Recall of three-item sequences by pigeons

H. S. TERRACE, SHAOFU CHEN, and VIKRAM JASWAL
Columbia University, New York, New York

Pigeons were trained to recall an arbitrary sequence on a delayed matching-to-successive-samples (DMTSS) task. Sample items were presented successively and then displayed simultaneously. Subjects were required to respond to them in the order in which they appeared. In Experiment 1, pigeons responded correctly on 75% of the trials on a two-item DMTSS task but at a chance level of accuracy on a three-item task. In Experiment 2, pigeons who learned to produce a three-item sequence prior to DMTSS training mastered a three-item DMTSS task at a 75% level of accuracy. Control groups, trained initially with the same items on nonserial tasks, performed as poorly on a three-item DMTSS task as the naive subjects of Experiment 1. It was hypothesized that pigeons that first learned to produce a three-item list were able to recall three-item samples in DMTSS because they had learned to represent three-item sequences.

In a typical experiment on human memory, subjects are asked to reproduce an arbitrary sequence of items, for example, the string of numbers 7, 2, 4, 9, 1, 5. The experimenter notes how many items the subject recalls correctly, how many of those items were reported in their original ordinal positions, how many were shifted, and so on. This deceptively simple procedure has given rise to a sizable literature on human serial recall (Crowder, 1976; Murdock, 1974). By contrast, there is a dearth of information about an animal's ability to recall a sequence of arbitrary items.

A major obstacle to our understanding of serial recall by animals is the lack of research on sequence discrimination and sequence production in the same subject. We do not know, for example, whether an animal who can discriminate a sequence can produce that sequence, and vice versa. Prima facia, the ability to recognize and to produce an arbitrary sequence is critical for its recall. The purpose of the present study was to train pigeons to recall an arbitrary sequence on a combined recognition–production task. At the beginning of each trial, three items were presented successively. Recall was evaluated by the pigeons' ability to respond to those items, in the appropriate order, when they were displayed simultaneously.

Sequence Discrimination

The simplest procedure for training an animal to discriminate sequences is to reinforce responding differentially following the presentation of a two-item sequence. A subject might, for example, be reinforced for responding after the sequence A→B is presented, but not after its converse, B→A (Wasserman, Nelson, & Larew, 1980). For the purposes of understanding how an animal represents a sequence, the task of discriminating A→B from B→A is too

simple. Subjects can respond correctly by noting the presence or absence of A at the start of a trial. This is true even when the complexity of the task is enhanced by requiring subjects to discriminate the sequence A→B from all possible two-item permutations: B→A, A→A, and B→B (Weisman & Dodd, 1979; Weisman, Wasserman, Dodd, & Larew, 1980). In that instance, differential responding to A→B and non–A→B sequences could result from the application of the following rules: (1) Determine whether the first item is A; (2) if A, maintain an overt response until the next item is presented; (3) if the next item is B, maintain the overt response until the choice stimulus appears. Maintaining an overt response could take many forms, for example, facing the stimulus display, standing on one leg, turning 45° from the stimulus display, and so on.

An association between Item B and an overt response provides the subject with an unambiguous basis for responding differentially to two-item sequences. Such associations are of no help, however, when it comes to discriminating three-item sequences. The association between Item B and an overt response could signify either of the following sequences: A→B or A→B→B, to consider but one example. Any rule that disambiguates these sequences is, by definition, a rule that orders the second and third items. To discriminate a three-item sequence, the subject must encode item *and* order information about each item.

A variety of experiments have shown that pigeons can discriminate three-item sequences (Roitblat, Bever, Helweg, & Harley, 1991; Roitblat, Scopatz, & Bever, 1987; Terrace, 1986). Consider, for example, Terrace's (1986) procedure for training a pigeon to discriminate a sequence composed of three colors (A, B, and C). A *yes* response was rewarded for identifying A→B→C sequences, and a *no* response for identifying non–A→B→C sequences. Non–A→B→C sequences were generated by a "without-replacement" rule for one group and by a "with-replacement" rule for another. Under the without-replacement condition (items were not repeated), subjects were trained to discriminate A→B→C from five of the six possible non–A→B→C

sequences. The sixth sequence was held in reserve for a generalization test. Under the with-replacement condition (items could be repeated), subjects were trained to discriminate A→B→C from 22 of the 26 possible non–A→B→C sequences. The remaining four non–A→B→C sequences were held in reserve for a generalization test. Both groups learned to discriminate all of the sequences on which they were explicitly trained. They also discriminated, with no decrement in accuracy, non–A→B→C sequences that were presented for the first time on generalization tests.

