JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2006, **85,** 349–369 NUMBER 3 (MAY)

EQUIVALENCE CLASSES WITH REQUIREMENTS FOR SHORT RESPONSE LATENCIES

Gerson Y. Tomanari, Murray Sidman, Adriana R. Rubio, and William V. Dube

UNIVERSIDADE DE SÃO PAULO (BRASIL), SARASOTA (FL),
AND UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL - SHRIVER CENTER

Five adult humans were tested for emergent conditional discriminations under rapid-responding contingencies. During four-comparison matching-to-sample baseline training (AB and AC), limited-hold contingencies for responding to samples and comparisons were gradually restricted to the shortest duration consistent with at least 95% accuracy and no more than 5% failures to respond. The final limited-hold values were 0.4–0.5 s for samples and 1.2–1.3 s for comparisons; mean response latencies were 0.15–0.28 s for samples and 0.59–0.73 s for comparisons; inter-trial intervals were 0.4 s. With these fast-responding requirements, test blocks presented 72 probe trials interspersed among 72 baseline trials, all without programmed differential consequences. Four equivalence test blocks (BC and CB probes, which tested simultaneously for both symmetry and transitivity) were followed by four symmetry (BA and CA probes) test blocks. Three subjects' results documented emergent performances indicative of equivalence classes despite fast-responding requirements that severely limited the time available for mediating vocal or subvocal responses. For these three subjects, mean latencies were slightly shorter in baseline trials than in probes, and shorter on symmetry than on equivalence probes. These differences, however, were usually less than the differences among mean latencies on the different types of trials within the baseline and probed performances.

Key words: stimulus equivalence, response latencies, naming, pointing, adult humans

Stimulus equivalence has largely been demonstrated in people with normally developed vocal repertoires. These findings have given rise to suggestions that there is a relation between the ability to form equivalence classes and language (Devany, Hayes, & Nelson, 1986), or that verbal responses may facilitate the emergence of equivalence relations (Devany et al., 1986; Green, 1990; Horne & Lowe, 1996; McIlvane & Dube, 1996). Similarly, Wilson and Milan (1995), in discussing the longer latencies shown by their elderly subjects in conditional-discrimination tests for equivalence relations, suggested that the differences from younger subjects might be due to ''age effects on covert mediating behavior, cognitive processes, or declines in processing capacity'' (p. 216).

Alternative explanations for facilitative effects of naming on equivalence relations, however, have been based on standard principles of stimulus discrimination. For example, naming the stimuli during a matching-to-sample task may improve a subject's performance by differentiating the ongoing contingencies and increasing stimulus control (McIlvane & Dube, 1996; K. Saunders & Spradlin, 1989; Sidman, 2000; Stromer & Dube, 1994; Stromer & Mackay, 1996; Torgrud & Holborn, 1989). Others, however, have hypothesized that verbal mediating responses may not only facilitate equivalence but may be necessary for it (Dugdale & Lowe, 1990; Horne & Lowe, 1996; for evidence of equivalence relations in nonhumans, however, see Kastak & Schusterman, 2002; Kastak, Schusterman, & Kastak, 2001; R. Saunders & Green, 1996; Schusterman & Kastak, 1993; Schusterman, Kastak, & Reichmuth, 1997). More extended discussions of verbal mediation and equivalence can be found in Stromer and Mackay (1996), and Stromer, Mackay, and Remington (1996).

In addition, it has been reported that comparison response latencies increase from

---

Some of the data were presented at the 26th and 30th Annual Conventions of the Association for Behavior Analysis (Washington, DC, 2000, and Boston, 2004). A *Research in Progress* report outlined the procedure under development at the beginning of this work (Tomanari, Sidman, Rubio, & Dube, 2000). The authors thank Miriam Garcia-Mijares for her contribution to this manuscript.

This research was partially supported by FAPESP and CNPq grants.

Correspondence may be addressed to Prof. Gerson Y. Tomanari. Department of Experimental Psychology, University of São Paulo. Av. Prof. Mello Moraes, 1721. São Paulo. SP, 05508-900, Brasil (e-mail: tomanari@usp.br).

doi: 10.1901/jeab.2006.107-04

directly taught baseline trials to derived symmetry, transitivity, and equivalence probe trials (Bentall, Dickins, & Fox, 1993; Spencer & Chase, 1996; Wulfert & Hayes, 1988), and when tests for emergent relations involve increasing numbers of nodes (Bentall, Jones, & Dickins, 1998; Fields, Adams, Verhave, & Newman, 1990; Fields, Landon-Jimenez, Buffington, & Adams, 1995; Imam, 2001; Kennedy, 1991; Kennedy, Itkonen, & Lindquist, 1994; Spencer & Chase, 1996). This positive relation between latencies and nodal number also has suggested to some that the formation of equivalence relations involves verbal mediating behavior in the form of stimulus names (e.g., Horne & Lowe, 1996) or problem-solving strategies (e.g., Arntzen & Holth, 2000; Holth & Arntzen, 2000). The hypothesis, sometimes implicit, as in the interpretation of the number of nodes separating stimuli in an emergent conditional relation as ''associative distance'' (Fields et al., 1990; Fields & Verhave, 1987), and sometimes explicit (as in, e.g., Bentall et al., 1993; Lowe & Horne, 1996), has been that the larger the number of nodes, the lengthier the mediating verbal chains (presumably subvocal) must be.

The present study investigated equivalence relations in people who had normally developed verbal repertoires but who were tested under conditions that limited the time available for any vocal or subvocal mediation. That is to say, the tests for emergent conditional discriminations included rapid-responding contingencies. During baseline training, intertrial intervals remained minimal, while the time available for the subjects to respond to sample and comparison stimuli was gradually restricted to the shortest possible values consistent with the maintenance of high accuracies. Imam (2001, 2003) also has reported a limited-hold restriction on the duration of comparison stimuli, but did not limit sample durations and made no attempt to establish minimal durations. In the present experiment, the imposition of limited-hold contingencies on sample and comparison stimulus durations generated the shortest possible response latencies for accurate baseline performances. Subsequent equivalence tests maintained those limited-hold durations. The experiment asked whether new performances indicative of equivalence would emerge even when the subjects had minimal opportunity to engage in subvocal mediating behavior.

METHOD

Subjects

Five subjects were recruited through personal contacts. Three females (CA, EMM, and SU) and 2 males (ALE and PLRA) ranged in age from 19 to 32 years. All had completed high school; ALE, CA, and SU were undergraduate students; EMM and PLRA were administrative assistants.

Setting, Apparatus, and Stimuli

Sessions were conducted in three different laboratories: Subject EMM's sessions at the E. K. Shriver Center, PLRA at the Universidade de São Paulo, and the 3 others at the Universidade Metodista de São Paulo. All experimental rooms were equipped with a chair, table, Apple Macintosh® computer, and a 14-inch color monitor with a touch-sensitive screen (MicroTouch). Computer software presented all experimental events and recorded responses (Dube & Hiris, 1999). The computer measured response latencies with an accuracy of plus-or-minus 0.008 s and reported latencies rounded to the nearest 0.01 s.

Subjects accumulated points by performing the experimental tasks. After sessions, points were exchanged for money at a previously defined rate of one cent per point. Subjects ALE, CA, and SU were paid an extra R$3.50 per session for transportation expenses. Subjects were paid in their native currencies and the amounts paid had roughly equivalent value within both countries. The exchange rate at the time of the study was approximately $1 US = R$3.50.

Experimental stimuli were twelve Greek letters (see Figure 1), approximately 2 cm × 2 cm in size, drawn in black lines and displayed on a white background. One letter was displayed as the sample in the center of the monitor screen. Four letters were displayed as comparisons, inset from the corners of the screen so that the center of each comparison was approximately 4.5 cm from the center of the screen. Adjacent comparison stimuli were approximately 2.5 cm center-to-center vertically and 10.0 cm horizontally. For

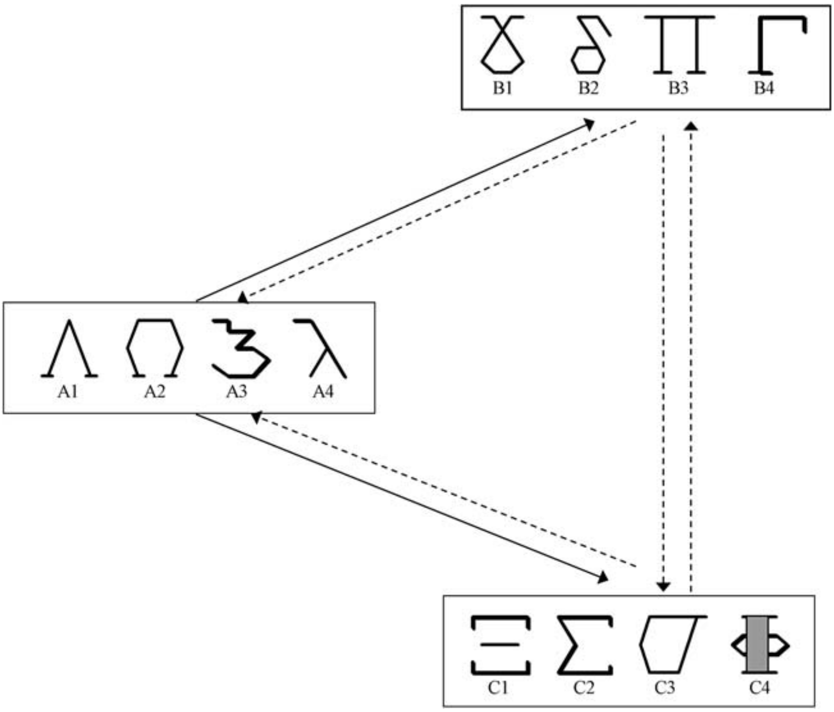

Fig. 1. Greek letters used as experimental stimuli. Solid arrows indicate directly trained AB and AC matching-to-sample tasks. Dashed arrows indicate potential emergent performances.

convenience, experimental stimuli will be designated in this report as three sets (A, B, and C) of four stimuli each (A1, A2, A3, A4; B1, B2, B3, B4; C1, C2, C3, C4). Subjects were not informed of these designations.

