

INTERTRIAL SOURCES OF STIMULUS CONTROL AND DELAYED MATCHING-TO-SAMPLE PERFORMANCE IN HUMANS

DEAN C. WILLIAMS, MARK D. JOHNSTON, AND KATHRYN J. SAUNDERS

UNIVERSITY OF KANSAS

Two experiments compared delayed matching-to-sample (DMTS) accuracy under 2 procedures in adults with mental retardation. In the trial-unique procedure, every trial in a session contained different stimuli. Thus, comparison stimuli that were correct on one trial were never incorrect on other trials in that session (or vice versa). In the 2-sample DMTS procedure, the same 2 comparison stimuli were presented on each trial, and their function changed quasi-randomly across trials conditional upon the sample stimulus. Across 2 experiments, 7 of 8 subjects showed the highest overall accuracy under the trial-unique procedure, and no subject showed consistently higher accuracy under the 2-sample procedure. Negative, exponential decay functions fit to logit p values showed that this difference was due largely to the steeper delay-mediated decline in sample control for the 2-sample procedure. Stimulus-control analyses indicated that, under the 2-sample procedure, the selection of the comparison stimulus on Trial N was often controlled by the comparison stimulus selection on Trial N-1 rather than the Trial-N sample stimulus. This source of competing stimulus control is not present in trial-unique procedures. Experiment 2 manipulated intertrial interval duration. There was a small but consistent increase in accuracy as a function of intertrial interval duration under the 2-sample procedure, but not under the trial-unique procedure.

Key words: delayed matching to sample, conditional discrimination, trial-unique matching, stimulus control, proactive interference, mental retardation, touch screen, humans

Delayed matching-to-sample (DMTS) procedures are widely used to assess, and to study variables related to, short-term memory in both human and nonhuman subjects. A delayed matching-to-sample trial begins with the presentation of a sample stimulus (e.g., a circle). A response to the sample removes it from the display and two or more comparison stimuli (e.g., a circle and a triangle) appear after a delay, or retention interval. A response to the comparison stimulus that is identical to the sample stimulus produces a reinforcer. In the most commonly used version of the procedure, the same two comparison stimuli are presented on each trial, and the function of the comparison stimuli can change on a trial-by-trial basis depending on which of the two

samples is presented. We will refer to this as 2-sample matching.

The duration of the retention interval can be varied across trials within a session, or held constant within a session and varied across sessions. The classic finding, shown in rats (e.g., Harper, McLean, & Dalrymple-Alford, 1994), pigeons (e.g., Blough, 1959; Cumming & Berryman, 1965), monkeys (e.g., Wright, Urcuioli, & Sands, 1986), and humans (e.g., Constantine & Sidman, 1975; Dalton & Crapper-McLachlan, 1984), is that matching accuracy decreases as a function of increases in the retention interval. The decreased accuracy shown at longer retention intervals indicates diminished sample-stimulus control of comparison selection. Studies have sought to elucidate the variables controlling the maintenance and loss of stimulus control as a function of retention-interval duration (see Wixted, 1989 for a behavioral account of memory). One target of study is the stimulus control that may replace control by the sample (e.g., Edhouse & White, 1988; Mackay & Gould, 1992).

Two sources of competing stimulus control over comparison selection have been studied, primarily using pigeons as subjects. *Intratrial* interference arises from stimuli present within the current trial, that is, concurrent with

Preparation of this manuscript was supported by the National Institutes of Health grants P30HD02528 and R01HD044731 awarded to the Schiefelbusch Institute for Life Span Studies, University of Kansas. Mark Johnston is now with the Boy Scouts of America. We thank Bill Dube for the software; Colleen Eisenbart and Pat White for their help with this work; Joe Spradlin, Jennifer O'Donnell, and Kathy Stewart for helpful comments. We appreciate the support of Parsons State Hospital and Training Center.

Direct correspondence to Dean Williams, University of Kansas, Parsons Research Center, 2601 Gabriel, Parsons, KS 67357 (email: Deanwms@ku.edu).

doi: 10.1901/jeab.2006.67-01

sample presentation, during the delay, or concurrent with the presentation of the comparison stimuli. *Intertrial* sources arise from stimulus conditions prior to the current trial. One key source of alternative stimulus control is the comparison stimulus selected on the previous trial. When the correct comparison on trial *N* differs from the comparison that was selected on trial *N-1*, accuracy is lower than when the two stimuli are identical (Edhouse & White, 1988; Roitblat & Harley, 1988; Wright *et al.*, 1986). Thus, instead of selecting the stimulus that matches the current sample, the subject selects the stimulus that was selected on the previous trial, particularly if that selection was reinforced (Edhouse & White, 1988; White, Parkinson, Brown, & Wixted, 2004).

Recognition difficulties over very brief delays may be said to indicate severely limited short-term memory at a fundamental process level (decay rate). Low DMTS accuracy, however, may not wholly reflect an underlying memory deficit. Based on a review of studies with nonhumans, Wright *et al.* (1986) concluded that the typical 2-sample DMTS procedure may underestimate short-term memory capabilities. Studies involving monkeys have shown that delayed-matching accuracy is higher with *trial-unique* procedures (Mishkin & Delacour, 1975; Overman & Doty, 1980). In trial-unique procedures, every trial in a session presents stimuli that have not been presented previously in that session. Wright *et al.* (1986) attributed lower accuracy with the 2-sample procedure to proactive interference that results when the same stimulus functions as both an S+ and an S- across trials (see also Dube, McIlvane, & Green, 1992). This across-trial change in comparison stimulus function provides the basis for intertrial interference to affect accuracy. By contrast, in the trial-unique procedure, stimuli never change functions across trials, thereby minimizing intertrial interference.

Wright *et al.* (1986) noted a need to assess the subject generality of performance differences shown across 2-sample and trial-unique matching to sample (MTS). There have been few relevant studies with human subjects. Studies using simultaneous matching procedures have shown mixed results. Weinstein (1941) reported little difference in trial-unique and 2-sample MTS in limited testing

with two 3-year-old children. A few studies have shown that children learned simultaneous trial-unique oddity tasks in fewer trials and to greater accuracy than 2-sample oddity (House, Brown, & Scott, 1974; Scott & House, 1978; Sugimura & Iyoda, 1982).

Although the trial-unique procedure has been used in studies of human memory (e.g., Park, Püschel, Sauter, Rentsch, & Hell, 2002), we know of only one published comparison of trial-unique and 2-sample delayed matching in human subjects. Williams, Dube, Johnston, and Saunders (1998) reported two experiments, one involving 4 individuals with mild retardation who were studied at delays ranging from 0 to 16 s, and one involving 5 individuals with moderate-severe retardation, who were studied at a single, brief delay. Six of the 9 subjects showed greater accuracy under the trial-unique procedure, and the other 3 subjects showed equivalent accuracy under the two procedures. No subject showed greater accuracy with the 2-sample procedures.

The present investigation had several goals. In the first experiment, trial-unique and 2-sample procedures were compared across a range of delays. Unlike in Williams *et al.* (1998), a detailed stimulus-control analysis was conducted to search for evidence of intertrial interference under the 2-sample procedure.

