JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1999, **71,** 395–415 NUMBER 3 (MAY)

EQUIVALENCE RELATIONS BETWEEN VISUAL STIMULI: THE FUNCTIONAL ROLE OF NAMING

TOM RANDELL AND BOB REMINGTON

UNIVERSITY OF SOUTHAMPTON

The functionality of verbal behavior in equivalence class formation was demonstrated by training 30 verbally able adults using different combinations of the same easily nameable, yet formally unrelated, pictorial stimuli. Match-to-sample baselines for four four-member classes were established sequentially (i.e., AB-BC-CD), with participants in the rhyme condition trained to select comparisons whose normative names rhymed with those of the samples. For the orthogonal condition, class rearrangement was such that on every trial all available comparisons' names rhymed with each other, but not with the name of the sample. In the diagonal condition, stimuli were allocated pseudorandomly as samples and comparisons. Although all participants maintained baseline discriminations prior to emergent testing, equivalence was confined almost exclusively to the rhyme condition, in which it was ubiquitous. These participants also required less training than those in the control conditions, among whom effects of nodal distance were observed most strongly. Subsequent testing presented participants with no-reinforcement trials involving novel pictorial stimuli, in which one of the available comparisons' names always rhymed with that of the sample. All rhyme participants consistently selected these comparisons. Results indicate that visual stimuli are named, that the phonological properties of those names can influence equivalence class formation, and that the emergence of untrained discriminations may, under certain circumstances, be rule governed.

Key words: stimulus equivalence, naming, generalization, nodal distance, match to sample, adult humans

---

Stimulus equivalence has come to occupy a prominent position within the experimental analysis of human behavior in recent years, and its definition is by now well known. When baseline training establishes relations among three or more behavioral or environmental events, new relations emerge, untrained, among those events. If testing without reinforcement reveals those relations to be reflexive, symmetric, and transitive, the events involved may be described as equivalent, or as participating in an equivalence relation (Sidman, 1990, 1997; Sidman & Tailby, 1982). An extensive literature attests to the generality of such emergent phenomena, and equivalence has been demonstrated using a diversity of stimuli (e.g., Dube, McIlvane, Mackay, & Stoddard, 1987; L. J. Hayes, Thompson, & Hayes, 1989; Markham & Dougher, 1993; Roche & Barnes, 1997), modalities (e.g., Annett & Leslie, 1995; Dube, Green, & Serna, 1993; L. J. Hayes, Tilley, & Hayes, 1988; O'Leary & Bush, 1996; Tierney, DeLargy, & Bracken, 1995), and experimental preparations (e.g., Dube, McIlvane, Maguire, Mackay, & Stoddard, 1989; Leader & Barnes, 1996; K. J. Saunders, Saunders, Williams, & Spradlin, 1993; Sigurdardottir, Green, & Saunders, 1990). Although the majority of studies have employed normal adults (e.g., Roche, Barnes, & Smeets, 1997; Stromer & Stromer, 1990; Wulfert & Hayes, 1988) or children as participants (e.g., Lipkens, Hayes, & Hayes, 1993; Pilgrim, Chambers, & Galizio, 1995), equivalence has also been demonstrated among people with mental retardation (e.g., K. J. Saunders & Spradlin, 1993; Sidman, Cresson, & Willson-Morris, 1974; Stromer & Osborne, 1982) and, more controversially, using nonhuman animals (e.g., McIntire, Cleary, & Thompson, 1987; Schusterman & Kastak, 1993; Vaughan, 1988).

From its inception, equivalence research has been closely linked with the study of verbal behavior (Sidman, 1971; Sidman & Cresson, 1973), and the relationship has remained intimate over the years (Hall & Chase, 1991; S. C. Hayes & Hayes, 1992; Sid-

---

The first author was supported during this research by a grant from the Department of Psychology, University of Southampton. We gratefully acknowledge William Dube and Eric Hiris' provision of the software used in this experiment, and offer our thanks to William Randell for his creation of many of the experimental stimuli.

Correspondence concerning this article should be addressed to Bob Remington, Department of Psychology, University of Southampton, Highfield, Southampton, Hampshire, SO17 1BJ, United Kingdom (E-mail: R.E.Remington@soton.ac.uk).

man, Willson-Morris, & Kirk, 1986). During the last decade, however, debate has intensified regarding the primacy of verbal behavior or equivalence, and over the suggestion that verbal behavior may provide the necessary and sufficient preconditions for equivalence class formation (Horne & Lowe, 1996, 1997).

Contrary to Sidman's conceptualization of equivalence as a behavioral primitive underpinning linguistic function (Sidman, 1994), Lowe and colleagues (Dugdale & Lowe, 1990; Horne & Lowe, 1996, 1997) have argued that equivalence and other higher order human behavior result from participants' overt or covert *naming* of stimuli, regardless of modality. In this view, naming is bidirectional stimulus-classifying behavior and describes the fusion of speaker and listener behavior resulting from an individual's history of reinforcement within a verbal community. The *name relation* is proposed as the basic unit of verbal behavior, and success on tests of equivalence is proposed to result either from participants' common naming of individual stimuli or their linking of individual stimulus names by intraverbal rules (Horne & Lowe, 1996).

Although the majority of relevant studies have reported no consistent evidence for either the necessity or sufficiency of naming in equivalence (e.g., Green, 1990; Lazar, Davis-Lang, & Sanchez, 1984; Sidman et al., 1986), acceptance of such a conclusion has been tempered by the methodological problems inherent in separating verbal and nonverbal experimental outcomes (Perone, 1988; Shimoff, 1984, 1986; Wulfert, Dougher, & Greenway, 1991), especially when the verbal processes implicated may be covert and unmeasurable (Skinner, 1969). Although some studies with children as participants (Barnes, McCullagh, & Keenan, 1990; Devany, Hayes, & Nelson, 1986) have suggested correlations between chronological age, development of verbal behavior, and equivalence class formation, criticisms have been made on methodological grounds (R. R. Saunders & Green, 1996) and because of the correlational nature of the data (Dugdale & Lowe, 1990). As Sidman (1994) has pointed out, causal data would be a step in the right direction toward resolving the naming debate.

With regard to this, Dugdale and Lowe (1990) noted that when normal children who had previously failed equivalence tests were taught a common name for visual stimuli congruent with the classes defining the experiment, they then proceeded to exhibit the pattern of behavior that defines equivalence. Likewise employing visual stimuli, Eikeseth and Smith (1992) reported the facilitative effects of common naming interventions in remediating failures on tests for equivalence among autistic children. Their findings did not, however, demonstrate the necessity of naming for class formation. Lowe and Beasty (1987) similarly demonstrated the efficacy of teaching intraverbal naming strategies to children in producing equivalence. Such studies have attracted criticism, however, for the potential confounding of verbal control by further exposure to the experimental contingencies (Mandell & Sheen, 1994).

Even very young human participants bring to experimentation an extensive verbal history, which in most paradigms has constituted an uncontrolled and uncontrollable variable, precluding definitive analyses. Although various studies have attempted to circumvent such difficulties by using language-disabled participants (e.g., Barnes et al., 1990; Devany et al., 1986) or abstract stimuli (e.g., Lazar et al., 1984; R. R. Saunders, Wachter, & Spradlin, 1988), other research has set out specifically to investigate the relationship between verbal behavior and equivalence by capitalizing on participants' verbal history, through the manipulation of visual stimuli among verbally able adults.

Mandell and Sheen (1994) employed three classes of textual stimuli to control potential naming among their undergraduate participants: phonologically correct, pronounceable pseudowords (e.g., SNAMB), phonologically incorrect pseudowords (e.g., NSJBM), and punctuation marks (e.g., +]*/\!). If, as they suggested, naming is an important determinant of class formation, then the pronounceability of presented stimuli should be indicative of the speed and accuracy with which classes form. Experiment 1 confirmed that participants exposed to pronounceable stimuli demonstrated equivalence more quickly and with greater consistency than participants in the other conditions, and also showed that participants in those conditions tended to produce spontaneously idiosyncratic names for the unpronounceable stimuli. A second experiment indicated that when participants were pretrained to apply names orally to phonologically incorrect pseudowords, their performance was enhanced in comparison to participants who received no such pretraining.

Participants' verbal behavior should not be regarded as a panacea for remediating equivalence failures, however, as Horne and Lowe (1996) have observed. In addition to evidence that some verbally able participants fail equivalence tests (e.g., Dugdale & Lowe, 1990; Eikeseth & Smith, 1992), other findings indicate that participants' verbal behavior can either facilitate or hinder class formation, depending on the congruence of the naming strategies employed with the experimenter-designated classes governing positive test outcomes. Dickins, Bentall, and Smith (1993) first taught three groups of participants baseline relations between sets of pictograms, after which the participants in two of the groups were taught paired associations between the names they had given those stimuli, forming verbal associations discordant with the stimulus classes already established by match-to-sample training. Participants in the third group were taught paired associations between neutral names and those of the experimental stimuli. Participants exposed to across-class paired associations were less successful on subsequent tests of equivalence than those in the latter group, with baseline discriminations repeatedly superseded by their subsequent verbal training.