General Procedures

Experimental sessions were conducted 5 days per week. At the beginning of the first session, subjects were given an information sheet stating that the general goals of the experiment were to study learning processes, and that the procedures were harmless, the approximate amount of money they would earn per session, and the payment schedule. After subjects read the information, they signed a consent form.

Sessions consisted of blocks of 0-second delayed matching-to-sample trials. Each trial began with a sample stimulus presentation. When the subject touched the sample, it disappeared and four comparison stimuli were immediately and simultaneously displayed. When the subject touched one of the four comparison stimuli, all four immediately disappeared. Differential reinforcement for correct and incorrect choices was programmed in some of the experimental conditions (details below). Following correct choices, the computer beeped and a counter in the top center of the screen advanced by one point (in training conditions) or two points (post-extinction blocks). A 0.4-s intertrial interval followed. During the intertrial interval, the monitor screen was completely white and any touches to the screen had no effect. Incorrect choices were followed only by the intertrial interval.

The number of trial blocks in a session varied, but most sessions consisted of four to six blocks of 144 trials each. Shorter blocks of 40 trials (or 36 or 48 for Subject EMM) were occasionally used to evaluate the subjects' performance, for example, just before a change in the experimental conditions. There was approximately 30–60 s between blocks while the experimenter loaded the computer file for the next block. Within all blocks of trials, each sample-comparison trial type for trained or tested conditional discriminations was presented before any was repeated. Comparison stimuli appeared equally often in each comparison location, and each comparison location was correct equally often. The trial sequence for the first block of trials in each session was always different from the first block in the previous session.

These general procedures were elaborated in four experimental phases, described below. Accuracy was evaluated at the end of every block of trials. The accuracy criterion to advance from one phase to the next was two consecutive blocks with at least 95% correct (137 correct in a 144-trial block, or 38 correct in a 40-trial block), and no more than one error on any type of trial (a specific sample and correct comparison).

Phase I: AB and AC Training

Immediately before the first block of trials in this phase, subjects read the following message on the screen: ''In this experiment, shapes will appear on the computer screen. You can earn points by touching them. The window at the top of the screen shows your score. Try to get the highest score you can. Later, you will be paid 1 cent for each point.'' (Subjects in Brazil read a Portuguese translation.) By pressing ''Continue'' on the touch screen, the subject started the first trial.

In Phase I, using the matching-to-sample procedure described above, subjects were taught by differential reinforcement to perform AB matching, that is to say, to select comparison stimuli B1, B2, B3, and B4 conditionally upon sample stimuli A1, A2, A3, and A4, respectively. When the accuracy criterion was met for AB, AC training was conducted (selecting C1, C2, C3, and C4 conditionally upon A1, A2, A3, and A4, respectively). When the accuracy criterion was met for AC, equal numbers of AB and AC trials were intermixed within the same block of trials until the accuracy criterion was again met.

During Phase I, there was no time restriction for responding, although response latencies always were recorded. Sample-response latencies were measured from the moment the sample appeared to the moment it was touched. Comparison-response latencies were measured from the moment the four comparison stimuli appeared to the moment one of them was touched.

Phase II: Introduction of Limited-Hold Contingencies and Gradual Reduction of Limited-Hold Durations

In the second training phase, a limited-hold contingency was added to the matching-to-sample procedure. Sample and comparison stimulus display durations were restricted, thus establishing a limit for response latencies. If the subject responded within the available time, the trial continued as described above, with differential consequences for correct and incorrect responses. If, however, the subject did not respond to the sample or to a comparison within the time available, then the trial was immediately ended by the removal of all stimuli and initiation of an intertrial interval, and scored as a failure to respond. Incomplete trials were followed by the next trial according to the sequence originally programmed (no correction procedure). Figure 2 outlines the matching-to-sample procedure with the limited-hold contingency.

To determine the shortest possible response latencies that would maintain accurate performance in Phase II, sample and comparison display durations were independently and gradually reduced across blocks of trials. The procedure included four steps:

(a) *Analyze response latencies.* After each block of trials, the sample and comparison response latencies (including both correct and incorrect responses) were analyzed separately. Frequency distributions were constructed with 0.10-s intervals, and cumulative percentages were calculated separately for sample and comparison distributions. For example, consider a hypothetical block of 144 trials that included 140 sample latencies, distributed as 20, 30, 35, 40 and 15 occurrences in 0–0.10 s, 0.11–0.20 s, 0.21–0.30 s, 0.31–0.40 s, and 0.41–0.50 s intervals, respectively. Cumulative percentage distributions for these intervals would be 14%, 36%, 61%, 89%, and 100%.

(b) *Calculate limited-hold time.* The cumulative distribution was used to identify the multiple of 0.10 s that was closest to the 90th percentile of latencies. This value was used for the limited hold in the next block of trials. In the example above, the limited hold for the sample stimulus in the next block would be 0.40 s.

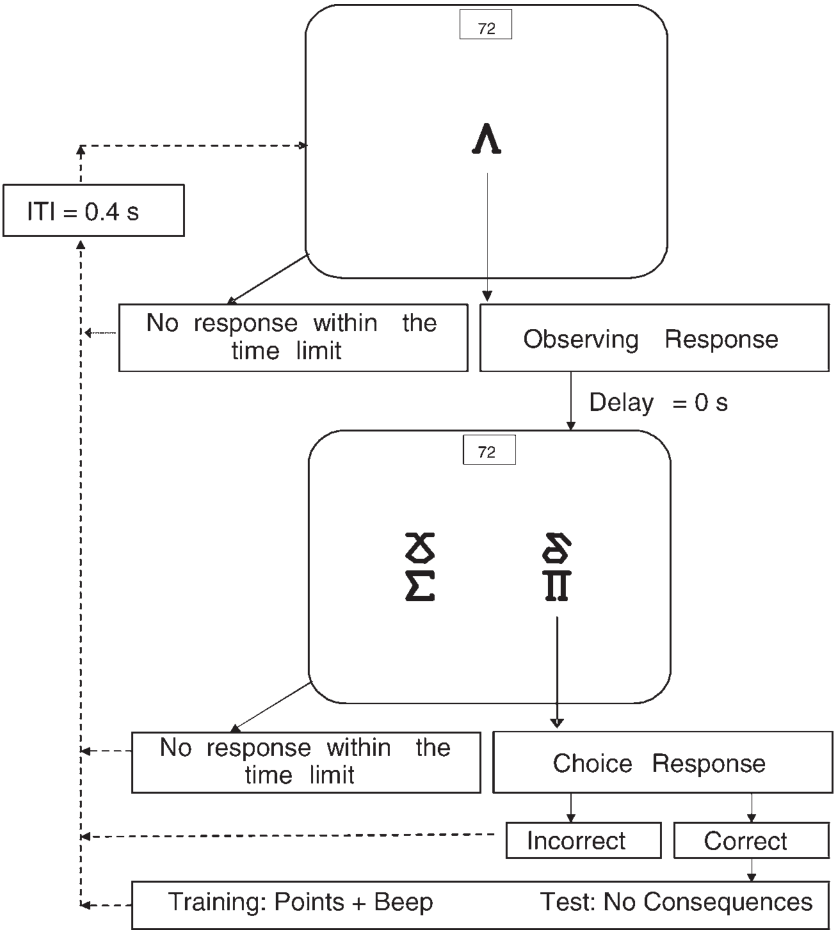

Fig. 2. Diagram of a matching-to-sample trial with limited-hold contingencies for sample and comparison responses.

(c) *Adjust limited-hold duration.* Reductions in the limited hold proceeded only if the subject met the accuracy criterion. Because the goal of Phase II was to produce the shortest response latencies possible without either too many incorrect choices or too many failures to respond, accuracy scores were calculated by dividing the number of correct responses by the sum of correct responses, incorrect responses, and trials with no response at all. If accuracy fell below the criterion for one block of trials, then the limited hold remained the same for the next block of trials. If the accuracy criterion was met in the second block, then limited-hold reduction continued. Otherwise, its duration was increased by 0.10 s.