The second experiment manipulated a variable that has been shown to influence intertrial interference in 2-sample DMTS in pigeons. Several studies have shown that delayed matching accuracy increases with increases in the duration of the intertrial interval. The increase is assumed to occur because conflicting control by stimuli presented on the previous trial diminishes over the longer ITI (Edhouse & White, 1988). Little is known about the relation of ITI duration to delayed matching accuracy in human subjects.

GENERAL METHOD

Subjects and Apparatus

The subjects were 8 adults with mild or moderate mental retardation, none of whom received psychotropic medication. For additional subject characteristics, see Table 1. Sessions were conducted in a small, sound-attenuating room. Subjects sat facing a 30-cm touch-sensitive monitor recessed into a wooden partition. A clear plastic cup protruded

Table 1
Subject Information.

Subject (Gender)	Chronological age in years	Mental age in years-months	Experiments
MW (M)	22	12-1	1
DA (M)	22	9-10	1
GC (M)	47	6-4	1 and 2
VS (F)	25	6-2	1
FJ (M)	21	6-8	1 and 2
CH (M)	19	8-1	2
JB (M)	19	4-8	2
SE (M)	18	9-5	2

Note. Mental age is based on the Peabody Picture Vocabulary Test-Revised.

through the partition to the lower right-hand side of the monitor. A universal feeder mounted behind the wall dispensed coins into the cup. Session events were presented, and responses recorded, using software developed by Dube (1991). The monitor screen was white and the stimuli were black arbitrary shapes, each approximately 3 cm². Example stimuli may be seen in Saunders and Spradlin (1993) and in Figure 1.

Procedure

One 60-trial session was presented 3 to 5 days each week. All trials began with the presentation of a sample stimulus in the center of the screen. Touching the sample removed it from the screen and initiated a delay of 0, 2, 4, 8, or 16 s. The delays were presented quasi-randomly with the stipulation that each delay occurred once in each block of five trials. Thus, there were 12 trials at each delay value per session. Following the delay, two comparison stimuli appeared simultaneously in two of the four corners of the screen. One stimulus was physically identical to the sample (S+); the second stimulus was not identical (S-). Within each session, the S+ and S- appeared an equal number of times in each of the four screen corners and neither the S+ nor S- occurred in the same position on more than three consecutive trials. Touching the S+ produced a 1.5-s series of tones and the delivery of a penny. Touching the S- produced a 0.5-s buzzer along with 3 s of black screen. An intertrial interval (ITI), the duration of which differed across the experiments, followed delivery of the consequences. Touches during the ITI (which occurred rarely) were recorded but had no programmed consequences.

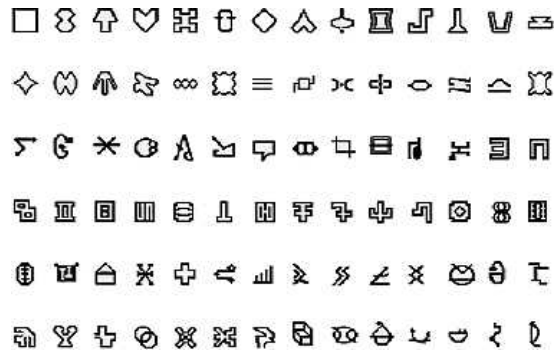


Fig. 1. Examples of stimuli used in the sessions.

Session types. There were two delayed matching-to-sample session types. The 2-sample procedure presented only two different stimuli in each session, with each stimulus serving as the sample on half of the trials. Trials were presented quasi-randomly with the stipulation that no more than three consecutive trials had the same S+. To reduce the potential for stimulus bias, each 2-sample session contained stimuli that were not used in any other 2-sample or trial-unique session. The trial-unique procedure presented a different S+ and S- on each trial (120 different stimuli per 60-trial session). Each trial-unique session differed in composition from other trial-unique sessions.

A total of 270 stimuli were used to create sessions. Thirty stimuli were randomly assigned to construct 15, 2-sample sessions (5 for Experiment 1 and 10 for Experiment 2). These 30 stimuli were not used in trial-unique session configurations. From the remaining stimuli, 15 different trial-unique sessions were constructed, as follows. The 240 stimuli were divided randomly into two, 120-stimulus sets. To create the first two sessions, stimuli within each set were randomly selected in pairs, and assigned as samples and comparisons. The third and fourth sessions were made from the first and second stimulus sets, respectively. So that the function of the stimulus in one session did not predict its function in another session, half of the former S+ stimuli became S- stimuli (and vice versa), and stimuli were never presented in the same pairs across sessions. An additional 11 sessions were made following this pattern. Session configurations were randomly assigned for daily sessions, and no

subject was exposed to the same configuration twice.

Pretraining. The 2-sample procedure was used in preliminary training sessions. Each session had 24 trials, and began with 12 trials of simultaneous MTS (i.e., the sample did not disappear when touched). In the remaining 12 trials, the sample disappeared when touched and the comparisons appeared immediately (zero delay). Before the session, subjects were told that they would be playing a matching game, that for each correct match they would receive a penny from the computer, and that at the end of the game, they would receive an additional penny for each penny earned. The experimenter remained in the room for the first four trials. When the sample appeared on the first trial, the subject was told to "Touch it" and, after the comparisons appeared, to "Touch the one that matches the picture in the middle." Preliminary training ended when accuracy was at least 90% on the 12 zero-delay trials, which occurred in one session for all participants.

EXPERIMENT 1

Experiment 1 compared accuracy across trial-unique and 2-sample DMTS sessions, as in Williams *et al.* (1998), and added a direct assessment of the contribution of the previous trial onto accuracy under the 2-sample procedure.

METHOD

Five adults with mental retardation participated (see Table 1). Each subject received five sessions of each type (trial-unique and 2-sample). Had there been an increasing or decreasing trend in accuracy for each session type, additional sessions would have been presented, but this did not occur. Each session type was presented once in each block of two sessions. A trial-unique session was presented first for all subjects. After that, session types were presented in random order within blocks. The ITI duration was 4 s.

RESULTS

Figure 2 (left column) shows mean accuracy from all five sessions under each procedure as a function of delay value. For all subjects, when the delay was zero or 2 s, accuracy was 90% or

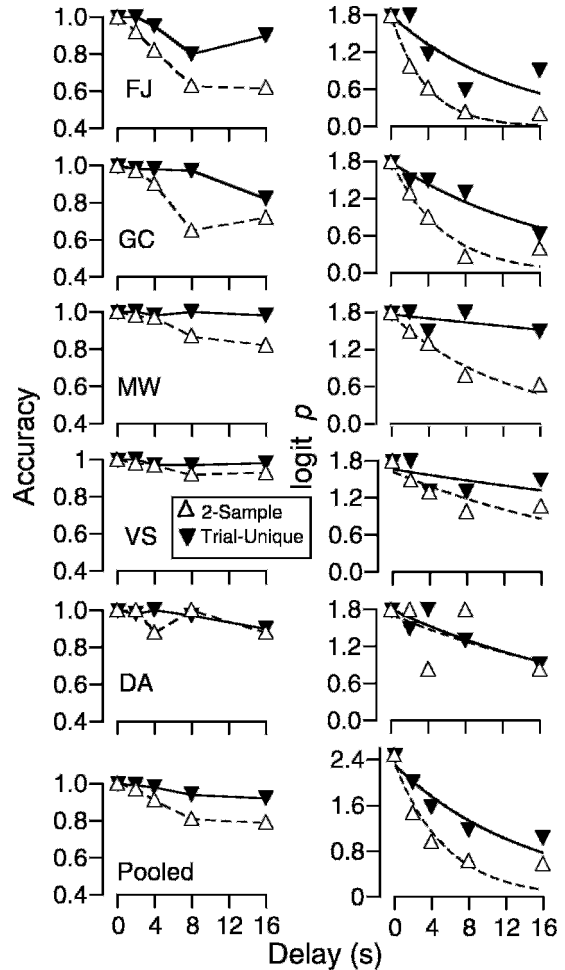


Fig. 2. Experiment 1: Accuracy (left column) and logit p values (right column) for individual subjects and for data pooled across subjects at each delay value. Logit p data were fitted to the negative exponential forgetting function of Equation 2; see text for details.

