JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2006, **86,** 337–354 NUMBER 3 (NOVEMBER)

EQUIVALENCE RELATIONS, CONTEXTUAL CONTROL, AND NAMING

TOM RANDELL AND BOB REMINGTON

UNIVERSITY OF SOUTHAMPTON

This paper reports two experiments that investigated the role of verbal behavior in the emergence and generalization of contextually controlled equivalence classes. During both experiments, participants were trained with two different combinations of the same easily nameable, yet formally unrelated, pictorial stimuli. Match-to-sample baselines for eight four-member classes were established under the contextual control of two colors. In the presence of one color, conditional relations were established between stimuli whose normative names rhymed. In the presence of the other color, conditional relations were established between stimuli whose normative names did not rhyme. Although, during Experiment 1, all participants demonstrated equivalence classes involving rhyming stimuli, none demonstrated the formation of nonrhyme equivalence classes. To investigate this finding, Experiment 2 evaluated whether participants would demonstrate both rhyme and nonrhyme equivalence classes given more extensive exposure to the experimental contingencies. All participants demonstrated contextually controlled rhyme and nonrhyme equivalence classes, although rhyme classes were demonstrated with greater facility than nonrhyme classes. Results indicate that visual stimuli are named, that verbal bases for stimulus classification can affect the emergence of contextually controlled equivalence classes, and that untrained contextually controlled conditional discriminations involving novel stimuli can emerge on the basis of participants' verbal behavior.

Key words: stimulus equivalence, contextual control, naming, verbal behavior, generalization, match-to-sample, adult humans

The emergence of untrained human behavior continues to provide a central focus for behavior analytic research, not least because of the insights it offers into key questions of conceptual learning and categorization (cf. Zentall, Galizio, & Critchfield, 2002). For many years, stimulus equivalence (Sidman, 1994, 2000) has provided the dominant paradigm for research into such phenomena, and a large number of studies have documented the diversity of participants, stimuli, and experimental preparations to which it can be applied (see Green & Saunders, 1998, for a review). Although Sidman (1994, 2000) has proposed equivalence relations to be a fundamental outcome of the environmental contingencies to which individuals are exposed, other researchers have sought to address the possibility that verbal behavior may, in fact, provide the prerequisites both for stimulus equivalence and other higher-order human responding (Horne & Lowe, 1996, 1997, 2000).

Although comparatively little research has, as yet, directly explored the role of verbal behavior in stimulus-class formation, a number of studies already have indicated that participants' verbal behavior can indeed be functional during equivalence research. Dugdale and Lowe (1990), for example, showed that when children who had previously failed to demonstrate equivalence classes of visual stimuli were taught a class-consistent common name for those stimuli, they were subsequently able to demonstrate equivalence. Similar findings have been reported by Lowe and Beasty (1987), who demonstrated that teaching intraverbal naming strategies to children facilitated the emergence of equivalence classes of visual stimuli, and by Eikeseth and Smith (1992) using visual stimuli presented to children with autism. Although such studies provided strong initial evidence of the functionality of verbal behavior in equivalence class formation, criticism has been made because the naming interventions employed were necessarily confounded with additional exposure to the match-to-sample contingencies of the experiments (Mandell & Sheen, 1994).

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

Correspondence concerning this article should be addressed to Bob Remington, School of Psychology, University of Southampton, Highfield, Southampton, Hampshire, SO17 1BJ, UK. (E-mail: rer1@soton.ac.uk).

doi: 10.1901/jeab.2006.82-05

Avoiding such concerns, more recent research has employed naming pretraining among very young children to demonstrate that teaching common tacting for arbitrary stimuli can result both in stimulus categorization and transfer of function when both speaker and listener behavior are present (Horne, Lowe, & Randle, 2004; Lowe, Horne, Harris, & Randle, 2002).

Complementing such findings, other researchers have suggested that adult participants' preexisting verbal repertoires can be used to provide an alternative means of investigating higher-order human behavior (e.g., Gagné, 2002). In accordance with this perspective, research already has shown that the emotionally charged, fear-related, and sexually explicit meanings of stimulus words can significantly affect the formation of equivalence classes (Plaud, 1995; Plaud, Gaither, Franklin, Weller, & Barth, 1998; Plaud, Gaither, Weller, et al., 1998) and that preexisting verbal relations also can affect the formation of equivalence relations involving Protestant and Catholic names among Irish participants (Roche, Barnes-Holmes, Barnes-Holmes, Stewart, & O'Hora, 2002). A number of studies have further demonstrated differences in equivalence-class formation as a result of adult participants' verbal behavior through direct manipulation of the naming strategies employed by participants during experimentation and by varying the nameability of the stimuli presented to them (e.g., Arntzen, 2004; Bentall, Dickins, & Fox, 1993; Dickins, Bentall, & Smith, 1993; Mandell & Sheen, 1994; Smith, Dickins, & Bentall, 1996).

Employing a related approach, Randell and Remington (1999) presented different class arrangements of the same easily nameable, yet formally unrelated, pictorial stimuli to groups of verbally able adults. No physical (i.e., perceptual) properties of the stimuli employed provided any consistent basis for categorization, and at no point during the experiment were participants instructed to name. If participants named stimuli, however, the class arrangements presented to one group of participants provided a ready verbal basis for categorization that the other arrangements did not (i.e., the names of the stimuli composing classes rhymed with each other). The results indicated firstly that all participants named the stimuli presented and, secondly, that when the names of stimuli composing classes rhymed, baseline learning and equivalence-class formation occurred more rapidly and accurately than when those names did not rhyme. Generalized class formation on the basis of participants' verbal behavior also was evaluated by presentation of extinction trials involving only novel pictorial stimuli, during which the name of one comparison always rhymed with that of its sample but never with those of the other comparisons. All participants previously trained with classes of rhyming stimuli met criterion for selecting only rhyme comparisons by the end of generalization testing. Surprisingly, nearly half of the participants previously trained with classes of nonrhyming stimuli performed likewise, indicating that when a ready verbal basis for stimulus categorization is available, that basis can become functional, even in the absence of specific training.

Likewise employing verbally able adult participants, Carr and Blackman (2001) reported a series of experiments that sought to investigate the relations between naming and contextually controlled equivalence classes, by establishing conflicting baseline control among classes of abstract visual stimuli. During Experiment 1, participants were simply exposed to the match-to-sample contingencies of the experiment but, during Experiment 2, receptive common names for the stimuli composing the experimental classes were additionally provided by the experimenter. During Experiment 3, participants were required to speak aloud the experimenter-provided common names during emergent testing. A high degree of variability was observed in participants' performances throughout the research, leading the authors to conclude that stimulus equivalence can result from a variety of sources of stimulus control, both verbal and nonverbal. It was noted, however, that because of the methodology employed, it had not been possible to evaluate the extent to which performance was under the control of idiosyncratic covert naming strategies, and whether or not these strategies had been at variance with the overt verbal and nonverbal contingencies of the experiment.

Because of the inconclusive nature of previous findings and the key theoretical and methodological role that contextual control has played in the development of equivalence research (e.g., Bush, Sidman, & de Rose, 1989;

Sidman, 1994; Sidman, Willson-Morris, & Kirk, 1986; Wulfert & Hayes, 1988), a principal aim of the present research was to investigate further the complex functional relations between naming, contextual control, and equivalence. The research additionally sought to provide a within-subjects replication of Randell and Remington's (1999) principal findings. As Remington (1996) has noted, such research may not have been optimal in that it relied on differences observed between groups of participants to demonstrate the functionality of verbal behavior. In the present study, therefore, one rhyming and one nonrhyming arrangement of stimulus classes (identical to those employed in the Rhyme and Diagonal conditions, respectively, by Randell & Remington, 1999) were presented to each participant under the contextual control of background colors. As many other methodological features as possible of the previous research were retained, including assessment of generalized class formation on the basis of participants' verbal behavior.