For Subject EMM, sample and comparison limited-hold durations were manipulated in series. First, the limited hold was applied to samples until the shortest latencies were obtained (see Step d, below). Then, it was applied to comparisons while maintaining the shortest limited hold with the sample. Finally, before ending this condition, the limited hold for the sample was again manipulated as described in Step d.

For all other subjects, limited-hold manipulations alternated between samples and comparisons over successive blocks of trials that met the accuracy criterion. That is to say, if the limited hold for the sample changed, then it was held constant for the comparison, and vice-versa.

(d) *Determine the limited-hold limit.* The limited-hold manipulation process continued for samples and comparisons until two consecutive attempts to reduce each of them resulted in failure to maintain the accuracy criterion. At this point, blocks of trials using the lowest values that previously had resulted in high accuracy were repeated. When the accuracy criterion was again reached, Phase II ended and Phase III began. The lowest limited-hold values determined in Phase II were used throughout the remaining experimental phases.

Phase III: AB and AC without Differential Consequences

In this phase, subjects were given blocks of AB and AC trials with the limited-hold values determined in Phase II, but with procedural extinction. There were no differential consequences for correct and incorrect responses; both correct and incorrect choices were followed immediately by the intertrial interval. Phase III began with one block under extinction. Immediately before any extinction block, subjects were told that ''points would not be delivered at that moment, but the computer would keep recording the responses, and a chance to make up for the points would follow sometime later.'' Extinction blocks were always immediately followed by a block with differential reinforcement, but with two points for each correct choice. Phase III ended when the accuracy criterion was met for four consecutive blocks, two with extinction and two with two-point reinforcers.

Phase IV: Equivalence and Symmetry Tests

Phase IV continued to alternate blocks of extinction trials immediately followed by blocks with two-point reinforcers. In the extinction blocks, however, 72 probe trials substituted for 36 AB and 36 AC trials, with a constraint of no more than three consecutive baseline trials or three consecutive probe trials. In one session, four equivalence-test blocks (BC and CB probes, which tested simultaneously for both symmetry and transitivity) were presented. In the following session, four symmetry-test blocks (BA and CA probes) were run. Each equivalence and symmetry test block consisted of a different sequence of trials.

(a) Combined equivalence test. Equivalence test blocks included 36 BC and 36 CB probes, in addition to 36 AB and 36 AC baseline trials, all with no differential consequences. Because the BC and CB probes tested simultaneously for both symmetry and transitivity, these have been called a combined test for equivalence (Sidman & Tailby, 1982). Choices consistent with conditional relations B1C1, B2C2, B3C3, B4C4, C1B1, C2B2, C3B3, and C4B4 would provide evidence for the emergence of four equivalence classes (A1B1C1, A2B2C2, A3B3C3, and A4B4C4).

(b) *Symmetry test.* Symmetry test blocks included 36 BA and 36 CA probes, in addition to 36 AB and 36 AC baseline trials, all with procedural extinction. In the symmetry probes, choices consistent with conditional relations B1A1, B2A2, B3A3, B4A4, C1A1, C2A2, C3A3, and C4A4 would demonstrate the emergence of symmetrical conditional discriminations consistent with equivalence classes.

Table 1 summarizes the sequence of experimental phases, the numbers of sessions, and the numbers of trials for each subject.

RESULTS

Limited-hold Reduction

Figure 3 shows the mean sample and comparison latencies for blocks of intermixed AB and AC trials. The leftmost set of lines shows Phase I, third step, with no limited-hold contingency. Latencies decreased in Phase I (except for Subject EMM), even though there were no time limits for responding.

The middle set of lines in Figure 3 shows Phase II, when the limited-hold contingency was introduced and gradually restricted. Latencies decreased immediately, as compared with those in Phase I, and continued to reduce gradually across the blocks of trials. Initial Phase-II latencies varied more widely across subjects (for example, compare Subjects EMM and SU) than did final latencies.

Table 2 includes the final sample and comparison limited-hold values for each subject in the last block of trials for Phase II, along with the actual mean latencies at each limited hold. The mean latencies were considerably shorter than the limited-hold values required.

Table 1

Sequence of experimental phases for each subject, including the sequence of sessions (''Ses'') and the total number of trials (''TR'') in each phase. Trials were presented in blocks of 36, 40, or 144. (The sequence of experimental conditions within each phase is not shown.)

	Subjects									
	ALE		CA		EMM		PLRA		SU	
Condition	Ses	TR	Ses	TR	Ses	TR	Ses	TR	Ses	TR
Phase I – AB–AC Training										
AB	1–3	432	1–2	576	1–4	432	1–5	1728	1–3	576
AC		576		432		360		576		432
AB–AC		576		328		432		656		432
Phase II – Limited-Hold Contingencies and Gradual Reduction of Limited-Hold Durations										
AB–AC	3–22	13872	3–19	12160	4–23	9144	6–27	14328	4–19 21–22	11832
Phase III – AB and AC without Differential Consequences										
AB–AC	23	40	20	80	24–25	144	28–32	760	20&23	120
AB–AC/Ext		432		288		288		1440		1008
AB–AC/2x		432		288		288		1440		1008
Phase IVa – Combined Equivalence Session										
AB–AC	24	288	21	512	26–27	96	33	144	24	80
AB–AC/Ext		144		–		–		144		–
BC–CB		576		576		576		576		576
AB–AC/2x		576		576		576		720		576
Phase IVb – Symmetry Session										
AB–AC	25	184	22	184	28–29	96	34	144	25	40
AB–AC/Ext		–		–		–		–		–
BA–CA		576		576		576		576		576
AB–AC/2x		576		576		576		576		576

The rightmost set of lines in Figure 3 shows data for Phase III, when training trials were presented in extinction. Latencies were maintained as in the previous phase.

Figure 4 shows sample and comparison latencies analyzed as frequency distributions within blocks of trials for two representative subjects, ALE and SU. The top row for each subject (''Without LH'') shows the final block of trials in Phase I (third step), just before the limited-hold manipulation started. The middle row (''With LH'') shows the final block of trials in Phase II, when the limited-hold manipulation ended. For each subject, a comparison of the top and middle rows shows that the distributions narrowed and shifted to the left. The bottom row (''With LH (Ext)'') shows the last block of trials in Phase III, with procedural extinction. A comparison of the middle and bottom rows shows that the distributions did not change appreciably with extinction.

Figure 4 also shows that the modes of the distributions in Phases II and III tended to be lower than the time available for responding (shown by the vertical, dashed line). For example, Subject SU's limited-hold limit for choosing a comparison stimulus was 1.3 s, but latencies occurred most frequently between 0.6 s and 0.8 s. Nevertheless, attempts to reduce the limited hold to less than 1.3 s resulted in increases in response failures and thus, failures to meet the accuracy criterion.

Despite the fast-responding requirements, performance during training often was perfectly accurate for entire blocks of trials. Although not instructed to do so, subjects almost always rested their elbow on the table, centered their hand on the computer screen, and moved only one finger to touch the stimuli. There were only occasional disruptions for minor events such as a cough or sneeze.

Performance Accuracy

Figure 5 shows individual subjects' accuracy scores on baseline and probe trials in all four equivalence and symmetry test blocks. All subjects maintained baseline (AB and AC

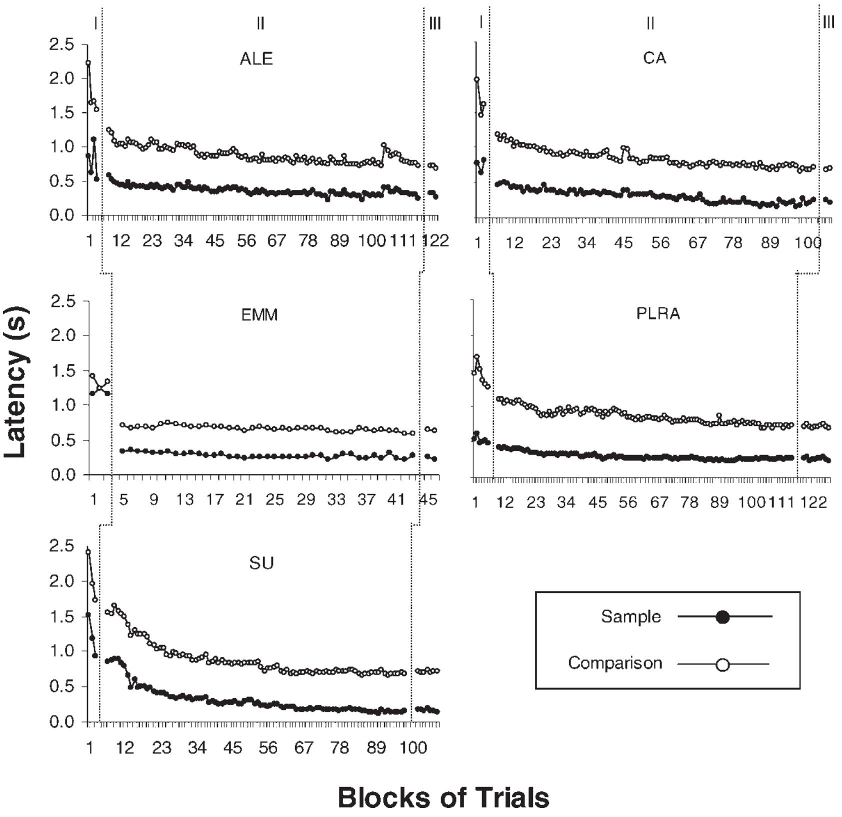

Fig. 3. Mean response latencies to sample (filled circles) and comparison stimuli (open circles) in blocks of AB and AC baseline trials. The leftmost isolated set of lines shows latencies from Phase I, third step, the middle set shows latencies from Phase II, and the rightmost set shows latencies from Phase III. Roman numerals identify phase numbers. See text for details.

trials) accuracy scores greater than 94% in all test blocks.