Bentall, Dickins, and Fox (1993, Experiment 1) presented three groups of undergraduate participants with different types of visual stimuli designated as nameable preassociated pictograms, nameable nonassociated pictograms, and hard-to-name abstract stimuli. As predicted, equivalence was demonstrated most quickly and with fewest errors by participants exposed to the preassociated stimuli and most slowly by those in the abstract condition, although postexperimental interviews failed to reveal the consistent use of class names. A methodological refinement (Experiment 2), employing only preassociated and abstract stimuli, supported the previous findings. Although both groups of participants reported naming nearly all the stimuli to which they were exposed, postexperimental interviews again failed to reveal consistent common naming of stimuli. A final experiment investigated the effects of pretraining different stimulus naming strategies on class formation. Prior to match-to-sample training, one group of participants was trained to name stimuli individually, and another group was taught to use class names. Although the latter group experienced more difficulty in learning stimulus names, criterion match-to-sample training took longer for the first group. Smith, Dickins, and Bentall (1996) further reported that the class-discordant associations trained orally between stimulus names prior to testing for emergent relations largely superseded previous match-to-sample baseline training, suggesting that participants' names for individual stimuli may play a role in the formation of equivalence classes but not in their maintenance.

The present research was conducted to elucidate the role of verbal behavior in equivalence class formation by presenting different arrangements of the same easily nameable, yet formally unrelated, visual stimuli to verbally able adult participants. As Remington (1996) has observed, if the properties of stimulus names, rather than the stimuli named, determine the ease with which equivalence classes form, it should be possible to identify groupings of stimuli that will be more or less easily related on the basis of their names. Research has suggested that a highly salient feature of words is their phonological characteristics, especially when those characteristics promote rhyme with other words (Goswami & Bryant, 1990). If participants name visual stimuli, then classes composed of stimuli whose names rhyme might be expected to become equivalent more readily than classes of stimuli whose names share no such similarity.

Three experimental conditions were compared; one in which the names of stimuli forming classes rhymed and two control conditions composed of different combinations of the same stimuli whose names were phonologically unrelated. Because the paradigm constituted a typical arbitrary visual–visual paradigm and participants in all conditions were exposed to the same performance-contingent training program, it was predicted that equivalence would be demonstrated by participants in all conditions. If the hypothesis was correct, however, equivalence would be demonstrated more quickly and with fewer

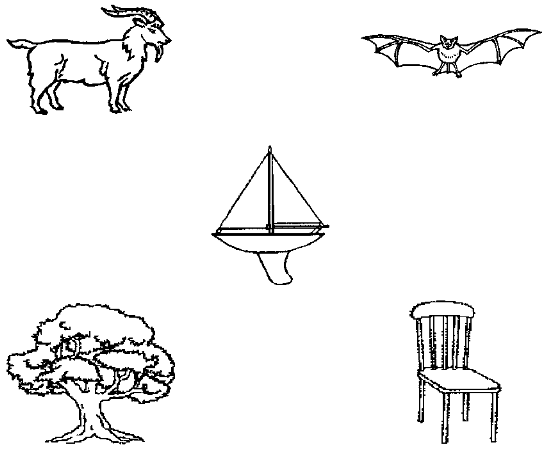

Fig. 1. Typical match-to-sample screen display, illustrating an A1B1 trial for the rhyme condition; *boat* (sample), *goat* (correct comparison).

errors by participants in the former condition. As a measure of the normativeness of stimulus names used, written postexperimental tests of naming were conducted. Effects of nodal distance (Fields, Adams, & Verhave, 1993; Fields, Adams, Verhave, & Newman, 1990; Fields & Verhave, 1987) and the possible generalization of rhyme-based equivalence classes to novel stimuli whose names rhymed were also assessed.

METHOD

Participants

Thirty students and staff at the University of Southampton (18 female, 12 male) volunteered to participate in the study and were assigned randomly, but in equal number, to three experimental conditions (rhyme, orthogonal, and diagonal). Aged between 18 and 40 years, all were native English speakers with no prior knowledge of the research. Participation was voluntary but paid at a rate of £2.50 (approximately $4.00) per 30 min, independent of experimental performance. Data from 1 participant, whose first language was Italian, are excluded from formal analyses and are considered separately.

Apparatus and Setting

Using software designed specifically for equivalence research (Dube & Hiris, 1996), a Power Macintosh® computer presented all stimuli and automatically recorded participants' responses and response latencies. During match-to-sample trials, its 15-in. (38-cm) monitor displayed five white "keys" (4.5 cm square) that were indiscernible against a white background. Sample stimuli were presented on the center key, and comparisons appeared on the four outer keys (see Figure 1). During generalization testing, one of the outer keys, its position varying from trial to trial, always remained blank. Participants were tested individually in a small windowless cubicle (1.5 m by 2.9 m) containing a desk on which were placed a sheet of written instructions, the computer, monitor, and mouse, and an envelope concealing a pen and posttest booklet for completion subsequent to match-to-sample testing. No keyboard was visible, and responses were made using the mouse. All participants completed the experiment in one sitting, which never exceeded 1.5 hr duration.

Table 1

Normative names of pictorial stimuli used in baseline training, emergent testing, and generalization testing.

Baseline training and emergent testing				Generalization testing		
boat	goat	note	coat	can	man	fan
flea	tree	bee	key	dog	frog	log
rat	bat	cat	hat	snake	cake	rake
bear	chair	hair	pear			

Stimuli and Class Arrangements

Stimuli were 25 black-and-white pictures of easily nameable items (some adapted from Snodgrass & Vanderwart, 1980), the normative names of which were each between three and five letters in length (see Table 1). Sixteen of these stimuli provided the potential for four four-member equivalence classes in each condition; the other nine, presented in the final testing phase only, provided the potential for three three-member generalized classes. Although all participants were exposed to the same stimuli throughout the experiment, the arrangements of stimuli composing potential equivalence classes differed among the three conditions. In the rhyme condition, classes were composed of stimuli whose names rhymed with each other, and trials always presented a sample whose name rhymed with that of the correct comparison but never with those of the incorrect com-

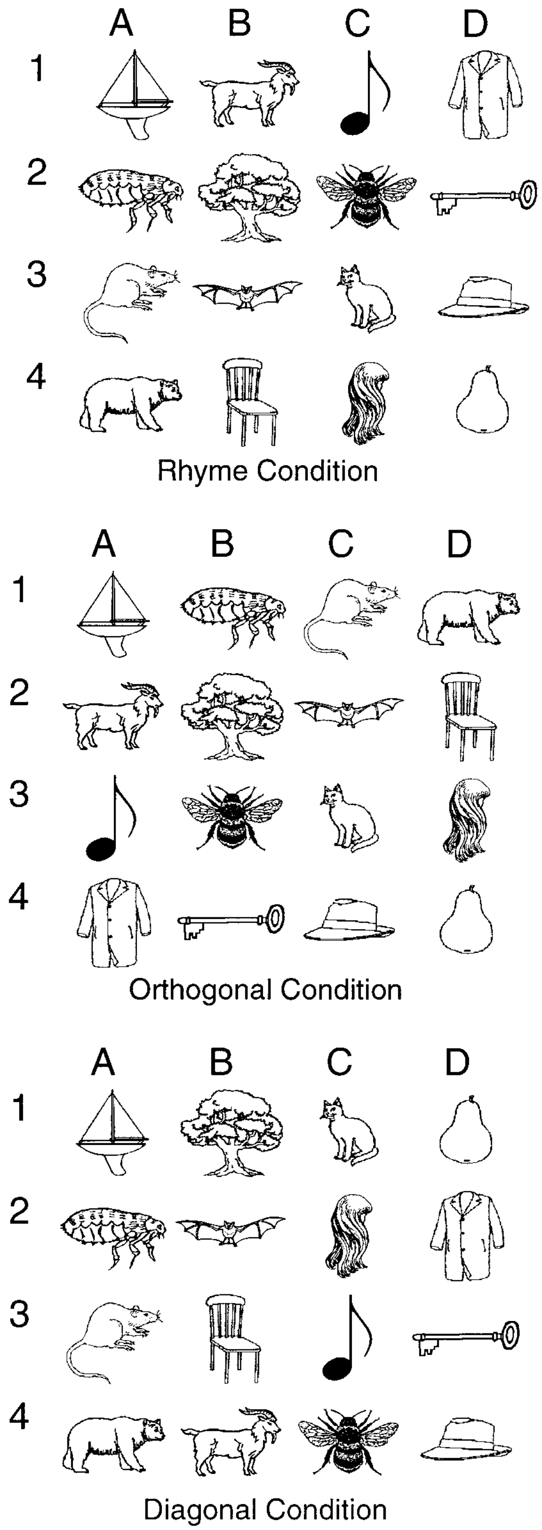

Fig. 2. Stimuli and class configurations used in baseline and emergent trials for all conditions. Numbered rows denote classes; lettered columns denote stimuli.