In recent years, a burgeoning literature has documented the emergence of generalized equivalence classes as a result of the merger of equivalence and open-ended (i.e., nonarbitrary) stimulus classes, indicating that equivalence classes can expand to include potentially limitless numbers of stimuli on the basis of physical similarity (e.g., Adams, Fields, & Verhave, 1993; Dougher, Perkins, Greenway, Koons, & Chiasson, 2002, Fields, Matneja, Varelas, & Belanich, 2003; Fields & Reeve, 1991; Fields, Reeve, Adams, Brown, & Verhave, 1997; Galizio, Stewart, & Pilgrim, 2004). Such findings have led to the suggestion that generalized equivalence classes may provide a basis for understanding the formation of natural categories (Fields, Adams, Buffington, Yang, & Verhave, 1996; Fields, Reeve, Adams, Verhave, 1991; Fields et al., 2002), an argument supported by research demonstrating that generalized equivalence classes can be brought under contextual control (Griffee & Dougher, 2002; Rehfeldt, 2003), and that specific contextual functions can generalize to arrangements of entirely novel stimuli on the basis of the procedural characteristics of previous contextual-control training (Pérez-Gonzáles & Serna, 2003; Serna & Pérez-Gonzáles, 2003).

Despite rapid growth in this area of research, however, no study has yet attempted to investigate whether contextually controlled equivalence classes can generalize to include novel stimuli solely on the basis of participants' verbal behavior. Within-subjects replication of Randell and Remington's (1999) generalization testing phase allowed this possibility to be addressed, additionally providing a methodological check on the function of the contextual stimuli previously employed during the research (cf. Meehan & Fields, 1995). A secondary phase of generalization testing, additional to that employed by Randell and Remington (1999), further allowed evaluation of whether previously employed contextual stimuli can, on the basis of participants' verbal behavior, participate in generalized classes composed of entirely novel compound stimuli (cf. Markham & Dougher, 1993; Schenk, 1995). In summary, therefore, the present research sought to investigate the effects of verbal behavior on the formation and generalization of contextually controlled equivalence classes of formally unrelated visual stimuli, additionally providing a constructive replication of Randell and Remington's (1999) principal findings. In fulfilment of these aims, the research sought further to increase our understanding of the relations between contingency-shaped and verbally controlled (Horne & Lowe, 1996) emergent behavior in adult humans.

EXPERIMENT 1

Method

Participants

Sixteen students and staff at the University of Southampton (9 female, 7 male) volunteered to participate in the experiment in exchange for payment at a rate of £2.50 (approximately $4.00) per 30 min. Aged between 18 and 30 years, all were native English speakers with no prior knowledge of the research.

Materials, Apparatus, and Setting

All stimuli were presented, and all participants' responses recorded, by a Power Macintosh® computer running software specifically designed for equivalence research (Dube & Hiris, 1997). Throughout baseline training, emergent testing, and stage 1 generalization testing, its 15-in. (38-cm) monitor presented five transparent keys (4.5 cm square) indis-

cernible over a red or blue background. During stage 2 generalization testing, the keys were presented over a white background. Throughout the experiment, sample stimuli were presented on the center key, and comparisons were presented on four outer keys, one in each corner of the screen. During stage 1 and 2 generalization testing, two of the outer keys, their positions varying pseudorandomly from trial to trial, always remained blank. Participants were tested individually in a windowless cubicle containing a desk, on which were placed a consent form, written experimental instructions, a 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 all responses were made using the mouse.

Stimuli and Class Arrangements

Thirty-one line drawings of easily nameable, yet formally unrelated, objects and animals (some adapted from Snodgrass & Vanderwart, 1980) were employed during the experiment, the normative names of which were all between three and five letters in length (see Figure 1). Contextual stimuli were provided by the red or blue background colors upon which the pictorial stimuli were presented. One red and one blue square (of the same colors as the contextual stimuli and the same size as the keys upon which the pictorial stimuli were presented) also were employed.

Four classes (rhyme classes), composed of stimuli whose names rhymed with each other, were trained, reviewed, and tested in the presence of one contextual stimulus (rhyme contextual stimulus). Trials involving these classes (rhyme trials) always presented a sample whose name rhymed with that of the correct comparison, but never with those of the incorrect comparisons. The other four classes (nonrhyme classes) were composed of the same stimuli rearranged into classes composed of stimuli whose names did not rhyme with each other. These classes were trained, reviewed, and tested in the presence of the other contextual stimulus (nonrhyme contextual stimulus). Trials involving these classes (nonrhyme trials) always presented a sample whose name rhymed with that of one of the incorrect comparisons, but never with those of either the correct or the other two incorrect comparisons, whose names also did not rhyme with each other.

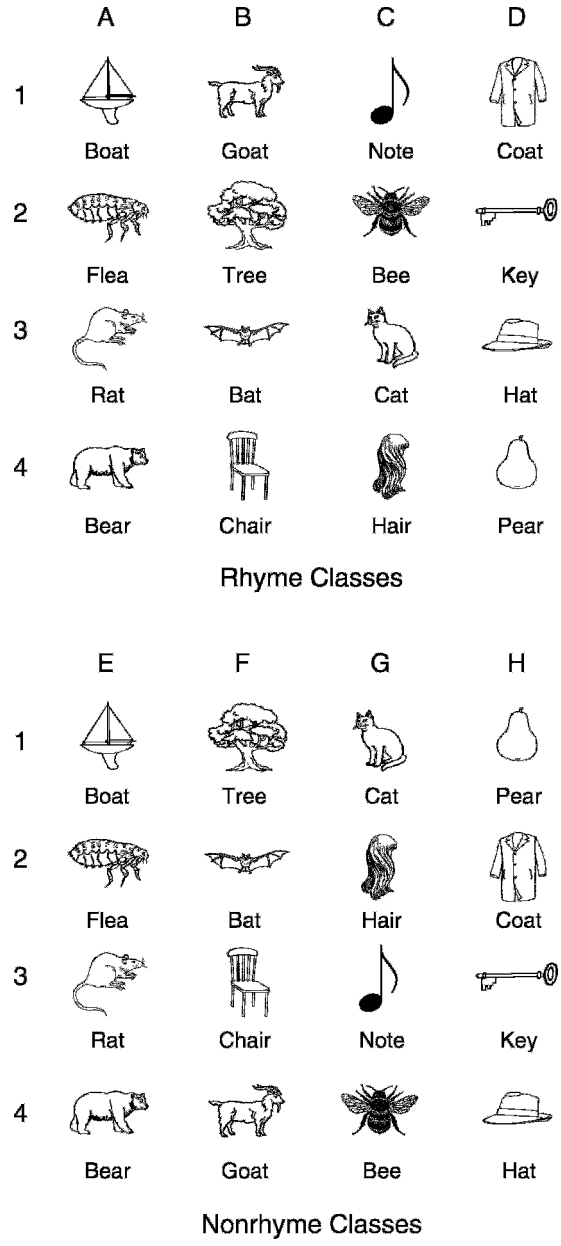

Fig. 1. Class configurations, stimuli, and normative names of stimuli presented to all participants during baseline training and review, and emergent testing in Experiments 1 and 2. Lettered columns denote stimuli, numbered rows denote classes.

Figure 2 presents stimuli and class arrangements employed during stage 1 and 2 generalization testing. During the former phase, a sample and two comparisons were presented

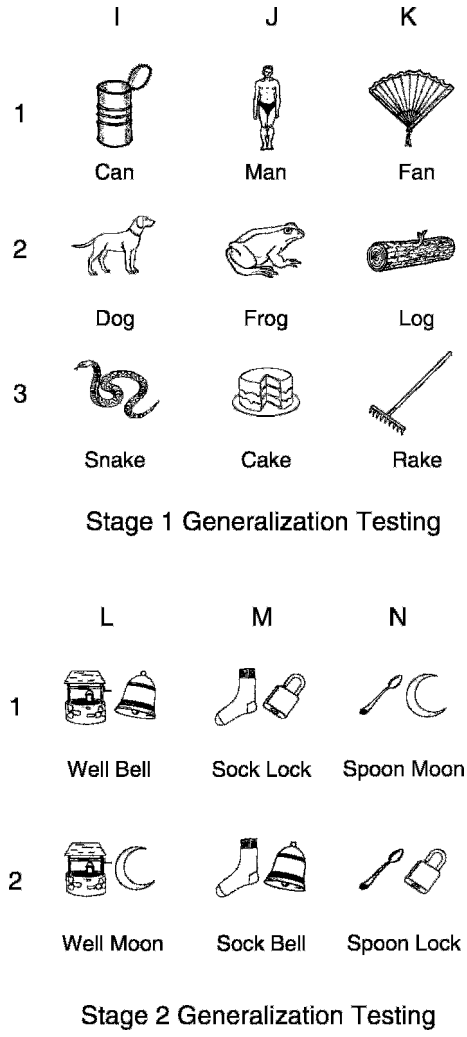

Fig. 2. Stimuli, stimulus compounds, and normative names of stimuli presented to all participants during stage 1 and 2 generalization testing in Experiments 1 and 2. Numbered rows denote rhyme classes or compounds; lettered columns denote stimuli.

on every trial, the name of one of which always rhymed with that of the sample and the name of the other of which did not. Half of these trials were presented in the presence of the rhyme contextual stimulus and half in the presence of the nonrhyme contextual stimulus. During the latter phase, a red and a blue square were presented as comparisons on every trial, and a stimulus compound, composed of a pair of pictorial stimuli whose names either rhymed (rhyme compound) or did not rhyme (nonrhyme compound) with each other, was presented as the sample.