higher for both session types. For 3 of the 5 subjects (FJ, GC, and MW), accuracy at the two longest delays clearly was higher in the trial-unique sessions than in the 2-sample sessions. Moreover, accuracy on trial-unique sessions was very high, falling below 90% in only two cases (the 16-s delay for Subject GC and the 8-s delay for Subject FJ). The bottom, left panel shows these data pooled across subjects (correct responses + pooled incorrect responses; see Hautus, 1997, for a discussion of using pooled data instead of means).

Figure 2 (right column) shows the same data converted to logit p scores using the

Table 2
Experiment 1: Parameters of Forgetting Function Fitted to Logit p Transformed Data.

Subject	Two-sample			Trial-unique		
	Logit $p-0$	b	R^2	Logit $p-0$	b	R^2
FJ	1.8	.266	.98	1.8	.076	.68
GC	1.8	.177	.94	1.8	.056	.93
MW	1.8	.080	.96	1.8	.009	.33
VS	1.6	.039	.73	1.7	.014	.26
DA	1.7	.036	.33	1.8	.040	.82
Pooled	2.3	.179	.91	2.3	.068	.89
Pooled ¹	2.3	.175	.93	2.2	.078	.92

¹ Pooled excluding data from subjects with R^2 below .6.

formula:

$$\text{logit } p = \text{Log}[(c + 1)/(e + 1)] \quad (1)$$

where c is the number of correct responses to all stimuli and e is the number of error responses to all stimuli. The addition of the constant 1 to the numerator and the denominator allowed ratio calculations for sessions with no errors, following the recommendation of Hautus (1995). For present purposes, the primary benefit is that, unlike accuracy, logit p provides an unbounded, linear upper range. This property makes logit p especially useful for examining performances with few errors, for example, in comparing the trial-unique and 2-sample outcomes for Subject VS. Moreover, performance differences at 2- and 4-s delays, where accuracy was high but not perfect for all subjects, are shown more clearly with this method. See White (1985) for a discussion of advantages of ratio-based measures over accuracy scores.

Following the recommendation of White (2001), curves were fitted to the data using the negative exponential function:

$$x = \text{logit } p0^{(-bt)} \quad (2)$$

which provides a value for the y intercept when delay (t) = 0 s (logit $p-0$) and a value (b) describing the rate of "forgetting" as a function of delay (t). The y intercept (logit p when $t = 0$ s) is considered to be an estimate of the discriminability of the stimuli with no memory requirement (see White, 1985, 2001). The curves shown in Figure 2 were fitted using Delta Graph 4.5 for Macintosh using least squares nonlinear regression. The values of logit $p-0$, b , and R^2 (an index of how well the

obtained data fit the function) are presented in Table 2.

The shape of the forgetting functions fit to the logit p scores show that the rate of forgetting is different across the two procedures. The estimated values of logit $p-0$ and b are shown in Table 2. There was little variation in the y -intercept (logit $p-0$) across subjects and across procedures. The value of b , however, was considerably higher under the 2-sample procedure for all subjects except Subject DA. That is, there was a greater decrease in sample control as a function of delay duration under the 2-sample procedure than under the trial-unique procedure.

Statistical analysis of the accuracy differences between the trial-unique and 2-sample procedures was conducted using the nonparametric Wilcoxon Matched-Pairs, Signed-Ranks Test, which compares each member of matched pairs using the sign and magnitude of the differences between the paired scores. The logic is that larger differences are weighted more than smaller differences, and the null hypothesis is that there should be an equal number of large and small difference scores of both signs in the sample (Siegel, 1956). Matched pairs consisted of each subject's pooled accuracy at each delay value under the trial-unique and 2-sample procedures. Pairs in which the difference is zero were excluded, and ties in ranked differences were each assigned the mean rank. The accuracy differences between the 2-sample and trial-unique conditions were significant ($p < .005$, sum of ranks = 11.5, Critical value = 32, $N = 19$).

The shape of the decay functions was evaluated using a t test for paired means (too

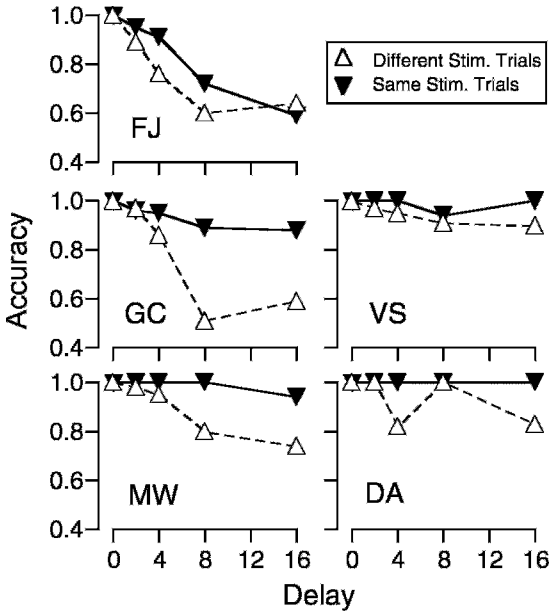


Fig. 3. Experiment 1: Accuracy as a function of delay on Trial N when S+ was the same as, or different from, the stimulus selected on Trial N-1.

few data points were available to conduct the Wilcoxon test). Again, each subject's data under the 2-sample and trial-unique conditions made up the pairs. The values of b for the trial-unique condition were significantly lower than under the 2-sample condition ($p = .04$, $df = 4$), but when the analysis included only values of b from cases where R^2 was greater than .60, the difference was not significant ($p = .1$).

Stimulus-control analysis. One potential source of stimulus control, other than the sample, is the comparison stimulus that was selected on the previous trial (e.g., Wright *et al.*, 1986). Given such interference, errors (i.e., selecting the S-) on Trial N should be more likely when a different comparison stimulus was selected on Trial N-1. Figure 3 shows the accuracy of S+ selections at each delay value when the same stimulus or a different stimulus was selected on the previous trial. Accuracy on same-stimulus trials was equal to or greater than accuracy on different-stimulus trials, with the only exception being Subject FJ under the 16-s delay. These differences were significant using the Wilcoxon Matched-Pairs, Signed-Ranks Test ($p < .005$, sum of ranks = 6, Critical sum = 28 with 18 pairs).

