JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1996, 65, 643–659 NUMBER 3 (MAY)

SPEED ANALYSES OF STIMULUS EQUIVALENCE

THOMAS J. SPENCER AND PHILIP N. CHASE

WEST VIRGINIA UNIVERSITY

The functional substitutability of stimuli in equivalence classes was examined through analyses of the speed of college students' accurate responding. After training subjects to respond to 18 conditional relations, subjects' accuracy and speed of accurate responding were compared across trial types (baseline, symmetry, transitivity, and combined transitivity and symmetry) and nodal distance (one- through five-node transitive and combined transitive and symmetric relations). Differences in accuracy across nodal distance and trial type were significant only on the first tests of equivalence, whereas differences in speed were significant even after extended testing. Response speed was inversely related to the number of nodes on which the tested relations were based. Significant differences in response speed were also found across trial types, except between transitivity and combined trials. To determine the generality of these comparisons, three groups of subjects were included: An instructed group was given an instruction that specified the interchangeability of stimuli related through training; a queried group was queried about the basis for test-trial responding; and a standard group was neither instructed nor queried. There were no significant differences among groups. These results suggest the use of response speed and response accuracy to measure the strength of matching relations.

Key words: response speed, stimulus equivalence, nodal relations, matching to sample, key press, college students

Research on stimulus equivalence has focused on the emergence of responding to reflexive, symmetric, and transitive relations among stimuli. Reflexivity is demonstrated by generalized identity matching. Symmetry is demonstrated by sample and comparison stimulus interchangeability. Transitivity is demonstrated by responding to the indirect relation between two or more stimuli through their direct relation to a common stimulus. For instance, a matching-to-sample (MTS) procedure might train subjects to select B in the presence of A and C in the presence of B, where A, B, and C are arbitrary stimuli. Given this training, responding would demonstrate symmetry if subjects selected A in the presence of B and B in the presence of C. Transitivity would be demonstrated if they selected C in the presence of A. Often subjects are also tested on trials that combine symmetric and transitive properties (combined trials). For example, subjects might select A in the presence of C, based on the BA and CB symmetric and the AC transitive properties.

The emergence of these performances is often marked by a transition from inaccurate to accurate or class-consistent test-trial responding. (For ease of exposition, the term *accuracy* will refer to both correct responses on trained match-to-sample trials and class-consistent responses on probe trials.) After subjects respond in accordance with the tested reflexive, symmetric, transitive, and combined relations, the stimuli are described as a class of functionally substitutable, equivalent stimuli.

Despite this description, potential limits to the functional substitutability of equivalent stimuli have been identified. For example, when the baseline conditional discriminations of established equivalence classes were altered in Pilgrim and Galizio's study (1990), subjects responded in accordance with new symmetric relations but not with new transitive relations; rather, they continued to respond in accordance with the transitive relations that were consistent with the originally established equivalence classes. Thus, changes in the relations that defined the class did not uniformly affect the derived symmetric and transitive relations.

This experiment was part of the first author's doctoral dissertation at West Virginia University. The research was supported by grants from the Psychology Alumni Fund and the University's Office of Academic Affairs and Research. Tom Spencer is now with The Continuous Learning Group, Inc. The software used to run the experiment was developed by Bill Dube at the E. K. Shriver Center, with support from NICHD Grants HD17445, HD24317, and HD22218.

Correspondence and requests for reprints should be sent to Philip N. Chase, Department of Psychology, West Virginia University, P.O. Box 6040, Morgantown, West Virginia 26506-6040.

Previous research has also shown differences in response accuracy on symmetry, transitivity, and combined trials during tests of stimulus control (e.g., Bush, Sidman, & de Rose, 1989). Accurate responding to symmetry trials sometimes emerges before responding to either transitivity or combined trials during repeated tests of matching to sample. Moreover, differences in accuracy of responding to multinode transitivity and combined trials have been shown to be inversely related to nodal distance (Fields, Adams, Verhave, & Newman, 1990). Nodes are stimuli linked by training to at least two other stimuli (Fields et al., 1990). For example, after AB, BC, and CD training, AC would be a one-node (B) transitive relation and AD would be a two-node (B and C) transitive relation. Thus, nodal distance refers to the number of stimuli that link two stimuli. As nodal distance increases, accuracy of responding decreases (Fields & Verhave, 1987). These results suggest that the degree of functional substitutability of equivalent stimuli is differentially distributed across trial types and nodal distance.

Wulfert and Hayes (1988) reported another difference between the substitutability of stimuli related through equivalence classes. Subjects' latencies of matching on baseline and symmetry trials differed significantly from their latencies on combined trials, even though accuracy on baseline, symmetry, and combined trials was similar. Response latencies also differentiated correct test-trial responses in a study by Bentall, Dickins, and Fox (1993). Subjects' response latencies were shorter when they were required to match symbols that had been directly related through training than when they were required to match stimuli that had not been related through training.

These analyses of response latency suggest that our ability to predict and control responding to the relations among stimuli in equivalence classes might be enhanced by measures that include time as a dimension of behavior. Measuring time can provide a more sensitive measure of performance than accuracy analyses alone, because temporal differences in responding may be evident even after response accuracy has stabilized (Johnson & Layng, 1992; Lindsley, 1960; Wulfert & Hayes, 1988). Like other detailed measures of variability (e.g., interresponse times, generalization gradients, frequency distributions), differences in latency of responding to stimuli in an equivalence class may provide a more precise description of the relations.

The purpose of this study, therefore, was to characterize baseline and test-trial responding through temporal analyses of responding. Speed of responding was compared across trial types (baseline, symmetry, transitivity, and combined, cf. Wulfert & Hayes, 1988) and nodal distance (one- through five-node transitive and combined relations, cf. Fields et al., 1990; Fields & Verhave, 1987; Fields, Verhave, & Fath, 1984). Speed was defined as the inverse of the latency to respond to the comparison stimuli (see Baron, 1985, for a justification of speed over latency).

In addition to analyzing response speed, this study looked at the effects of instructions across groups of subjects. The literature suggests that instructional control may eliminate differences in accuracy of emergent performance (Ayllon & Azrin, 1964), and therefore may eliminate differences in speed. Consequently, one group received an instruction that specified the substitutability of stimuli related through training. This instruction was designed to facilitate immediate accuracy on tests for emergent relations and potentially eliminate any concomitant differences in response speed. Self-instructions have been shown to facilitate performance in much the same way as experimenter-provided instructions (Rosenfarb, Newland, Brannon, & Howey, 1992). Therefore, during each test block, a second group of subjects was queried intermittently about the basis for test-trial responding (cf. Rosenfarb et al., 1992). A third group of subjects served as a standard MTS group and was neither instructed nor queried.

METHOD

Subjects

Twelve female college students served as subjects. They were hired through a subject recruitment board in the Psychology Department at West Virginia University. Interested students wrote their names and telephone numbers on the announcement.

Subjects were required to sign an informed consent agreement that specified the frequency and duration of their participation in the experiments and the method of payment for such participation. Daily sessions were broken into blocks of trials separated by short breaks, and sessions lasted between 1 and 2 hr (three to seven blocks of trials).

Subjects earned 1 point for each trial correctly completed. At the conclusion of the subjects' participation, points were exchanged for money at the rate of $0.01 per point. After each training block of trials, subjects were told the total number of points earned. During test sessions, however, subjects were not told how many points they had earned, but they were told that those points would be exchanged for money at the conclusion of their participation. Subjects also earned a bonus of $2.00 for every hour of participation.

Apparatus

Stimuli were displayed on the screen (19 cm by 14 cm) of a Macintosh® computer. The display consisted of five locations of white squares (4.5 cm by 4.5 cm) on a gray field. One location, designated for sample stimuli, was the center of the screen. The other four locations, designated for comparison stimuli, were the four corners of the screen.

Response Mode

Subjects selected sample and comparison stimuli by moving the computer's mouse to place the cursor on a stimulus location. Then one press of the mouse button registered the subject's choice. There were no programmed consequences for button presses that occurred when the cursor was outside the boundaries of a stimulus location. The computer automatically recorded the comparison choice and its latency. Latency was defined by the interval between selecting the sample stimulus and selecting a comparison stimulus. The computer then calculated whether the response was correct, and the experimenter converted the latency to a speed score.