Procedure and Match-to-Sample Contingencies

Each participant took part in one experimental session that lasted no longer than 1.5 hr. Subsequent to providing informed consent, participants were asked to read the instructions before them and then were 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 choices, 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 onscreen instructions and match-to-sample training commenced.

Each trial began with the presentation of a sample stimulus, an observing response with the mouse causing the comparison stimuli additionally to be displayed. All stimuli remained in view until selection of a comparison caused them to disappear, followed, after a 1-s interval, by presentation of the next trial. Positions of correct and incorrect comparisons varied pseudorandomly from trial to trial and, throughout baseline training and review, and emergent testing, comparisons were always the members of all other stimulus classes sharing the same alphabetic designation (e.g., B1, B2, B3, B4 in Figure 1). Throughout stage 1 generalization testing, for every sample, two comparisons sharing the same alphabetic designation were selected quasirandomly from one of the other two stimulus classes (e.g.,

Sample I1, Comparisons J1 and J3 in Figure 2). At no point during the experiment 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. During reinforcement trials, selection of a class-consistent comparison was followed by a beep and the word ''Correct'' displayed onscreen. Other choices resulted in a buzz and a darkened screen. During extinction trials, however, the only consequence of a response was presentation of the next trial.

All participants were exposed to the same performance-contingent linear-series (i.e., AB–BC–CD) training and testing program. Order of baseline training (rhyme or nonrhyme classes first) and color of rhyme and nonrhyme contextual stimuli (red or blue) were counterbalanced across participants. The training order presented below is for participants trained with rhyme classes first. For participants trained with nonrhyme classes first, Phases 3 and 4 were presented prior to Phases 1 and 2.

Phase 1: Train rhyme baseline relations. Initially, AB relations were trained, with each of the four relations appearing once every four trials in pseudorandom order. When a criterion of 12 consecutive correct responses had been achieved, BC relations were trained in identical fashion. When criterion for these relations had been met, CD relations were trained, again to the same criterion.

Phase 2: Review rhyme baseline relations. When the above criteria had been met, rhyme baseline relations were reviewed with reinforcement in 12-trial blocks composed of all AB, BC, and CD trials intermixed in pseudorandom order. Samples from the same class were never presented consecutively. On completion of one reinforcement review block with 100% accuracy, all rhyme baseline relations were reviewed in extinction. If performance remained at 100% accuracy over the first 12-trial block, training of nonrhyme baseline relations commenced (Phase 3). If criterion was not achieved during the second (extinction) review block, rhyme baseline relations were again reviewed with reinforcement until criterion was re-attained, permitting review in extinction. Review of rhyme baseline relations continued in this way until 100% accuracy was demonstrated over one extinction review block.

Phases 3 and 4: Train and review nonrhyme baseline relations. All nonrhyme (i.e., EF, FG, and GH) baseline relations were trained and reviewed in identical fashion to rhyme baseline relations during Phases 1 and 2, above.

Phase 5: Full review of rhyme and nonrhyme baseline relations. Initially, a maximum of two blocks of 24 extinction trials was presented in which all baseline relations established during Phases 1 to 4 were reviewed in pseudorandom order. No more than two rhyme or nonrhyme trials were presented consecutively. If participants satisfied a minimum criterion of 22 of 24 correct responses during either block, emergent testing commenced (Phase 6). If criterion was not met, however, all baseline relations were reviewed in the same manner, but with reinforcement. On completion of one block of reinforcement trials to the same 22 of 24 criterion, all baseline relations were again reviewed in extinction. Review of baseline relations continued in this way until criterion was met in extinction.

Phase 6: Emergent testing. All possible rhyme and nonrhyme emergent relations except reflexivity (see Bush et al., 1989) were presented intermixed, in pseudorandom order, in a maximum of two blocks of 72 extinction trials. Each block was composed of 12 symmetric, 12 transitive, and 12 equivalence (i.e., combined transitivity and equivalence) rhyme and nonrhyme trials, each presented once. No more than two rhyme or nonrhyme trials were ever presented consecutively. If, during the first block of emergent testing, participants satisfied a minimum criterion of 68 of 72 class-consistent responses across rhyme and nonrhyme trials, stage 1 generalization testing (Phase 7) commenced. If criterion was not achieved, however, full review of rhyme and nonrhyme baseline relations (Phase 5) was repeated until criterion was again met in extinction. The second block of emergent testing was then presented, followed, regardless of performance, by stage 1 generalization testing.

Phase 7: Stage 1 generalization testing. One block of 36 extinction trials was presented, involving only novel pictorial stimuli (Figure 2, top). Eighteen trials presented stimuli in the presence of the rhyme contextual stimulus (rhyme context trials), and 18 trials presented the same combinations of stimuli in the presence of the nonrhyme contextual stimulus (nonrhyme context trials). Each

stimulus was presented twice as a sample in the presence of both contextual stimuli. All trials were intermixed in pseudorandom order and the same sample or contextual stimulus was never presented on more than two consecutive trials. Subsequent to completion of this phase, stage 2 generalization testing (Phase 8) commenced.

Phase 8: Stage 2 generalization testing. A maximum of two 12-trial blocks was presented in extinction, involving only novel visual stimuli (Figure 2, bottom). Each rhyme and nonrhyme compound was presented twice as sample during both blocks. If, at the end of the first block, participants had selected only rhyme keys (i.e., keys of the same color as the rhyme contextual stimulus) in response to rhyme compounds, and nonrhyme keys (i.e., keys of the same color as the nonrhyme contextual stimulus) in response to nonrhyme compounds, then match-to-sample testing was concluded by an onscreen message requesting participants to complete the written posttest in the envelope before them (Phase 9). If participants did not respond in this way, then the second block was presented, subsequent to which the same message was displayed, regardless of performance.

Phase 9: Written posttesting. Participants completed a written posttest that firstly requested them to indicate, for each stimulus, any name or names by which they had referred to it during the experiment and, secondly, to describe briefly any strategies by which they had selected stimuli during each phase of the experiment.

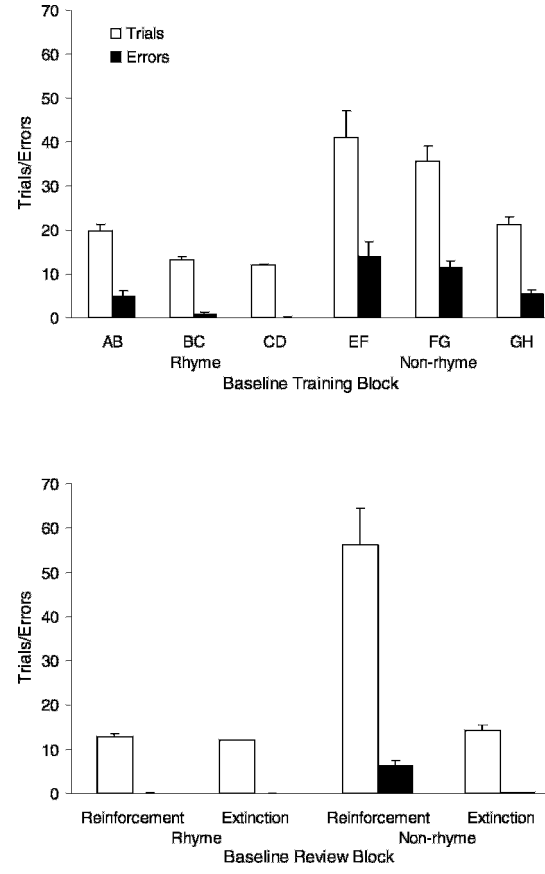

Fig. 3. Mean number of trials and errors (+*SE*) during rhyme and nonrhyme baseline training (Phases 1 and 3, top graph) and baseline review (Phases 2 and 4, bottom graph). Accuracy criteria were 12 consecutive correct trials for each baseline training block and 12/12 trials correct for each baseline review block.