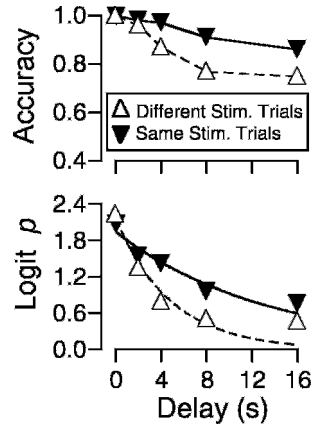


Fig. 4. Experiment 1: Data pooled across subjects at each delay value. Top: Accuracy as a function of delay on Trial N when S+ was the same as, or different from, the stimulus selected on Trial N-1. Bottom: Logit p transformation of data fitted to the negative exponential forgetting function of Equation 2.

Figure 4 presents the pooled accuracies and the forgetting functions for pooled data from same- and different-stimulus trials. The parameters of the forgetting functions for individual subjects and the pooled data are presented in Table 3. For all subjects, fits of Equation 2 showed that the decay parameter, b , was larger for different-stimulus trials than for same-stimulus trials (Table 3). The result of a t -test for paired means also showed that this difference was significant ($p = .02$, $df = 4$).

DISCUSSION

Four of the 5 subjects showed greater accuracy and discriminability under the trial-unique procedure than under the 2-sample procedure. There was no subject for whom accuracy was higher with the 2-sample procedure. Similar outcomes have been reported previously for monkeys (Mishkin & Delacour, 1975; Overman & Doty, 1980) and humans (Williams *et al.*, 1998). When negative exponential decay functions were fit to logit p scores, there was no difference across procedures in initial discriminability (i.e., in the y-intercept, or Logit $p=0$), but there was a substantial difference in the rate of forgetting (value of b) between the two procedures. Discriminability across delays decreased more as a function of delay duration under the 2-sample procedure than under the trial-unique procedure, and this difference cannot be

Table 3

Experiment 1: Parameters of Forgetting Function Fitted to Logit p Transformed Data from Stimulus Control Analysis.

Subject	Different-stimulus trials			Same-stimulus trials		
	Logit p_0	b	R^2	Logit p_0	b	R^2
FJ	1.6	.295	.97	1.3	.135	.97
GC	1.7	.225	.94	1.3	.046	.83
MW	1.6	.101	.97	1.4	.021	.62
VS	1.4	.037	.79	1.4	.011	.80
DA	1.5	.039	.27	1.4	.005	.27
Pooled	2.2	.203	.93	1.9	.074	.93
Pooled ¹	2.1	.21	.95	1.9	.083	.93

¹ Pooled excluding data from subjects with R^2 below .6.

accounted for by differences in the discriminability (without delay factors) of the stimuli used in the two procedures.

The stimulus-control analysis showed that the majority of errors occurred when the stimulus chosen on the previous trial was different from the correct choice on the current trial (i.e., different-stimulus trials). There was a disparity between the data and interpretation of the proportion-based, accuracy analyses and the ratio-based, logit p analysis at 0 s delay, however. Accuracy was 100% on both same- and different-stimulus trials for all subjects. Both the point and interpolated values of logit p_0 were lower for same-stimulus trials for 4 of 5 subjects. Lower logit p_0 values are interpreted as decreased discriminability without the influence of delay (White, 1985).

The ratio-based analysis complements the accuracy analysis in that it better shows differences when accuracy is high, as is often the case with human subjects on DMTS procedures. Some readers may have preferred the log d transformation (see White, 1985), because it can separate discriminability from stimulus bias in a two-choice procedure. Log d cannot be computed for the trial-unique procedure, however, because the calculations assume that Stimulus 1 and Stimulus 2 are the same across sessions. Moreover, our 2-sample procedures were designed to minimize stimulus bias by using a different pair of stimuli in each of the five, 2-sample sessions. To reduce position bias, we used four screen positions for comparison presentation. Log d and logit p are considered to be interchangeable when stimulus bias is minimized (White, 1985).

It should be noted that the fit of the forgetting function using Equation 2 to the obtained data was quite poor in several cases (i.e., R^2 below .60). The poor fit occurred in three cases within the primary data shown in Table 2 (Subject DA under the 2-sample and Subjects MW and VS under the trial-unique procedures), and two cases in the stimulus-control analysis shown in Table 3 (Subject DA). A more conservative approach might be to exclude these data from the quantitative analysis. This would not change the description of the logit p data, except to reduce the number of cases, and in the case of the stimulus-control analysis, the difference in b on same-stimulus trials vs. different-stimulus trials still would attain statistical significance ($p = .03$, $df = 3$). Given that accuracy was unaffected, the primary results remain despite the poor fit of Equation 2 in some cases. However, the poor fit of the forgetting function to the obtained data may be of theoretical importance relating to the effects of interference on such forgetting functions. That is, other equations may fit DMTS performance better than negative exponential decay function in the absence of intertrial interference.

Although ratio-based quantitative measures have advantages, there also are some limitations. The value of logit p is affected by the number of trials in the calculation, which can introduce difficulties in comparing across conditions or studies with different numbers of trials. This effect is compounded when the correction for 100% accuracy (adding a constant of .5 or 1) is used. This correction causes discriminability to be underestimated, al-

though the magnitude of this problem decreases as the total number of trials increases. One advantage of accuracy scores is that they provide a common metric by which to compare results across conditions, subjects, and studies with different numbers of trials.

The results of the stimulus-control analysis for the 2-sample procedure usually were consistent with results of nonhuman studies of proactive interference (see Edhouse & White, 1988; Wright *et al.*, 1986). When sample control of comparison selection was diminished, the comparison selected on the previous trial often controlled selection. Thus, much of the accuracy difference between the trial-unique and 2-sample procedures was due to errors on different-stimulus trials, indicating that intertrial interference factors account for the lower accuracy and the higher rate of forgetting in the 2-sample procedure.

EXPERIMENT 2

Studies with nonhuman subjects have shown that accuracy on the 2-sample DMTS task is a function of ITI duration, with longer ITIs producing higher accuracy (Dunnett & Martel, 1990; Edhouse & White, 1988; Holt & Shafer, 1973; Roberts & Kraemer, 1982; White, 1985). Edhouse and White, for example, showed increases in accuracy as the ITI increased across 5, 10, and 20 s. Under a delayed position-matching procedure, Mackay and Gould (1992) varied ITI duration from 0.5 to 10 s in 6 human subjects with moderate mental retardation. The retention interval was held constant, and different ITI durations were presented in blocks of 10 consecutive trials per session. Accuracy was highest with the longest ITI. The present experiment varied the ITI duration under DMTS procedures similar to those used with nonhuman subjects (and in Experiment 1). ITI durations were 2 s and 8 s.

Both the 2-sample and trial-unique tasks were presented. The trial-unique task was included to examine effects of ITI duration under conditions that should be relatively free of proactive interference. A proactive interference perspective predicts especially low accuracy and discriminability in the 2-sample task under the 2 s ITI, with different-stimulus trials affected more than same-stimulus trials. If, however, ITI duration affects accuracy and

discriminability similarly under both procedures, then processes other than proactive interference would be involved.