Stimuli

Figure 1 illustrates the stimuli (approximately 2.5 cm by 2.5 cm) that were used. Each of the stimuli was assigned a letter and a number. The letters designate sets of comparison stimuli. For example, when B1 served

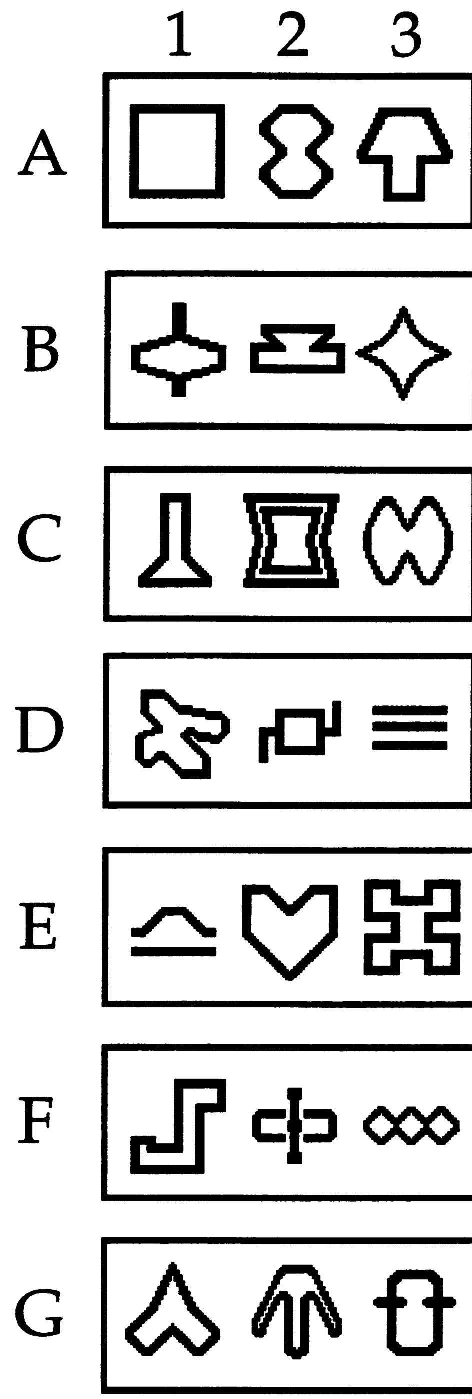

Fig. 1. The stimuli used. The letters designate sets of comparison stimuli. The numbers designate stimulus classes from which symmetric, transitive, and equivalence relations may emerge.

as a comparison stimulus, B2 and B3 also served as comparison stimuli on that trial. The numbers designate stimulus classes from which symmetric, transitive, and combined relations may emerge (e.g., teaching A1B1 may yield B1A1, and teaching A1B1 and B1C1 may yield A1C1 and C1A1). Subjects were not given these letters and numbers.

Teaching Procedures

Matching to sample. Each trial began with the presentation of a sample stimulus in the center of the screen. Selecting the sample stimulus resulted in the presentation of three comparison stimuli. All other responses had no programmed consequences. Across trials, the locations of the comparison stimuli were varied unsystematically in the four corners of the screen. If reinforcement was not precluded on a trial (see Testing below), selecting the correct comparison stimulus produced a 1-s display of the word "correct," advanced a point counter, and initiated a 1.5-s intertrial interval (ITI). Selecting an incorrect comparison stimulus, including the blank stimulus location, terminated the trial, darkened the screen for 1 s, and initiated the ITI. The point counter was not visible to the subject while she was responding to a block of trials; rather, the subject's earnings were displayed on the screen at the end of each block. Responses made during the ITI reset the ITI timer so that the next trial did not start until 1.5 s had passed without a response.

Pretraining. The subjects were initially exposed to the apparatus via 26 trials of three-choice MTS. The task involved matching upper case to lower case English letters. At the start of the first trial, the experimenter asked the subject if she had ever used a mouse. If she had not, the experimenter demonstrated how to move the cursor around the screen with the mouse and click the mouse button. Then the experimenter demonstrated how to complete the correct sequence of responses for a trial; that is, the experimenter first selected the sample stimulus and then selected the correct comparison stimulus. This produced the word "correct" on the screen. The experimenter informed the subject that when "correct" was displayed, a point counter advanced 1 point, which was equivalent to $0.01.

On the second trial, the experimenter demonstrated an incorrect choice. The experimenter first selected the sample stimulus and then selected an incorrect comparison stimulus, which resulted in the darkening of the screen. At this time, the experimenter informed the subject that the screen would turn black for 1 s following incorrect choices, and no points would be earned for incorrect choices.

Throughout the training, instructions were kept to a minimum (see description of the instructed group for an exception). If the subject asked the experimenter a question, the experimenter either repeated the demonstration or instruction or stated that it was up to the subject to figure out how to earn points.

The subjects were required to complete the remaining 24 trials with no errors. All 12 subjects did so.

Training. Subjects learned six sets of one-to-one relations between stimuli via matching-to-sample procedures (A1B1, A2B2, A3B3; B1C1, B2C2, B3C3; C1D1, C2D2, C3D3; D1E1, D2E2, D3E3; E1F1, E2F2, E3F3; F1G1, F2G2, F3G3). One set of three relations was taught at a time, beginning with A1B1, A2B2, and A3B3. Subsequent relations were taught in the order listed above. Each block of training trials consisted of 48 trials. All 48 of these trials were training trials during AB training. Twenty-four training trials and 24 baseline trials (trials with previously trained relations) were randomly intermixed in each subsequent training stage (see Table 1).

Table 1

Number and type of training trials during each stage of training.

Training stage	Number of trials per block					
	AB	BC	CD	DE	EF	FG
1. AB	48					
2. BC	24	24				
3. CD	12	12	24			
4. DE	8	8	8	24		
5. EF	6	6	6	6	24	
6. FG	3	3	3	6	9	24

Standard accuracy criterion. The accuracy criterion necessary to advance from one stage of training to another was 90% or greater accuracy on each of two consecutive blocks of trials with no more than one error on any one relation per block.

Baseline maintenance. Because testing was conducted under extinction conditions, subjects were required to demonstrate accurate performance on the 18 trained relations under extinction conditions before testing began (baseline maintenance). The criterion necessary to advance to testing was 90% or greater accuracy on five consecutive blocks of extinction trials with no more than one error on any one trial type per block.

Before the onset of each baseline maintenance block (and each test block, as described below), the following note was displayed on the computer:

> Today you will be given no feedback regarding the accuracy of your responses. "Correct" will not be displayed when you make a correct choice, and the screen will not turn black when you make an incorrect choice. You will not be told how many points you've earned, but I'll still keep track of your points and exchange them for money at the end of the study.

Testing Procedures

Before the onset of testing, subjects were assigned to one of the three groups. The first 11 subjects were matched on the basis of the number of training blocks required to complete the training and were then randomly assigned to the groups. The last subject to finish training (KP) was assigned to the query group to produce an equal number of subjects in each group.

Instructed group. The 4 subjects in the instructed group were given an instruction that prescribed accurate test-trial performance. This instruction was adapted from a description of MTS performance in Fields et al. (1984):

> Here is a general strategy that may help you to (a) decide what to choose and (b) earn as many points as possible.
>
> If you learn that DOG = CANIS, and CANIS = PERRO, then DOG, CANIS, and PERRO are interchangeable: DOG, CANIS, and PERRO all can be matched to each other on the computer, even though you may have never matched DOG to PERRO or PERRO to DOG. In the same way, if you learn that DOG = CANIS, and CANIS = PERRO, and PERRO = HUND, then DOG, CANIS, PERRO, and HUND are interchangeable: DOG, CANIS, PERRO, and HUND all can be matched to each other on the computer, even though you may have never matched DOG to PERRO or HUND and HUND to DOG or PERRO.
>
> The rule to remember and follow is: The number of symbols (e.g., word, object, picture) that can be in a group of equal or interchangeable symbols is unlimited. Each group of equal symbols consists of symbols related directly by being matched to each other, or related indirectly by both being matched to another symbol or symbols.
>
> Press "enter" to continue.

The computer then displayed the following text:

> Although you will not be tested with the words DOG, CANIS, PERRO, and HUND on the computer, the general strategy described here can be applied to what you have already learned.
>
> Please write a summary of the rule on the card provided.