Results

Although all 16 participants completed the experiment, none demonstrated the formation of contextually controlled rhyme and nonrhyme equivalence classes. The effects of participants' verbal behavior were strongly implicated in their match-to-sample performance throughout the experiment and confirmed by written posttesting.

Phases 1 and 3: Train Rhyme and Nonrhyme Baseline Relations

The upper graph of Figure 3 shows, across all participants, the mean trials and errors required to reach the baseline criteria. All participants but one made fewer errors and required fewer trials to establish rhyme baseline relations than to establish nonrhyme baseline relations. Three participants required only 36 trials (0 errors) to meet the three consecutive 12 of 12 criteria for rhyme baseline training and the greatest number required was 54 (13 errors). By contrast, the smallest number of trials required by any participant to meet the same criteria for nonrhyme baseline relations was 54 (13 errors) and the greatest was 170 (79 errors).

Phases 2 and 4: Review Rhyme and Nonrhyme Baseline Relations

The lower graph of Figure 3 shows mean trials and errors to criteria during baseline review. All participants met the 12 of 12

criterion for rhyme baseline relations during the first reinforcement review block except one, who did so during the second. By contrast, only 1 participant met the same criterion for nonrhyme baseline relations during the first reinforcement review block and only 3 others did so during the second. The greatest number of nonrhyme trials required by any participant to meet criterion with reinforcement was 120 (17 errors). All participants met the 12 of 12 criterion for rhyme baseline relations during the first extinction review block, and all performed likewise during extinction review of nonrhyme baseline relations except 3, who made one error each during the first block but met criterion during the second.

Phases 5 and 6: Full Review of Rhyme and Nonrhyme Baseline Relations and Emergent Testing

Because participants alternated between Phases 5 and 6 depending upon their performance in Phase 6, the data in these phases should be considered together.

Review prior to emergent testing block 1. Only 4 participants met the 22 of 24 criterion for full baseline review during the first extinction review block, although 2 others made only three errors each. None of the 12 participants who failed to meet criterion during the first extinction review block required more than two additional reinforcement review blocks to re-attain criterion, and all 12 met the criterion for full baseline review in extinction during their second review block.

Emergent testing block 1. The upper graph of Figure 4 shows each participant's total errors on rhyme and nonrhyme trials during the first block of emergent testing. No participant met both of the 34 of 36 criteria for contextually controlled rhyme and nonrhyme equivalence classes during the first block of emergent testing. Twelve participants (JW2, AW, HL, JP, AA2, AA1, MS, RW, JW1, AC, CB, AT) met criterion for formation of rhyme equivalence classes, however, and the remaining 4 participants (AH, FT, RR, AJ) respectively made only four, three, three, and six errors each. No participant met criterion for nonrhyme equivalence classes, and only 2 participants made fewer than 25 errors. Ten participants made 31 errors or more. On average, 81.7% of the errors made by participants on nonrhyme trials were other-context (i.e., rhyme) selections.

Review prior to emergent testing block 2. Six participants met criterion for full baseline review in extinction immediately subsequent to the first block of emergent testing, and 2 others made only three errors each. Only 2 of the other 8 participants required more than a single reinforcement review block to re-attain criterion (two blocks each), and all met criterion for full baseline review in extinction during their second review block and proceeded to the second block of emergent testing.

Emergent testing block 2. The lower graph of Figure 4 shows each participant's total errors on rhyme and nonrhyme trials during the second block of emergent testing. No participant met both criteria for contextually controlled equivalence classes during the second block of emergent testing. Thirteen participants (JW2, AW, HL, JP, FT, RR, AA2, AA1, MS, RW, JW1, AC, CB) met criterion for rhyme equivalence classes, however, and the remaining 3 participants (AH, AJ, AT) respectively made only four, three, and six errors each. As during block 1, no participant met criterion for nonrhyme equivalence classes, and only 1 participant (RR) made fewer than 25 errors. Nine participants made 31 errors or more. On average, 87.7% of the errors made by participants on nonrhyme trials were other-context selections.

Phase 7: Stage 1 Generalization Testing

One participant (AW) selected only rhyme comparisons on rhyme trials and 17 of 18 nonrhyme comparisons on nonrhyme trials. Another 3 participants (JW2, RW, JW1) selected 16 rhyme comparisons or more on rhyme trials, and only 1 of these (JW2) selected fewer than 15 nonrhyme comparisons on nonrhyme trials. Nine participants selected only rhyme comparisons regardless of contextual stimulus, and 3 others (AH, AA2, AA1) selected two nonrhyme comparisons or fewer. On average, participants made 17.6 (SD = .8) rhyme selections on rhyme trials and 3.7 (SD = 6.6) nonrhyme selections on nonrhyme trials.

Phase 8: Stage 2 Generalization Testing

Three participants (AW, AA2, JW1) met the 12 of 12 criterion for differential selection of rhyme keys in response to rhyme compounds and nonrhyme keys in response to nonrhyme

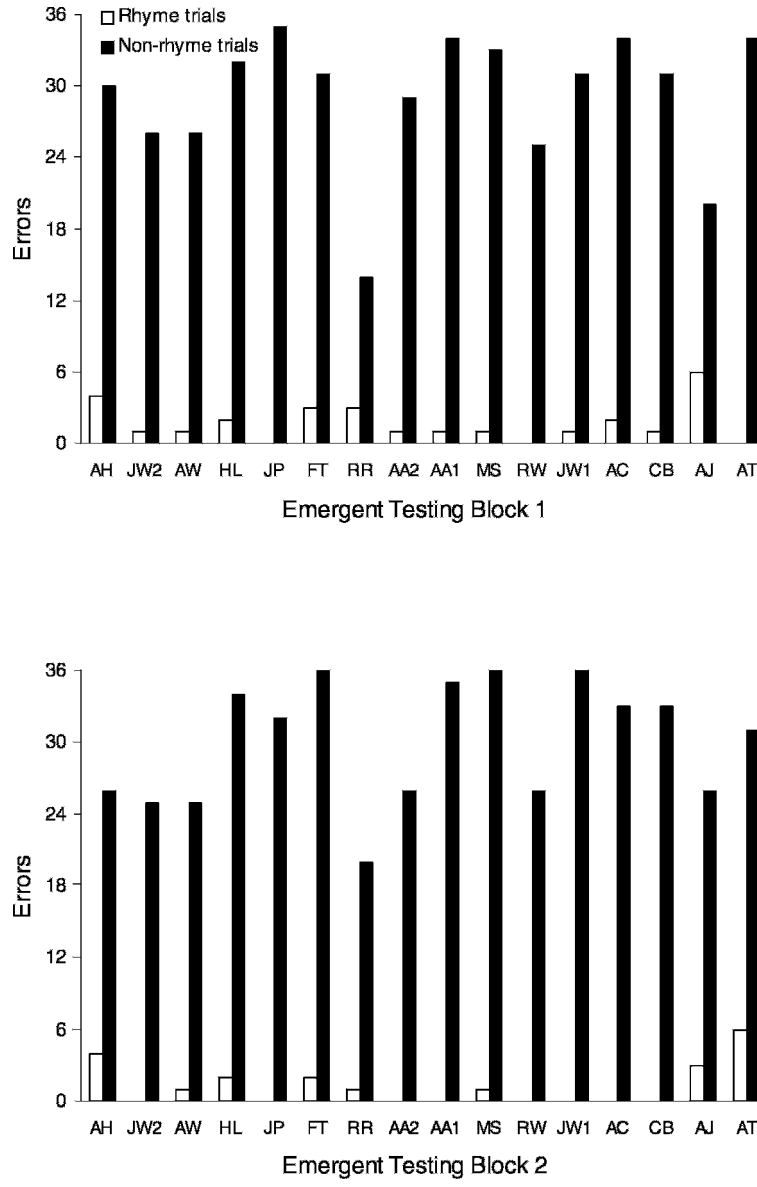

Fig. 4. Number of errors on rhyme (open bars) and nonrhyme (filled bars) trials for each participant during emergent testing (Phase 6) Block 1 (top graph) and Block 2 (bottom graph). Minimum accuracy criterion for each block was 68/72 trials correct.

compounds during the first testing block. A further 3 met criterion during the second block, during which a 4th made only a single other-context selection. The 10 participants who did not meet criterion in either testing block made, on average, 8.9 (*SD* = 1.5) context-appropriate selections during the first block, and 9.3 (*SD* = 1.5) context-appropriate selections during the second.