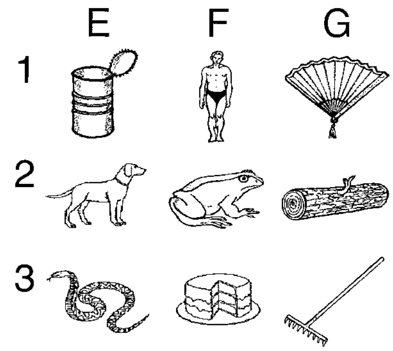

Fig. 3. Stimuli and class configurations used in generalization testing for all conditions. Numbered rows denote classes; lettered columns denote stimuli.

parisons. For two control conditions, classes consisted of stimuli whose names did not rhyme: In the orthogonal condition, class rearrangement was such that all available comparisons' names always rhymed with each other but never rhymed with that of the sample. In the diagonal condition, one of the incorrect comparisons' names always rhymed with the sample's name but never rhymed with those of either the correct or the other two incorrect comparisons, whose names also did not rhyme with each other (see Figure 2 for stimuli and class arrangements for all conditions). The other nine stimuli (presented in identical class configurations for all conditions) permitted use of only three comparisons per trial, the name of one of which always rhymed with the sample's name but never with those of the other two comparisons (see Figure 3).

Procedure

Participants were asked to familiarize themselves with the instructions before them, and were then left to complete the experiment.

Instructions. Initially, the following text was displayed on the computer's monitor: "When you are familiar with the written instructions, please click on 'Continue' to start the experiment." The written instructions were as follows:

> When the experiment begins, and at the start of each subsequent trial, you will see a picture in the middle of the screen in front of you. Use the mouse to click on it. More pictures will now appear in the corners of the screen. Use the mouse to click on one of these. At first, you will receive feedback on your choic-

es, a "beep" for correct and a "buzz" for incorrect. During later stages of the experiment, you will no longer receive feedback on your choices—the computer will tell you when. Keep on going however, and continue to do the best you can! Please aim to complete the experiment as quickly and accurately as possible. The computer will record your performance throughout, and a message on screen will tell you when the experiment is over. When you are ready to start, please click on "Continue." Thank you for participating in this experiment. You are free to leave at any point.

The specified action removed the on-screen instructions and match-to-sample training commenced.

General procedure and match-to-sample contingencies. Each trial began with presentation of a sample stimulus, an observing response with the mouse causing the comparison stimuli to be displayed. All stimuli remained in view until selection of a comparison caused them immediately to disappear, followed, after a 1-s interval, by presentation of the next trial. Comparison selections made within 0.5 s of presentation had no such consequence, however, and all stimuli remained in view. There was no limit to trial duration. Positions of correct and incorrect comparisons varied pseudorandomly from trial to trial, and, throughout training and testing, comparisons were always the members of all other stimulus classes sharing the same alphabetic designation (e.g., B1, B2, B3, B4). At no point did the location of the correct comparison remain constant for more than two consecutive trials, nor did the same sample stimulus appear for more than two trials consecutively.

All participants were exposed to the same performance-contingent training and testing program. In reinforced training, selection of class-consistent comparison stimuli was followed by a beep and the word "CORRECT" displayed on the screen. Other choices resulted in a buzz and a darkened screen. During testing without reinforcement, however, the only consequence of a response was the presentation of the next trial. To assess effects of nodal distance, training was sequential (i.e., AB-BC-CD), allowing the potential emergence of 12 symmetric (BA, CB, DC), eight one-node transitive (AC, BD), eight one-node equivalence (CA, DB), four two-node transitive (AD), and four two-node equivalence relations (DA). The overall procedure for all conditions was designed as a series of successive training and testing blocks, the details of which are given below.

Phase 1: Establish AB, BC, and CD baseline relations. Initially, AB relations were trained, with each of the four relations (i.e., A1B1, A2B2, etc.) presented in pseudorandom order once every four trials. When a criterion of 12 consecutive correct responses had been achieved, BC relations were trained in identical fashion. When the same criterion had been attained for these relations, CD relations were established, again to the same criterion.

Phase 2: Review baseline relations, with feedback. Consequent to fulfillment of the above criteria, all baseline relations were reviewed in 12-trial blocks, with all AB, BC, and CD trials intermixed in pseudorandom order. Samples from the same class were never presented consecutively. On completion of one reinforced trial block with 100% accuracy, the next review phase commenced, assessing baseline maintenance in extinction.

Phase 3: Review baseline maintenance without feedback. Otherwise identical to Phase 2, all trials during this phase were completed in the absence of explicit reinforcement. If performance remained at 100% accuracy over the first 12-trial block, emergent testing commenced (Phase 4). If criterion was not achieved by the end of the second block, however, baseline relations were reviewed, again with feedback (Phase 2). Review of baseline and its maintenance in extinction continued in this way until 100% accuracy was demonstrated over one block of test trials (i.e., trials in which no reinforcement was delivered). Table 2 shows the conditional discriminations established in Phases 1 to 3.

Phase 4: Emergent testing. All possible emergent relations except reflexivity were presented, in pseudorandom order, in a maximum of four 36-trial blocks (see Table 3). Generalized identity-matching repertoires were assumed (Bush, Sidman, & de Rose, 1989). If participants satisfied the criterion of a minimum of 35 of 36 class-consistent responses in any one emergent testing block, generalization testing commenced (Phase 5). If, however, criterion had not been achieved at the end of two consecutive blocks of emergent

Table 2

Baseline trial configurations, using single-sample and four-comparison displays.

Trained relations								
AB			BC			CD		
Sa	Co+	Co−	Sa	Co+	Co−	Sa	Co+	Co−
A1	B1	B2, B3, B4	B1	C1	C2, C3, C4	C1	D1	D2, D3, D4
A2	B2	B1, B3, B4	B2	C2	C1, C3, C4	C2	D2	D1, D3, D4
A3	B3	B1, B2, B4	B3	C3	C1, C2, C4	C3	D3	D1, D2, D4
A4	B4	B1, B2, B3	B4	C4	C1, C2, C3	C4	D4	D1, D2, D3

Note. Sa: sample stimulus; Co+: positive comparison stimulus; Co−: negative comparison stimuli.

testing, baseline relations were again reviewed, first without feedback (Phase 3) and, if 100% accuracy had not been achieved at the end of two trial blocks (24 trials), again with feedback (Phase 2). Baseline review continued in this way until all relations were again demonstrated with 100% accuracy over one block of test trials. The two final blocks of emergent testing were then presented. Generalization testing followed completion of the fourth emergent block, regardless of performance.

Phase 5: Generalization testing. Two consecutive blocks of 18 test trials each were presented, involving previously unseen stimuli. Each trial presented a novel sample stimulus followed, after an observing response, by three novel comparisons (see Table 4). If, at the end of the first block, participants had selected only comparisons whose names rhymed, match-to-sample testing ended. An on-screen message automatically informed participants of this and asked them to complete the posttest in the envelope before them. Following the second block of generalization trials, the same message was displayed, regardless of performance.

Phase 6: Naming posttest. Subsequent to match-to-sample testing, participants completed a written posttest that was designed to indicate their naming responses during the experiment. The booklet was headed by the following instructions:

> Printed below are the pictures that you have seen during the experiment. Did you mentally name any of them, or refer to them in any way during testing? If you did, please write under each picture the name, or names, you used for

Table 3

Emergent trial configurations, using single-sample and four-comparison displays.

Emergent relations												
	Sa	Co+	Co−	Sa	Co+	Co−	Sa	Co+	Co−	Sa	Co+	Co−
Symmetry												
BA	B1	A1	A2, A3, A4	B2	A2	A1, A3, A4	B3	A3	A1, A2, A4	B4	A4	A1, A2, A3
CB	C1	B1	B2, B3, B4	C2	B2	B1, B3, B4	C3	B3	B1, B2, B4	C4	B4	B1, B2, B3
DC	D1	C1	C2, C3, C4	D2	C2	C1, C3, C4	D3	C3	C1, C2, C4	D4	C4	C1, C2, C3
One-node transitivity												
AC	A1	C1	C2, C3, C4	A2	C2	C1, C3, C4	A3	C3	C1, C2, C4	A4	C4	C1, C2, C3
BD	B1	D1	D2, D3, D4	B2	D2	D1, D3, D4	B3	D3	D1, D2, D4	B4	D4	D1, D2, D3
One-node equivalence												
CA	C1	A1	A2, A3, A4	C2	A2	A1, A3, A4	C3	A3	A1, A2, A4	C4	A4	A1, A2, A3
DB	D1	B1	B2, B3, B4	D2	B2	B1, B3, B4	D3	B3	B1, B2, B4	D4	B4	B1, B2, B3
Two-node transitivity												
AD	A1	D1	D2, D3, D4	A2	D2	D1, D3, D4	A3	D3	D1, D2, D4	A4	D4	D1, D2, D3
Two-node equivalence												
DA	D1	A1	A2, A3, A4	D2	A2	A1, A3, A4	D3	A3	A1, A2, A4	D4	A4	A1, A2, A3

Note. Sa: sample stimulus; Co+: positive comparison stimulus; Co−: negative comparison stimuli.

Table 4

Generalization trial configurations, using single-sample and three-comparison displays.

	Generalized emergent relations								
	Sa	Co+	Co−	Sa	Co+	Co−	Sa	Co+	Co−
EF	E1	F1	F2, F3	E2	F2	F1, F3	E3	F3	F1, F2
FE	F1	E1	E2, E3	F2	E2	E1, E3	F3	E3	E1, E2
FG	F1	G1	G2, G3	F2	G2	G1, G3	F3	G3	G1, G2
GF	G1	F1	F2, F3	G2	F2	F1, F3	G3	F3	F1, F2
EG	E1	G1	G2, G3	E2	G2	G1, G3	E3	G3	G1, G2
GE	G1	E1	E2, E3	G2	E2	E1, E3	G3	E3	E1, E2

Note. Sa: sample stimulus; Co+: positive comparison stimulus; Co−: negative comparison stimuli.

it during the experiment. If you did not refer to a picture in any such way, please leave the space underneath it blank.