Subjects were required to write a summary of the rule that included a reference to the interchangeability of stimuli related indirectly during training. All 4 of these subjects did so. The subjects were also asked to write a summary of the rule several times during testing. When the combined trials appeared on the test, "Please write a summary of the rule" was displayed on the computer screen after the 9th, 18th, 27th, 36th, and 45th test trials of each block. This procedure was also used when transitive relations were tested. The same prompt was also displayed when symmetric relations were tested, except that it was displayed only three times per block, after the 6th, 12th, and 18th test trials.

The total number of times each subject was prompted to write a summary of the rule depended on the number of trials required to reach criterion performance (see Standard Accuracy Criteria During Testing below), but was at least 15 times during combined and transitivity testing and nine times during symmetry testing. The prompt occurred after each kind of emergent performance tested. During combined and transitivity tests, the prompt occurred once after a one-, two-, three-, four-, and five-node relation in each test block. The one- through five-node relations were varied unsystematically as the 9th, 18th, 27th, 36th, and 45th

Table 2

Symmetric, transitive, and combined relations that may emerge after AB, BC, CD, DE, EF, and FG training.

Symmetric	Transitive	Number of nodes	Combined	Number of nodes
B1A2, B2A2, B3A3	A1C1, A2C2, A3C3	1	C1A1, C2A2, C3A3	1
C1B1, C2B2, C3B3	A1D1, A2D2, A3D3	2	D1A1, D2A2, D3A3	2
D1C1, D2C2, D3C3	A1E1, A2E2, A3E3	3	E1A1, E2A2, E3A3	3
E1D1, E2D2, E3D3	A1F1, A2F2, A3F3	4	F1A1, F2A2, F3A3	4
F1E1, F2E2, F3E3	A1G1, A2G2, A3G3	5	G1A1, G2A2, G3A3	5
G1F1, G2F2, G3F3	B1D1, B2D2, B3D3	1	D1B1, D2B2, D3B3	1
	B1E1, B2E2, B3E3	2	E1B1, E2B2, E3B3	2
	B1F1, B2F2, B3F3	3	F1B1, F2B2, F3B3	3
	B1G1, B2G2, B3G3	4	G1B1, G2B2, G3B3	4
	C1E1, C2E2, C3E3	1	E1C1, E2C2, E3C3	1
	C1F1, C2F2, C3F3	2	F1C1, F2C2, F3C3	2
	C1G1, C2G2, C3G3	3	G1C1, G2C2, G3C3	3
	D1F1, D2F2, D3F3	1	F1D1, F2D2, F3D3	1
	D1G1, D2G2, D3G3	2	G1D1, G2D2, G3D3	2
	E1G1, E2G2, E3G3	1	G1E1, G2E2, G3E3	1

Note. The letters designate sets of comparison stimuli. The numbers designate stimulus classes from which symmetric, transitive, and combined relations may emerge.

test trials. During symmetry testing, the prompt occurred only after trials testing symmetry. The subjects were required to write a summary of the rule on sequentially numbered index cards. After completing each card, the subjects deposited the card into a can beside the computer. No differential reinforcement was programmed for the content of their replies.

Query group. The 4 subjects in the query group were questioned about the rationale of their test-trial responses. "How do you know which symbol to choose?" was displayed on the computer screen at the same points in the test blocks as the instructed group's prompt for the rule.

Standard group. The 4 subjects in the standard group were not instructed or queried about their test-trial responding.

Testing. The composition of test blocks varied across the following three stages of testing: (a) combined tests, (b) transitivity tests, and (c) symmetry tests, presented in that order. The number of baseline trials in each test block was fixed at 36, two trials for each of the 18 relations trained. In the combined test stage, 81-trial blocks included one test trial for each of the 45 potential combined relations (see Table 2 for a list of combined, transitivity, and symmetry trials). In the transitivity stage, 96-trial blocks included one trial for each of the 45 potential transitive relations and 15 combined trials (one for each sample-comparison set). Each combined trial, therefore, was presented once every three blocks during transitivity testing. Finally, in the symmetry stage, 84-trial blocks included one trial for each of the 18 potential symmetric relations and 15 trials each for the combined and transitivity trials (one for each sample-comparison set). Each combined and transitivity trial, therefore, was presented once every three blocks during symmetry testing. In each test stage, trial types were randomly intermixed.

Test blocks were conducted under extinction conditions with no programmed reinforcement on any trial type. Before the start of each test block, subjects were informed about the absence of differential feedback as described above.

Retraining. If a subject's accuracy on the baseline trials within a test block fell below 90%, remedial training blocks with the baseline relations were given until the subject performed at or above 90% accuracy within a block.

Standard accuracy criterion during testing. Each stage of testing continued until the subject made accurate or class-consistent choices on 90% or more of the trials on each of three consecutive blocks, with no more than one error on any one type of relation per block (e.g., GA, CB).

Table 3
Number of training blocks for each subject.

Group	Subject	Number of training blocks
Standard	TH	16
	KS	15
	TV	15
	SW	13
Query	EM	19
	HR	15
	LS	13
	KP	21
Instructed	BG	16
	RS	13
	SV	18
	KC	13

Note. The minimum number of training blocks required was 12.

RESULTS

Training

Two analyses were conducted to determine whether the baseline conditional discrimination performances of the subjects were similar across groups. Group comparisons were based on the number of training blocks required to meet the standard accuracy criterion and the mean speed of correct responses during the baseline maintenance blocks.

Accuracy. Table 3 lists the number of training blocks to reach criterion performance for each subject. Subjects required a minimum of 12 blocks with a range between 13 and 21. A one-way analysis of variance (ANOVA) revealed no significant difference among the groups, $F(2, 9) = 0.87$, $p > .05$. Thus, the groups appeared to be similar on accuracy measures prior to the experimental manipulations.

Speed. All subjects met the standard accuracy criterion for baseline maintenance in the minimum number of five blocks. Speed analyses of correct responding were then conducted for each subject for these five training blocks. Here, as in all analyses of speed that follow, speed data were analyzed only for highly stable, accurate performance. Figure 2 presents the mean speed of correct responses per subject. The query group's response speeds were more variable across subjects than either the standard or instructed group. However, a 3 $\times$ 5 (Group $\times$ Block) repeated measures ANOVA did not reveal a significant main effect of group, $F(2, 9) = 0.45$, $p > .05$. Thus, the groups were also similar on speed of responding before the experimental manipulations. These data also show that for 10 of the 12 subjects, speed of correct responding tended to increase from the first to the last baseline maintenance block.

Testing

Analyses of accuracy and speed of test performance were conducted to determine differences among groups, across trial types, and across nodal distances. These analyses were conducted independently for each phase of testing (combined, transitivity, and symmetry).

Combined testing. During the combined testing stage, subjects were tested on both baseline and combined trials. Because of the nature of the training, there were five sets of one-node (GE, FD, EC, DB, CA), four sets of two-node (GD, FC, EB, DA), three sets of three-node (GC, FB, EA), two sets of four-node (GB, FA), and one set of five-node (GA) combined relations tested. To minimize problems of unequal numbers of nodal relations tested and the order and differential amount of training on each baseline conditional discrimination, all analyses of nodal distance were performed using data from only those trial types with G stimuli as samples (GE, GD, GC, GB, GA) or A stimuli as comparisons (CA, DA, EA, FA, GA). Consequently, there were two sets of trials analyzed for each of the one- through four-node relations and one set for five-node relations. These trials represented the combined relations involving the least trained baseline conditional discriminations (those required for combined trial types involving G stimuli as samples) and those involving the most trained conditional discriminations (those required for combined trial types involving A stimuli as comparisons). In addition, each of the first four nodal distances had one trial type related to the baseline conditional discriminations trained first and one trial type related to the baseline conditional discriminations trained last. The five-node relations were made up of GA trials, which involved all the baseline conditional discriminations.