Phase 9: Written Posttesting

Fourteen participants reported using normative stimulus names for every pictorial stimulus presented during the experiment, although 2 of these indicated that they had initially used nonnormative names for one or more baseline stimuli. Of these 2 participants, 1 additionally indicated that he had continued to use nonnormative names for some stimuli,

Table 1

Strategies reported during written posttesting, showing number of participants who indicated using a given strategy on rhyme and nonrhyme trials during baseline training and emergent testing, on rhyme and nonrhyme context trials during stage 1 generalization testing, and on rhyme and nonrhyme compound trials during stage 2 generalization testing.

Baseline training and emergent testing	Rhyme trials	Nonrhyme trials
Selected on basis of rhyme	15	8
Remembered pairs of stimuli	0	12
Intraverbal strategies	0	11
Preexperimental associations	0	10
Imagery to link stimuli	0	9
Spatio-temporal strategies	0	4
Stage 1 generalization testing	**Rhyme context trials**	**Nonrhyme context trials**
Selected on basis of rhyme	15	10
Selected on basis of nonrhyme	0	4
Spatio-temporal strategies	0	2
Stage 2 generalization testing	**Rhyme compound trials**	**Nonrhyme compound trials**
Selected on basis of rhyme	9	1
Selected on basis of nonrhyme	1	9
Preexperimental associations	3	3
Spatio-temporal strategies	1	1
Did not name stimuli	1	1
Did not use any strategy	1	1

but only on nonrhyme trials. The remaining 2 participants either did not indicate names for stimuli, or noted the names of other nonrhyme class members (e.g., ''tree'' for ''boat''). One participant made no indication of using any strategy during any phase of the experiment. All 15 other participants, however, reported that they had selected comparisons on the basis of rhyme on rhyme trials during baseline training and emergent testing, but that on nonrhyme trials, they had employed a variety of strategies to select stimuli (see Table 1).

Discussion

The effects of participants' verbal behavior were strongly evident during all match-to-sample phases of the experiment. Although every participant met all criteria for establishment and maintenance of contextually controlled baseline relations prior to emergent testing, none demonstrated the emergence of both rhyme and nonrhyme equivalence classes under contextual control. Effects of verbal behavior were particularly salient during emergent testing, with participants selecting rhyme stimuli throughout the majority of emergent trials, regardless of contextual cue. The functionality of verbal behavior also was strongly implicated in participants' performance throughout generalization testing. During stage 1 generalization testing, two specific patterns of verbally controlled responding predominated: differential selection of rhyme and nonrhyme comparisons in the presence of the rhyme and nonrhyme contextual stimuli, respectively, and selection of rhyme comparisons regardless of contextual cue. By the end of stage 2 generalization testing, likewise, nearly half of the participants differentially selected rhyme keys in response to rhyme compounds and nonrhyme keys in response to nonrhyme compounds—a performance that, for these participants, would only appear explicable as a result of verbal control. Written posttesting further confirmed the effects of verbal behavior observed during match-to-sample training and testing, with all participants indicating that they named the pictorial stimuli, and that this naming was almost ubiquitously normative. Participants' posttests also indicated that although they almost always selected stimuli on the basis of rhyme on rhyme trials during baseline training and emergent testing, they used a variety of strategies on nonrhyme trials.

Although Experiment 1 succeeded in demonstrating the functionality of verbal behavior within a contextually controlled equivalence paradigm, it nevertheless failed to demonstrate

the contextually controlled emergence of nonrhyme equivalence classes. A substantial literature attests to the delayed emergence of equivalence within a variety of experimental preparations (e.g., Bush et al., 1989; Harrison & Green, 1990; Holth & Arntzen, 1998; Sigurdardottir, Green, & Saunders, 1990), however, and to investigate whether more extensive exposure to the experimental contingencies would permit nonrhyme equivalence classes to emerge, a second experiment was conducted. If successful, Experiment 2 would confirm the functionality of participants' verbal behavior in both the emergence and generalization of contextually controlled equivalence classes.

EXPERIMENT 2

Method

Four students (CH, FU, MW, and MG) at the University of Southampton (2 female, 2 male) aged between 23 and 26 years volunteered to participate in a maximum of five experimental sessions on consecutive days (Sessions 1 to 5). No session exceeded 1.5 hr duration and payment for participation, at the same rate as for Experiment 1, was made at the end of each session. Stimuli, class arrangements, and procedure were the same as during Experiment 1 except that if, during Sessions 1 to 4, participants did not meet criterion during emergent testing (Phase 6), the session was concluded with onscreen instruction prompting the participant to arrange an appointment for the next session. Baseline training and review (Phases 1 to 4) were not presented during Sessions 2 to 5, each of which commenced with full review of baseline relations in extinction, and with reinforcement if criterion was not met (Phase 5).

Results

All participants completed the experiment, each demonstrating the formation of contextually controlled rhyme and nonrhyme equivalence classes to criterion prior to or during Session 4. As during Experiment 1, effects of participants' verbal behavior were evident throughout the experiment.

Phases 1 and 3: Train Rhyme and Nonrhyme Baseline Relations

All participants made fewer errors and required fewer trials to establish rhyme baseline relations than to establish nonrhyme baseline relations. The smallest number of trials required by any participant (CH) to meet the three criteria for rhyme baseline training was 36 (0 errors) and the greatest (FU) was 51 (10 errors). By contrast, the smallest number of trials required by any participant (MG) to meet the same criteria for nonrhyme baseline relations was 66 (13 errors) and the greatest (CH) was 441 (381 errors).

Phases 2 and 4: Review Rhyme and Nonrhyme Baseline Relations

Although all participants met criterion for rhyme baseline relations during the first reinforcement review block, only MW performed likewise during the first nonrhyme reinforcement baseline review block. FU and MG met criterion during their third and fourth blocks, respectively, and CH required 12 blocks. All participants met criterion during the first extinction review block of rhyme baseline relations, and all performed likewise during the first extinction review block of nonrhyme baseline relations except MW, who made a single error during the first block and met criterion during the second.

Phases 5 and 6. Full Review of Rhyme and Nonrhyme Baseline Relations and Emergent Testing

Figure 5 shows the number of extinction (top graph) and reinforcement (bottom graph) full baseline review trials required by each participant to meet criterion prior to each block of emergent testing completed during the experiment. Figure 6 shows the number of errors on rhyme (top graph) and nonrhyme (bottom graph) trials during each of those emergent testing blocks for each participant. Because participants alternated between Phases 5 and 6, depending upon their performance in Phase 6, the data in Figures 5 and 6 should be considered together. Each participant's performance in these two phases is presented individually.