All experimental stimuli were presented, each followed by a blank space and dotted line.

RESULTS

All participants completed the experiment, although the accuracy with which they did so strongly differentiated participants in the rhyme condition from those in the two control conditions. Individual participants' trials and errors during all phases of match-to-sample training and testing are presented in Appendix A, and their latency data appear in Appendix B.

Phase 1: Establish Baseline Relations

Acquisition of baseline relations was easiest for participants in the rhyme condition and was most difficult for those in the orthogonal condition. The mean number of trials required by rhyme participants to meet all three criteria for this phase was 53.6 (SD = 24.7) with a mean error score of 9.9 (SD = 12.3), whereas the mean number of trials required by participants in the orthogonal condition was 115.7 (SD = 25.1), with a mean of 42.6 errors (SD = 18.6). diagonal participants required a mean of 98 trials (SD = 36), with a mean of 31.2 errors (SD = 16.7). With one exception (Participant SG), all participants in the rhyme condition required fewer trials to establish baseline relations than any participant in either of the two control conditions. Participant SG aside, the greatest number of trials required by a participant in the rhyme condition was 57 (nine errors). By contrast, the smallest number of trials required by any participant in the orthogonal condition was 87 (29 errors), and was 67 (21 errors) in the diagonal condition: The smallest number of trials required by a participant in the rhyme condition was 36 (0 errors)—the minimum to meet the criteria. Apart from Participant SG, the maximum number of errors made by anyone in the rhyme condition was 14 and, although the most accurate diagonal participant made only 12 errors, no one else in that condition made fewer than 20 errors. In the orthogonal condition, the minimum error score was 25. Figure 4 shows the mean number of trials required and errors made by participants in all conditions during this phase of training.

The trial and error data were subjected to mixed design analyses of variance, in both of which the between-subject factor was condition (rhyme, orthogonal, diagonal) and the within-subject factor was training block (AB, BC, CD). Regarding the number of trials required to meet the initial criteria, there was a significant main effect of condition, $F(2, 27) = 12.11$, $p < .0001$, and a significant interaction of Condition × Training Block, $F(4, 54) = 4.96$, $p < .01$. Regarding errors, there was a significant main effect of condition, $F(2, 27) = 10.67$, $p < .0001$, and a significant interaction of Condition × Training Block, $F(4, 54) = 3.73$, $p < .01$. Scheffé tests (at an alpha level of .05) indicated that the number of trials required and errors made during AB training did not differ significantly between conditions. During BC training, however, the number of trials and errors in the rhyme condition was significantly smaller than in the control conditions, which did not differ significantly from each other. Again, during CD

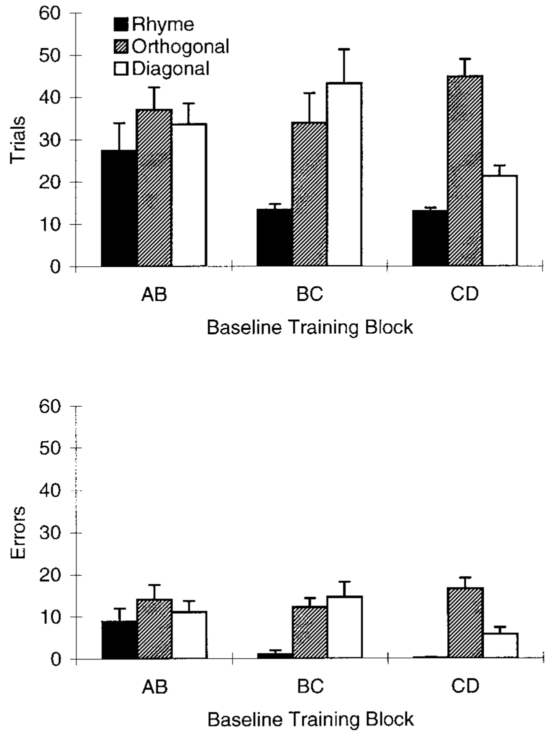

Fig. 4. Mean trials and errors (+*SE*) in all conditions to establish component baseline relations (Phase 1).

training, significantly fewer trials were required in the rhyme condition than in the orthogonal condition, and significantly fewer errors were made in the former condition than in the orthogonal and diagonal conditions (the latter at an alpha level of $p = .084$). There were also significantly fewer trials required and errors made in the diagonal condition than in the orthogonal condition.

Phases 2, 3, and 4: Baseline Review and Emergent Testing

Review prior to initial emergent testing. All 10 participants in the rhyme condition achieved complete accuracy during their first block of reinforced review, but only 3 participants in the orthogonal condition and 1 in the diagonal condition performed likewise. The top half of Figure 5 shows the total number of trials and errors for participants in each condition during this review phase.

Subsequent to meeting criterion for reinforced review, all participants but 2 in each condition performed without error during their first block of extinction review and proceeded to emergent testing. Of the remaining participants, both of those in the rhyme condition and 1 in the orthogonal condition performed errorlessly during their second block of extinction review. The other orthogonal participant required one additional block of reinforced review before meeting criterion in extinction. In the diagonal condition, 1 participant required two additional reinforced blocks; the other required an additional 27 blocks.[1] The lower half of Figure 5 shows the total number of trials and errors for all participants during extinction review.

Initial emergent testing (Blocks 1 and 2). Although all participants showed errorless baseline maintenance in extinction immediately prior to emergent testing, equivalence was confined almost exclusively to participants in the rhyme condition.

Within the first block of emergent testing, all rhyme participants but 1 fulfilled the 35 of 36 criterion for equivalence; 7 performed errorlessly. The remaining participant (SG) performed without error during the second testing block. By contrast, only 3 participants from the control conditions achieved criterion during initial testing, one from each condition in the first block and another from the diagonal condition in the second block. From the 18th trial of the first block and throughout the second block, 1 participant (RC) in the diagonal condition selected only comparisons whose names rhymed with those of the samples. The left sections of Figure 6 show means of percentage error scores made by participants in all conditions during these two blocks. Percentages have been presented because the numbers of emergent relations in each block were unequal.

These data were subjected to a mixed-design analysis of variance in which the between-subject factor was condition (rhyme, orthogonal, diagonal) and the within-subject factor was relation type (symmetry, one-node transitivity, one-node equivalence, two-node

[1] Of these blocks, three were completed without error. Nine blocks contained a single incorrect response to Sample A3, and another four blocks contained a single incorrect response to Sample B4. A further six blocks contained incorrect responses to both of these samples. None of the remaining five blocks contained more than four errors, and only two of these blocks did not contain an incorrect response to Samples A3, B4, or both.