Results on equivalence and symmetry probes include three types of trial outcomes: touching the comparison that was consistent with the same experimental equivalence class as the sample, touching a comparison that was inconsistent with equivalence, and failures to touch a stimulus during the limited time it was displayed. For convenience, the terms ''correct'' and ''incorrect'' will refer to responses consistent or inconsistent, respectively, with the experimental equivalence classes.

Only correct and incorrect selections are included in the calculation of accuracy scores. Data for failures to respond within the time constraints will be included in the analyses of response distributions presented below (and

Table 2

Sample and comparison limited-hold values for each subject in the last block of trials for Phase II, along with the actual mean latencies at each limited hold.

	Limited-hold values (s)		Mean latencies (s)	
Subject	Sample	Comparisons	Sample	Comparisons
ALE	0.50	1.30	0.25	0.71
CA	0.50	1.30	0.25	0.71
EMM	0.40	1.20	0.28	0.59
PLRA	0.40	1.20	0.26	0.73
SU	0.50	1.30	0.15	0.67

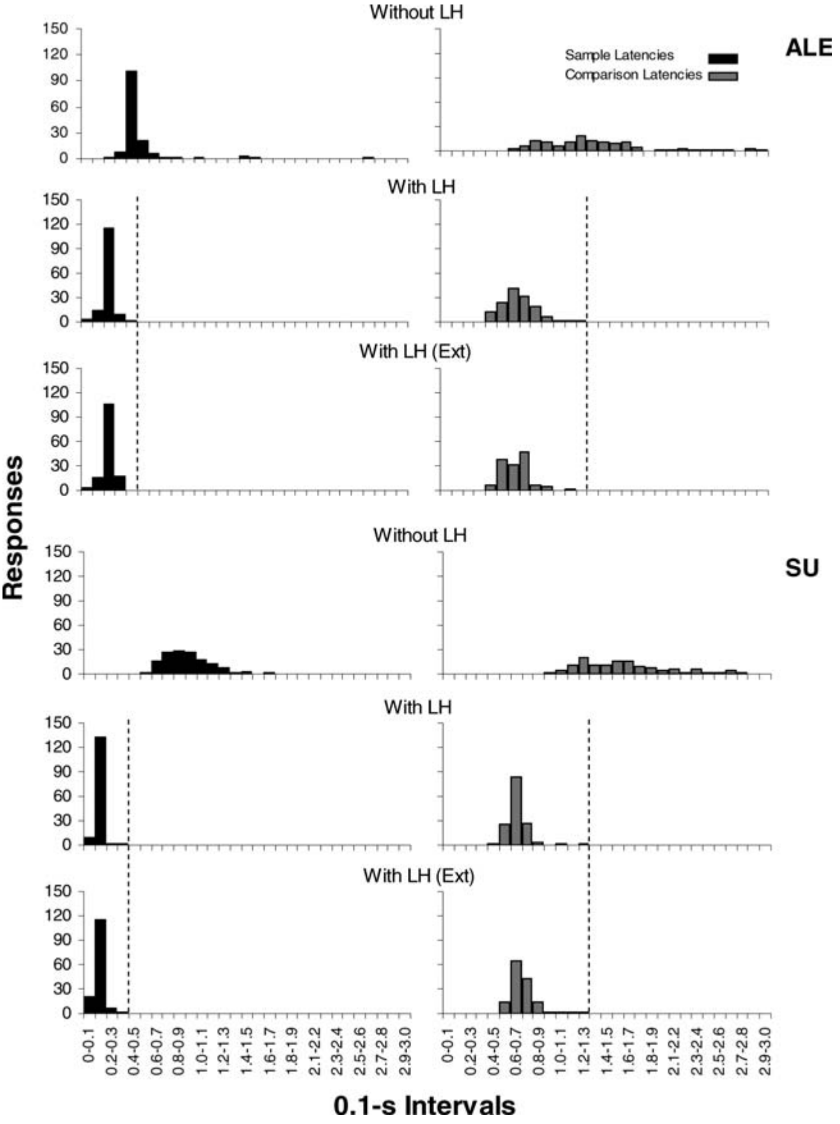

Fig. 4. Frequency distributions of sample and comparison response latencies in the last block of baseline trials for two representative subjects: Without limited hold (top row of data for each subject, Phase I, third step), with limited hold (identified by the vertical dashed lines) and differential reinforcement (middle rows, Phase II), and with limited hold and no differential reinforcement (bottom rows, Phase III).

in Figures 7 and 8). Accuracy scores were calculated on the basis of actual comparison selections because the limited-hold threshold data (above) do not support an assumption that responding and not responding were controlled by the same variables. Rather, the increases in the number of no-response trials that occurred within otherwise highly accurate baselines when the limited-hold contingency became too restricted indicate that nonresponses were controlled by variables that differentiate them from incorrect comparison selections.

For equivalence probes (BC and CB trials), accuracy scores were at least 98% and 94% for Subjects EMM and SU, respectively. For Sub-

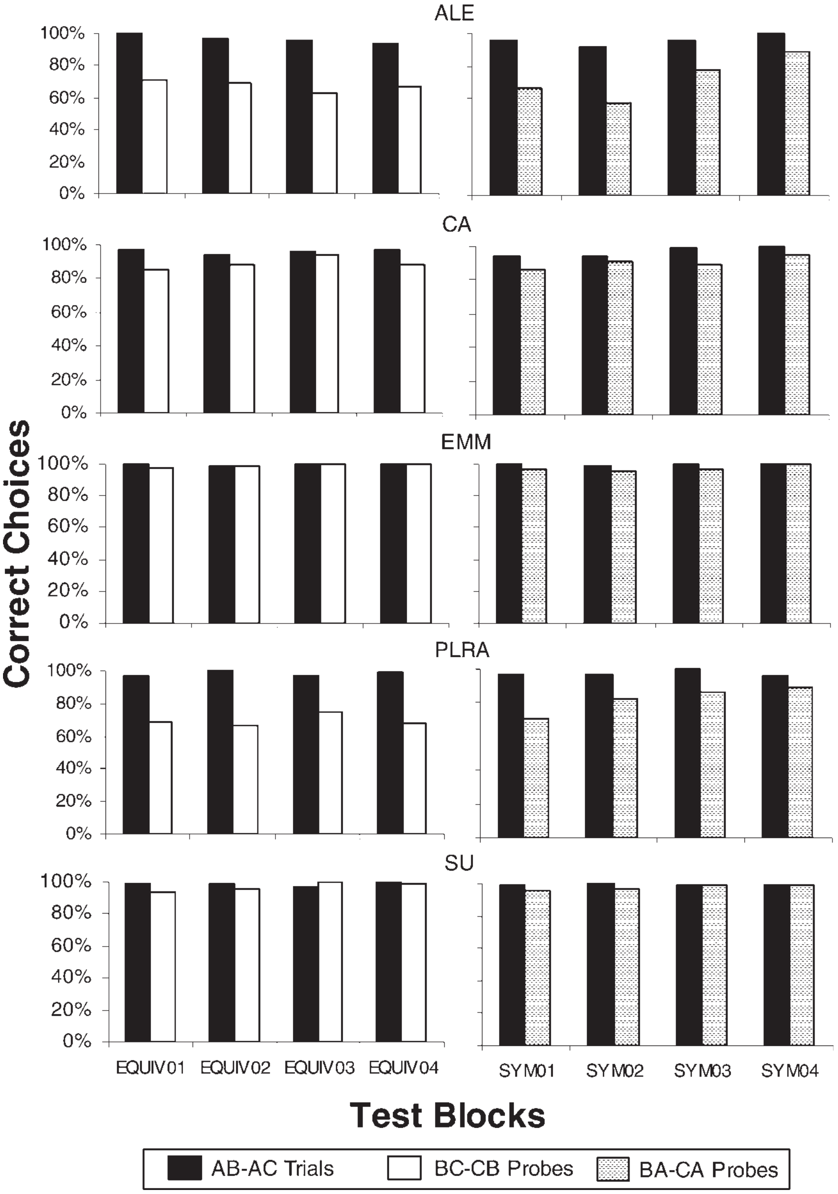

Fig. 5. Accuracy on baseline and probe trials in the four equivalence test blocks and four symmetry test blocks. Equivalence data appear in the left column of graphs and symmetry data appear in the right column of graphs.

ject CA, scores were slightly lower, from 85% (first block) to 94% (third block). For Subjects ALE and PLRA, scores varied within the range of 63% (ALE, third block) to 74% (PLRA, third block).