Participant CH. This participant completed three experimental sessions, meeting criterion for equivalence during her fifth emergent testing block. CH required only one full baseline review block to meet criterion in extinction immediately prior to every emergent testing block that she completed and consequently received no full baseline review

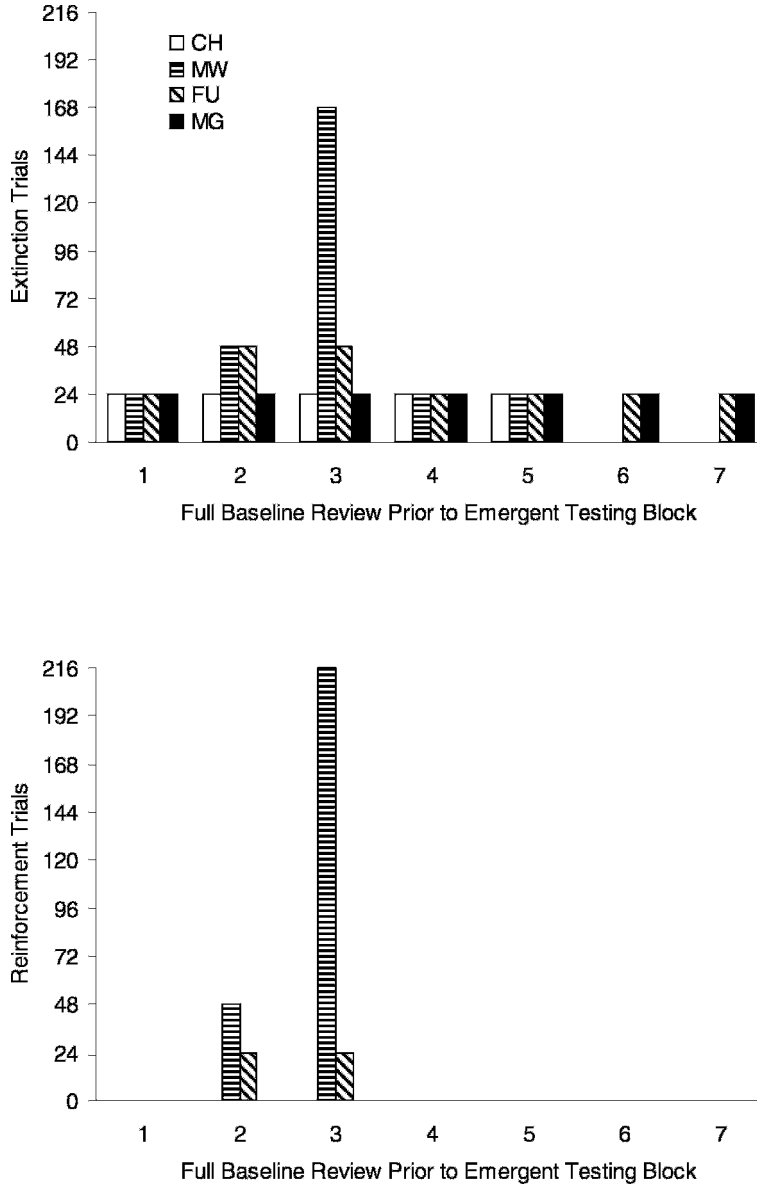

Fig. 5. Number of trials required by each participant to meet criteria for reinforcement (bottom graph) and extinction (top graph) during full baseline review (Phase 5) prior to each block of emergent testing. Minimum accuracy criterion for each block was 22/24 trials correct.

reinforcement trials. During a total of 360 emergent trials, CH made 81 errors, three of which were on rhyme trials (one other-context selection) and 78 of which were on nonrhyme trials (45 other-context selections).

Participant MW. This participant also completed three experimental sessions and met criterion for equivalence during his fifth emergent testing block. MW required only one full baseline review block to meet criterion in extinction prior to emergent testing blocks 1, 4, and 5, but required two and nine reinforcement blocks to re-attain criterion in extinction prior to emergent testing blocks 2 and 3, respectively. During a total of 360 emergent trials, MW made 109 errors, three of which were on rhyme trials (no other-context selections) and 106 of which

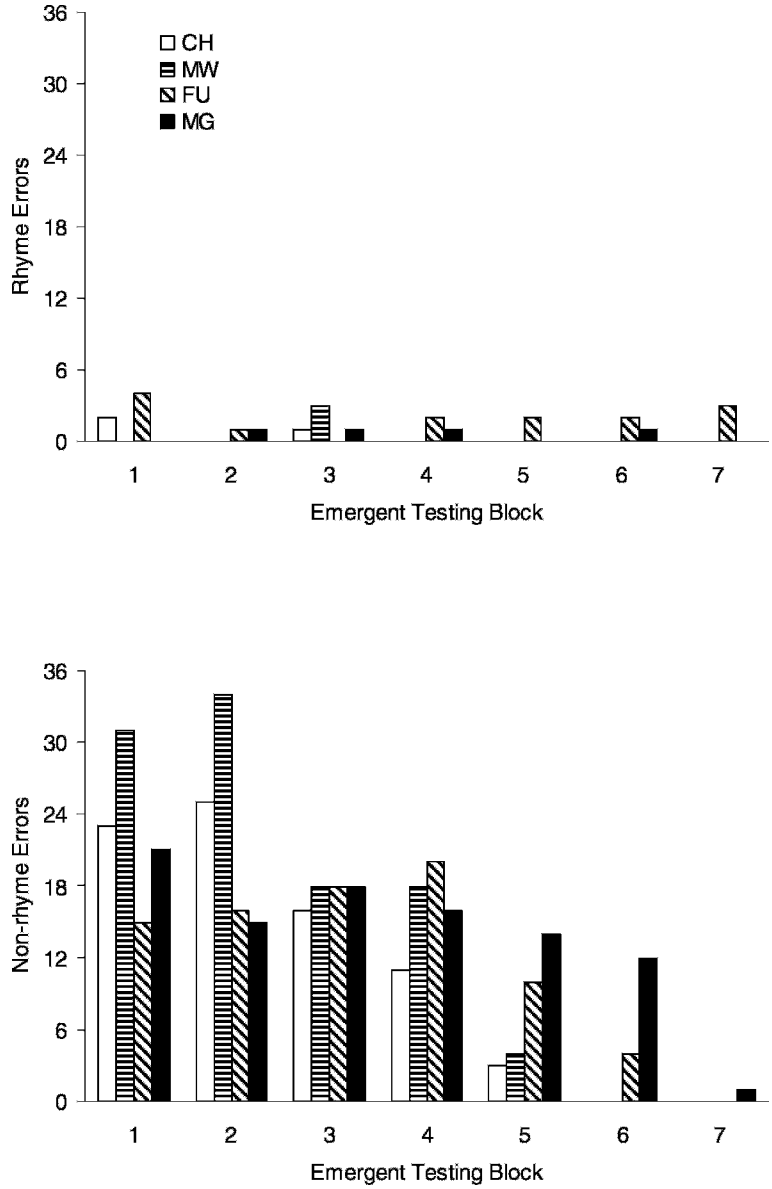

Fig. 6. Number of errors for each participant on rhyme (top graph) and nonrhyme (bottom graph) trials during emergent testing (Phase 6). Minimum accuracy criterion for each block was of 68/72 trials correct.

were on nonrhyme trials (80 other-context selections).

Participant FU. This participant completed four experimental sessions, meeting criterion for equivalence during her seventh emergent testing block. FU required only one full baseline review block to meet criterion in extinction prior to all emergent testing blocks except 2 and 3, prior to both of which she required only a single reinforcement block to re-attain criterion in extinction. During a total of 504 emergent trials, FU made 97 errors, 14 of which were on rhyme trials (two other-context selections) and 83 of which were on nonrhyme trials (22 other-context selections).

Participant MG. This participant also completed four experimental sessions and met criterion for equivalence during his seventh

emergent testing block. MG required only one full baseline review block to meet criterion in extinction prior to each emergent testing block. During a total of 504 emergent trials, MG made 101 errors, four of which were on rhyme trials (one other-context selection) and 97 of which were on nonrhyme trials (33 other-context selections).

Phase 7: Stage 1 Generalization Testing

Participants predominantly selected rhyme comparisons on rhyme context trials and nonrhyme comparisons on nonrhyme context trials. CH, FU, MW, and MG selected 16, 14, 13, and 15 rhyme comparisons on the 18 rhyme context trials, respectively, and 15, 15, 17, and 16 nonrhyme comparisons on the 18 nonrhyme context trials.

Phase 8: Stage 2 Generalization Testing

All participants selected only rhyme keys in response to rhyme compounds and nonrhyme keys in response to nonrhyme compounds during the first block of stage 2 generalization testing.