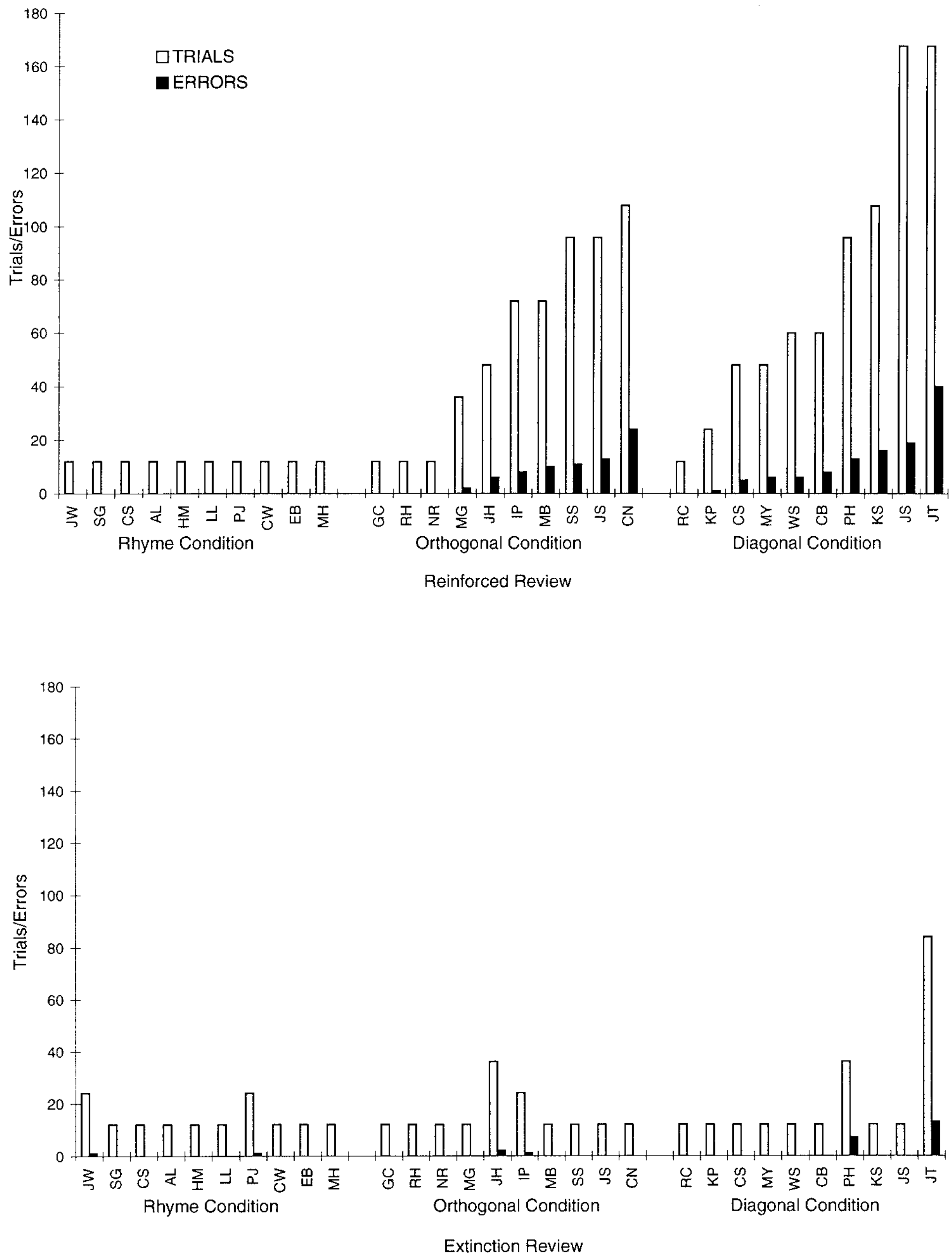

Fig. 5. All participants' trials and errors in reinforced and extinction baseline reviews (Phases 2 and 3) prior to initial emergent testing.

transitivity, two-node equivalence). Regarding the number of errors made during the first block of emergent testing, there was a significant main effect of condition, $F(2, 27) = 14.86$, $p < .0001$, of relation type, $F(4, 108) = 14.41$, $p < .0001$, and a significant interaction of Condition × Relation Type $F(8, 108) = 4.54$, $p < .0001$. Scheffé tests (at an alpha level of .05) further indicated that although significantly fewer errors were made in the rhyme condition than in either of the control conditions, those conditions did not differ significantly from each other. Post hoc means comparisons (at an alpha level of .05)

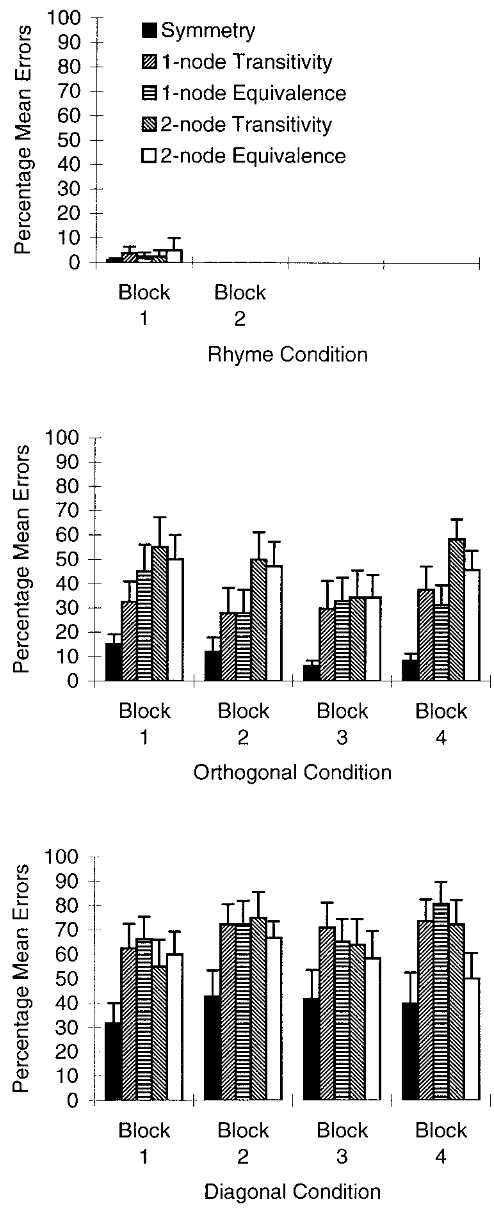

Fig. 6. Means of percentage error scores (+*SE*) of all participants in the rhyme, orthogonal, and diagonal conditions during testing of emergent relations (Phase 4).

indicated that participants in both control conditions made significantly fewer errors on symmetry trials than on any other trial type, that participants in the orthogonal condition made significantly fewer errors on trials of one-node transitivity than on any other trial type, and that participants in the diagonal condition made significantly fewer errors on trials of two-node transitivity than on trials of two-node equivalence. No significant differences were observed, however, in the data of the rhyme condition.

Response latencies also differentiated the rhyme condition from the control conditions during emergent testing. Mixed-design analysis of variance of latencies during the first block of testing—the only block to which all participants were exposed—showed significant main effects of condition, $F(2, 27) = 5.24$, $p < .05$, and of relation type, $F(4, 108) = 8.48$, $p < .0001$. An interaction of Condition × Relation Type was also observed, $F(8, 108) = 1.86$, $p < .07$.[2] Post hoc means comparisons (at an alpha level of .05) further indicated that latencies in the rhyme condition were significantly shorter than in the control conditions, and that latencies in the latter conditions did not differ significantly from each other. Post hoc means comparisons also indicated that overall, mean latencies were significantly longer on two-node equivalence trials than on any other relation type. Mean latencies for symmetry trials were also significantly shorter than on two-node transitivity trials. Figure 7 shows mean response latencies in each condition during Block 1 for each type of emergent relation tested.

Review prior to final emergent testing. Because all participants in the rhyme condition had shown equivalence during initial testing, none received further baseline review. Of the 8 orthogonal and 9 diagonal participants who had failed to meet criterion for equivalence, 5 from each condition demonstrated continued baseline maintenance, performing errorlessly in extinction within two consecutive review blocks. Two of the 3 remaining orthogonal participants required two additional reinforcement blocks each before meeting criterion in extinction; the other participant required a total of six blocks. Two of the 4 diagonal participants received a sin-

[2] These differences were confirmed by latency data averaged across all testing blocks to which participants were exposed, analysis of variance again indicating significant main effects of condition, $F(2, 27) = 6.54$, $p < .01$, and relation type, $F(4, 108) = 15.1$, $p < .0001$, and a significant interaction of Condition × Relation Type, $F(8, 108) = 2.93$, $p < .01$. Because of the small proportion of class-consistent responses outside the rhyme condition, analyses were performed on latencies for both correct and incorrect responses throughout.

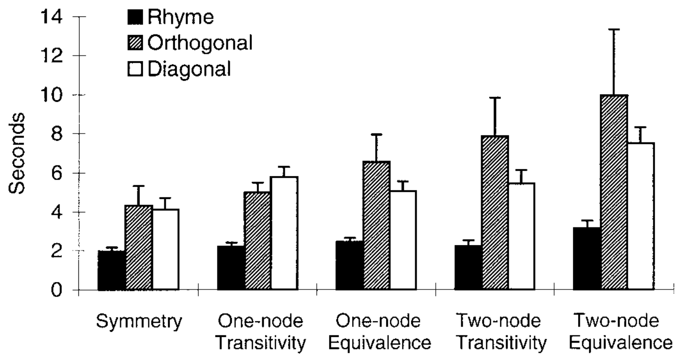

Fig. 7. Mean response latencies (+*SE*) in all conditions during the first block of emergent testing (Phase 4).

gle additional review block with reinforcement; the remaining 2 received a total of four blocks each.

Final emergent testing (Blocks 3 and 4). Of the 20 participants in the control conditions, 17 (8 orthogonal and 9 diagonal) received further emergent testing. Of these, only 2, from the orthogonal condition, showed equivalence—both in their third testing block. The remaining 15 participants who never showed equivalence (including Participant RC, who again selected only rhyming comparisons throughout both final blocks) had each received a total of 144 emergent testing trials and had made an average of 71.8 errors (SD = 9).

The means of percentage error scores of these participants during the final block of emergent testing were subjected to a mixed-design analysis of variance in which the between-subject factor was condition and the within-subject factor was trial type. This indicated a significant main effect of condition, $F(1, 13) = 5.02$, $p < .04$, the remaining participants in the orthogonal condition having made fewer errors than those in the diagonal condition. A significant main effect of relation type was also observed, $F(4, 52) = 7.77$, $p < .0001$, as well as a significant interaction of Condition × Relation Type, $F(4, 52) = 2.58$, $p < .05$. Post hoc means comparisons (at an alpha level of .05) indicated that participants in the orthogonal condition made significantly fewer errors on symmetry trials than on any other trial type, and that participants in the diagonal condition made significantly fewer errors on symmetry trials than on any other trial type except two-node equivalence. Participants in the orthogonal condition also made significantly more errors on trials of two-node transitivity than on trials of one-node transitivity or one-node equivalence, whereas participants in the diagonal condition made significantly fewer errors on trials of two-node equivalence than on any other trial type except symmetry. The right sections of Figure 6 show means of percentage errors made in the orthogonal and diagonal conditions during this phase of testing.