Comparison revealed no significant differences among groups on either accuracy or speed of correct responding, $F(2, 9) = 0.89$,

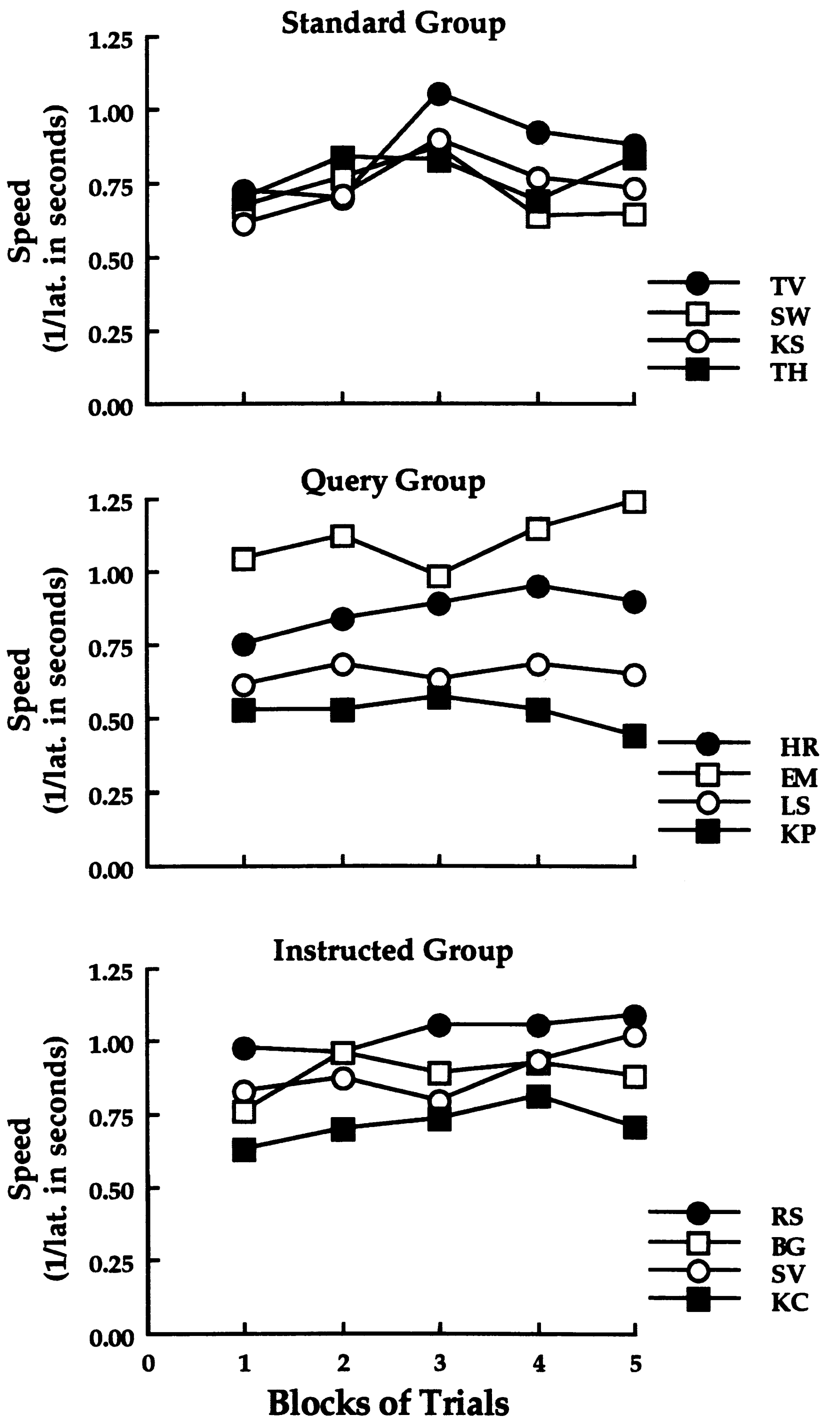

Fig. 2. Baseline maintenance: Mean speed of correct responses per subject for the five baseline maintenance blocks. The number of trials on which the means were based ranged from 45 to 48.

Table 4

Standard group: Percentage of correct responses on combined and baseline trials during combined trial testing.

	TV		SW		KS		TH	
Block	Com	Base	Com	Base	Com	Base	Com	Base
1	87	100	91	100	98	100	22	100
2	98	100	96	100	98	94	53	92
3	100	100	96	97	96	100	67	94
4	98	100					67	97
5							58	100
6							76	97
7							80	100
8							73	100
9							87	100
10							91	92
11							93	97
12							98	100

$p > .05$, and $F(2, 9) = 1.47$, $p > .05$, respectively, or any interactions between group and trial type. Differences were found between trial types, however, with both accuracy and speed measures.

Tables 4, 5, and 6 present the percentage of correct responding on combined and baseline trials during the combined tests. Six of the 12 subjects required more than the minimum number of test blocks (three) to reach the standard accuracy criterion. In each case repeated blocks were necessary because combination performance was below 90%, whereas baseline performance was usually at 100%. A two-way ANOVA revealed that these differences between baseline and combined trial types were significant, $F(1, 9) = 6.1$, $p < .05$, with trial type accounting for 18% of the variance ($eta^2 = .178$).

Analyses of speed of correct responding showed further differences between trial types. Figure 3 shows the mean speed of correct responding during the last three blocks of combined testing, when each subject had met the standard accuracy criterion. All 12 subjects continued to show faster correct responding on baseline trials than on combined trials. A two-way ANOVA revealed that these differences were significant, $F(1, 9) = 86.72$, $p < .01$, and that trial type accounted for 58% of the variance ($eta^2 = .58$).

Subjects were also compared on responses to combination trials of varying nodal distance for the first three trial blocks of combined testing. Figure 4 presents the mean percentage correct for each subject across

Table 5

Query group: Percentage of correct responses on combined and baseline trials during combined trial testing.

	HR		EM		LS		KP	
Block	Com	Base	Com	Base	Com	Base	Com	Base
1	93	100	91	100	58	97	33	94
2	98	100	76	100	96	94	36	91
3	100	100	87	92	91	100	47	91
4			96	100	96	100	42	94
5			96	100			40	91
6			96	94			49	91
7							40	91
8							38	94
9							40	91
10							53	91
11							73	91
12							71	97
13							76	91
14							88	97
15							100	94
16							98	97
17							100	97

one through five nodes. A two-way ANOVA was performed on percentage correct as a function of group and nodal distance. Even though the instructed group showed less variability across subjects, there was no significant main effect of group, $F(2, 9) = 1.54$, $p > .05$, and no Group × Nodal Distance interaction, $F(8, 36) = 0.72$, $p > .05$. Most subjects, however, showed a decrease in accuracy as nodal distance increased, and a significant main effect of nodal distance, $F(4, 36) = 4.86$, $p < .01$, was shown. Nodal distance, however, accounted for only 6% of the variance ($eta^2 = .06$) on accuracy measures. A trend analysis revealed a significant linear trend, $F(1, 36) = 16.55$, $p < .01$, and no higher order trends were significant.

Figure 5 shows the mean speed of correct responding on the one- through five-node

Table 6

Instructed group: Percentage of correct responses on combined and baseline trials during combined trial testing.

	RS		BG		SV		KC	
Block	Com	Base	Com	Base	Com	Base	Com	Base
1	96	100	100	100	80	100	91	100
2	98	100	100	100	96	100	100	100
3	100	100	100	100	98	100	100	100
4					96	100		

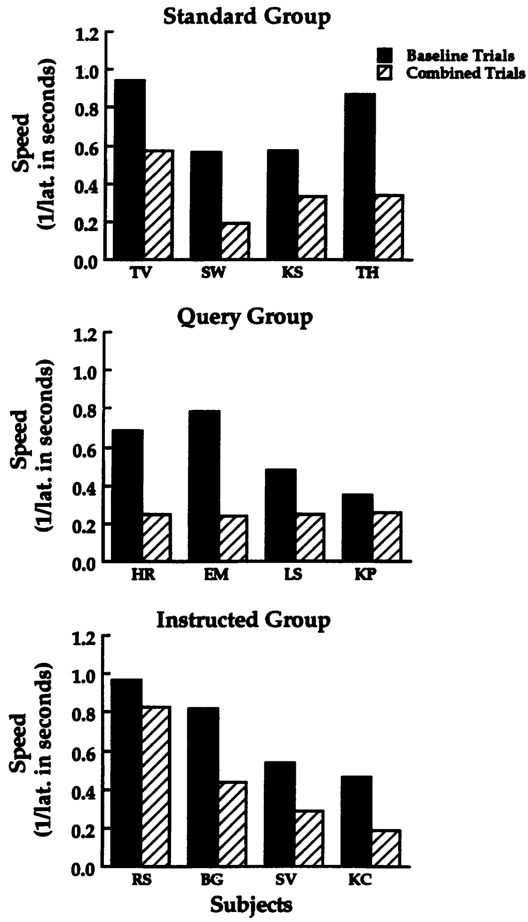

Fig. 3. Combined test phase: Mean speed of correct responses on baseline and combined trials per subject from the last three blocks of combined trial testing. The number of trials on which the means were based ranged from 104 to 108 (baseline) and 127 to 135 (combined).