For symmetry probes (BA and CA trials), the ranges of scores for all subjects were generally similar to those for equivalence probes. One difference between equivalence and symmetry tests was a more pronounced tendency for scores to increase across blocks in the latter.

Initial Probe Results

The first test block in which subjects were exposed to probe trials is particularly important for evaluating the potential for verbal mediation in emergent performances. The

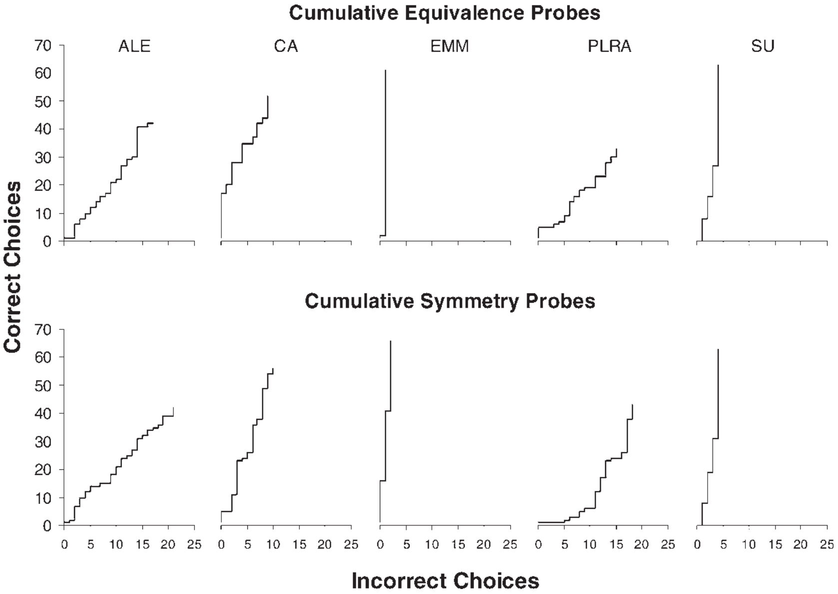

Fig. 6. Cumulative frequency of correct choices (ordinate) plotted against the cumulative frequency of incorrect choices (abscissa) for initial blocks of equivalence (top panel) and symmetry (bottom panel) probes.

initial probe block allowed little or no opportunity for verbal mediation to become elaborated across trials; in the second, third, and fourth probe blocks, it might have been possible for subjects to have rehearsed briefly between blocks. The remaining data, therefore, will come only from the first equivalence and symmetry tests.

Figure 6 shows a trial-by-trial analysis of responses on probe trials for each subject's first equivalence and symmetry test. The cumulative frequency of correct choices on these probe trials is plotted against the cumulative frequency of incorrect choices.

On the first exposure to equivalence probes, Subject EMM made one single incorrect choice very early in the test, followed by correct choices for the rest of the block. Subject SU made only three incorrect choices, all in the first half of the block. Subject CA began the equivalence test with a series of 17 consecutive correct choices (including at least two presentations of each type of probe trial), followed by sequences of correct choices interspersed among incorrect choices. It is on the basis of her early test performance (only 3 incorrect choices in the first 35 comparison responses) that Subject CA's results are included among the positive outcomes for equivalence. Subjects ALE and PLRA made correct and incorrect choices throughout the test block.

The corresponding data for the first exposure to symmetry probes is similar to that for the equivalence probes in that correct choices were cumulatively more frequent for Subjects EMM and SU; intermediate for CA; and less frequent for ALE and PLRA.

The matrices in Figures 7 and 8 analyze responses on each type of probe trial, along with failures to respond, for the first block of equivalence (Figure 7) and symmetry (Figure 8) tests. Each trial type was presented nine times. The matrices show the number of times each comparison stimulus was selected (columns) on trials with each sample stimulus (rows). The highlighted diagonal cells show responses that were scored as correct (consistent with the experimental equivalence classes). Responses in all other cells are incorrect. Trials on which there was no response are listed in the NR column at the right of each matrix. These are trials in which the subject did not respond to the sample (a rare occurrence) or comparison within the time limit.

Most often, Subjects CA, EMM, and SU responded correctly and missed few opportu-

Equivalence

Comparisons

ALE Sample	C1	C2	C3	C4	NR		B1	B2	B3	B4	NR
B1	8	0	0	0	1	C1	8	0	0	0	1
B2	0	5	0	0	4	C2	1	5	0	0	3
B3	1	0	0	7	1	C3	0	1	1	6	1
B4	0	0	0	7	2	C4	0	0	1	8	0

CA	C1	C2	C3	C4	NR		B1	B2	B3	B4	NR
B1	7	0	1	1	0	C1	7	1	0	0	1
B2	0	6	0	0	3	C2	2	3	1	0	3
B3	0	2	6	0	1	C3	0	1	6	0	2
B4	0	0	0	9	0	C4	0	0	0	8	1

EMM	C1	C2	C3	C4	NR		B1	B2	B3	B4	NR
B1	6	0	0	0	3	C1	6	0	0	0	3
B2	0	9	0	0	0	C2	0	6	0	0	3
B3	0	1	7	0	1	C3	0	0	9	0	0
B4	0	0	0	9	0	C4	0	0	0	9	0

PLRA	C1	C2	C3	C4	NR		B1	B2	B3	B4	NR
B1	1	1	1	0	6	C1	6	0	0	1	2
B2	0	1	1	1	6	C2	0	8	0	0	1
B3	0	0	0	6	3	C3	0	1	1	3	4
B4	0	0	0	8	1	C4	0	0	0	8	1

SU	C1	C2	C3	C4	NR		B1	B2	B3	B4	NR
B1	8	0	0	0	1	C1	9	0	0	0	0
B2	0	7	0	0	2	C2	2	6	0	0	1
B3	1	0	7	1	0	C3	0	0	9	0	0
B4	0	0	0	8	1	C4	0	0	0	9	0

Fig. 7. Response matrices showing the number of responses for each type of probe trial during the first block of equivalence tests. Rows correspond to sample stimuli and columns to comparison stimuli. Each type of probe was presented for nine trials. Highlighted diagonal cells show choices consistent with experimental equivalence classes. The column designated NR (No Response) shows number of trials on which the subject failed to respond to the sample or comparisons within the limited-hold duration.

nities to respond, whereas Subjects ALE and PLRA made many errors and often failed to respond. For all subjects, however, errors and/or failures to respond appeared to be controlled by particular stimuli. For example, both in equivalence (Figure 7) and symmetry (Figure 8) tests, Subject CA failed to respond most often when B2 and C2 were samples and was most likely to make errors on symmetry probes when C3 was the sample; Subject EMM most often failed to respond when B1 and C1 were samples; and most of Subject SU's errors occurred when C2 was the sample. In the equivalence test (Figure 7), Subjects ALE and PLRA chose comparison stimulus C4 following samples B4 and B3, and comparison B4 following samples C4 and C3. In the symmetry test (Figure 8), Subject PLRA chose A4 following samples B4 and B3; Subject ALE chose A4 following samples C4 and C3, as

Symmetry

Comparisons

Sample

ALE	A1	A2	A3	A4	NR		A1	A2	A3	A4	NR
B1	7	0	0	0	2	C1	6	0	0	1	2
B2	1	6	1	1	0	C2	0	7	1	0	1
B3	1	0	0	8	0	C3	0	1	2	5	1
B4	0	0	0	7	2	C4	1	0	0	7	1

CA	A1	A2	A3	A4	NR		A1	A2	A3	A4	NR
B1	6	1	1	1	0	C1	9	0	0	0	0
B2	0	6	0	0	3	C2	0	6	0	0	3
B3	0	0	8	0	1	C3	3	1	4	1	0
B4	0	0	0	9	0	C4	1	0	0	8	0

EMM	A1	A2	A3	A4	NR		A1	A2	A3	A4	NR
B1	8	0	0	0	1	C1	8	0	0	0	1
B2	0	9	0	0	0	C2	1	7	1	0	0
B3	0	0	8	0	1	C3	0	0	9	0	0
B4	0	0	0	8	1	C4	0	0	0	9	0

PLRA	A1	A2	A3	A4	NR		A1	A2	A3	A4	NR
B1	5	1	3	0	0	C1	7	0	0	0	2
B2	0	5	2	1	1	C2	0	8	1	0	0
B3	0	1	0	4	4	C3	0	0	6	2	1
B4	1	1	1	4	2	C4	0	0	0	8	1

SU	A1	A2	A3	A4	NR		A1	A2	A3	A4	NR
B1	9	0	0	0	0	C1	8	0	0	0	1
B2	0	8	0	0	1	C2	1	6	0	1	1
B3	0	0	7	0	2	C3	0	0	7	1	1
B4	0	0	0	9	0	C4	0	0	0	9	0

Fig. 8. Response matrices showing the number of responses for each type of probe trial during the first block of symmetry tests. See Figure 7 caption for details.

well as B4 and B3; and Subject PLRA failed to respond most often following samples B1 and B2 on equivalence probes (Figure 7) and B3 on symmetry probes (Figure 8).

