-matching session in this phase, their color-matching performance in subsequent sessions typically remained above 90% correct.) In subsequent tests of symmetry, we therefore began probing for symmetry with unreinforced trials (because positive results would be most convincing when obtained

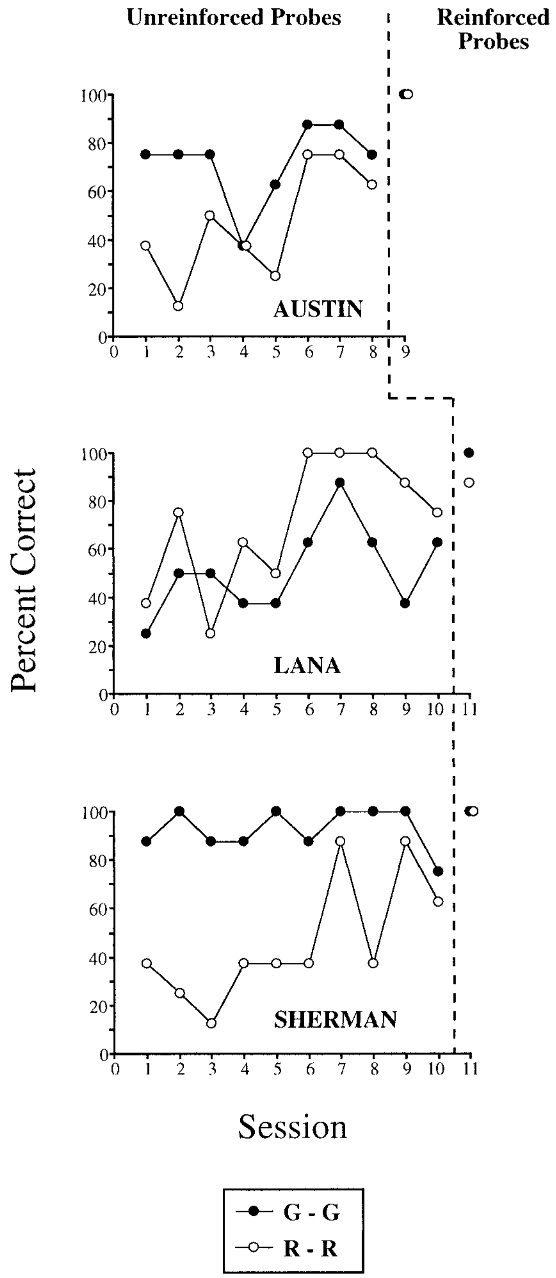

Fig. 1. Sherman's, Austin's, and Lana's percentage of correct responses on the red and green identity matching trial types in the test sessions of Phase 1. Each data point represents eight trials.

without reinforcement; see Sidman, 1981) and then, if the outcome of these unreinforced tests was at all equivocal, we introduced reinforced probes and took immediate criterion accuracy as evidence of symmetry.

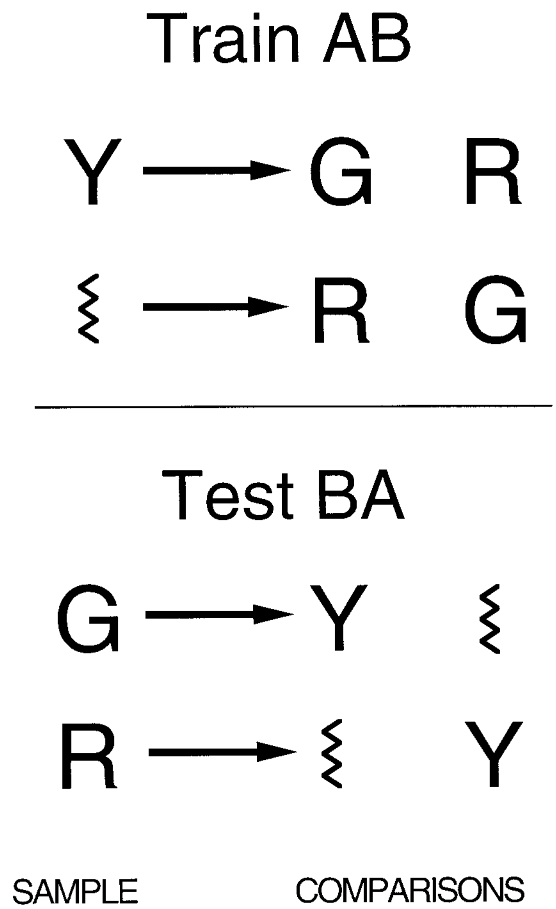

Fig. 2. Stimulus relations presented to the chimps during AB training trials and BA symmetry test trials. Arrows point from sample stimuli (only one presented at a time) to corresponding (correct) comparison stimuli (G = green, R = red).

PHASE 2: AB TRAINING

Method and Results

The subjects were next presented with AB arbitrary matching trials in which the Y and zig-zag were Set A samples and the green hue and red hue were Set B comparisons. When the sample was a Y, reinforcers were contingent upon choice of the green comparison, and when the zig-zag was the sample, reinforcers were contingent upon choice of the red (see Figure 2).

After trial-and-error teaching procedures failed, several training procedures were used in an attempt to accelerate learning of the AB task. Table 1 lists the number of 48-trial ses-

Table 1

The number of 48-trial sessions presented to each chimp in each Phase 2 stage of AB training and preparation for testing.