METHOD

There were 5 subjects (see Table 1), 2 of whom (FJ and GC) had participated in Experiment 1. With the exception of the ITI manipulation, session procedures were identical to those in Experiment 1. Each subject received five blocks of four sessions each. Each block contained one of each combination of session type (trial-unique or 2-sample) and ITI duration (2 or 8 s). The order of the four session types was randomized within blocks and across subjects, except that (a) all subjects began the first block with a trial-unique session, and (b) no more than two consecutive sessions were of the same type or had the same ITI (across blocks).

RESULTS AND DISCUSSION

Figure 5 shows, for each subject, accuracy (left panels) and logit p values (right panels) as a function of delay duration under both the trial-unique and 2-sample procedures with both 2-s and 8-s ITIs. Logit p values were calculated using Equation 1. Accuracy and logit p scores usually were higher under the trial-unique procedure for both ITI durations. Forgetting functions were fitted to the data using Equation 2. Forgetting functions for trial-unique sessions appeared to be more linear and to have shallower slopes than functions fitted to 2-sample sessions. As shown in Table 4, values of b were consistently smaller for trial-unique sessions. The smaller b values indicate that discriminability decreased less at comparable delays under the trial-unique condition than under 2-sample sessions regardless of ITI duration.

Effects of ITI duration can be seen by comparing the open and closed symbols in Figure 5. Under the 2-sample procedure, accuracy (left panels) was generally greater with the 8-s ITI than under the 2-s ITI with delays of 4 s and longer. A similar pattern was apparent in the logit p transformations and fitted forgetting functions (right panels). For all subjects, the values of b (given in Table 4) were smaller under the 8-s ITI than under the 2-s ITI, indicating a slower rate of forgetting in the former case.

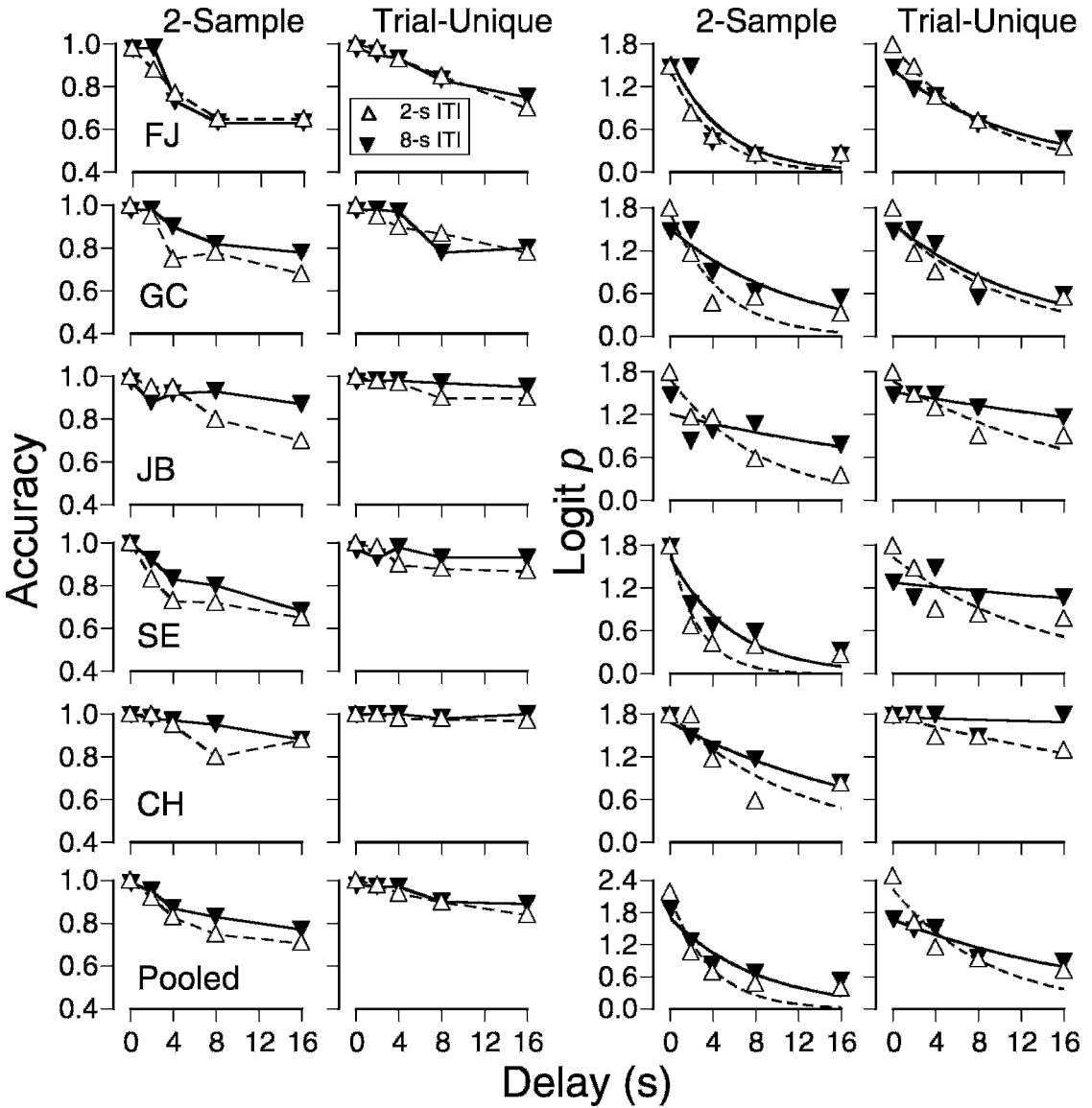


Fig. 5. Experiment 2: Accuracy (two left columns) and logit p values (two right columns) as a function of delay, for individual subjects and pooled across subjects, under the 2-sample and trial-unique procedures, under two durations of intertrial interval. Logit p data were fitted to the negative exponential forgetting function of Equation 2; see text for details.

Accuracy measures for 2-s and 8-s ITI durations at each delay value were compared using the Wilcoxon Matched-Pairs, Signed-Ranks Test. Each subject's accuracy at each delay at 2-s and 8-s ITI durations made up the matched pairs. As shown in Table 5, the differences between 2-s and 8-s ITI durations were significant for the 2-sample sessions, but not significant for the trial-unique sessions.

This result is consistent with increased interference by the stimulus selected on the previous trial (proactive interference) with shorter ITI durations, as there was no significant effect of this variable under conditions designed to minimize such interference.

Differences between the values of b as a function of ITI were evaluated using a simple t test for paired means. Again, each subject's

Table 4

Experiment 2: Parameters of Forgetting Function Fitted to Logit p Transformed Data from Trial-unique and Two-sample Sessions with 2-s or 8-s ITIs.