Phase 5: Generalization Testing

Figure 8 shows mean errors made by participants in all conditions during generalization testing, with errors defined as selection of a comparison stimulus whose name did not rhyme with that of the sample. All but 2 participants in the rhyme condition selected only rhyming comparisons in the first block of testing, with the remaining 2 doing so in their second block (1 of these being Participant SG). Only 1 participant from each of the control conditions made no errors in the first block, although 6 participants in the orthogonal condition and 3 in the diagonal condition selected only rhyming comparisons during the second block.

Phase 6: Naming Posttests

All participants' posttests suggested a high degree of normative stimulus naming, although some stimuli were so named less consistently than others (e.g., *louse, bug, gnat,* or *mite* for *flea, yacht* for *boat, mouse* for *rat,* and *tin* for *can*). No rhyme participant indicated using a nonnormative stimulus name throughout the experiment, although 6 indicated that they had changed the names

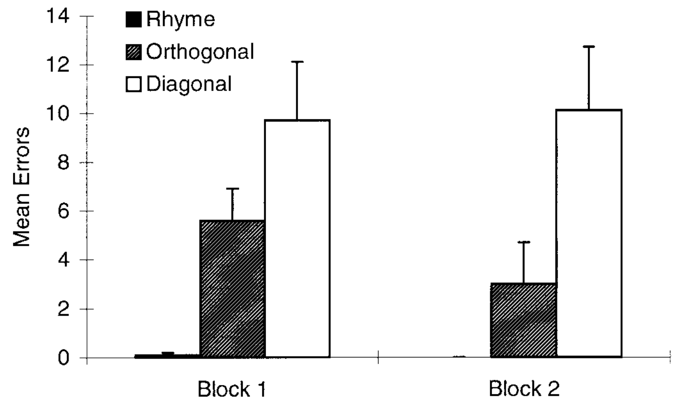

Fig. 8. Mean errors (+*SE*) in all conditions during generalization testing (Phase 5).

used to the normative ones during the course of the experiment.

One participant in the orthogonal condition named every stimulus normatively, as did 3 participants in the diagonal condition. Another orthogonal participant reported changing from a nonnormative to a normative stimulus name on the basis of rhyme (*oak* to *tree*), and 4 participants from the control conditions noted using intraverbal phrases (e.g., *there's a bee in your hair, the cat eats the pear, cat stuck in a tree,* etc.) during training and testing.

Effects of Language Experience

The data from 1 participant (PM) who was originally assigned to the rhyme condition were rejected from analysis because his first language was Italian. Posttesting revealed his use of a variety of idiosyncratic stimulus names, in both Italian and English (e.g., *sorcio* [mouse] for *rat, topa* [beautiful woman] in response to *hair, cap* for *hat,* etc.). This was the only participant exposed to rhyme condition training who never met the criterion for equivalence, making a total of 34 errors over the first two emergent blocks and 39 during the latter two. This participant also made a total of 27 errors during generalization testing, and reported 11 nonnormative naming responses on the posttest.

Although her stated first language was English and her data are included in the above analyses, it is perhaps worthy of note that Participant SG—whose performance was also exceptional in the rhyme condition—was a fluent speaker of both English and Hindi.

DISCUSSION

The results suggest that the participants named the experimental stimuli without instruction, and that the phonological properties of the names given (specifically their rhyming characteristics) influenced match-to-sample performance. When the names of stimuli composing classes rhymed with each other, baseline learning and maintenance, equivalence, and generalized class formation occurred more quickly and reliably than when those names did not rhyme. Less variability was also observed in the data of participants who were trained with rhyming combinations of stimuli. Because the facilitative effects of rhyme necessarily depended on the naming of stimuli, albeit covertly, the results observed could not have occurred in the absence of naming. It can therefore be inferred that naming—a result of participants' individual verbal histories (Horne & Lowe, 1996, 1997)—was functional during the experiment.

Baseline establishment (Phase 1) consisted of three consecutively presented training blocks (AB, BC, and CD), each composed of newly introduced stimuli, yet all participants but 3 in the rhyme condition performed without error throughout BC and CD training. Only 1 participant in the control conditions performed without error during either of these blocks (diagonal Participant CS during CD training). Because the baseline stimuli presented in all conditions bore no consistent formal resemblances to each other, it seems plausible that rhyme participants' consistently class-congruent selection of previously unseen BC and CD comparisons was verbally controlled (Horne & Lowe, 1996) or rule governed (Skinner, 1969), in that during AB training they had learned that selection of any comparison whose name rhymed with that of its sample would be correct. Participants in the control conditions, learning a series of purely arbitrary discriminations, had no such straightforward verbal basis for selection available. The virtually errorless maintenance of baseline demonstrated by participants in the rhyme condition during review (Phases 2 and 3) did not undermine this interpretation.

It seems equally plausible that the rapid and accurate demonstration of equivalence by participants in the rhyme condition (Phase 4) was a product of the same verbal control: By the end of the second block of 36 trials, all 10 participants in the rhyme condition had met criterion for equivalence, whereas only 3 of the 20 participants in the control conditions had performed likewise. The data from 1 participant in the diagonal condition further indicated the functionality of verbal behavior during the experiment: Participant RC selected only rhyming comparisons throughout his last 137 emergent trials, despite having received no previous reinforcement for selecting such comparisons, and despite having mastered the diagonal condition's baseline. This performance (a flawless demonstration of the emergent relations of equivalence had he been in the rhyme condition) indicates that baseline training may be superseded by verbal control during testing without reinforcement, if a ready verbal basis for the categorization of stimuli is available.

During the first block of emergent testing, no significant effects of nodal distance were observed in the accuracy with which participants in the rhyme condition responded. In accordance with previous research (Kennedy, 1991; Kennedy, Itkonen, & Lindquist, 1994), however, both control conditions produced significantly greater accuracy on symmetry trials than on those of any other relation type. Also in accordance with previous findings (Fields et al., 1990), participants in the orthogonal condition showed significantly greater accuracy on trials of one-node transitivity than on any other trial type except symmetry. That significantly longer latencies were observed in the responses of all conditions on trials of two-node equivalence than on any other trial type also supported previous reports of the transitivity latency effect (Dickins et al., 1993) in verbally able humans. Although the error scores of the remaining participants in both control conditions differed significantly during the final block of emergent testing, the pattern of errors evident in those scores was similar to those reported for Block 1, suggesting that more extensive exposure to the experimental contingencies would have been required for effects of nodal distance to be minimized (Fields et al., 1993) and for full equivalence to emerge (Lazar et al., 1984; Sidman et al., 1986; Spradlin, Cotter, & Baxley, 1973).

During generalization testing (Phase 5), all participants in the rhyme condition met the errorless criterion for selection of comparisons whose names rhymed with their sample's; all but 2 did so within the first block of testing. More surprisingly, 7 participants in the orthogonal condition and 4 in the diagonal condition also met this criterion, although only 1 participant from each condition did so during the first testing block. Not only did the stimuli involved bear no formal resemblance to each other or to the stimuli used in baseline training and emergent testing, but their names were also phonologically unrelated to the names of the baseline stimuli.

Such a finding is perhaps relevant to discussion of the functional relationships among equivalence classes, generalized classes, and natural categories (Fields, Adams, Buffington, Yang, & Verhave, 1996; Fields, Reeve, Adams, Brown, & Verhave, 1997; Fields, Reeve, Adams, & Verhave, 1991; Haring, Breen, & Laitinen, 1989). Equivalence classes have been described as arbitrary classes, in that their constituents need not bear the formal resemblance to each other definitive of feature classes (Stromer & Mackay, 1996). Although physical similarity has been suggested as a basis for the merger of equivalence and feature classes (Adams et al., 1993), the majority of participants in the present experiment showed the consistent selection of novel comparisons indicative of class generalization, even though the stimuli involved shared no such similarities. Once again, the only basis for such consistency was the rhyme between sample and comparison names, and it would therefore again seem plausible that for participants in the rhyme condition, generalized class formation was the result of the verbal control engendered during baseline training. That some participants in the control conditions also showed consistent selection of rhyming comparisons during generalization testing would again indicate that when a ready verbal basis for categorization is available during testing without reinforcement (as it was for Participant RC earlier in the experiment), that basis may become functional.

Additional confirmation that stimuli were named was provided by written posttests, which also indicated that the naming of the experimental stimuli was usually normative. Although some participants in the control conditions reported using intraverbal phrases to link stimulus names (as previously reported by Horne & Lowe, 1996), no participants in the rhyme condition reported having used such strategies, although 4 indicated that they had changed to normative names during the experiment on the basis of rhyme. Anecdotal evidence from participants in the rhyme condition further suggested that they had simply selected ''pictures whose names rhymed'' early in the experiment.

It is perhaps interesting also to note a correlation between the language experience of 2 participants exposed to rhyme condition training and testing and their performance during the experiment. By far the most errors made by any participant in the rhyme condition included in the above analyses were by Participant SG who, as noted above, was a fluent speaker of both English and Hindi. Although the posttest suggested her use of mostly normative English names, it does not seem unreasonable to suggest that some of the names she had actually used during the experiment may have been Hindi (cf. Perone, 1988; Shimoff, 1984, 1986). Participant PM, whose first language was Italian, performed at a level akin to the low-accuracy participants in the control conditions. The mixture of normative and nonnormative naming responses in both English and Italian that he reported on the posttest strongly suggested that, for him, the task was not one of simply matching stimuli whose names rhymed.