AB session type	Number of sessions		
	Lana	Sherman	Austin
Training			
Standard AB	4	26	27
Compound samples		9	
Enlarged baseline	4	3	5
Interspersed identity trials	2	2	3
Fading	30[a]	23[a]	82
Preparation for testing			
Stage 1: Fading, maximum cue at start, reinforcement probability reduced from 1.0 to .2	8[b]	6[b]	
Stage 2: Fading, no cue at start, reinforcement probability = .2	2[b]	5[b]	
Stage 3: Standard AB (no fading), reinforcement probability = .2	2[b]	2	
Stage 4: Standard AB (no fading), reinforcement probability = 1.0		2[a]	

[a] Criterion accuracy reached at the end of this stage.
[b] Criterion accuracy maintained throughout this stage.

sions each chimpanzee received on each AB training procedure. Unsuccessful procedures included (a) enlarging the arbitrary matching baseline by two more trial types comprised of Set A samples and triangle and cross comparisons (sample–correct comparison combinations were Y–triangle and zig-zag–cross); (b) interspersing AA and BB identity matching trials among the AB trials; and (c) compound samples (Sherman only): Each Set A sample shape was colored the same as its corresponding Set B comparison (i.e., the Y was colored green, and the zig-zag was colored red), and then, once responding had reached criterion, the color was removed from the samples.

Sherman and Lana eventually learned the AB relations via a fading program. This program capitalized on their preexisting color-matching skills. Each sample shape was initially drawn on a background square of color that matched the color of its corresponding comparison; the Y sample appeared on a green background, whereas the zig-zag sample appeared on a red background. At the start of the session, the background square was identical in size and shape to the comparisons (i.e., 4 cm square). Thereafter, when the subject responded correctly to a given trial type, then the next time that trial type appeared, its sample shape was centered on a slightly smaller color square. Conversely, an incorrect response to a given trial type made the sample color square slightly larger on the next presentation of that trial type (up to a maximum of 4 cm square). There were five sizes of background color cue for each sample shape, ranging from 4 cm square to 0.5 cm square. Trials containing the smallest background color cue represented the final step of the fading program; a correct response to these trials resulted in the disappearance of the sample color cue (i.e., the presentation of a standard AB trial) on the next occurrence of that trial type. From that point on, standard AB trials were presented if the subjects continued to respond correctly; errors caused the sample color cue to reappear in ascending order of size. To reach criterion the subjects had to make no more than a single error for at least 18 trials per trial type with no sample color cue. Sherman and Lana reached criterion in 23 and 30 sessions, respectively. Austin, however, never reached criterion (see Table 1) and was therefore dropped from the study.

Next, Sherman and Lana went through several stages in preparation for BA testing (see Table 1). In Stage 1, Sherman and Lana received sessions in which the probability of feedback on each AB trial was gradually lowered from 1.0 to .2. Fading was still in effect and each AB session began with sample color cues at their maximum size. After reaching the .2 level, additional sessions in Stage 2 began with no sample color cue, although fading was still in effect (i.e., the cue reappeared on the first trial following an incorrect choice). In Stage 3, sessions were conducted at the .2 level with the color cues totally absent (no fading). Lana's AB performance remained above criterion throughout these stages, so she proceeded to unreinforced BA testing. In contrast, Sherman's AB accuracy decreased to 79% in the first Stage 3 session (Y–green = 87.5%, zig-zag–red = 71%) and 58% in the second (Y–green = 58%, zig-zag–red = 58%). He therefore went through an additional stage (Stage 4) consisting of two sessions in which the probability of reinforcement was restored to 1.0 (every correct response was reinforced). Sherman's AB performance in Stage 4 recovered to 89% in the first session (Y–green = 100%, zig-zag–red = 79%) and 96% in the second (Y–green = 96%, zig-zag–red = 96%). We decided to let Sherman proceed directly to reinforced symmetry testing (see Phase 5 below).

PHASE 3: UNREINFORCED BA (SYMMETRY) TESTS

Method and Results

Each session in this phase consisted of 32 AB baseline trials (16 per trial type) and 16 BA symmetry test trials (eight per trial type). On symmetry test trials, the stimuli that had formerly been samples and comparisons were interchanged; that is, a green or red sample was presented with Y and zig-zag comparisons (see Figure 2). Correct responses on BA symmetry test trials were not reinforced, so the probability of feedback on AB baseline trials was increased sufficiently to maintain an overall probability of .2.

Lana was given 12 test sessions; the percentage of correct responses for each trial type is shown in Figure 3. Lana's AB baseline performance was about 90% correct throughout testing, and in each session she never made more than three errors (out of 16 trials) per trial type. Despite this, her accuracy on the symmetry test trials was at or around the chance level of 50% correct throughout testing.

Discussion

Lana's failure during the first phase of the symmetry test may have been due to the test procedure. First, Lana's performance may have deteriorated on BA test trials simply because on those trials no reinforcement was forthcoming. Second, on BA test trials Lana was presented with the Set A shapes as comparisons for the first time since Phase 1 identity matching. These relatively novel stimulus arrangements might have disrupted her test performance. The next phase sought to eliminate these possibilities.