Sequence Production

The traditional method for studying sequence production in animals is to train an animal to execute a *successive* chain, for example, a succession of turns in a maze. In the present experiments, however, subjects were trained to produce *simultaneous* chains (Terrace, 1984). A key feature of a simultaneous chain is the simultaneous presentation of list items throughout each trial (typically on a touch-sensitive video monitor). A variety of species have been trained to execute simultaneous chains containing as many as six items: pigeons (Straub & Terrace, 1981; Terrace, 1987), monkeys (Chen, Swartz, & Terrace, in press; D'Amato & Colombo, 1988; Swartz, Chen, & Terrace, 1991), young children (McGonigle & Neapolitan, in press), and human adults (Stromer & MacKay, 1993).

A familiar example of a simultaneous chain is the sequence of numbers a customer executes when entering a personal identification code to obtain money from a cash machine. Typically, an array of arabic numerals is displayed simultaneously on a touch-sensitive video monitor. Unlike a cash machine display, however, the physical configuration of the items used in experiments with animals varies from trial to trial. Variation of the location of list items prevents subjects from executing the required sequence as a stereotyped chain of responses.

No cues are provided regarding the required sequence of responses while the subject responds to successive items. Thus, none of the items of a simultaneous chain can function as a cue to direct the subject's responses from one item to the next. In a successive chain, punctate cues (exteroceptive and proprioceptive) can direct a subject through a sequence, for example, the choice points of a maze and feedback from turning left or right at a particular choice point. Because punctate cues are absent in a simultaneous chain, the required sequence must be guided by the subject's moment-to-moment *representation* of its place in the sequence. The ability of an animal to represent its position in a simultaneous chain provides a basis for comparing animal and human memory that is lacking in the case of successive chains.

Taken together, studies of sequence discrimination and sequence production provide abundant evidence that animals can represent the order of three or more list items. In each instance, the animal evaluates its position in a representation of the sequence it is discriminating or producing. On a sequence discrimination task, in which sequences are generated by a with-replacement rule, a subject has to match its representation in working memory of each of the

items displayed on a given trial with a representation of the positive sequence in reference memory. When sequences are generated by a without-replacement rule, the matching of representations of items in working and long-term memory is necessary only in the case of the first two items. If those items match, the subject can earn reinforcement by responding *yes*; if not, by responding *no*. On a production task, changes in the physical configuration of list items from one trial to the next ensure that the subject cannot execute the required sequence as a fixed motor pattern. The only way a subject can produce the sequence is to locate its position in a representation of the sequence when moving from one item to the next.

Memory of Fixed Versus Variable Sequences

The ability to recognize regularly occurring environmental sequences and to learn to produce fixed sequences has obvious adaptive value. Because of their regularity, such sequences are best stored in long-term memory. It is also important, however, for animals to respond appropriately to serially organized information the order of which is variable. Under these circumstances, the animal must bridge the gap between a sequence it perceives and the behavior it is subsequently required to execute. The only way to bridge that gap is to rely on a representation of the sequence.

There is ample evidence that pigeons can discriminate sequences of their own behavior (Grayson & Wasserman, 1979; Jitsumori & Sugimoto, 1982; Shimp, 1976a, 1976b; Shimp & Moffitt, 1974). For example, consider Shimp's procedure for evaluating working memory of three-item sequences. During each trial, three stimuli were displayed successively on either of two response keys, left (L) or right (R). The order in which each key was illuminated varied randomly from trial to trial (e.g., LLR, LRL, LRR, and so forth). The pigeon's task was to peck whichever key was illuminated. At the end of each trial, a probe stimulus directed the pigeon to report which key it pecked first, second, or third. Following a red probe, for example, the pigeon was rewarded for pecking the key it had pecked first. The other two probes directed the pigeon to peck the key it had pecked second and third.