As Goswami and Bryant (1990, p. 3) note, ''rhyme is an extremely important part of our everyday lives. Rhymes are to be found practically everywhere—in poems, in songs, in advertisements and in political slogans,'' and rhyme was chosen as a potential verbal basis for stimulus classification because of this saliency. That training and testing often lead to the contraction of intraverbally linked names employed by subjects (e.g., *green-cross* to *cross*) has been noted previously (Horne & Lowe, 1996), and might offer a further behavioral explanation of the effects observed in the present experiment. If, precurrently, participants in the rhyme condition repeated the names of the stimuli during the experiment, the salience of the phonetic similarity between those names might be increased, facilitating class formation. The same behavior in relation to stimuli with phonologically unrelated names would be unlikely to have such striking effects, as evidenced perhaps by the performances of participants in the control conditions. Similarly, from this perspective, exposure to pictures whose names rhyme might provide grounds for the common naming of the stimuli involved (e.g., *cat, hat, rat,* and *bat* all share the common phonetic element *-at*).

It might also be argued that, for participants in the rhyme condition, these common phonetic elements functioned as contextual stimuli, and that this second-order control could have facilitated the emergence of equivalence.[3] Whether such contextual control actually occurred, however, remains to be determined.

In summary, therefore, the findings of the present research suggest strongly that verbally able humans' performance on equivalence tasks can be influenced by their naming of stimuli and by the phonological properties of the names thus given. Although the present study is clearly not a demonstration of the necessity or even the sufficiency of naming for equivalence or generalized class formation, it nevertheless provides an unequivocal demonstration that the emergence of untrained relations can be substantially affected both by

[3] We are grateful to an anonymous reviewer for this suggestion.

participants' verbal histories and by their verbal behavior during experimentation. The finding that verbal behavior can be functional in equivalence is perhaps reminiscent of previous analyses of the determinants of human schedule-maintained performance (Horne & Lowe, 1993; Lowe, 1983; Rosenfarb, Newland, Brannon, & Howey, 1992).

Acceptance of the conclusion that naming can be functional in the emergence of stimulus equivalence carries a further implication: An unequivocal demonstration of nonhuman equivalence would not mean that all demonstrations of equivalence in humans could be interpreted without reference to their verbal abilities. The potential role of naming in previous demonstrations of equivalence with verbally able participants should not therefore be disregarded, and it might be suggested that future experimentation using this population would benefit from more careful analysis of the verbal behavior involved.

REFERENCES

Adams, B. J., Fields, L., & Verhave, T. (1993). Formation of generalized equivalence classes. *The Psychological Record, 43,* 553–566.

Annett, J. M., & Leslie, J. C. (1995). Stimulus equivalence classes involving olfactory stimuli. *The Psychological Record, 45,* 439–450.

Barnes, D., McCullagh, P. D., & Keenan, M. (1990). Equivalence class formation in non-hearing impaired and hearing impaired children. *The Analysis of Verbal Behavior, 8,* 19–30.

Bentall, R. P., Dickins, D. W., & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *The Quarterly Journal of Experimental Psychology, 46B,* 187–214.

Bush, K. M., Sidman, M., & de Rose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior, 51,* 29–45.

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.

Dickins, D. W., Bentall, R. P., & Smith, A. B. (1993). The role of individual stimulus names in the emergence of equivalence relations: The effects of interpolated paired-associates training of discordant associations between names. *The Psychological Record, 43,* 713–724.

Dube, W. V., Green, G., & Serna, R. W. (1993). Auditory successive conditional discrimination and auditory stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior, 59,* 103–114.

Dube, W. V., & Hiris, E. (1996). Match to sample program (version 10.18) [Computer software]. Waltham, MA: E. K. Shriver Center.

Dube, W. V., McIlvane, W. J., Mackay, H. A., & Stoddard, L. T. (1987). Stimulus class membership established via stimulus–reinforcer relations. *Journal of the Experimental Analysis of Behavior, 47,* 159–175.

Dube, W. V., McIlvane, W. J., Maguire, R. W., Mackay, H. A., & Stoddard, L. T. (1989). Stimulus class formation and stimulus–reinforcer relations. *Journal of the Experimental Analysis of Behavior, 51,* 65–76.

Dugdale, N., & 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). Hove, England: Erlbaum.

Eikeseth, S., & Smith, T. (1992). The development of functional and equivalence classes in high-functioning autistic children: The role of naming. *Journal of the Experimental Analysis of Behavior, 58,* 123–133.

Fields, L., Adams, B. J., Buffington, D. M., Yang, W., & Verhave, T. (1996). Response transfer between stimuli in generalized equivalence classes: A model for the establishment of natural kind and fuzzy superordinate categories. *The Psychological Record, 46,* 665–684.

Fields, L., Adams, B. J., & Verhave, T. (1993). The effects of equivalence class structure on test performances. *The Psychological Record, 43,* 697–712.

Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior, 53,* 345–358.

Fields, L., Reeve, K. F., Adams, B. J., Brown, J. L., & Verhave, T. (1997). Predicting the extension of equivalence classes from primary generalization gradients: The merger of equivalence classes and perceptual classes. *Journal of the Experimental Analysis of Behavior, 68,* 67–91.

Fields, L., Reeve, K. F., Adams, B. J., & Verhave, T. (1991). Stimulus generalization and equivalence classes: A model for natural categories. *Journal of the Experimental Analysis of Behavior, 55,* 305–312.

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48,* 317–322.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read.* Hove, England: Erlbaum.

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

Hall, G. A., & Chase, P. N. (1991). The relationship between stimulus equivalence and verbal behavior. *The Analysis of Verbal Behavior, 9,* 107–119.

Haring, T. G., Breen, C. G., & Laitinen, R. E. (1989). Stimulus class formation and concept learning: Establishment of within- and between-set generalization and transitive relationships via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior, 52,* 13–25.

Hayes, L. J., Thompson, S., & Hayes, S. C. (1989). Stimulus equivalence and rule following. *Journal of the Experimental Analysis of Behavior, 52,* 275–291.

Hayes, L. J., Tilley, K. J., & Hayes, S. C. (1988). Extending equivalence class membership to gustatory stimuli. *The Psychological Record, 38,* 473–482.

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

Horne, P. J., & Lowe, C. F. (1993). Determinants of human performance on concurrent schedules. *Journal of the Experimental Analysis of Behavior, 59,* 29–60.

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.

Kennedy, C. H. (1991). Equivalence class formation influenced by the number of nodes separating stimuli. *Behavioral Processes, 24,* 219–245.

Kennedy, C. H., Itkonen, T., & Lindquist, K. (1994). Nodality effects during equivalence class formation: An extension to sight-word reading and concept development. *Journal of Applied Behavior Analysis, 27,* 673–683.

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

Leader, G., & Barnes, D. (1996). Establishing equivalence relations using a respondent-type training procedure. *The Psychological Record, 46,* 685–706.

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

Lowe, C. F. (1983). Radical behaviorism and human psychology. In G. Davey (Ed.), *Animal models of human behavior: Conceptual, evolutionary, and neurobiological perspectives* (pp. 159–192). New York: Wiley.

Lowe, C. F., & Beasty, A. (1987). Language and the emergence of equivalence relations: A developmental study. *Bulletin of the British Psychological Society, 40,* A49.

Mandell, C., & Sheen, V. (1994). Equivalence class formation as a function of the pronounceability of the sample stimulus. *Behavioral Processes, 32,* 29–46.

Markham, M. R., & Dougher, M. J. (1993). Compound stimuli in emergent stimulus relations: Extending the scope of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 60,* 529–542.

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

O'Leary, C. A., & Bush, K. M. (1996). Stimulus equivalence in the tactile modality. *The Psychological Record, 46,* 509–517.

Perone, M. (1988). Laboratory lore and research practices in the experimental analysis of human behavior: Use and abuse of subjects' verbal reports. *The Behavior Analyst, 11,* 71–75.

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

Remington, B. (1996). The evolution of naming—just so! [commentary]. *Journal of the Experimental Analysis of Behavior, 65,* 243–245.

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

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

Rosenfarb, I. S., Newland, M. C., Brannon, S. E., & Howey, D. S. (1992). Effects of self-generated rules on the development of schedule-controlled behavior. *Journal of the Experimental Analysis of Behavior, 58,* 107–121.

Saunders, K. J., Saunders, R. R., Williams, D. C., & Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record, 43,* 725–744.

Saunders, K. J., & Spradlin, J. E. (1993). Conditional discrimination in mentally retarded subjects: Programming acquisition and learning set. *Journal of the Experimental Analysis of Behavior, 60,* 571–585.

Saunders, R. R., & Green, G. (1996). Naming is not (necessary for) stimulus equivalence [commentary]. *Journal of the Experimental Analysis of Behavior, 65,* 312–314.

Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior, 49,* 95–115.

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

Shimoff, E. (1984). Post-session questionnaires. *Experimental Analysis of Human Behavior Bulletin, 2,* 1.

Shimoff, E. (1986). Post-session verbal reports and the experimental analysis of behavior. *The Analysis of Verbal Behavior, 4,* 19–22.