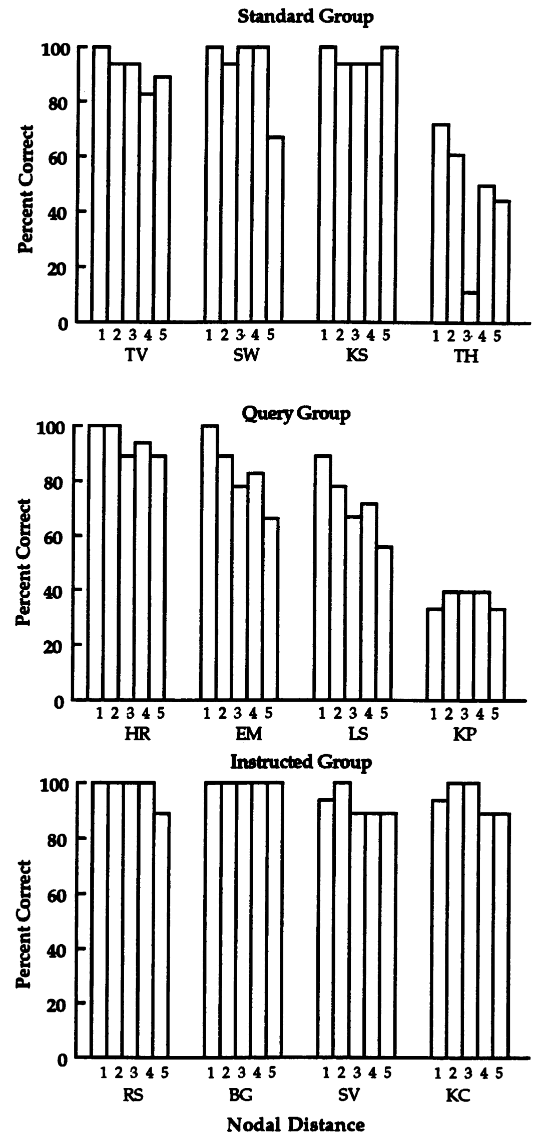

Fig. 4. Combined test phase: Mean percentage correct on combined trials per subject. Each bar for one- through four-node relations represents 18 trials from the first three blocks of combined testing; the bars for five-node relations represent nine trials from the first three blocks of combined trial testing.

combined trials during the last three blocks of combined trial testing. These blocks were selected because the overall performance had met the standard accuracy criterion. For most subjects, regardless of group, the mean speed of responding decreased as nodal distance increased. A two-way ANOVA revealed a main effect of nodal distance, $F(4, 36) = 17.04$, $p < .01$, which accounted for 18% of the variance ($\text{eta}^2 = .18$). There was no significant effect of group, $F(2, 9) = 1.47$, $p > .05$, or Group $\times$ Nodal Distance interaction, $F(8, 36) = 29$, $p > .05$. A trend analysis revealed a significant linear trend, $F(1, 36) = 65.73$, $p < .01$. The linear trend accounted for 17% of the variance ($\text{eta}^2 = .17$). No higher order trends were significant.

Transitivity testing. Transitivity test blocks were comprised of transitivity, combined, and baseline trials. Analysis of performances during transitivity testing was conducted to determine whether differences could be attributed to the groups, trial types, and nodal

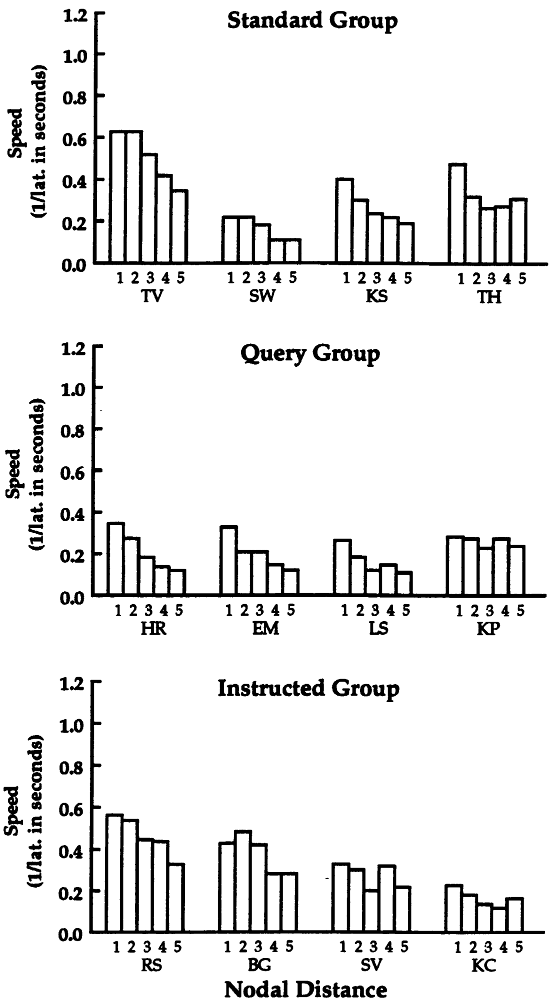

Fig. 5. Combined test phase: Mean speed of correct responding on one- through five-node combined trials during the last three blocks of combined trial testing. The number of trials on which the means were based ranged from 6 (five-node) to 18.

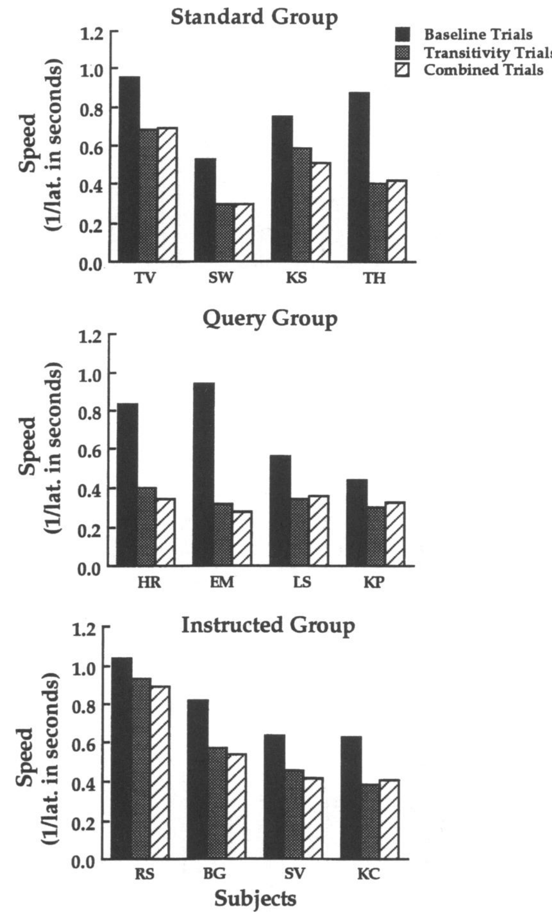

Fig. 6. Transitivity test phase: Mean speed of correct responses on baseline, transitivity, and combined trials per subject. The number of trials on which the means were based ranged from 105 to 108 (baseline), 130 to 135 (transitivity), and 43 to 45 (combined).

distance. As with the combined relations, all analyses of nodal distance were performed using data from only those trial types with A stimuli as samples (AC, AD, AE, AF, AG) or G stimuli as comparisons (EG, DG, CG, BG, AG).

By this point in the testing, no differences were found on accuracy measures among groups, among trial types (baseline, combined, and transitivity), or across nodal distances. All subjects performed at or around 100% correct on all trials. The lowest performance was 92% correct on baseline trials for 1 subject. Therefore, the remaining analyses were limited to speed of correct responding.