Experiments with rats in a runway apparatus have also provided evidence that an animal can discriminate sequences of its own behavior (Capaldi, Miller, Alptekin, & Barry, 1990; Haggbloom, Birmingham, & Scranton, 1992). The basic procedure was to divide trials into study and test phases that were separated by intervals as long as 15 min. During the study phase, the rat was given a series of rewarded (R) and nonrewarded (N) trials (e.g., RRN or RN). On RRN sequences, rats would run fast, fast, then slow; on RN sequences, fast and then slow. In order to run at the appropriate speed during the second part of each trial, the rat had to recall the entire series on which it had run during the study phase.

Paradigms like these in which pigeons and rats are trained to recall an arbitrary sequence of responses they had produced during the study phase of a trial are limited, however, by the type of events that can function as cues. Another limitation is the absence of analogous experiments

on human memory. Both limitations can be overcome by a paradigm that combines the basic features of production and recognition tasks. That paradigm, which will be referred to as delayed-matching-to-successive-samples (DMTSS), is an elaboration of a single-sample delayed matching-to-sample (DMTS) task (Skinner, 1957). In the simplest form of a DMTSS task, two items are first presented successively (S1→S2). Following the termination of S2, both items of the sample are displayed simultaneously. The subject is then required to reconstruct the order in which the sample items appeared by responding first to S1 and then to S2. In some instances, one or more stimuli are added to the display as distractors. Distractors are chosen from the set of stimuli that are not included in the sample on a particular trial.

The DMTSS paradigm requires subjects to recognize the sequence of items that are presented as the sample during the first phase of each trial and to recall that sequence during the second phase of the trial. The measure of recall is the accuracy with which the sequence of sample items is produced. Performance on a DMTSS task provides an assay of working memory for both item and order information because subjects must recall which sample items were presented on each trial and the order in which the items appeared.

None of the handful of experiments in which animals have been trained on a DMTSS task used samples containing more than two items (Devine, Burke, & Rohack, 1979; MacDonald, 1993). It is relatively easy to execute a two-item sequence, because the start and the end items can be readily associated with the beginning and the end of a trial, respectively (Straub & Terrace, 1981; Swartz et al., 1991). By contrast, three-item sequences like those used here require the subject to identify an interior item that does not lend itself to an association with any trial events. The purpose of Experiment 1 was to evaluate a pigeon's ability to recall a three-item sample on a DMTSS task.

EXPERIMENT 1

To minimize the difficulty of the DMTSS task, we incorporated features of training procedures used in previous studies of DMTSS performance. Devine and Jones (1975) showed that accuracy of performance on a DMTSS task was highest when each item of the sample was drawn from a different category. Thus, color→shape and shape→color sequences were recalled more accurately than were color→color and shape→shape sequences. In an experiment in which pigeons were trained on a two-item DMTSS task (MacDonald, 1993), the items differed only with respect to color. MacDonald attributed the failure of some of her subjects to exceed a chance level of performance to interference between sample items. In the present experiment, each sample item was drawn from a different category (color, shape, and line orientation) to maximize item discriminability.

In MacDonald's (1993) experiment, accuracy of the first and second choices during the recall phase of each trial

was influenced by the relative durations of S1 and S2. The longer an item's duration, the greater the accuracy of choosing that item. In the present experiment, the duration of S1 was longer than that of S2 during the phase of training in which two-item samples were presented. Similarly, when three-item samples were presented, the duration of S1 was longer than that of S2, which, in turn, was longer than the duration of S3.

In a pilot experiment, MacDonald (1993) found that pigeons performed more accurately when they were required to respond in the order S2→S1 than in the order S1→S2. Because of that result, and evidence that animals have an innate tendency to respond first to the most recent of a series of stimuli (Roberts & Grant, 1976), MacDonald used a "backward" procedure in which her pigeons were required to respond to S2 before responding to S1. A pilot experiment conducted prior to the present experiment, however, failed to reveal any difference in the difficulty of a two-item DMTSS task with respect to the order in which pigeons were required to report the items of the sample. To the extent that differences were observed, they favored the forward order of reporting. For that reason, and also to be consistent with procedures used in studies of DMTSS performance with monkeys (Devine & Jones, 1975) and recall with human subjects, the subjects of the present experiment were required to recall samples in a forward order.

Method

Subjects

The subjects were 8 experimentally naive male White Carneaux pigeons (*Columba livia*). Their age at the start of training was approximately 12 months. Each pigeon was maintained at 80% of its free-feeding weight throughout the experiment and was provided with constant access to water and grit. The temperature of the windowless room in which the birds were housed was maintained at 22°C. Overhead fluorescent lights were turned on daily at 8:30 a.m. and turned off 12 h later.