PHASE 4: REINFORCED BA (SYMMETRY) TESTS WITH IDENTITY TRIALS

Method and Results

Two changes were made for Lana's second symmetry test phase. First, reinforcement was given for each correct (i.e., class consistent) response on baseline and test trials. Second, identity matching trials requiring Lana to match each Set A shape and each Set B color to itself were added to the baseline. These identity trials ensured that Lana responded discriminatively to sample colors *and* comparison shapes during testing. Eight test sessions were conducted, each consisting of 12 BA symmetry test trials interspersed randomly among a baseline of 24 AB trials and 24 identity matching trials.

Figure 4 shows that accuracy was consistently high on all the baseline trial types. In contrast, accuracy on the BA symmetry trials did not reach criterion when reinforcement was introduced, and it actually decreased as testing progressed. Lana increasingly chose the zig-zag comparison irrespective of the sample hue, doing so on all but two of the 24 BA trials in each of the final two sessions. Lana thus failed the symmetry test even with reinforced test trials.

Discussion

On BA trials, the red sample and green sample were presented successively, never appearing on screen together. Perhaps Lana had failed the symmetry test because she had not discriminated between red and green hues when presented separately. Lana performed well on baseline hue–hue identity trials, but because those trials presented red or green samples alongside red and green comparisons, they required only simultaneous discrimination of the hues.

PHASE 5: REINFORCED BA (SYMMETRY) TESTS WITH ZERO-DELAY IDENTITY TRIALS

Method

This phase was identical to the previous one except that zero-delay identity matching trials replaced simultaneous identity match-

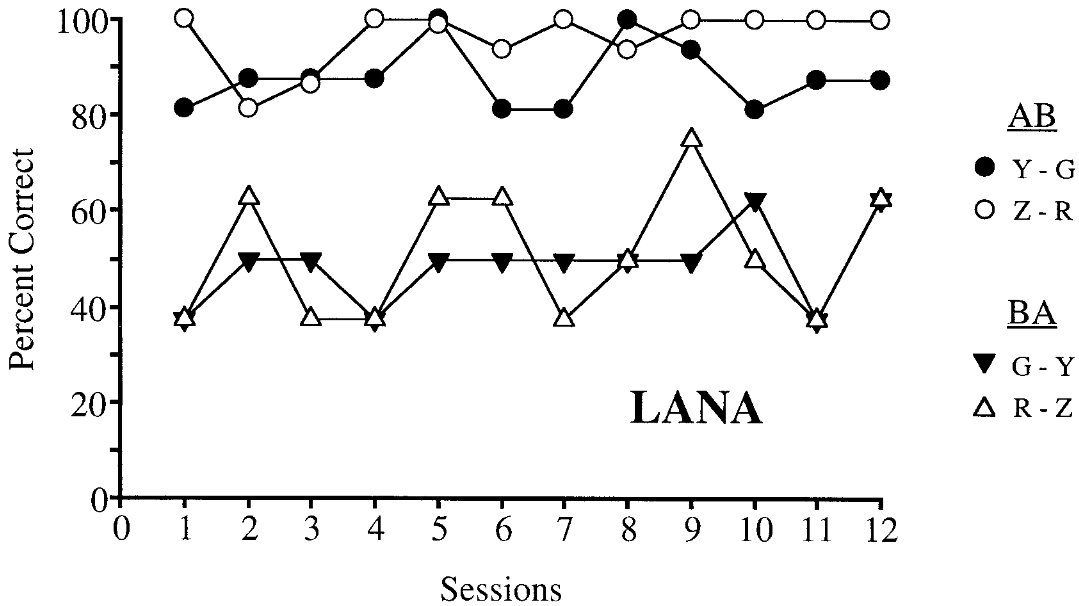

Fig. 3. Lana's percentage of correct responses on each trial type in all Phase 3 sessions. Each AB data point represents 16 trials, and each BA point represents eight trials. (Y = Y shape, Z = zig-zag, G = green, R = red)

ing trials. On zero-delay trials the sample disappeared when it was pressed and then the comparisons appeared immediately. High accuracy on these baseline trials would show that the subject was able to discriminate successively the same B samples as in symmetry probes at the time of testing (cf. D'Amato, Salmon, Loukas, & Tomie, 1985). Lana had 15 sessions under these procedures, and Sherman had 12. Every correct response was reinforced.

Results and Discussion

Lana. The procedural change made little difference to Lana's test performance. Figure 5 shows Lana's accuracy on trial types in symmetry test sessions and in intervening baseline-only sessions that were introduced following baseline deterioration in symmetry tests. When performance on a baseline trial type fell below 80% correct in a test session, then the BA probes were removed within one or two sessions from the onset of deterioration and were reinstated (thus resuming testing) only when performance in a baseline-only session was above 80% correct per trial type. In the first two tests, Lana still showed a marked preference for the zig-zag comparison on BA trials, while her score on all baselines was excellent (only three errors from 72 baseline trials, and never more than a single error per trial type per session). Lana still failed the tests, even though she could discriminate successively between the hues, and even though reinforcers continued to be given for every correct choice on the symmetry probes.

Lana's preference for the zig-zag comparison on BA probes diminished with each test session, and by the end of the phase her accuracy on the green–Y trial type had risen to match those previously achieved on the red–zig-zag trials. Her BA score averaged 85% correct over the last four sessions (48 trials) of testing. By that time, however, Lana had been given a total of 276 reinforced BA trials, presumably sufficient for the relations to be learned via direct training and reinforcement.