Phase 9: Written Posttesting

All participants reported having named every pictorial stimulus normatively during the experiment and, during baseline training and emergent testing, having selected comparisons on the basis of rhyme on rhyme context trials. On nonrhyme context trials, however, CH and FU reported having used preexperimental associations between stimuli as a basis for comparison selection, and MW and MG reported having used ''mental imagery'' (both gave the example of a ''cat eating a pear in a boat''). All participants indicated that during stage 1 generalization testing they had selected stimuli whose names rhymed on rhyme context trials and stimuli whose names did not rhyme on nonrhyme context trials—although FU additionally noted that she had made a number of errors in so doing. All participants indicated that during stage 2 generalization testing they had selected rhyme keys in response to rhyme compounds and nonrhyme keys in response to nonrhyme compounds.

Discussion

As in Experiment 1, all participants met criterion for mastery and maintenance of contextually controlled baseline relations in extinction prior to emergent testing, but, in contrast to the previous experiment, all participants also demonstrated the emergence of contextually controlled rhyme and nonrhyme equivalence classes. The effects of participants' verbal behavior remained strongly evident throughout match-to-sample training and testing in that all participants established rhyme baseline relations more rapidly and accurately than nonrhyme baseline relations and, throughout emergent testing, all participants made fewer errors on rhyme trials than on nonrhyme trials. During stage 1 generalization testing, all participants selected, on a majority of trials, rhyme comparisons on rhyme context trials and nonrhyme comparisons on nonrhyme context trials, and during the first block of stage 2 generalization testing, all participants selected only rhyme keys in response to rhyme compounds and nonrhyme keys in response to nonrhyme compounds. As in Experiment 1, the match-to-sample data were supported by written posttesting, during which all participants indicated having named every pictorial stimulus normatively and that, in the presence of the rhyme or nonrhyme contextual stimulus respectively, they had selected comparisons on the basis of rhyme or a variety of idiosyncratic strategies. Experiment 2 therefore succeeded in demonstrating the functionality of verbal behavior in the emergence and generalization of contextually controlled equivalence classes, additionally indicating that the nonemergence of nonrhyme equivalence classes observed in Experiment 1 was a result of participants' preexisting verbal repertoires and their verbal behavior during the experiment.

GENERAL DISCUSSION

The present research sought firstly to investigate the relations between verbal behavior and the formation and generalization of contextually controlled equivalence classes, and secondly to provide a within-subjects replication and extension of Randell and Remington's (1999) principal findings. Two experiments were carried out, throughout both of which the functionality of participants' verbal behavior was strongly evident. Both experiments confirmed Randell and Remington's findings that all participants named the

stimuli presented without instruction to do so, and that when those names provided a ready verbal basis for categorization (i.e., rhyme) congruent with the experimentally defined stimulus classes, baseline establishment and equivalence class formation occurred more rapidly and accurately than when such a basis was absent. The results of both experiments further indicated that previously established contextual control can generalize to classes of novel stimuli on the basis of participants' verbal behavior and that, on the same basis, previously established contextual stimuli can enter into classes composed of novel stimulus compounds.

During Experiment 1, all participants met criterion for maintenance of contextually controlled rhyme and nonrhyme baseline relations in extinction immediately prior to emergent testing but none demonstrated the emergence of contextually controlled rhyme and nonrhyme equivalence classes: all participants but three met criterion for rhyme equivalence class formation but none met the same criterion for nonrhyme equivalence classes, participants predominantly selecting rhyme comparisons throughout emergent testing, regardless of contextual stimulus. This finding supports the conclusion that baseline training can be superseded by verbal control during emergent testing if a ready verbal basis for categorization is available (Randell & Remington, 1999), and further demonstrates that such a basis can supersede previously established contextual control during testing without reinforcement. Although Experiment 1 indicated the functionality of participants' verbal behavior during match-to-sample training and testing, it nevertheless did not succeed in its primary aim of investigating the role of verbal behavior in the emergence of contextually controlled equivalence classes. To investigate whether more extensive exposure to the experimental contingencies employed during Experiment 1 would permit the emergence of contextually controlled nonrhyme equivalence classes, a second experiment was carried out.

During Experiment 2, all participants demonstrated contextually controlled rhyme and nonrhyme equivalence classes, although all participants established baseline relations and demonstrated equivalence more rapidly and accurately when classes were composed of stimuli whose names rhymed than when they were composed of stimuli whose names did not rhyme. The rapid emergence of rhyme equivalence classes observed also supports the suggestion that participants' performance was verbally controlled in that, in the presence of the rhyme contextual stimulus, they had simply learned to select rhyme stimuli. In the presence of the nonrhyme contextual stimulus, however, participants had to learn a series of arbitrary discriminations, having no such ready verbal basis for selection available. Written posttesting confirmed these suggestions, further indicating that although participants employed a variety of idiosyncratic verbal strategies on nonrhyme trials, these did not promote the emergence of equivalence as effectively as selection on the basis of rhyme.

The finding that contextually controlled equivalence classes emerged only after repeated training and testing during Experiment 2 (i.e., that their emergence was delayed) also may suggest interactions between participants' contingency-shaped and verbally controlled behavior during the experiment: immediate session termination consequent to the nonemergence of equivalence may have served as a cue to participants that they had responded incorrectly during emergent testing. Participants' consistently high levels of accuracy during both baseline and emergent rhyme trials also may have indicated that errors had been made on emergent nonrhyme trials. In combination, therefore, across sessions, these sources of control may have produced a shift away from indiscriminate rhyme-based responding to allow the emergence of contextually controlled equivalence classes. This interpretation would be supported by the finding that participants' performance during emergent testing changed little within sessions but quite markedly between sessions.[1]

As Zentall et al. (2002) point out, categories can be defined as classes of stimuli that occasion common responses in a given context. Although a number of studies have evaluated equivalence-class generalization on the basis of formal similarities between stimuli or procedural characteristics (e.g., Dougher et al., 2002; Fields et al., 1996; Fields et al., 1991; Fields et al., 2002; Meehan & Fields, 1995; Pérez-Gonzáles & Serna, 2003; Serna & Pérez-

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

Gonzáles, 2003), none has investigated the generalization of contextually controlled equivalence classes on the basis of verbal behavior. An additional aim of the present research was therefore to investigate this possibility. During both experiments, stage 1 generalization testing investigated whether previously established contextual stimuli would permit categorization (i.e., differential selection) of formally unrelated novel pictorial stimuli in the absence of additional reinforcement training. During Experiment 1, a majority of participants selected rhyme comparisons regardless of contextual stimuli suggesting that, for these participants, the rhyme basis for responding that had controlled performance during emergent testing was still functional. A number of other participants during Experiment 1, however, and all participants during Experiment 2, differentially selected rhyme and nonrhyme comparisons in the presence of the rhyme and nonrhyme contextual stimuli, respectively, suggesting that contextually controlled categorization of novel stimuli occurred as a result of participants' verbal behavior (i.e., the rhyme contextual stimulus controlled selection of rhyme stimuli and the nonrhyme contextual stimulus controlled selection of nonrhyme stimuli).

Stage 2 generalization testing suggested further links between participants' verbal behavior and generalized class formation. During both experiments, stimulus compounds were presented as samples, each composed of a pair of novel pictorial stimuli whose names either rhymed with each other or did not. Comparison stimuli were always two colored keys (the rhyme key and the nonrhyme keys), each of the same color as one of the contextual stimuli. Although, during Experiment 1, only a minority of participants differentially selected rhyme and nonrhyme keys in response to rhyme and nonrhyme compounds respectively, all participants performed in this way during Experiment 2. This finding indicates firstly that previously established contextual stimuli can enter into classes of novel compound stimuli on the basis of participants' verbal behavior, and additionally provides further evidence of the links between verbal behavior, categorization, and contextual control (cf. Pérez-Gonzáles & Serna, 2003; Rehfeldt, 2003; Serna & Pérez-Gonzáles, 2003).