Subject	ITI	Two-sample			Trial-unique		
		Logit p_0	b	R^2	Logit p_0	b	R^2
FJ	2 s	1.4	.243	.96	1.8	.112	.99
	8 s	1.6	.205	.82	1.4	.082	.92
GC	2 s	1.7	.210	.86	1.6	.095	.86
	8 s	1.5	.087	.86	1.6	.079	.83
JB	2 s	1.7	.120	.95	1.7	.053	.85
	8 s	1.2	.030	.40	1.5	.017	.92
SE	2 s	1.7	.366	.91	1.6	.071	.76
	8 s	1.6	.174	.89	1.3	.012	.21
CH	2 s	1.8	.082	.73	1.8	.021	.84
	8 s	1.7	.048	.95	1.8	.002	.33
Pooled	2 s	2.1	.259	.92	2.2	.112	.87
	8 s	1.7	.122	.88	1.7	.047	.87
Pooled ¹	8 s	1.8	.145	.91	1.65	.066	.92

¹ Pooled excluding data from subjects with R^2 below .6.

b values under the 2-s and 8-s ITI durations made up the pairs. The values of b for 2-s ITIs were significantly greater than those for 8-s ITIs under both 2-sample ($p = .016$, $df = 4$) and trial-unique conditions ($p = .007$, $df = 4$). These results support the conclusion that accuracy was consistently lower with the shorter ITI duration, and the rate of accuracy decline was higher with the shorter ITI duration.

Stimulus-control analysis. Figure 6 shows, for each subject and pooled across subjects, accuracy on same-stimulus and different-stimulus trials under the 2-s and 8-s ITI durations. For the most part, accuracy was higher for same-stimulus trials than for different-stimulus trials at both ITI durations, and, for 3 subjects at the 16-s delay duration, the difference between same- and different-stimulus trials was considerably larger under the 2-s compared with the 8-s ITI duration trials.

The differences in accuracy were assessed using the Wilcoxon Matched-Pairs Rank-Order test as in Experiment 1. Results are presented in Table 6. In the comparisons of same- versus

different-stimulus trials, the pairs were composed of accuracy on same- versus different-stimulus trials for each subject at each delay value, yielding 25 potential pairs for analysis. Separate analyses were conducted for 2-s and 8-s ITI durations. As predicted, accuracy was significantly higher on same-stimulus trials than on different-stimulus trials, reproducing the effect reported in Experiment 1, and the difference was significant for both ITI durations. The bottom half of Table 6 presents results from the Wilcoxon test for 2- versus 8-s ITI duration trials on different-stimulus trials and same-stimulus trials. Accuracy differed across the two ITI durations only for different-stimulus trials, under which intertrial or proactive interference should decrease accuracy. The finding that the ITI duration affected different-stimulus trials, but not same-stimulus trials, supports the interpretation that interfering stimulus control by the stimulus selected in the previous trial is greater with short ITI durations.

Data from same- and different-stimulus trials were transformed to logit p values, and

Table 5

Experiment 2: Results of Wilcoxon Matched-Pairs, Signed-Ranks Test Comparing Accuracy on 2- vs. 8-s ITI Durations Under Trial-unique and Two-sample Procedures.

Comparisons	Significance level*	Number of pairs	Critical sum of ranks	Sum of ranks
2-s ITI vs. 8-s ITI				
Two-sample	$p < .01$	21	49	47.5
Trial-unique	$p > .05$	20	52	68.5

* One-tailed.

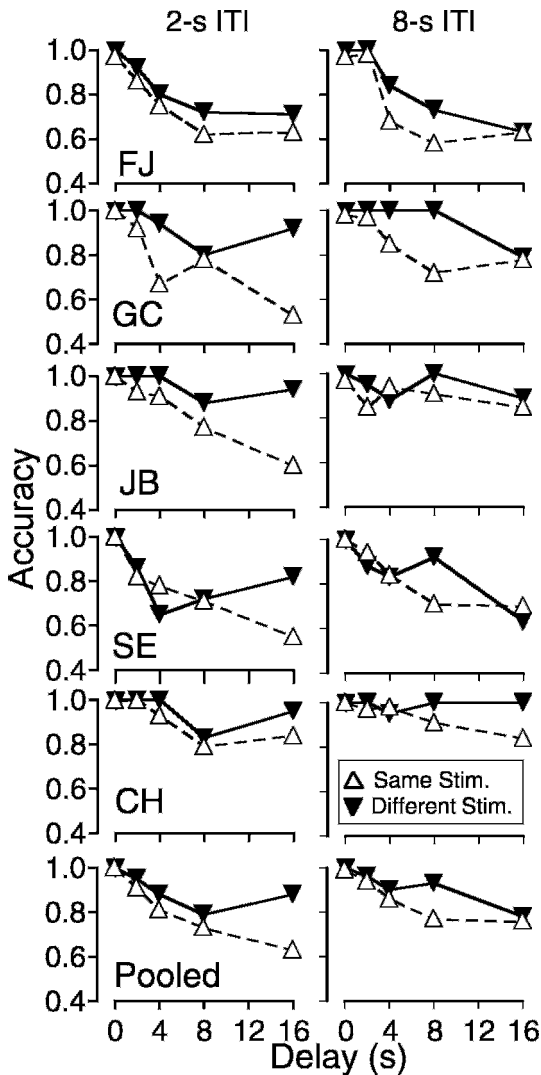


Fig. 6. Experiment 2: Accuracy as a function of delay and intertrial interval duration when S+ was the same as, or different from, the stimulus selected on Trial N-1.

forgetting functions were fitted to individual-subject and pooled data using Equation 2. The parameters of the forgetting functions are presented in Table 7. For all subjects, the values of b were higher for different-stimulus trials than for same-stimulus trials at both ITI durations. This indicates a greater rate of decline in discriminability with increased delay duration when the sample was different from the previous comparison choice than when the sample was the same as the previously chosen stimulus. Accuracy and discriminability were

lower for different-stimulus trials, with the effect more pronounced and consistent under the 2-s ITI than under the 8-s ITI.

Table 7 shows the results of fitting Equation 2 to data from same- and different-stimulus trials pooled across subjects. Values of the b parameter associated with same-stimulus and different-stimulus trials were more discrepant under the 2-s ITI than under the 8-s ITI, an effect consistent with the notion that longer ITI durations reduce interference experienced during different-stimulus trials. Values of the b parameter at the 2-s and 8-s ITI durations were more discrepant for different-stimulus trials than for same-stimulus trials, an effect consistent with the notion that different-stimulus trials generate intertrial interference. Differences in the b values were tested using a repeated-measures, two-way analysis of variance. There was a significant main effect of ITI duration $F(1, 16) = 8.14, p = .011$ indicating that the value of b (decay rate) was significantly larger under the 2-s ITI ($M = .17, SD = .10$) than under the 8 s ITI ($M = .09, SD = .08$), and of trial type, $F(1, 16) = 14.03, p = .002$, with the value of b significantly larger for different-stimulus trials ($M = .19, SD = .10$) than for same-stimulus trials ($M = .08, SD = .06$). There was no significant interaction effect $F(1, 16) = 3.55, p = .130$. Figure 7 shows the mean values of b to illustrate this analysis. The significant main effect of ITI duration is due primarily to a much higher b value on the different-stimulus trials. A t -test was conducted comparing 2-s vs. 8-s ITI durations on same-stimulus trials. This difference was not significant $t(8) = 1.03, p = .330$. The significant main effect of trial type is due primarily to higher b values under the 2-s ITI duration. A t -test comparing same- and different-stimulus trials under the 8-s ITI duration was not significant $t(8) = 1.38, p = .204$. These results are also in accordance with the conceptualization of intertrial interference as the source of the competing stimulus control on delayed matching procedures.