Figure 6 shows the mean speed of correct responses on baseline, transitivity, and combined trials for each subject by group. All subjects performed faster on baseline trials than on either transitivity or combined trials. Most subjects showed similar speeds on transitivity and combined trials. A two-way ANOVA revealed a main effect of trial type, $F(2, 18) = 38.78$, $p < .01$, but no main effect of group, $F(2, 9) = 1.40$, $p > .05$, or Group × Trial Type interaction, $F(4, 18) = 0.91$, $p > .05$. Trial type accounted for 36.5% of the variance. Planned comparisons revealed that differences between baseline and transitivity tri-

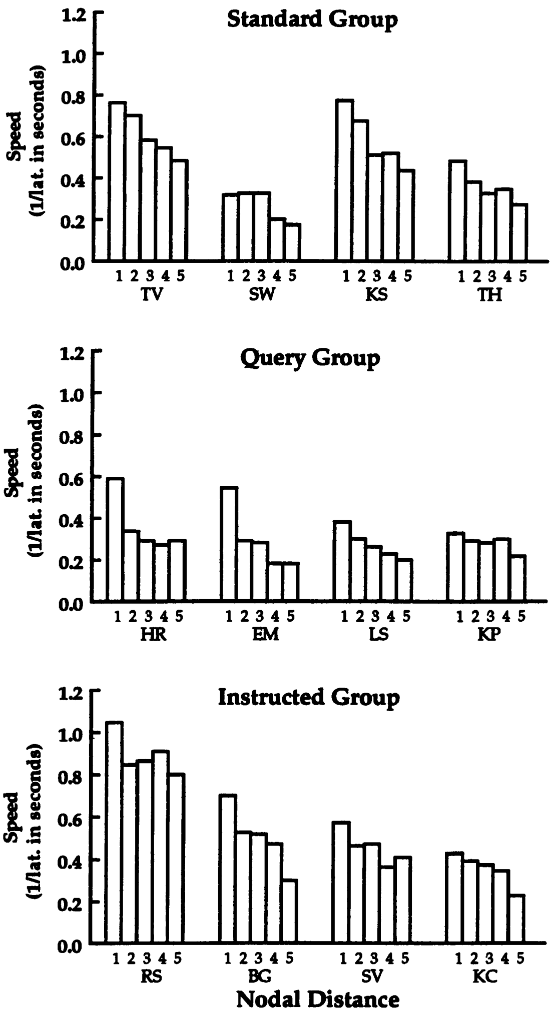

Fig. 7. Transitivity test phase: Mean speed of correct responding on one- through five-node transitivity trials. The number of trials on which the means were based ranged from 9 (five-node) to 18.

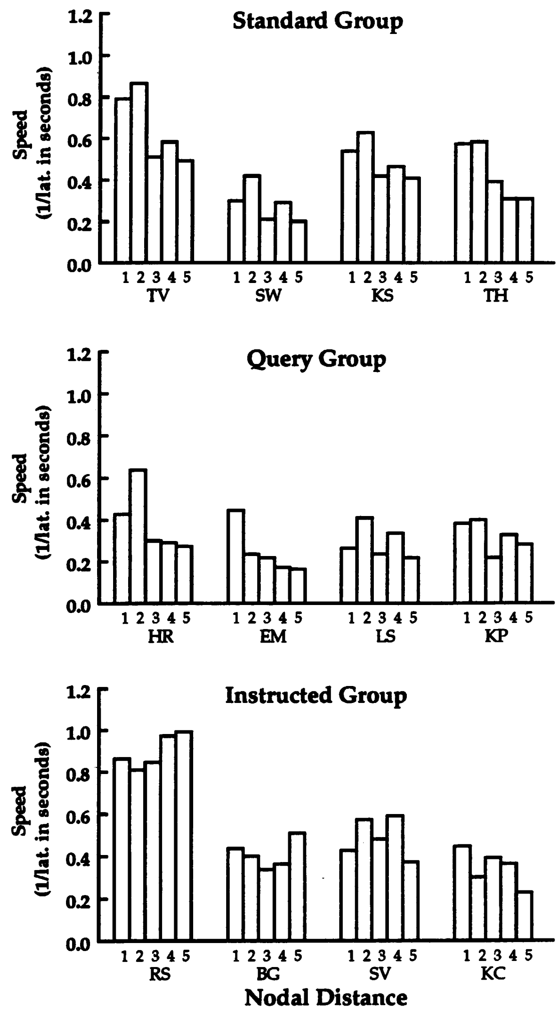

Fig. 8. Transitivity test phase: Mean speed of correct responding on one- through five-node combined trials. The number of trials on which the means were based ranged from three (five-node) to six.

als as well as between baseline and combined trials were significant at $p < .01$, $F(1, 18) = 55.21$ and 60.98, respectively. There was no significant difference in response speed on transitivity and combined trials, $F(1, 18) = 12.04$, $p > .05$.

Figure 7 shows the speed of correct responses on one- through five-node transitivity trials for each group. Most subjects' mean speed of responding decreased as nodal distance increased. A two-way ANOVA revealed a main effect of nodal distance $F(4, 36) = 34.22$, $p < .01$. Nodal distance accounted for 17% of the variance. Trend analyses revealed significant ($p < .05$) linear, quadratic, and cubic trends, $F(1, 36) = 127.94$, 4.31, and 4.5, respectively. There was no significant main effect of group, $F(2, 9) = 2.22$, $p > .05$, or Group $\times$ Nodal Distance interaction, $F(8, 36) = .81$, $p > .05$. These results are consistent with that of the combined trials in the combined test blocks.

Figure 8 shows the speed of correct responding on one- through five-node combined trials for each group. For most subjects in the standard and query groups, the mean speed of responding decreased as nodal distance increased. However, the response speed of the subjects in the instructed group did not vary systematically as a function of nodal distance. Two subjects showed an increase in response speed as nodal distance increased, 1 subject's speed decreased as nodal distance increased, and 1 subject's speed varied. A two-way ANOVA revealed a significant Group × Nodal Distance interaction, $F(8, 36) = 2.56$, $p < .05$. Simple main effect analyses revealed a significant ($p < .01$) effect of nodal distance for the query and standard groups, $F(4, 36) = 4.76$ and 9.1, respectively. There was also a significant main effect of nodal distance, $F(4, 36) = 9.06$, $p < .01$. Trend analyses revealed significant ($p < .05$) linear and quartic trends in the standard group's speeds, $F(1, 36) = 24.65$ and 8.30, respectively. Trend analyses also revealed significant linear and quartic trends in the query group's speeds, $F(1, 36) = 12.64$ and 5.19, $p > .05$, respectively.

Symmetry testing. Symmetry test blocks were comprised of symmetry, transitivity, combined, and baseline trials. As with tests during the transitivity stage of testing, most subjects performed at 100% accuracy on all trial types, so further analyses of accuracy of responding were not conducted.

Figure 9 shows the speed of correct responding on baseline, symmetry, transitivity, and combined trials for each group. For all subjects, baseline and symmetry responding were faster than transitivity and combined responding. For most subjects, baseline responding was still faster than symmetry responding. No systematic differences were found between combined and transitivity responding. A two-way ANOVA revealed a main effect of trial type, $F(3, 27) = 43.84$, $p < .01$, but no significant main effect of group, $F(2, 9) = 1.54$, $p > .05$, or Group × Trial Type interaction, $F(6, 27) = 1.37$, $p > .05$. Trial type accounted for 31% of the variance. Planned comparisons revealed differences between response speed on the following: baseline and symmetry trials, $F(1, 27) = 5.17$, $p < .05$; baseline and transitivity trials, $F(1, 27) = 83.77$, $p < .01$; baseline and combined trials, $F(1, 27) = 81.31$, $p < .01$; symmetry and transitivity trials $F(1, 27) = 47.30$, $p < .01$; and symmetry and combined trials $F(1, 27) = 47.30$, $p < .01$. There were no significant differences between transitivity and combined trials, $F(1, 27) = 0.02$, $p > .05$.

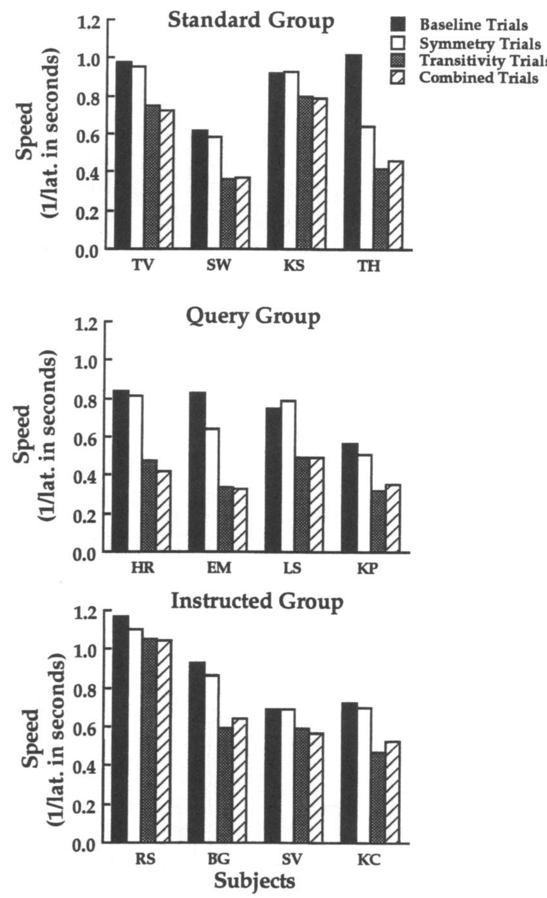

Fig. 9. Symmetry test phase: Mean speed of correct responses on baseline, symmetry, transitivity, and combined trials per subject. The number of trials on which the means were based ranged from 42 to 54 (symmetry, transitivity, and combined) and from 105 to 108 (baseline).