Table 3 shows the distribution of trials with failures to respond in probe and baseline trials during subjects' first equivalence probe block. For two subjects, EMM and PLRA, the greatest number of response failures occurred during the first quarter of the block and the majority of those failures were on probe trials. Subjects occasionally failed to respond for two consecutive trials (column labeled ''Two Consecutive''), but consecutive response failures were equally likely to occur at any point within the block, with the exception of Subject EMM, who had two such failures within the first quarter of the trial block. No subject failed to respond for three consecutive trials.

Latencies

Figure 9 shows sample (open circles) and comparison response (filled circles) latencies in the first block of equivalence and symmetry tests. Mean latencies are shown for correct responses in baseline trials during the equiv-

Table 3

Distribution of baseline and probe trials with failures to respond during the first equivalence block.

Subject	Trials	Baseline	Probe	Two consecutive
ALE	1–36	1	4	0
	37–72	3	5	0
	73–108	1	0	0
	109–144	0	4	1
CA	1–36	1	3	1
	37–72	0	0	0
	73–108	0	6	0
	109–144	1	2	1
EMM	1–36	2	6	2
	37–72	0	2	0
	73–108	0	1	0
	109–144	1	1	0
PLRA	1–36	2	7	1
	37–72	3	4	1
	73–108	0	8	2
	109–144	1	5	0
SU	1–36	1	2	0
	37–72	2	2	0
	73–108	2	1	1
	109–144	0	0	0

Note: ‘‘Two Consecutive’’ indicates the number of times the subject failed to respond for two consecutive trials.

alence and symmetry test blocks (‘‘BL E’’ and ‘‘BL S’’, respectively), and in the equivalence and symmetry probe trials (‘‘EQUIV’’ and ‘‘SYM’’, respectively) within those same test blocks. Also shown are the ranges of the latency values.

Each individual subject’s mean sample latencies during the test blocks were virtually identical in baseline trials, equivalence probes, and symmetry probes. Also, mean comparison latencies for baseline trials in equivalence and symmetry test blocks were nearly the same. Within those same test blocks, however, mean comparison latencies for equivalence and symmetry probe trials were slightly longer than those for baseline trials. The differences between baseline and probe latencies for Subject EMM were 0.03 s and 0.07 s on symmetry and equivalence tests, respectively; 0.09 s and 0.07 s for Subject CA; and 0.16 s and 0.14 s for Subject SU.

Mean comparison latencies were slightly longer on equivalence than on symmetry probes for two of the subjects whose data were consistent with equivalence relations; the difference was 0.08 s for Subject EMM and 0.03 s for Subject CA. Subject SU’s mean latency on symmetry probes was 0.03 s longer than on equivalence probes. The range of comparison latencies in baseline and probe trials overlaps for Subjects EMM and CA, but not for SU; symmetry and equivalence probe latencies overlapped considerably for all three subjects.

In summary, comparison latency differences between baseline and probe trials were small but consistent among the subjects whose data were consistent with equivalence relations;

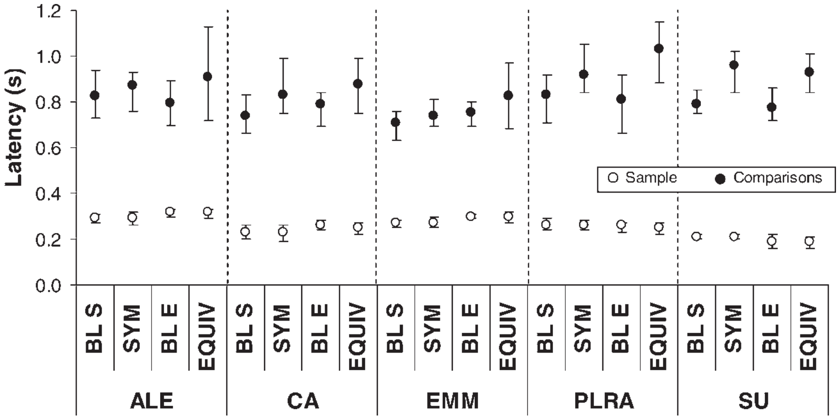

Fig. 9. Sample (open circles) and comparison (filled circles) latencies in the first block of equivalence and symmetry tests. ‘‘BL S’’ indicates correct-choice latencies on baseline trials during the symmetry test block, and ‘‘BL E’’ in the equivalence block. ‘‘SYM’’ indicates mean latencies on correct symmetry probe trials, and ‘‘EQUIV’’ on correct equivalence probes. Vertical lines through each point show the range of latencies for the various trial types within each block.

differences between symmetry and equivalence probes were small and inconsistent; and there was considerable overlap in the ranges of latencies on baseline, equivalence, and symmetry trials.

DISCUSSION

After five subjects learned the baseline matching-to-sample tasks (AB and AC), limited-hold contingencies that reduced sample and comparison durations succeeded in establishing very fast matching performances[1]. Baseline sample and comparison latencies came to be distributed almost completely below the limited-hold values (Figure 4).

Tests for equivalence (BC and CB; combined symmetry and transitivity) and symmetry (BA and CA) were then carried out with the limited-hold contingencies remaining in effect. In the initial probe block, three of the five subjects, CA, EMM, and SU, demonstrated immediate emergence of conditional discriminations consistent with equivalence classes (Figure 5). Although equivalence classes were not demonstrated in all subjects, the positive results for some subjects indicate that the longer time intervals of the typical laboratory procedure are not necessary conditions to produce equivalence (although they may be facilitative).

Covert Naming

Because of the fast-responding requirements, these emergent performances occurred under conditions that severely limited the time available for mediating vocal or subvocal responses. Although naming and other forms of tacting contingencies may undoubtedly exert a facilitating role (Dugdale & Lowe, 1990; Eikeseth & Smith, 1992; Green, 1990; Kendler, 1972; McIlvane & Dube, 1996; Stromer & Mackay, 1996), the present results, for reasons to be elaborated below, support an interpretation that vocal or subvocal mediation is not necessary for equivalence.

Verbal mediation accounts have described covert problem-solving behavior (e.g., ''If A is related to B and A is also related to C, then B must be related to C.'') and two types of mediation by naming responses, common naming and intraverbal naming (Horne & Lowe, 1996). In common naming, it is proposed that subjects learn to produce overt or covert names that are the same for all stimuli related to each other. For example, during baseline training the subject might name A1, B1, and C1 all ''George''; A2, B2, C2 all ''Jerry''; and so forth. When a sample stimulus is presented, the subject names it, and the stimulation produced by this naming response is the controlling stimulus for the comparison selection response. The proposed process is similar for intraverbal mediation, except that each stimulus is assigned a different name and the stimulus–stimulus relations are mediated by overt or covert intraverbal chains (e.g., ''red square, green triangle, blue circle, …'').

Because common naming seems to be the least time-consuming of the mediational accounts, we will consider what would be necessary for it to provide a reasonable explanation for the behavior of CA, EMM, and SU during their initial blocks of probe trials. The sum of Subject CA's average sample and comparison latencies was 1.11 s in the first set of equivalence-probe trials and 1.06 s during the first set of symmetry probes. The sums for Subject EMM were 1.12 s per trial for equivalence probes and 1.01 s per trial for symmetry probes, and for Subject SU they were 1.11 s for equivalence and 1.16 s for symmetry. (These response latencies were briefer than the limited-hold values. A puzzling finding was the inability to maintain the subjects' performances with further reduction of the limited-hold values, even though the mean latencies were less than the possible maximum.)

According to a common naming account, the following series of events must occur, and in the following order: First, the subject observes the sample stimulus and subvocally produces the correct common name. Equivalence probe trials presented sample stimuli that the subject had seen before only as comparisons, in company with three other comparisons (see R. Saunders & Green, 1999, for a more complete discussion of simultaneous versus successive discriminations in conditional discriminations). In symmetry probe trials, the subject previously

---

[1] A real-time video clip showing the performance of one subject in the conditions established at the end of his fast-responding training (Phase II) (i.e., with 0.5 s and 1.3 s as the maximum time to respond to the sample and to one of the four comparisons, respectively) is posted online at: http://www.behavior.org/video/Tomanari_Sidman_VideoClip_40s.ram

had seen those same stimuli as samples, but only in the equivalence probes. In spite of the unfamiliar context and function of the probe samples, subjects did not respond more slowly to those stimuli than to the familiar baseline samples (Figure 9). Thus, subjects must not only produce the name that had preceded and controlled selection of the probe sample stimulus when it had appeared previously as part of a comparison array, they must do so with no alteration of observing duration and manual response latency relative to baseline trials.