Figure 5 shows that Lana's accuracy on the green–green zero-delay identity trials fell below 80% correct on Test Sessions 3, 4, 5, and 6. It might be argued that these sessions should be discounted because we could not tell that Lana was discriminating the hues successively in these sessions. However, one should examine Lana's pattern of responding on hue–hue trials during the four disrupted sessions (Tests 3 through 6). When the sam-

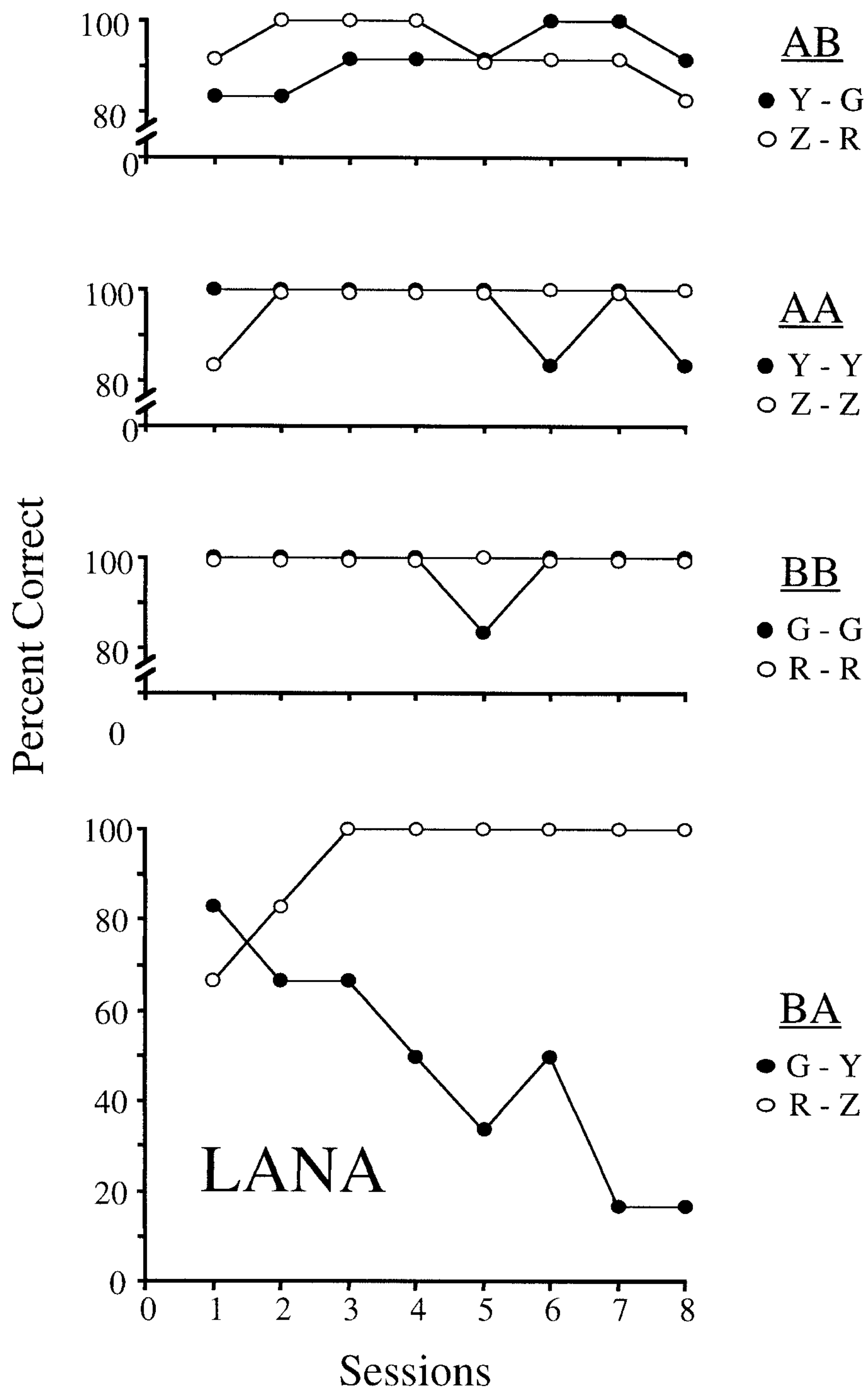

Fig. 4. Lana's percentage of correct responses on each trial type in all Phase 4 sessions. The top panel shows data from both AB baseline trial types, the middle two panels show data from the identity matching baseline trial types (AA and BB), and the bottom panel shows data from symmetry test trials (BA). Each AB data point represents 12 trials; all other points represent six trials.