Figure 10 shows the speed of correct responding on one- through five-node transitivity trials for each group. Although individual subjects' speeds varied, most subjects continued to respond faster on the shorter nodal distances than on the larger ones. Seven of the subjects, however, responded faster on five-node relations than on four-node relations. Statistical comparisons revealed no significant main effect of group, $F(2, 9) = 1.70$, $p > .05$, or Group × Nodal Distance interaction, $F(8, 36) = 0.48$, $p > .05$. However, a two-way ANOVA revealed a main effect of nodal distance, $F(4, 36) = 10.00$, $p < .01$, and nodal distance accounted for 14% of the variance ($eta^2 = .14$). Trend analysis revealed a significant linear trend, $F(1, 36) = 24.76$, $p <$

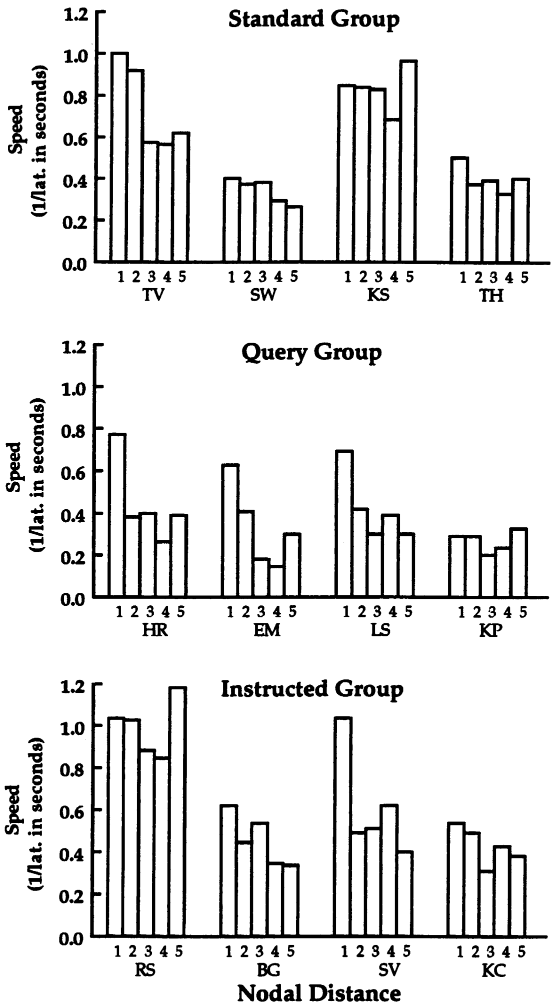

Fig. 10. Symmetry test phase: Mean speed of correct responding on one- through five-node transitivity trials. The number of trials on which the means were based ranged from three (five-node) to six.

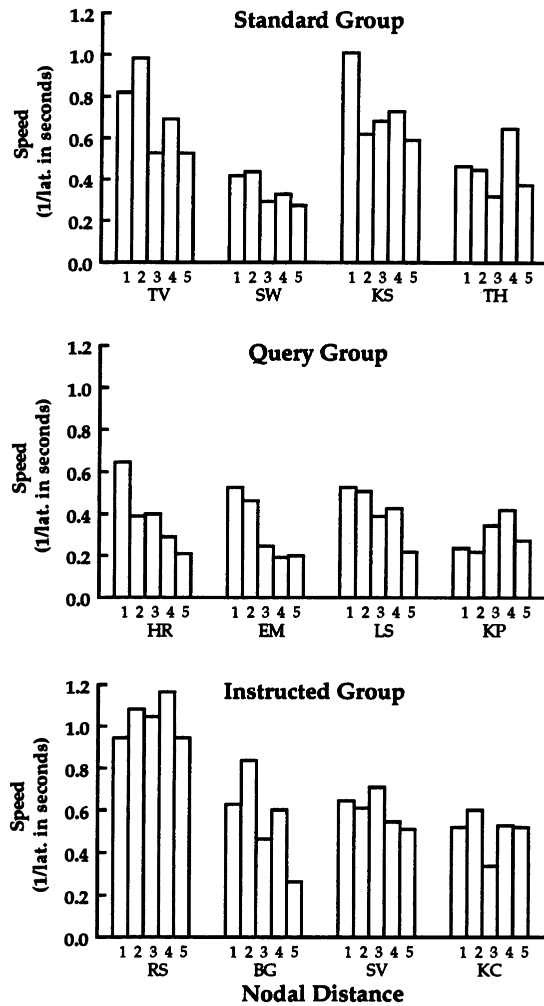

Fig. 11. Symmetry test phase: Mean speed of correct responding on one- through five-node combined trials. The number of trials on which the means were based ranged from three (five-node) to six.

.05, and a significant quadratic trend, $F(1, 36) = 15.21$, $p < .05$.

Figure 11 shows the speed of correct responses on one- through five-node combined trials for each group. Again, most subjects' response speeds decreased as nodal distance increased, with 2 subjects' speeds increasing across nodal distance. Statistical comparisons revealed a significant main effect of nodal distance, $F(4, 36) = 6.59$, $p < .01$ ($\text{eta}^2 = .10$), but no main effect of group, $F(2, 9) = 3.17$, $p > .05$ ($\text{eta}^2 = .30$), or Group × Nodal Distance interaction, $F(8, 36) = 0.67$, $p > .05$ ($\text{eta}^2 = .02$). Trend analyses revealed a significant linear trend, $F(1, 36) = 19.43$, $p < .05$ ($\text{eta}^2 = .08$). No higher order trends were present. Note that although the response speed of the instructed group did not decrease as a function of nodal distance on combined trials during transitivity testing (see Figure 8), it did during symmetry testing.

Table 7 shows subjects' mean speed of correct responding to baseline, combined, and transitivity trials during the three phases of testing. All subjects' speed of responding on baseline, transitivity, and combined trials increased from the combined test phase to the symmetry test phase. Sample ANOVAs re-

Table 7

Mean speed scores of baseline, combined, and transitivity relations across combined, transitivity, and symmetry test phases.

Subject	Baseline trials			Combined trials			Transitivity trials	
	C	T	S	C	T	S	T	S
TV	0.94	0.95	0.98	0.57	0.69	0.72	0.68	0.75
SW	0.56	0.53	0.62	0.19	0.30	0.37	0.30	0.36
KS	0.57	0.75	0.92	0.33	0.51	0.79	0.58	0.80
TH	0.87	0.87	1.02	0.34	0.42	0.46	0.40	0.42
HR	0.69	0.83	0.84	0.25	0.34	0.42	0.40	0.48
EM	0.79	0.94	0.83	0.24	0.28	0.33	0.32	0.34
LS	0.48	0.56	0.75	0.25	0.36	0.49	0.34	0.49
KP	0.35	0.44	0.57	0.26	0.33	0.35	0.30	0.32
RS	0.97	1.04	1.17	0.53	0.89	1.04	0.93	1.05
BG	0.82	0.82	0.93	0.44	0.54	0.64	0.57	0.59
SV	0.54	0.64	0.69	0.29	0.42	0.57	0.46	0.59
KC	0.46	0.63	0.72	0.19	0.41	0.53	0.38	0.47

Note. C = combined test phase, T = transitivity test phase, S = symmetry phase.

vealed a significant main effect of test stage for combined relations, $F(2, 18) = 49.76$, $p < .01$, which accounted for 30% of the variance in speed (eta^2 = .30) and for baseline relations, $F(2, 18) = 21.28$, $p < .01$, (eta^2 = .14). Thus, performance continued to change with repeated testing in ways that were not apparent from accuracy measures.