In summary, Experiment 2 reproduced the effects of trial-unique versus 2-sample matching under both a longer and a shorter ITI duration relative to the ITI duration observed in Experiment 1. In answer to the primary question, findings were consistent with the idea that increasing the ITI can decrease intertrial interference (e.g., White, 1985).

Table 6

Experiment 2: Results of Wilcoxon Matched-Pairs, Signed-Ranks Test for Accuracy on Same-Stimulus and Different-Stimulus Trials.

Comparisons	Significance level*	N	Critical sum of ranks	Sum of ranks
Same vs Different trials				
2-s ITI	$p < .01$	20	43	16
8-s ITI	$p < .01$	23	62	49
2-s ITI vs. 8-s ITI				
Different-stimulus trials	$p < .025$	23	73	73
Same-stimulus trials	$p > .05$	18	40	68.5

* One-tailed.

Effects of the ITI manipulation were especially clear in the higher accuracy under the 2-sample procedure, both overall and on different-stimulus trials, with the longer ITI. Effects of the ITI were not limited to the 2-sample procedure, however, but were seen also in the values of b under the trial-unique conditions. The latter finding suggests that the ITI effect may reflect more than diminished control by the stimulus selected on a previous trial.

GENERAL DISCUSSION

Compared to 2-sample delayed-matching procedures, trial-unique procedures produce higher accuracy and discriminability. Although this has been a robust finding in nonhumans (e.g., Wright et al., 1986), the effect has been studied rarely in humans. Along with Williams et al. (1998), the present

study extends the generality of this effect to humans. Across the two experiments in the present study, all 8 subjects showed high accuracy with a delay of 0 s under both procedures. At greater delays, accuracy usually was higher in the trial-unique procedure, and no subject responded more accurately on the 2-sample task than on the trial-unique task. Stimulus-control analyses indicated that errors were more likely when the S+ on Trial N was different from the comparison selected on Trial N-1, thus replicating findings from the nonhuman literature on proactive interference (e.g., Wright et al., 1986).

In extending the proactive-interference effect to humans, the present study also extends the demonstration to subjects who show generalized identity matching. One might consider nonhuman subjects to be especially likely to develop multiple sources of stimulus

Table 7

Experiment 2: Characteristics of Forgetting Functions for Same-Stimulus and Different-Stimulus Trials at Each ITI Duration.

Sub	ITI	Same-stimulus Trials			Different-stimulus Trials		
		Logit $p-0$	b	R^2	Logit $p-0$	b	R^2
CH	2 s	1.3	.160	.90	1.3	.245	.96
	8 s	1.3	.003	.20	1.5	.055	.95
FJ	2 s	1.3	.045	.89	1.6	.267	.89
	8 s	1.4	.153	.93	1.4	.216	.78
GC	2 s	1.4	.032	.46	1.5	.150	.97
	8 s	1.4	.039	.58	1.4	.109	.80
JB	2 s	1.0	.125	.40	1.6	.319	.91
	8 s	1.2	.023	.26	1.1	.026	.37
SE	2 s	1.7	.108	.65	1.9	.274	.95
	8 s	1.6	.067	.62	1.6	.195	.94
Pooled	2 s	1.7	.110	.65	1.9	.270	.95
	8 s	1.7	.090	.81	1.6	.130	.88
Pooled ¹	2 s	1.5	.142	.65	1.9	.274	.95
	8 s	1.7	.087	.81	1.7	.162	.90

¹ Pooled excluding data from subjects with R^2 below .6.

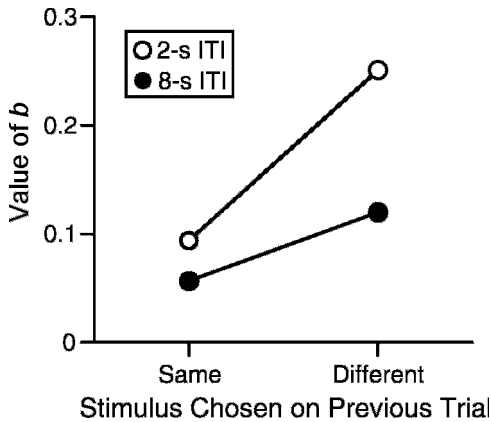


Fig. 7. Mean values of b when the current trial was preceded by the subject selecting, on the previous trial, the same or different stimulus as the sample on the current trial. Open and closed symbols are means from sessions with 2-s and 8-s ITI durations, respectively.

control, including proactive interference, given that nonhumans (a) begin training at chance levels, (b) require many training trials to reach high accuracy levels, during which stimulus control presumably is fluctuating, and (c) are exposed to the same two stimuli throughout a study. Unlike the nonhuman subjects of previous studies, our subjects responded to new pairs of stimuli errorlessly (when the delay was 0 s). Thus, regardless of whether the matching baseline involves generalized identity matching, or is essentially equivalent to arbitrary matching (Carter & Eckerman, 1975; Sidman et al., 1982), proactive interference can be shown in 2-sample DMTS procedures.

In Experiment 2, accuracy was higher with the longer of two ITIs, reproducing findings in pigeons (e.g., Edhouse & White, 1988). Moreover, the ITI-duration effect was most prominent on different-stimulus trials, as would be expected if interference is increased with shorter ITI durations. Accuracy and forgetting functions showed little effect of ITI duration under the trial-unique procedure, in which intertrial interference is minimized.

Both experiments demonstrated that the proportion-based measure of accuracy and ratio-based analyses (logit p) can provide complementary information, as discussed following Experiment 1. The present data support White's (1985) suggestion that fitting a negative exponential forgetting function

(Equation 2) to data transformed via logit p enhances the visual representation of highly accurate performance data. In addition, the curve-fitting process yields formal indices, both of discriminability without remembering requirements, and of the rate of discriminability decline as a function of delay, which are not available with accuracy data alone.

Stimulus-control analyses revealed that, as accuracy decreased with delay duration, responses on Trial N often were controlled by the comparison that was selected on Trial N-1. A high proportion of errors under the 2-sample procedure can be attributed to this source of stimulus control. The stimulus-control analysis presented here, however, does not allow an estimate of true accuracy and forgetting functions independent of interference. Logic suggests that an equivalent portion of correct responses on same-stimulus trials also might be under the stimulus control of the comparison selected on the previous trial. Supplementary sources of stimulus control that promote correct responses might be called proactive facilitation. Thus, on both same-stimulus and different-stimulus trials, accuracy and logit p scores could reflect a mixture of stimulus control by the sample on Trial N and by the comparison selected on Trial N-1. If such facilitation occurs, same-stimulus trials (under the 2-sample procedure) might sometimes produce higher accuracy than the trial-unique procedure does. In the present study, however, very high accuracy introduced ceiling effects.