The second common-naming event takes place after the comparisons have appeared (and, in this experiment, after the sample has disappeared). Now, the stimulus properties of the covertly produced sample response control observing behavior that locates the one comparison stimulus in the current array of four that had previously been selected conditionally upon the name (equivalence probes), or had previously controlled producing the name (symmetry probes). The subject would have to observe one to four comparisons before hitting the one with the same name as the sample.

Third, the subject responds manually, touching the screen at the location of the named comparison stimulus. All three of these events must happen within approximately 1.1 s, and the subject must be ready to do it again immediately after the intertrial interval, 0.4 s later.

Three questions seem central to the naming issue. First, do the results of this experiment provide convincing evidence that subvocal naming did not occur? No. We present no data showing that such unobservable events did or did not occur.

Second, did the fast-responding contingencies leave enough time for subjects to produce covert verbal behavior? Possibly. We have no data relevant to this question, and thus we can neither confirm nor rule out the occurrence of covert naming. Well-rehearsed subjects might be able to name familiar stimuli at a rate of approximately one per second while performing a conditional-discrimination task. Future research might begin to approach this question by repeating the limited-hold training described in the present paper with an added requirement for overt sample naming. In the present experiment, however, there was certainly not enough time for more elaborate behavior such as problem solving and, with one possible exception, probably not enough time for intraverbal chains.

The possible exception was Subject EMM who failed to respond to the comparison stimuli within the limited-hold duration on 2 baseline trials and 6 probe trials within the first 36 trials of her first probe block. These failures produced eight gaps of 1.6 s each, the maximum comparison display duration plus the ITI that followed. (Because Subject EMM twice failed to respond on two consecutive trials, she may have experienced one or two longer gaps. If, on the second trial of the pair, she had touched the sample but otherwise ignored it, the resulting gap would be 3.5 s, two maximum comparison display durations plus 2 intertrial intervals plus one sample latency of approximately 0.3 s.) EMM was the only subject with a positive outcome on the equivalence tests who failed to respond early in testing on a sufficient number of trials (eight) to allow use of the extra time to engage in a covert intraverbal mediational chain for each of the eight emergent relations that were tested. The possibility that Subject EMM did this cannot be ruled out, although it seems a remote possibility. Figure 7 shows that most of her response failures occurred during trials with samples B1, C1, or C2, and there were no failures at all on trials with samples B2, B4, C3, or C4. Thus, if Subject EMM had used the extra time during response failures to engage in covert verbal behavior, and if that behavior were related to the emergent relations tested on probe trials, then at least some of that behavior had to be unrelated to the sample stimulus for the current trial. Table 3 shows that the two other subjects with positive equivalence outcomes, CA and SU, did not have a sufficient number of response failures early in testing to allow even the remote possibility that they might have used the extra time for covert intraverbal mediation related to the eight emergent relations.

The third and most critical question for a common-naming account is: Did the fast-responding contingencies leave enough time for stimulation produced by any covert verbal naming to control the comparison selection responses? The answer to this question also is not clear. It seems to depend on whether the covert name was generated early enough in the trial to exert conditional stimulus control. If naming behavior occurred late in the trial, concurrently with the comparison selection, then any covert stimulation it produced could not enter into control of the comparison selection response. Thus, a covert naming account seems to require that the sum of three values is not greater than 1.1 s: the covert sample naming response latency, covert naming response duration, and the latency (conditional reaction time) for observing the comparison array and making the comparison selection response.

We were unable to find data in the literature directly relevant to covert sample naming. ERP data in Experiment 1 of Dehaene et al. (2001, suggested by an anonymous reviewer) are consistent with an interpretation that adult subjects instructed to ''name [printed words] in their head'' (p. 752) were able to do so in a passive viewing task with a presentation rate of 0.5 s. Perhaps more relevant, Shatzman and Schiller (2004) reported latencies to the onset of overt naming responses for pictures of high-frequency words with no homonyms. These latencies reached an asymptote of approximately 0.58 s with repeated presentations. Subjects in both the Dehaene et al. and Shatzman and Schiller studies were not engaged in behavior other than naming the stimuli. It seems reasonable to speculate that a context with additional behavioral demands, such as the present experiment's four-choice matching-to-sample task, would produce response latencies at least as long as these. By estimating conservatively that any sample-naming latencies in the present experiment were not shorter than those of Shatzman and Schiller, and, for the sake of argument, estimating a covert naming response duration of zero, the time available for observing behavior and the manual response to the comparison stimulus is approximately 0.52 s (1.1 s minus 0.58 s).

At this point the discussion becomes more difficult. On what basis does one evaluate conditional reaction times to unobservable stimuli? For example, Experiment 2 of Dehaene et al. (2001), conducted with subjects different from those in Experiment 1, presented a semantic conditional reaction time task. Printed words were displayed and subjects pressed a left vs. right hand-held button depending on whether the word was equivalent to a natural or manufactured object. Conditional reaction times were approximately 0.62 s. Relative to the Dehaene et al. procedure, the matching-to-sample task in the present experiment had the additional complexity of a four-stimulus comparison array with stimuli that changed from trial to trial and a manual response location that varied from trial to trial among the four comparison positions. As in the naming discussion above, it seems reasonable to conclude that the additional demands of the present experiment would result in reaction times at least as long as those of Dehaene et al.'s Experiment 2.

Because the Dehaene et al. (2001) reaction time of 0.62 s is greater than the 0.52 s available in the present experiment, one might conclude that there was not enough time for covert names to exert conditional stimulus control. However, proponents of naming accounts seem likely to argue that the controlling stimulus in the Dehaene et al. study was not the printed word, but rather a covert naming response to the printed word and thus the reaction times in that experiment included the time for the covert naming. In fact, we are unable to provide any benchmark for conditional reaction times in adult humans for which unobservable covert behavior can be unambiguously discounted. Response latencies and reaction times become meaningless if the event that initiates the response cannot be observed or measured.

Finally, a common-naming account of the present results would have to assert that it was not necessary to name the comparisons covertly. There was certainly not enough time for comparison naming responses to exert conditional stimulus control over the manual comparison selections. Because three of the stimuli would have names associated with equivalence classes different from that of the sample, subjects would have to name more than one comparison on most trials before producing the name that could serve as a conditional stimulus for a correct response.

Given the discussion above, is it more reasonable to conclude that extremely rapid covert naming behavior exerted stimulus control on each trial in an extended series of associative mediational chains, or is it more reasonable to describe any such activity as collateral behavior that may have accompanied comparison selections? For a behavioral analysis, we think that the latter interpretation is the more reasonable one: Exposure to the contingencies of reinforcement is a sufficient explanation both for the subjects' behavior of selecting comparisons conditionally upon samples, and for any overt or covert verbal behavior that may have accompanied selection responses.

It must be emphasized that the subjects had never seen the probe trials before these tests. They experienced eight new trial types—eight new conditional-discrimination ''problems''—nine times each in their first equivalence test block, and eight more new trial types in their first symmetry test. Even if reinforcement had been programmed during the tests, would the rapid rate of trial presentations have permitted the subjects to learn those new conditional discriminations at all? It certainly would be safe to predict that they could not have learned them so quickly, that is to say, within the first test block. Particularly in the equivalence probes (CB and BC trials), the rapidity of the trial presentations, along with the absence of programmed differential consequences, would perhaps have prevented and would certainly have delayed the development of naming and/or fully detailed vocalization of the baseline and derived stimulus relations that otherwise might have facilitated learning (and perhaps did facilitate learning early in baseline development).

Response Latencies

Although there was no difference in response latencies to sample stimuli on baseline and probe trials, mean comparison latencies on probe trials were slightly longer than those for baseline trials (Figure 9). Might this difference reflect subvocal mediation in these unfamiliar trial types? Previous analyses of emergent performances have shown that mean latencies tend to be longer on probe trials than on baseline trials, increasing with nodal number or with the complexity of the derived relations (Bentall et al., 1993; Dymond & Rehfeldt, 2001; Imam, 2001; O'Hora, Roche, Barnes-Holmes, & Smeets, 2002; Spencer & Chase, 1996; Wulfert & Hayes, 1988). The present findings are in accord with these reports.

The differences obtained here between baseline and probe latencies were, however, quite small: for Subject CA, 0.09 s and 0.07 s on the first symmetry and equivalence tests, respectively; for Subject EMM, 0.03 s and 0.07 s; and for Subject SU, 0.16 s and 0.14 s. In the equivalence probes, especially, these differences seem too small to account for the initial development of any effective subvocal mediation, and perhaps even for any fully developed mediation. In most of the other reported studies, too, the latency differences were small. In those studies that converted latency to speed, the reciprocal of latency (for example, Imam, 2001, 2003; Spencer & Chase, 1996), that conversion exaggerated small latency differences, particularly when the latencies being compared fell below 1.5 s (speed accelerates more and more rapidly as latency decreases, as shown in Figure 10).