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

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

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

Sidman, M. (1997). Equivalence: A theoretical or a descriptive model? *Mexican Journal of Behavior Analysis, 23,* 125–145.

Sidman, M., & Cresson, O. (1973). Reading and the cross-modal transfer of stimulus equivalences in severe retardation. *American Journal of Mental Deficiency, 77,* 515–523.

Sidman, M., Cresson, O., & Willson-Morris, M. (1974). Acquisition of matching-to-sample via mediated-transfer. *Journal of the Experimental Analysis of Behavior, 22,* 261–273.

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

Sidman, M., Willson-Morris, M., & Kirk, B. (1986). Matching-to-sample procedures and the development of equivalence relations: The role of naming. *Analysis and Intervention in Developmental Disabilities, 6,* 1–19.

Sigurdardottir, Z. G., Green, G., & Saunders, R. R. (1990). Equivalence classes generated by sequence training. *Journal of the Experimental Analysis of Behavior, 53,* 47–63.

Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis.* New York: Appleton-Century-Crofts.

Smith, A. B., Dickins, D. W., & Bentall, R. P. (1996). The role of interfering stimulus names in the emergence of equivalence relations II: The effects of interfering tasks prior to and after tests for equivalence relations. *The Psychological Record, 46,* 109–130.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Learning and Memory, 6,* 174–215.

Spradlin, J. E., Cotter, V. W., & Baxley, N. (1973). Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. *American Journal of Mental Deficiency, 77,* 556–566.

Stromer, R., & Mackay, H. A. (1996). Naming and the formation of stimulus classes. In T. R. Zentall & P. M. Smeets (Eds.), *Advances in psychology: 117. Stimulus class formation in humans and animals* (pp. 221–252). Amsterdam: Elsevier.

Stromer, R., & Osborne, J. G. (1982). Control of adolescents' arbitrary matching-to-sample by positive and negative stimulus relations. *Journal of the Experimental Analysis of Behavior, 37,* 329–348.

Stromer, R., & Stromer, J. B. (1990). The formation of arbitrary stimulus classes in matching to complex samples. *The Psychological Record, 40,* 51–66.

Tierney, K. J., De Largy, P., & Bracken, M. (1995). Formation of an equivalence class incorporating haptic stimuli. *The Psychological Record, 45,* 431–437.

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

Wulfert, E., Dougher, M. J., & Greenway, D. E. (1991). Protocol analysis of the correspondence of verbal behavior and equivalence class formation. *Journal of the Experimental Analysis of Behavior, 56,* 489–504.

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

Received July 15, 1998
Final acceptance January 7, 1999

APPENDIX A

Trials (Tr) required and errors (Er) made by all participants in the rhyme, orthogonal, and diagonal conditions during each stage of match-to-sample training and testing: Phase 1, baseline establishment (AB, BC, CD); Phase 2, reinforced baseline review (Br+); Phase 3, extinction baseline review (Br−); Phase 4, emergent testing (Emgt); Phase 5, generalization testing (Gn). Blank cells indicate that no training or testing occurred during that phase. Data rejected from analysis (Participant PM) are presented separately.

Condition	Partici-pant	AB		BC		CD		Br+		Br−		Emgt 1	
		Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
Rhyme	JW	32	8	13	1	12	0	12	0	24	1	36	0
	SG	83	33	25	9	12	0	12	0	12	0	36	7
	CS	15	2	12	0	12	0	12	0	12	0	36	0
	AL	13	1	12	0	12	0	12	0	12	0	36	0
	HM	31	13	12	0	12	0	12	0	12	0	36	1
	LL	18	5	12	0	12	0	12	0	12	0	36	0
	PJ	12	0	12	0	12	0	12	0	24	1	36	0
	CW	18	4	12	0	12	0	12	0	12	0	36	1
	EB	31	14	12	0	12	0	12	0	12	0	36	0
	MH	20	8	12	0	21	1	12	0	12	0	36	0
Orthogonal	JH	26	7	22	8	55	18	48	6	36	2	36	20
	MG	61	31	44	22	60	32	36	2	12	0	36	21
	MB	14	1	44	18	29	10	72	10	12	0	36	18
	CN	29	12	54	18	36	15	108	24	12	0	36	4
	SS	57	34	24	5	27	9	96	11	12	0	36	22
	JS	25	7	32	11	62	20	96	13	12	0	36	7
	IP	51	21	44	18	58	23	72	8	24	1	36	3
	GC	24	4	29	11	43	14	12	0	12	0	36	16
	RH	43	13	20	7	33	5	12	0	12	0	36	11
	NR	40	10	26	3	45	19	12	0	12	0	36	0
Diagonal	JS	27	9	22	8	27	10	168	19	12	0	36	17
	KS	48	19	96	37	14	2	108	16	12	0	36	17
	JT	61	27	79	30	21	4	348	40	84	13	36	9
	PH	53	17	30	7	40	17	96	13	36	7	36	21
	RC	33	13	19	7	15	1	12	0	12	0	36	29
	CS	24	4	45	8	12	0	48	5	12	0	36	23
	MY	35	10	25	5	20	5	48	6	12	0	36	31
	CB	16	2	33	11	21	9	60	8	12	0	36	24
	WS	20	2	51	23	18	3	60	6	12	0	36	5
	KP	19	7	32	9	24	6	24	1	12	0	36	1
	PM	37	15	27	13	32	10	12	0	12	0	36	16

APPENDIX A

(Extended)

Emgt 2		Br−		Br+		Emgt 3		Emgt 4		Gn 1		Gn 2	
Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
										18	0		
36	0									18	1	18	0
										18	0		
										18	0		
										18	0		
										18	0		
										18	0		
36	1									18	0		
										18	1	18	0
										18	0		
36	21	24	1			36	19	36	16	18	12	18	6
36	18	12	0			36	18	36	17	18	8	18	14
36	4	36	2	24	1	36	4	36	5	18	11	18	7
36	5	12	0			36	1			18	0		
36	13	36	2	24	1	36	6	36	5	18	2	18	0
36	3	60	7	72	6	36	4	36	4	18	3	18	0
36	0									18	1	18	0
36	18	24	1			36	17	36	16	18	7	18	0
36	6	24	1			36	0			18	4	18	0
										18	8	18	0
36	21	12	0			36	22	36	22	18	11	18	13
36	16	12	0			36	22	36	19	18	14	18	11
36	21	36	3	12	0	36	20	36	23	18	15	18	15
36	22	24	1			36	18	36	22	18	16	18	17
36	36	48	28	48	4	36	36	36	36	18	0		
36	24	36	2	12	0	36	13	36	19	18	2	18	0
36	32	12	0			36	32	36	32	18	18	18	18
36	23	36	5	48	3	36	21	36	25	18	18	18	17
36	6	12	0			36	3	36	3	18	1	18	0
										18	2	18	0
36	20	24	1			36	21	36	18	18	13	18	14

APPENDIX B

Mean response latencies (in seconds) of all participants in the rhyme, orthogonal, and diagonal conditions on trials of symmetry, one-node transitivity (1-node tr), one-node equivalence (1-node eq), two-node transitivity (2-node tr), and two-node equivalence (2-node eq) during the first block of emergent testing. Data rejected from analysis (Participant PM) are presented separately.

		Mean response latencies				
Condition	Participant	Symmetry	1-node tr	1-node eq	2-node tr	2-node eq
Rhyme	JW	2.08	2.76	2.31	2.86	3.43
	CS	1.22	1.64	2.26	1.52	1.87
	AL	1.53	1.57	1.76	1.7	1.81
	HM	2.45	3.34	3.54	2.57	3.85
	LL	1.55	1.4	1.4	1.13	2.24
	PJ	2.83	2.45	3.07	2.53	4.14
	EB	1.64	1.98	2.22	2.64	2.15
	MH	1.6	1.97	1.62	1.31	4.01
	SG	2.66	2.86	3.53	3.78	5.66
	CW	1.98	2.05	2.66	2.06	2.2
Orthogonal	NR	2.69	5.02	3.87	4.12	6.41
	IP	4.46	4.44	4.85	9.78	9.29
	CN	4.64	5.91	6.07	7.91	11.52
	RH	3.12	5.36	5.69	9.17	6.86
	JH	5.21	5.72	5.82	5.18	5.67
	MG	2.48	2.8	3.91	2.86	2.29
	MB	3.19	6.01	7.88	5.07	8.55
	SS	12.93	7.56	18.75	24.07	39.76
	JS	2.14	1.77	2.33	2.75	2.87
	GC	2.33	5.28	6.25	7.64	6.46
Diagonal	KP	5.36	4.67	5.8	7.35	9.26
	JS	3.66	8.7	8.06	8.32	10.31
	KS	3.46	6.8	6.61	3.21	8.49
	JT	2.98	5.52	2.8	2.48	6.83
	PH	4.93	6.09	5.86	5.74	6.36
	RC	3.51	4.77	4.6	5.77	4.2
	CS	3.21	5.82	3.81	4.54	6.26
	MY	8.38	7.12	5.29	5.06	11.86
	CB	2.33	3.4	2.46	2.79	3.77
	WS	3.3	4.99	5.29	9.12	7.94
	PM	2.73	4.48	4.7	5.14	7.36