Although a number of studies have investigated categorization resulting from the merger of equivalence and open-ended stimulus classes on the basis of shared physical similarities (e.g., Adams et al., 1993; Fields et al., 1997; Galizio et al., 2004), primary stimulus generalization alone is not sufficient to account for many examples of categorization (Griffee & Dougher, 2002). As these authors point out, category labelling can permit class expansion in the absence of training with specific exemplars. For example, once a stimulus is labelled as a category member, it can take on the functions shared by other members of that category. A similar view has been put forward by researchers who have suggested that naming, both common and intraverbal, is a prerequisite for categorization to occur (Dugdale & Lowe, 1990; Horne & Lowe, 1996, 1997; Horne et al., 2004; Lowe et al., 2002)). Although the present results do not indicate the necessity of naming for categorization, the findings of stage 1 and 2 generalization testing do provide evidence of its sufficiency, demonstrating that, in the absence of reinforcement, differential selection of formally unrelated novel stimuli can occur on the basis of participants' verbal behavior.

The extent to which emergent human behavior is verbally controlled remains a key theoretical issue within the experimental analysis of behavior, as Arntzen (2004) has observed. Although, according to Horne & Lowe (2000, p. 142), ''the naming account has crucial implications for the distinction drawn between contingency-shaped and rule-governed behavior'', a number of studies have concluded that verbal behavior is neither necessary nor sufficient for equivalence-class formation (e.g., Green, 1990; Lazar, Davis-Lang, & Sanchez, 1984; Sidman et al., 1986). Nevertheless, a growing body of evidence supports the view that verbal behavior, and naming in particular, is of critical importance in understanding the phenomena of equivalence and other higher-order human behavior (e.g., Bentall et al., 1993; Dickins et al., 1993; Eikeseth & Smith, 1992; Hayes & Hayes, 1992; Horne et al., 2004; Lowe et al., 2002; Mandell & Sheen, 1994; Randell & Remington, 1999; Smith et al., 1996). The results of the present research both support and extend these findings, demonstrating the functionality of participants' verbal behavior during the for-

mation of contextually controlled equivalence classes and their generalization to include novel stimuli, stimulus compounds, and previously established contextual stimuli. The results also indicate that the interaction of preexisting verbal repertoires and experimental procedures can result not only in rapid and extensive stimulus-class expansion but also in interference with the intended effects of experimental procedures commonly employed to establish stimulus classes.

Finally, the present findings imply an important consideration in the interpretation of existing research into emergent human behavior. Although the majority of such studies have employed verbally able adults as participants, only a few have attempted to evaluate their participants' verbal behavior, covert or otherwise, during experimentation. Although the difficulties of assessing such behavior have been widely noted (e.g., Perone, 1988; Shimoff, 1984, 1986), these should not preclude its consideration. The current research illustrates the extent to which participants' verbal behavior can affect both the formation and generalization of contextually controlled equivalence classes. Although rhyme was chosen as a salient means by which to highlight these effects (Goswami & Bryant, 1990), it would seem unwarranted to assume that verbally able participants do not name some, or even all, of the ''hard-to-name'' stimuli with which they are commonly presented, and that this naming is not functional. It would therefore seem prudent to suggest that future research would benefit from an awareness that participants' verbal behavior can act as a powerful moderator, and indeed source, of higher-order human behavior.

REFERENCES

Adams, B. J., Fields, L., & Verhave, T. (1993). Effects of test order on intersubject variability during equivalence class formation. *The Psychological Record, 43*, 133–152.

Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record, 54*, 275–291.

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.

Carr, D., & Blackman, D. E. (2001). Relations among equivalence, naming, and conflicting baseline control. *Journal of the Experimental Analysis of Behavior, 65*, 245–247.

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.

Dougher, M., Perkins, D. R., Greenway, D., Koons, A., & Chiasson, C. (2002). Contextual control of equivalence-based transformation of functions. *Journal of the Experimental Analysis of Behavior, 78*, 63–93.

Dube, W. V., & Hiris, E. (1997). Match to sample program (version 10.18) [Computer software]. Waltham, MA: E. K. Shriver Center.

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: Lawrence Erlbaum Associates.

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., Matneja, P., Varelas, A., & Belanich, J. (2003). Mutual selection and membership in open-ended classes: Variant-to-base and base-to-variant testing. *The Psychological Record, 53*, 287–311.

Fields, L., & Reeve, K. F. (1991). A methodological integration of generalized equivalence classes, natural categories, and cross-modal perception. *The Psychological Record, 51*, 67–87.

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., Reeve, K. F., Matneja, P., Varelas, A., Belanich, J., Fitzer, A., & Shamoun, K. (2002). The formation of generalized categorization repertoire: Effect of training with multiple domains, samples, and comparisons. *Journal of the Experimental Analysis of Behavior, 78*, 291–313.

Gagné, C. L. (2002). The competition-among-relations-in-nominals theory of conceptual combination: Implications for stimulus class formation and class expansion. *Journal of the Experimental Analysis of Behavior, 78*, 551–565.

Galizio, M., Stewart, K. L., & Pilgrim, C. (2004). Typicality effects in contingency-shaped generalized equivalence classes. *Journal of the Experimental Analysis of Behavior, 82*, 253–273.

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.

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229–262). New York: Plenum Press.

Griffee, K., & Dougher, M. J. (2002). Contextual control of stimulus generalization and stimulus equivalence in hierarchical categorization. *Journal of the Experimental Analysis of Behavior, 78*, 433–447.

Harrison, R. J., & Green, G. (1990). Development of conditional and equivalence relations without differential consequences. *Journal of the Experimental Analysis of Behavior, 54*, 225–237.

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

Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record, 48*, 81–110.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65*, 185–243.

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. (2000). Putting the naming account to the test: Preview of an experimental program. In J. C. Leslie & D. E. Blackman (Eds.), *Experimental and applied analyses of human behavior* (pp. 127–148). Reno, NV: Context Press.

Horne, P. J., Lowe, C. F., & Randle, V. R. L. (2004). Naming and categorization in young children: II. Listener behavior training. *Journal of the Experimental Analysis of Behavior, 81*, 267–288.

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.

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

Lowe, C. F., Horne, P. J., Harris, F. D. A., & Randle, V. R. L. (2002). Naming and categorization in young children: Vocal tact training. *Journal of the Experimental Analysis of Behavior, 78*, 527–549.

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.

Meehan, E. F., & Fields, L. (1995). Contextual control of new equivalence classes. *The Psychological Record, 45*, 165–182.

Pérez-Gonzáles, L. A., & Serna, R. W. (2003). Transfer of specific contextual functions to novel conditional discriminations. *Journal of the Experimental Analysis of Behavior, 79*, 395–408.

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.

Plaud, J. J. (1995). The formation of stimulus equivalences: Fear-relevant versus fear-irrelevant stimulus classes. *The Psychological Record, 45*, 207–222.

Plaud, J. J., Gaither, G. A., Franklin, M., Weller, L. A., & Barth, J. (1998). The effects of sexually explicit words on the formation of stimulus equivalence classes. *The Psychological Record, 48*, 63–79.

Plaud, J. J., Gaither, G. A., Weller, L. A., Bigwood, S. J., Barth, J., & von Duvillard, S. P. (1998). Rational-emotive behavior therapy and the formation of stimulus equivalence classes. *Journal of Clinical Psychology, 54*, 597–610.

Randell, T., & Remington, B. (1999). Equivalence relations between visual stimuli: The functional role of naming. *Journal of the Experimental Analysis of Behavior, 71*, 395–415.

Rehfeldt, R. A. (2003). Establishing contextual control over generalized equivalence relations. *The Psychological Record, 53*, 415–428.

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

Roche, B., Barnes-Holmes, Y., Barnes-Holmes, D., Stewart, I., & O'Hora, D. (2002). Relational Frame Theory: A new paradigm for the analysis of social behavior. *The Behavior Analyst, 25*, 75–91.

Schenk, J. J. (1995). Complex stimuli in nonreinforced simple discrimination tasks: Emergent simple and conditional discriminations. *The Psychological Record, 45*, 477–494.

Serna, R. W., & Pérez-Gonzáles, L. A. (2003). An analysis of contextual control of conditional discriminations. *Journal of the Experimental Analysis of Behavior, 79*, 383–393.

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

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*, 127–146.

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.

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.

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.

Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorization, concept learning, and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior, 78*, 237–248.

Received: September 23, 2005
Final acceptance: May 24, 2006