The presumed elimination of vocal mediation as a necessary component of the emergent performances that define equivalence relations raises questions about the significance of the small latency increases that seem to occur as derived relations require more and more complex histories. One alternative account, which does not postulate vocal mediation, is that the latency differences reflect different degrees of relatedness among the members of an equivalence class (for example, Fields et al., 1995), and/or that latencies measure the degree of substitutability of stimuli in an equivalence class (Fields et al., 1990; Spencer & Chase, 1996).

The suggestion that equivalence classes can contain elements that are unequally related or differentially substitutable (Fields, 1993) calls into question the logical status of the class concept itself. If presumed class members could be differentially related, what would be the utility of the class concept, whether equivalence or any other type of classes? The concept of classes cannot be thrown away so casually; it provides the pathway through which a contingency-oriented science of behavior can account for the innumerable observations that our behavior often is controlled by physically dissimilar events we have never actually experienced. When differential latencies, or even differences in substitutability are observed, it therefore becomes relevant to question the origin of such differences. Several considerations are pertinent.

First, the members of any class need not share—and probably never do share—all of their features in common. Class membership requires only that they share whatever characteristic defines the class. Elements in any class

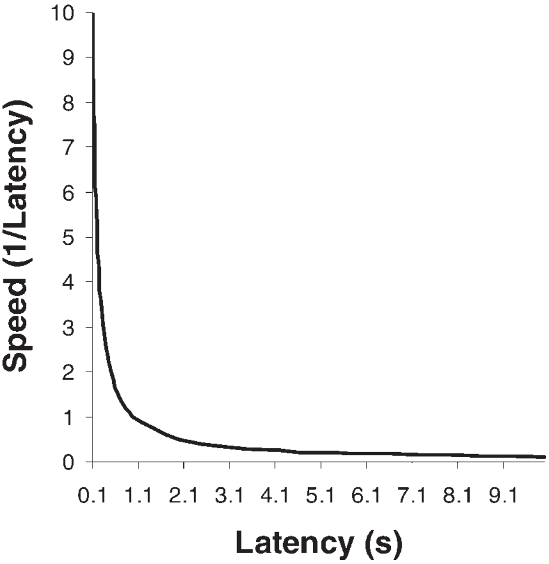

Fig. 10. Speed (1/latency) as a function of latency; speed accelerates more rapidly as latency decreases.

can be expected to be members of other classes as well, with the prevailing membership at any particular time being controlled by contextual features (see Sidman, 1994, for a lengthier discussion of the contextual control of equivalence relations). Whenever known members of an equivalence class seem to be differentially substitutable, control by seemingly irrelevant contextual features is to be suspected (e.g., Stikeleather & Sidman, 1990). In this same direction, with respect to latency differences, Imam (2003) has presented data indicating the possibility that class membership may be established on the basis of response speeds.

Irrelevant differences among class members may also be artifactual, in the sense that they arise from uncontrolled features of the experiment. For example, Imam (2001, 2003, 2006) reported that increasing the number of nodes failed to bring about latency increases (in this instance, speed decreases) when the number of presentations of baseline trial types was carefully equalized. Although he did observe decreases in speed as the complexity of the tested relations increased (that is to say, from baseline to symmetry to transitivity), he also pointed out that those comparisons did not equate the number of trials in each tested relation.

Also to be considered is the number and types of discriminations required. As the complexity, and/or the number of nodes and prerequisite conditional discriminations for tested relations increases, a subject will have to learn and remember more successive discriminations among sample stimuli and simultaneous discriminations among comparisons (R. Saunders & Green, 1999). Even with the relatively small-sized classes (three members) in the present experiment, the analysis of error frequencies and failures to respond (Figures 7 and 8) indicated clearly that the subjects had problems with specific stimulus discriminations. With each additional node multiplying the number of required discriminations, more and more discrimination failures, and possibly accompanying latency increases, are to be expected.

Furthermore, not only were mean latency differences in the present study small, but they usually were less than the differences between the mean latencies on the different types of trials that made up the baseline, symmetry probes, and equivalence probes (Figure 9 summarizes the latency ranges). The present findings suggest, therefore, that latencies may be a fragile measure to distinguish baseline trials from equivalence and symmetry probes. Differences among latencies averaged across different individual types of trials must be carefully examined in relation to the variability within those individual trial-type latencies.

REFERENCES

Arntzen, E., & Holth, P. (2000). Equivalence outcome in single subjects as a function of training structure. *The Psychological Record*, *50*, 603–628.

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.

Bentall, R. P., Jones, R. M., & Dickins, D. W. (1998). Errors and response latencies as a function of nodal number in five-member equivalence classes. *The Psychological Record*, *48*, 93–115.

Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J. F., Poline, J. B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, *4*, 752–758.

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.

Dube, W. V., & Hiris, E. J. (1999). *MTS software documentation.* Waltham, MA: E. K. Shriver Center.

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

Dymond, S., & Rehfeldt, R. A. (2001). Supplemental measures of derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin, 19*, 8–12.

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. (1993). Are stimuli in equivalence classes equally related to each other? *The Psychological Record, 43*, 85–105.

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., Landon-Jimenez, D. V., Buffington, D. M., & Adams, B. J. (1995). Maintained nodal-distance effects in equivalence classes. *Journal of the Experimental Analysis of Behavior, 64*, 129–145.

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48*, 317–332.

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

Holth, P., & Arntzen, E. (2000). Reaction times and the emergence of class consistent responding: A case for precurrent responding? *The Psychological Record, 50*, 305–337.

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.

Imam, A. A. (2001). Speed contingencies, number of stimulus presentations, and the nodality effect in equivalence class formation. *Journal of the Experimental Analysis of Behavior, 76*, 265–288.

Imam, A. A. (2003). Assessing transfer of response speed and nodality via conditional discriminations. *Experimental Analysis of Human Behavior Bulletin, 21*, 1–7.

Imam, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior, 85*, 107–124.

Kastak, C. R., & Schusterman, R. J. (2002). Sea lions and equivalence: Expanding classes by exclusion. *Journal of the Experimental Analysis of Behavior, 78*, 449–465.

Kastak, C. R., Schusterman, R. J., & Kastak, D. (2001). Equivalence classification by California sea lions using class-specific reinforcers. *Journal of the Experimental Analysis of Behavior, 76*, 131–158.

Kendler, T. S. (1972). An ontogeny of mediational deficiency. *Child Development, 43*, 1–17.

Kennedy, C. H. (1991). Equivalence class formation influenced by the number of nodes separating stimuli. *Behavioural 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.

Lowe, C. F., & Horne, P. J. (1996). Reflections on naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65*, 315–353.

McIlvane, W. J., & Dube, W. V. (1996). Naming as a facilitator of discrimination. *Journal of the Experimental Analysis of Behavior, 65*, 267–272.

O'Hora, D., Roche, B., Barnes-Holmes, D., & Smeets, P. M. (2002). Response latencies to multiple derived stimulus relations: Testing two predictions of relational frame theory. *The Psychological Record, 52*, 51–75.

Saunders, K. J., & Spradlin, J. E. (1989). Conditional discrimination in mentally retarded adults: The effect of training the component simple discriminations. *Journal of the Experimental Analysis of Behavior, 52*, 1–12.

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

Saunders, R. R., & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior, 72*, 117–137.

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

Schusterman, R. J., Kastak, D., & Reichmuth, C. J. (1997). What's in a name? Equivalence by any other name would smell as sweet. *Journal of the Experimental Analysis of Behavior, 68*, 252–258.

Shatzman, K. B., & Schiller, N. O. (2004). The word frequency effect in picture naming: Contrasting two hypotheses using homonym pictures. *Brain and Language, 90*, 160–169.

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

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

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

Spencer, T. J., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65*, 643–659.

Stikeleather, G., & Sidman, M. (1990). An instance of spurious equivalence relations. *The Analysis of Verbal Behavior, 8*, 1–11.

Stromer, R., & Dube, W. V. (1994). Differential observing of complex sample stimuli and delayed matching performance: A brief report. *Experimental Analysis of Human Behavior Bulletin, 12*, 17–20.

Stromer, R., & Mackay, H. A. (1996). Naming and the formation of stimulus classes. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 221–250). New York: Elsevier.

Stromer, R., Mackay, H. A., & Remington, B. (1996). Naming, the formation of stimulus classes, and applied behavior analysis. *Journal of Applied Behavior Analysis, 29*, 409–431.

Tomanari, G. Y., Sidman, M., Rubio, A., & Dube, W. V. (2000). Effects of fast-responding requirements on the development of equivalence classes in a matching-to-sample task. *Experimental Analysis of Human Behavior Bulletin, 18*, 21–22.

Torgrud, L. J., & Holborn, S. W. (1989). Effectiveness and persistence of precurrent mediating behavior in delayed matching to sample and oddity matching with children. *Journal of the Experimental Analysis of Behavior*, *52*, 181–191.

Wilson, K. M., & Milan, M. A. (1995). Age differences in the formation of equivalence classes. *Journal of Gerontology*, *50B*, 212–218.

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 September 22, 2004
Final acceptance October 31, 2005