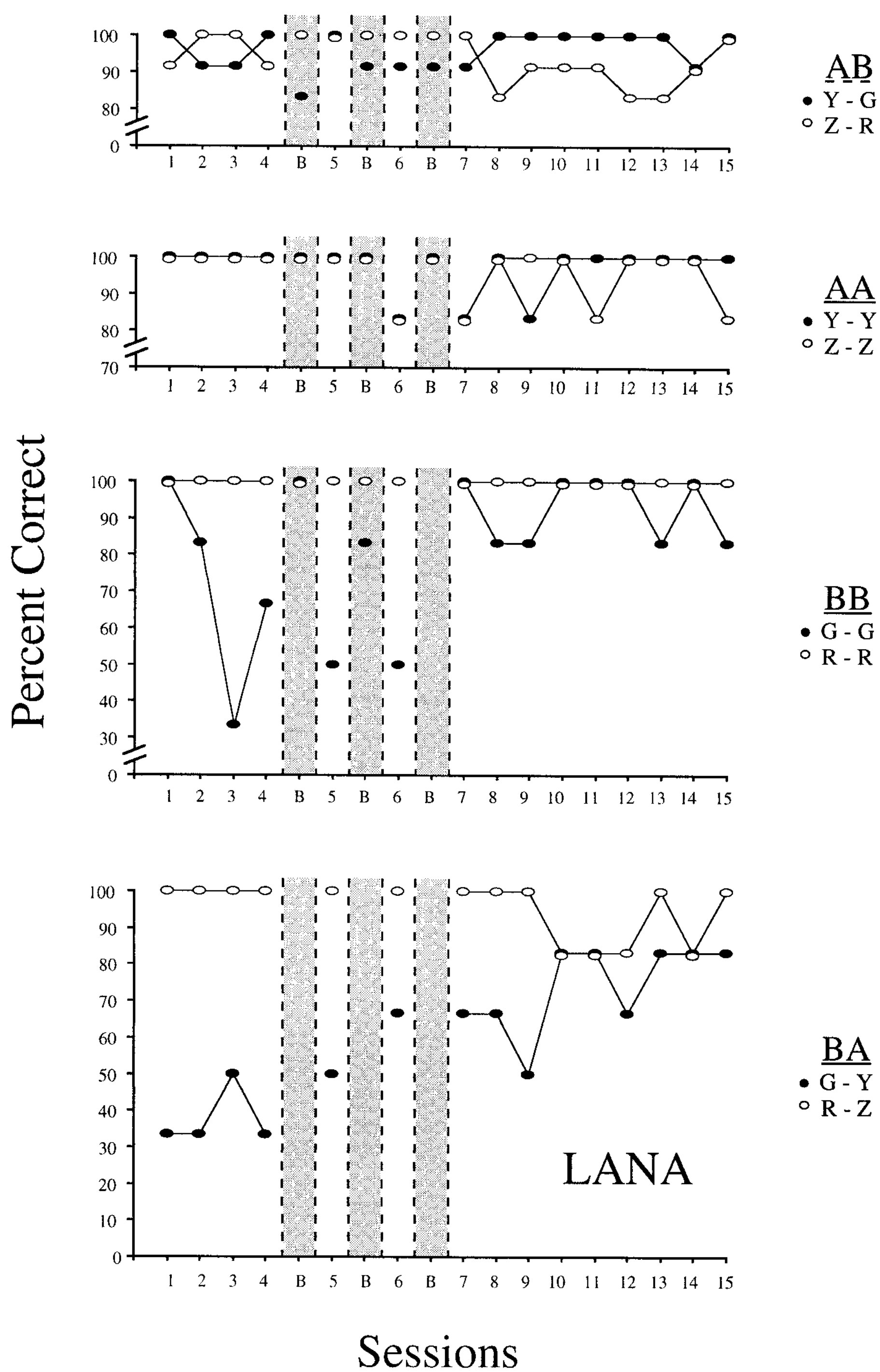

Fig. 5. Lana's percentage of correct responses on each trial type in all Phase 5 sessions. The top panel shows data from both AB trial types, the middle two panels show data from the zero-delay identity matching trial types (AA and BB), and the bottom panel shows data from symmetry test trials (BA). Symmetry test sessions are numbered consecutively along the abscissa and occupy unshaded areas of the figure. Sessions marked with a B on the abscissa (occupying shaded areas) are those in which baseline trials were presented alone (BA symmetry trials were removed) in an attempt to reverse deteriorating trends in baseline performance. Each AB data point represents 12 trials; all other points represent six trials.

ple was red Lana chose the red comparison every time, but when the sample was green she chose red the same number of times as green (12 times in 24 trials). It is unlikely that such a pattern could occur in the complete absence of successive hue discriminations. It might also be argued that Tests 3 through 6 should be discounted because the disrupted baselines may have adversely affected Lana's performance on the BA probe trials. Figure 5, however, suggests that the reverse was true. Lana's green–green identity baseline deteriorated in the presence of the probes and recovered each time they were removed. Such probe-induced disruption was cited by Sidman et al. (1982) as additional evidence against symmetry, and therefore equivalence. If the stimuli were equivalent, then reinforcement on symmetry test trials should, if anything, strengthen the baseline relations, and certainly should not weaken them.

Sherman. Figure 6 shows that Sherman's performance on the reinforced symmetry probes was at or around chance levels in all 12 test sessions. His scores on the zero-delay hue–hue baseline remained excellent throughout, ruling out sample-discrimination failure as the source of his poor symmetry score. He also made no more than a single error per session on either the Y–green or Y–Y baselines. However, baseline disruption occurred on trials with zig-zag samples, particularly on zig-zag–red AB trials (which fell below 80% correct on eight of the 12 sessions). These disrupted baselines recovered when symmetry probes were removed (albeit less quickly than was the case with Lana), only to deteriorate again when probes were reinstated. If zig-zag and red had been equivalent (and thus symmetrically related), then reinforcement of correct responses on red–zig-zag trials should have strengthened Sherman's performance on zig-zag–red trials, not weakened it. Further evidence against symmetry comes from the four sessions in which Sherman's accuracy in all baseline trial types was uniformly high, yet his accuracy on the symmetry probes (48 trials in total) remained at chance level.