Taken as a whole, the data suggest that procedural differences across studies involving DMTS can affect the amount of intertrial interference, thus changing the shape of the function relating accuracy to delay duration. Important differences might be found not only in comparing trial-unique and 2-sample procedures, but also across 2-sample procedures that have different values of certain parameters. One procedural variable affecting 2-sample outcome is the duration of the ITI, as shown in the present study and others (e.g., Edhouse & White, 1988). Another possibility, suggested by the stimulus-control analyses, is that more subtle differences in session configuration might affect the delay function. For example, it is common practice for trial sequences to be randomized, but with a specified limit on the number of consecutive trials

with the same sample. That limit is usually set at two or three trials. Relative to a three-trial limit, a two-trial limit would have a somewhat larger number of different-stimulus trials. Given factors that weaken the control of the current sample stimulus within a trial (e.g., delay duration), the two-trial-limit procedure might produce lower accuracy than the three-trial limit.

These procedural factors—the diminishing of proactive interference in the trial-unique procedure and the modulation of proactive interference across variations in the 2-sample procedure—must be taken into consideration when arranging baselines for the study of nonbehavioral variables on DMTS performance. For example, researchers who use DMTS procedures in the study of neurological (e.g., progressive disorders such as Alzheimer's disease, see White & Ruske, 2002) or pharmacological variables should be alert to differences in the stimulus control shown under the conditional versus trial-unique procedures. The trial-unique procedure minimizes one source of interference, and may provide a purer measure of short-term remembering (stimulus control across a temporal distance) than does the more commonly used 2-sample procedure. Moreover, the difference in accuracy levels produced by the two procedures may be taken as a measure of the amount of proactive interference. Thus, using the procedures conjointly provides a way of assessing accuracy with and without proactive interference, thereby making it possible to characterize both the direct memory-altering effects of factors such as drugs and their interactions with interference.

REFERENCES

- Blough, D. S. (1959). Delayed matching in the pigeon. *Journal of the Experimental Analysis of Behavior*, *2*, 151–160.
- Carter, D. E., & Eckerman, D. A. (1975, February 21). Symbolic matching by pigeons: Rate of learning complex discriminations predicted from simple discriminations. *Science*, *187*, 662–664.
- Constantine, B., & Sidman, M. (1975). Role of naming in delayed matching-to-sample. *American Journal of Mental Deficiency*, *79*, 680–689.
- Cumming, W. W., & Berryman, R. (1965). The complex discriminated operant: Studies of matching-to-sample and related problems. In D. I. Mostofsky (Ed.), *Stimulus generalization* (pp. 284–330). Stanford, CA: Stanford University Press.
- Dalton, A. J., & Crapper-McLachlan, D. R. (1984). Incidence of memory deterioration in aging persons with Down's syndrome. In J. M. Berg (Ed.), *Perspectives and progress in mental retardation: Biomedical aspects* (Vol. II, pp. 55–62). Baltimore, MD: University Park Press.
- Dube, W. V. (1991). Computer software for stimulus control research with Macintosh computers. *Experimental Analysis of Human Behavior Bulletin*, *9*, 28–30.
- Dube, W. V., McIlvane, W. J., & Green, G. (1992). An analysis of generalized identity matching-to-sample test procedures. *The Psychological Record*, *42*, 17–28.
- Dunnett, S. B., & Martel, F. L. (1990). Proactive interference effects on short-term memory in rats: I. Basic parameters and drug effects. *Behavioral Neuroscience*, *104*, 655–665.
- Edhouse, V. W., & White, K. G. (1988). Sources of proactive interference in animal memory. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 56–70.
- Harper, D. N., McLean, A. P., & Dalrymple-Alford, J. C. (1994). Forgetting in rats following medial septum or mammillary body damage. *Behavioral Neuroscience*, *108*, 691–702.
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, and Computers*, *27*, 46–51.
- Hautus, M. J. (1997). Calculating estimates of sensitivity from group data: Pooled versus averaged estimators. *Behavior Research Methods, Instruments and Computers*, *29*, 556–562.
- Holt, G. L., & Shafer, J. M. (1973). Function of the intertrial interval in matching to sample. *Journal of the Experimental Analysis of Behavior*, *19*, 181–186.
- House, B. J., Brown, A. L., & Scott, M. S. (1974). Children's discrimination learning based on identity or difference. In H. W. Reese (Ed.), *Advances in child development and behavior* (pp. 1–45). New York: Academic Press.
- Mackay, H. A., & Gould, D. D. (1992). Relative sample recency and proactive interference in adults with mental retardation. *Experimental Analysis of Human Behavior Bulletin*, *10*, 1–5.
- Mishkin, M., & Delacour, J. (1975). An analysis of short-term visual memory in the monkey. *Journal of Experimental Psychology: Animal Behavior Processes*, *1*, 326–334.
- Overman, W. H., Jr., & Doty, R. W. (1980). Prolonged visual memory in macaques and man. *Neuroscience*, *5*, 1825–1831.
- Park, S., Püschel, J., Sauter, B. H., Rentsch, M., & Hell, D. (2002). Visual object working memory function and clinical symptoms in schizophrenia. *Schizophrenia Research*, *59*, 261–268.
- Roberts, W. A., & Kraemer, P. J. (1982). Some observations of the effects of intertrial interval and delay on delayed matching to sample in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *8*, 342–353.
- Roitblat, H. L., & Harley, H. F. (1988). Spatial delayed matching-to-sample performance by rats: Learning, memory, and proactive interference. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 71–82.

- Saunders, K. J., & Spradlin, J. E. (1993). Conditional discrimination in mentally retarded subjects: Programming acquisition and learning set. *Journal of the Experimental Analysis of Behavior*, *60*, 571-585.
- Scott, M. S., & House, B. J. (1978). Repetition of cues in children's oddity learning and transfer. *Journal of Experimental Child Psychology*, *25*, 58-70.
- Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior*, *37*, 23-44.
- Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- Sugimura, T., & Iyoda, Y. (1982). Children's oddity learning as a function of age and stimulus repetition. *Psychologia*, *25*, 244-250.
- Weinstein, B. (1941). Matching-from-sample by rhesus monkeys and by children. *Journal of Comparative Psychology*, *31*, 195-213.
- White, K. G. (1985). Characteristics of forgetting functions in delayed matching to sample. *Journal of the Experimental Analysis of Behavior*, *44*, 15-34.
- White, K. G. (2001). Forgetting functions. *Animal Learning & Behavior*, *29*, 193-207.
- White, K. G., Parkinson, A. E., Brown, G. S., & Wixted, J. T. (2004). Local proactive interference in delayed matching to sample: The role of reinforcement. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*, 83-95.
- White, K. G., & Ruske, A. C. (2002). Memory deficits in Alzheimer's disease: The encoding hypothesis and cholinergic function. *Psychonomic Bulletin & Review*, *9*, 426-437.
- Williams, D. C., Dube, W. V., Johnston, M. D., & Saunders, K. J. (1998). Conditional vs. trial-unique delayed matching to sample. *American Journal on Mental Retardation*, *103*, 186-192.
- Wixted, J. T. (1989). Nonhuman short-term memory: A quantitative reanalysis of selected findings. *Journal of the Experimental Analysis of Behavior*, *52*, 409-426.
- Wright, A. A., Urcuioli, P. J., & Sands, S. F. (1986). Proactive interference in animal memory. In D. F. Kendrick, M. E. Rilling, & M. R. Denny (Eds.), *Theories of animal memory* (pp. 101-125). Hillsdale, NJ: Lawrence Erlbaum.

Received: August 10, 2001

Final acceptance: June 7, 2006