The self-reports of the query subjects and the rule summaries of the instructed subjects revealed that within the first two combined test blocks all query subjects accurately described a rule comparable to that given to the instructed group. (Copies of these self-reports are available from the second author.)

DISCUSSION

Response Speed and Equivalence

Differences in response speed were found across trial types (baseline, symmetry, transitivity, and combined) for most subjects throughout all stages of testing, with trial type accounting for a large portion of the variance in speed (36% to 58%). During the first stage of testing, all subjects had faster response speeds on baseline trials than on combined trials. The second stage of testing also resulted in all subjects continuing to respond faster on baseline trials than on either combined or transitivity trials. These differences continued during the third stage of testing, with all subjects responding faster on baseline trials than on either combined or transitivity trials. In addition, 9 of the 12 subjects responded faster on baseline trials than on symmetry trials. These latter differences, although statistically significant, were smaller than the differences between baseline and combined or transitivity responding. Systematic differences were not found, however, between transitivity and combined trials. In contrast to the difference found when speed of responding was measured, the relation between accuracy and trial type was apparent only during the first stage of testing, and even then trial type accounted for only 18% of the variance.

The differences due to trial type extend the results from Wulfert and Hayes (1988) and Bentall et al. (1993). Those studies found significant differences in latency to respond to baseline and transitivity trials but no significant difference between response speeds (latencies were reported) on baseline and symmetry trials. The difference between the present results and previous results might be due to a difference in how time was measured. Neither Wulfert and Hayes nor Bentall et al. required subjects to make an overt response to the sample stimulus to produce the comparison stimuli. Rather, the comparison stimuli were presented 2 s after the sample stimulus. Thus, the latencies did not describe the time between two overt responses, as did the speed measures in this study, but instead described the time between the presentation of the comparison stimuli and the selection of a comparison stimulus. We speculate that the 2 s provided between the sample and the comparison in these other studies washed out the differences between baseline and symmetry responding. Two seconds may have been sufficient for the subjects to respond in unmeasured ways that prepared them to respond to the comparisons when they appeared.

Analyses of the speed of accurate responding also revealed that for most subjects response speed was inversely related to the number of nodes on which the tested relations depended. Nodal distance accounted for 14% to 18% of the variance. For example, during combined testing all subjects responded faster to one-node than to five-node combined relations, and data from half of the subjects indicated an inverse relation. During transitivity testing, all but 2 of the subjects again responded faster on one-node than on five-node combined relations, and all subjects responded faster on one-node than on five-node transitive relations. Six subjects' data showed an inverse relation on transitivity trials as well. During symmetry testing, 9 subjects responded faster on one-node than on five-node combined and transitive relations. These speed data contrast with the accuracy data, in which the relation between accuracy and nodal distance was apparent only during the first phase of testing, and nodal distance accounted for only 6% of the variance. Thereafter, all subjects responded accurately to all combined and transitivity trials regardless of nodal distance.

These results support Fields et al.'s (1984) prediction that there is an inverse relation between nodal distance and measures of response strength (accuracy and speed). The data also replicate the effects of nodal distance found by Fields et al. (1990) and extend those effects with analyses of response speed. Further, the current results show the effects of nodal distance with three-, four-, and five-node relations.

Instructions and Equivalence

There were no significant differences among the accuracy or speed of the different groups' test-trial responding. Thus, neither the experimenter-provided instructions nor the content of the query group's self-reports significantly reduced the need for repeated testing or eliminated the effects of nodal distance.

Despite the lack of effect with instructions and queries, this study illustrates a method for asking subjects to describe their MTS performance. This was accomplished by querying the subjects throughout the test blocks, rather than waiting until after the experiment was finished (Hayes & Hayes, 1989; Wearden, 1988). Although this method of obtaining self-reports is promising, there was one drawback. Querying the subjects during testing may have decreased the strength of the subjects' MTS performance: The test-trial responding of query subjects was consistently (although not significantly) slower and less accurate than the responding of the standard subjects. Thus, the query may have prompted a response that competed with fast, accurate test-trial responding (Critchfield & Perone, 1990).

Conclusions

Although stimuli have been traditionally viewed as equivalent after accurate test-trial performance is demonstrated, the analyses of trial type and response speed conducted in this experiment suggest that stimuli in an equivalence class can be categorized further by speed of responding. Long after the accuracy data suggested that the stimuli in the equivalence classes were functionally substitutable, speed data indicated that subjects were responding differently to the stimuli. Describing these differences through the measurement of response speed may be useful at a number of levels of analysis.

Such data show that responding to the stimuli within a given equivalence class involves considerable variability. This finding is important because it provides a more thorough understanding of the substitutability of stimuli in an equivalence class than do accuracy measures alone. According to Fields, Adams, Verhave, and Newman (1993), accurate performances on emergent relations tests are sufficient to demonstrate class membership, but they are weak measures of substitutability because emergent relations tests can be passed even if the stimuli in a class are not equally related. They suggest using tests for transfer as a measure of substitutability. The current results support Fields et al. (1993) by suggesting that the more similar the speed of accurate responding on baseline, symmetry, transitivity, and combined trials, the more substitutable the stimuli. It follows from this view that the stimuli in a class may not be equally substitutable, even after traditional measures of equivalence class formation have been obtained.

Responses that vary in speed may also show variation in other measures of response strength. For example, perhaps MTS performances that have the quickest speeds will be maintained longer under conditions of extinction or will occur more readily under conditions of stress or distraction than will slower MTS performances. Such differences in strength, if found, would be fundamental to the precision with which the relations among stimuli and responding could be predicted and controlled.

If better measures of substitutability and strength help us to predict and control the kinds of relations involved in equivalence classes like those tested in the current study, they may also help us to apply equivalence to understanding verbal behavior (e.g., Hall & Chase, 1991; Hayes & Hayes, 1992; Sidman, 1986). Verbal behavior can involve relations among stimuli that are even more extended than the five-node relations tested in the current experiment. Fluent verbal behavior requires quick and accurate responding to a variety of questions, comments, and demands from other individuals as well as to events and objects that are observed. Given the results reported here, measuring both the accuracy and speed of verbal behavior seems more likely to lead to precise prediction and control of these complex relations.

The results of this study indicate that the speed of accurate responding may provide a more fine-grained analysis of subjects' learning. These results are important for a number of reasons. They suggest methods of data collection that might identify variables that affect MTS performance. The results support the view that the substitutability of stimuli should not be defined by the accuracy of match-to-sample performance alone and suggest measuring speed to provide additional information on the relations among stimuli in a class. Finally, the results have applied implications. Training that emphasizes both speed and accuracy may lead to more predictable performance under conditions of retention and application (Johnson & Layng, 1992).

REFERENCES

Ayllon, T., & Azrin, N. H. (1964). Reinforcement and instructions with mental patients. *Journal of the Experimental Analysis of Behavior, 7,* 327–331.

Baron, A. (1985). Measurement scales and the age-complexity hypothesis. *Experimental Aging Research, 11*(4), 193–199.

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.

Critchfield, T. S., & Perone, M. (1990). Verbal self-reports of delayed matching to sample by humans. *Journal of the Experimental Analysis of Behavior, 53,* 321–344.

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., Adams, B. J., Verhave, T., & Newman, S. (1993). Are stimuli in equivalence classes equally related to each other. *Psychological Record, 43,* 85–106.

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

Fields, L., Verhave, T., & Fath, S. (1984). Stimulus equivalence and transitive associations: A methodological analysis. *Journal of the Experimental Analysis of Behavior, 42,* 143–157.

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

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

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

Johnson, K. R., & Layng, T. V. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American Psychologist, 47,* 1475–1490.

Lindsley, O. (1960). Characteristics of the behavior of chronic psychotics as revealed by free-operant conditioning methods. *Diseases of the Nervous System, 21,* 66–78.

Pilgrim, C., & Galizio, M. (1990). Relations between baseline contingencies and equivalence probe performances. *Journal of the Experimental Analysis of Behavior, 54,* 213–224.

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

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Erlbaum.

Wearden, J. H. (1988). Some neglected problems in the analysis of human operant behavior. In G. Davey & C. Cullen (Eds.), *Human operant conditioning and behavior modification* (pp. 197–224). New York: Wiley.

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

Received June 21, 1994
Final acceptance January 24, 1996