GENERAL DISCUSSION

The chimpanzees we tested are unique among all nonhumans in that they have a history of conditional discrimination and symmetry exemplar training that is quite unprecedented in its extent and complexity. Yet Lana failed the symmetry tests despite having 468 BA test trials, 276 of which were reinforced, and Sherman failed despite having 144 (all reinforced). Negative results such as these are most convincing when produced with reinforced probes. Evidence that the reinforcers were effective came from the fact that (a) when they were first introduced in the tests of identity matching in Phase 1 of the study, each chimp's color-matching accuracy immediately rose from around chance level to above criterion, (b) the high accuracy on all baseline trial types (including color matching) in symmetry tests were subsequently maintained in the vast majority of sessions (falling below 80% on only 10 of Lana's 138 baseline data points and on only 20 of Sherman's 72 baseline data points; see Figures 3 through 6), and (c) as was the case with Sidman et al. (1982) in their symmetry tests of rhesus monkeys and baboons, probe-induced disruption of baselines occurred only during reinforced tests, never when probes were unreinforced. In this respect at least, our reinforcers functioned in a manner indistinguishable from those used by Sidman et al. (which were food reinforcers established via standard food-deprivation procedures). In sum, we are confident that the chimps' test failure occurred in spite of presenting strong positive consequences for correct responding.

It is also relevant to note that Sherman's and Lana's rate of learning the AB relations via fading compares well with that of 2-year-old normal children in studies by Beasty (1987). Using the same matching-to-sample procedure, essentially the same apparatus, and the same learning criterion, Sherman required 1,104 trials, Lana 1,440 trials, and the children 1,068 trials on average (range, 912 to 1,224) to learn to match shape samples to red and green comparisons via the same stimulus fading procedure (presented in each case after shape and hue identity matching had been established). The difference between the children's and the chimps' MTS performances came not during this training phase, but on later testing; unlike the chimps, all these children passed their subsequent symmetry tests, even with unreinforced

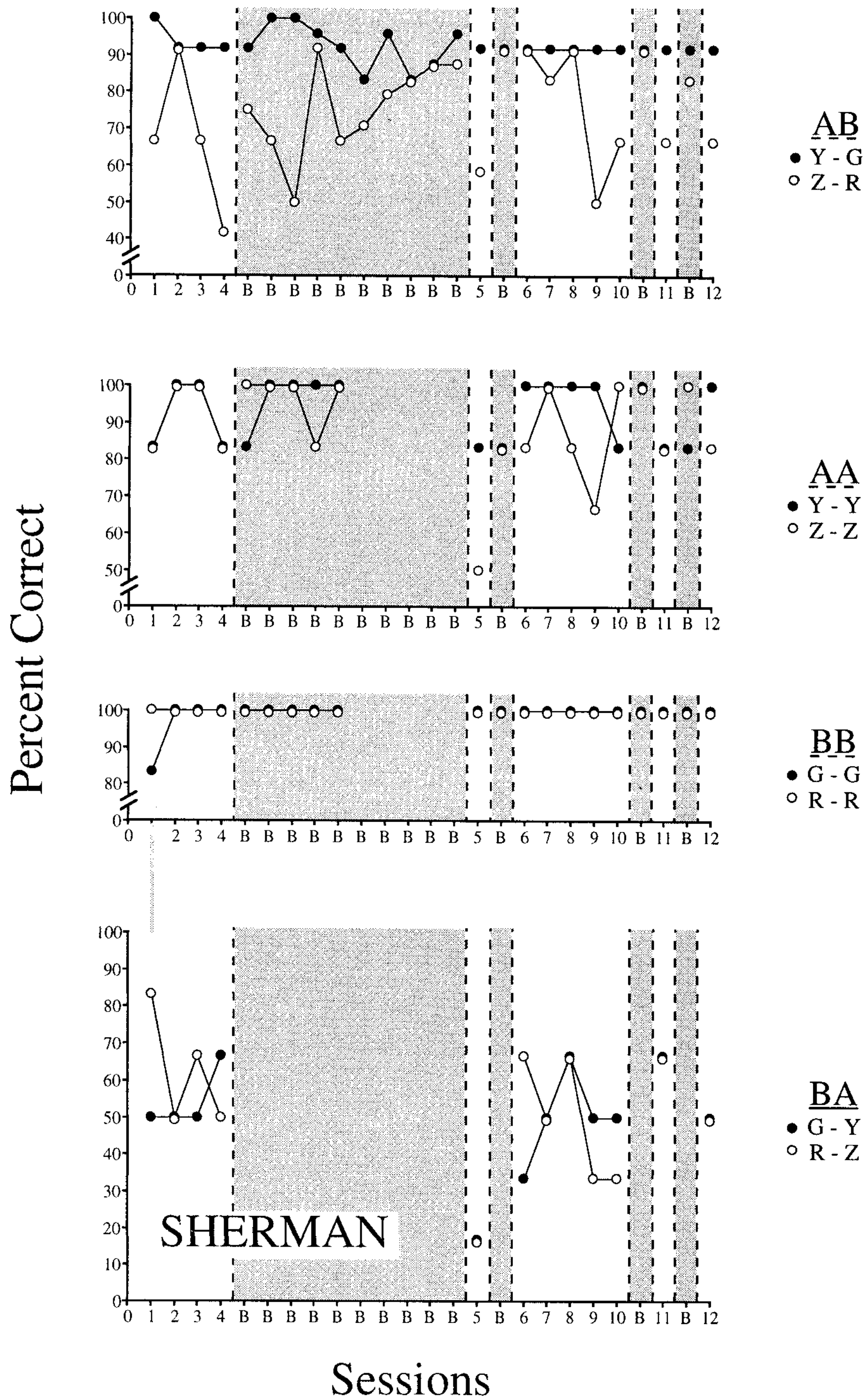

Fig. 6. Sherman's percentage of correct responses on each trial type in all sessions of Phase 5 (the only phase in which he was tested). The top panel shows data from both AB baseline trial types, the middle two panels show data from the zero-delay identity matching baseline trial types (AA and BB), and the bottom panel shows data from symmetry test trials (BA). Symmetry test sessions are numbered consecutively along the abscissa and occupy unshaded areas of the figure. Sessions marked with a B on the abscissa (occupying shaded areas) are those in which baseline trials were presented alone (BA symmetry trials were removed) in an attempt to reverse deteriorating trends in baseline performance. Each AB data point represents 12 trials; all other points represent six trials.

probes. Such rates of learning do not therefore preclude passing symmetry tests.

Given their extensive history of symmetry exemplar training, why were the chimps unable to pass our standard symmetry test? One possibility is that their exemplar training with lexigrams did give rise to the ability to pass symmetry tests but that this potential was not realized on our tests because the context in which they were conducted was too far removed from the original exemplar-training context. The chimps' tests were conducted in the absence of a lexigram keyboard, although it was present during all of their symmetry exemplar training prior to this study. Their responding in accordance with symmetry exemplars had always previously been reinforced in the presence of a lexigram keyboard and never in its absence. Under such conditions the keyboard, the lexigrams, or any other stimuli closely associated with lexigram training (e.g., the three-dimensional food or tool objects used as opposed to the exclusively two-dimensional stimuli used in our tests) could come to function as contextual stimuli signaling the appropriateness of responding in accordance with symmetrical relations, and their presence during testing might therefore set the occasion for the subjects to pass symmetry tests (cf. Bush, Sidman, & de Rose, 1989; Gatch & Osborne, 1989; Kennedy & Laitinen, 1988; Wulfert & Hayes, 1988).

It is, however, important not to lose sight of the procedural similarity of the chimps' exemplar training history and our tests. Both were comprised exclusively of trials in which visual comparisons were selected conditional upon visual samples (i.e., visual–visual ''selection-based'' conditional discriminations; see Michael, 1985). This is not the case with normal human children; their exemplar training consists of auditory–visual listener or speaker relations, not visual–visual relations as in MTS symmetry tests, and the speaker relations involve making ''topographically distinct'' verbal responses (Michael, 1985), not selection-based responses as in symmetry tests. The task remains for exemplar training theory to explain why children as young as 2 years of age can pass symmetry and equivalence tests with a format (arbitrary visual–visual matching) that they have probably never encountered before (Green, 1990; Horne & Lowe, 1996, in press), whereas the chimps failed even though they had never encountered anything but that format.

Perhaps the chimps failed our tests because their earlier exemplar training had simply not given rise to any symmetrical responding with lexigrams. This merits serious consideration. Our review of the chimps' lexigram training (see the introduction) revealed no evidence of symmetry.

We fully acknowledge that the negative results of the present study cannot be considered conclusive. It is, of course, the case that although these chimpanzees have not passed symmetry tests thus far, it does not mean that they never could given modifications to the training or test procedure. Perhaps the chimps were not adequately prepared for our symmetry tests or were not adapted to the novelty of our apparatus. Although these possibilities remind one of the maxim, ''absence of proof is not proof of absence,'' we wish to stress that absence of proof is not proof of presence either! Taken together with the negative findings from tests of symmetry with a variety of nonhuman species, the present data from chimpanzees support suggestions (Dugdale & Lowe, 1990; Horne & Lowe, 1996) that equivalence classes may only be observed in the repertoires of linguistically competent organisms. Further experimentation with a range of nonverbal subjects, both human and nonhuman, will, of course, need to be conducted before this issue can be conclusively resolved. We eagerly await the outcome of that inquiry and hope that it is stimulated in part by the study reported here.

REFERENCES

Beasty, A. (1987). *The role of language in the emergence of equivalence relations: A developmental study.* Unpublished doctoral thesis, University of Wales.

Boelens, H. (1994). A traditional account of stimulus equivalence. *The Psychological Record, 44,* 587–605.

Brodigan, D. L., & Peterson, G. B. (1976). Two-choice conditional discrimination performance of pigeons as a function of reward expectancy, prechoice delay, and domesticity. *Animal Learning & Behavior, 4,* 121–124.

Bush, K. M., Sidman, M., & de Rose, T. (1989). Conditional control of stimulus classes. *Journal of the Experimental Analysis of Behavior, 51,* 29–45.

Cerutti, D. T., & Rumbaugh, D. M. (1993). Stimulus relations in comparative primate perspective. *The Psychological Record, 43,* 811–821.

D'Amato, M. R., Salmon, D. P., Loukas, E., & Tomie, A. (1985). Symmetry and transitivity of conditional relations in monkeys (*Cebus apella*) and pigeons (*Columba livia*). *Journal of the Experimental Analysis of Behavior, 44,* 35–47.

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

Dugdale, N. A. (1988). *The role of naming in stimulus equivalence: Differences between humans and animals.* Unpublished doctoral thesis, University of Wales.

Dugdale, N. A., & Lowe, C. F. (1990). Naming and stimulus equivalence. In D. E. Blackman & H. Lejeune (Eds.), *Behaviour analysis in theory and practice: Contributions and controversies* (pp. 115–138). Brighton, UK: Erlbaum.

Epstein, R. (1982). ''Representation'' in the chimpanzee. *Psychological Reports, 50,* 745–746.

Essock, S. M. (1977). Color perception and color classification. In D. M. Rumbaugh (Ed.), *Language learning by a chimpanzee* (pp. 207–224). New York: Academic Press.

Essock, S. M., Gill, T. V., & Rumbaugh, D. M. (1977). Language relevant object- and color-naming tasks. In D. M. Rumbaugh (Ed.), *Language learning by a chimpanzee* (pp. 193–206). New York: Academic Press.

Gatch, M. B., & Osborne, J. G. (1989). Transfer of contextual stimulus function via equivalence class development. *Journal of the Experimental Analysis of Behavior, 51,* 369–378.

Gill, T. V., & Rumbaugh, D. M. (1974). Mastery of naming skills by a chimpanzee. *Journal of Human Evolution, 3,* 483–492.

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

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

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

Hayes, S. C., & Hayes, L. J. (1992). Verbal relations, cognition, and the evolution of behavior analysis. *American Psychologist, 47,* 1383–1395.

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.

Horne, P. J., & Lowe, C. F. (1997). Toward a theory of verbal behavior. *Journal of the Experimental Analysis of Behavior, 68,* 271–296.

Horne, P. J., & Lowe, C. F. (in press). Putting the naming account to the test: Preview of an experimental program. In J. C. Leslie & D. E. Blackman (Eds.), *Issues in experimental and applied analyses of human behavior.* Reno, NV: Context Press.

Kennedy, C. H., & Laitinen, R. (1988). Second-order conditional control of symmetric and transitive relations: The influence of order effects. *The Psychological Record, 38,* 437–446.

Lowe, C. F., & Horne, P. J. (1996). Reflections on naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65,* 315–340.

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

Michael, J. (1985). Two kinds of verbal behavior plus a possible third. *The Analysis of Verbal Behavior, 3,* 1–3.

Rumbaugh, D. M. (1977). *Language learning by a chimpanzee: The LANA project.* New York: Academic Press.

Rumbaugh, D. M., & Gill, T. V. (1976). Language and the acquisition of language-type skills by a chimpanzee (*Pan*). *Annals of the New York Academy of Sciences, 280,* 90–123.

Rumbaugh, D. M., & Gill, T. V. (1977). Lana's acquisition of language skills. In D. M. Rumbaugh (Ed.), *Language learning by a chimpanzee* (pp. 165–192). New York: Academic Press.

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

Saunders, K. J., Williams, D. C., & Spradlin, J. E. (1996). Derived stimulus control: Are there differences among procedures and processes? In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 93–109). Amsterdam: Elsevier.

Savage-Rumbaugh, E. S. (1981). Can apes use symbols to represent their world? *Annals of the New York Academy of Sciences, 364,* 35–59.

Savage-Rumbaugh, E. S. (1984). Verbal behavior at a procedural level in the chimpanzee. *Journal of the Experimental Analysis of Behavior, 41,* 223–250.

Savage-Rumbaugh, E. S. (1986). *Ape language: From conditioned response to symbol.* New York: Columbia University Press.

Savage-Rumbaugh, E. S., Pate, J. L., Lawson, J., Smith, S. T., & Rosenbaum, S. (1983). Can a chimpanzee make a statement? *Journal of Experimental Psychology: General, 112,* 457–492.

Savage-Rumbaugh, E. S., & Rumbaugh, D. M. (1978). Symbolization, language and chimpanzees: A theoretical reevaluation based on initial language acquisition processes in four young *Pan troglodytes. Brain and Language, 6,* 265–300.

Savage-Rumbaugh, E. S., Rumbaugh, D. M., & Boysen, S. (1978). Linguistically mediated tool-use and exchange by chimpanzees (*Pan troglodytes*). *Behavioral and Brain Sciences, 4,* 539–554.

Savage-Rumbaugh, E. S., Rumbaugh, D. M., Smith, S. T., & Lawson, J. (1980). Reference: The linguistic essential. *Science, 210,* 922–925.

Schusterman, R. J., & Kastak, D. (1993). A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *The Psychological Record, 43,* 823–839.

Sidman, M. (1981). Remarks. *Behaviorism, 9,* 127–129.

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

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

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

Tomonaga, M., Matsuzawa, T., Fujita, K., & Yamamoto, J. (1991). Emergence of symmetry in a visual condition-

al discrimination by chimpanzees (*Pan troglodytes*). *Psychological Reports, 68,* 51–60.
Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50,* 125–144.
Yamamoto, J., & Asano, T. (1995). Stimulus equivalence in a chimpanzee (*Pan troglodytes*). *The Psychological Record, 45,* 3–21.

Received July 13, 1998
Final acceptance September 17, 1999