Apparatus

All training was conducted in three identical experimental chambers that measured 30.5 × 29.8 cm (floor dimensions) × 30.5 cm (height) (Ferster & Skinner, 1957). A ground-glass screen (23.4 × 9.5 cm), mounted behind the front wall of each chamber, provided a translucent surface on which stimuli could be presented in any of eight response locations. These were arranged in two rows of four each. Those in the top row were numbered, from left to right, 1–4; those in the bottom row were numbered 5–8. The centers of the bottom and top rows of response locations were positioned 18.5 and 23.5 cm, respectively, above the floor of the chamber. The edges of each row were equidistant from the vertical edges of the panel (9.0 cm). Each response location, which was 2.5 cm in diameter, was surrounded on top and bottom by three pairs of infrared emitters and detectors (Models TIL 31 and 81, respectively), the centers of which were located 0.95 cm in front of the ground glass (Clauson, Izatt, & Shimp, 1985). Any interruption of an emitter–detector pair at a particular response location (lit or unlit) was recorded as a peck to that location. Two feeding trays were centered below Response Locations 6 and 7, 1.66 cm above the grid floor on which the pigeon stood. Only the right-hand feeding tray was used in this experiment. Reward, a standard 45-mg pigeon pellet (produced by Bio-Serv of Frenchtown, NJ), was dispensed by a Gerbrands pellet feeder (Model G5110).

The stimuli were colored fields of light (red, green, or blue); achromatic geometric shapes (a white equilateral triangle, a white circle, and a white square, each presented on a black background); and a pair of black lines on a white background presented in one of three orientations (+45°, 90°, and −45°). All of the stimuli were generated on a Sony KB-1530 color monitor by a graphics program and then projected, via a 1:1 optical system, onto the rear of the ground-glass screen. The monitor was situated 64.2 cm behind the experimental chamber. The chamber was illuminated by a 15-W house-light located above the left rear corner of the compartment in which the pigeon was housed. During a 5-sec interval that followed each operation of the pellet dispenser, the only source of illumination in the chamber was a 2-W bulb located 2.5 cm above the feeding tray. A cut-out in the upper right rear corner of the chamber housed the lens of a color video camera. White noise was presented continuously in the chamber to mask extraneous sounds. All experimental events were programmed and recorded by Apple II computers (one for each chamber) that were located in an adjacent room.

Procedure

Subjects were given preliminary training, followed by a series of MTS, two-item DMTSS, and three-item DMTSS tasks. The subject initiated each trial by pecking a gray cross on a black background. A peck to the cross produced one or more items in Position 2. Each trial consisted of a sample and a recall component. The recall component began immediately after the offset of the last item of the sample sequence. During the recall phase, all of the items of the sample were displayed simultaneously. The subject's task was to respond to the simultaneously displayed items in the order in which they were presented during the sample phase. In all instances, a subject was advanced from one phase of training to the next when it satisfied the following acquisition criterion: two successive sessions in which 75% of the choice responses were correct for *each* type of sample on *non-correction* trials, that is, on the first trial on which the stimulus was presented. The arrangement of the sample and comparison stimuli within each trial and the number of trials used during each phase of training are shown in Table 1.

Preliminary Training

Following a brief period during which the pigeon was allowed to eat freely from the food tray, the pecking response was conditioned by an autoshaping procedure (Brown & Jenkins, 1968). All subjects were autoshaped to Peck Response Location 2. The stimulus projected onto this response location was a light gray cross on a black background. The maximum duration of each stimulus was 5 sec. The first peck to the stimulus terminated the trial and was reinforced by the delivery of a food pellet. The intertrial interval (ITI) was deter-mined by a variable time schedule ($M = 15$ sec, range = 7–30 sec). After four successive noncontingent trials on which a peck occurred, 40 trials were presented in which food was contingent on pecking. This procedure was followed until each subject responded on at least 36 of the programmed presentations of the stimulus. No subject required more than two sessions to satisfy this criterion.

Phase 1. Matching-to-Sample

Phase 1A. MTS with two comparison stimuli displayed in three positions. At the start of each trial, one of three types of item was presented as a sample in Position 2: a color, a shape, or a line orientation. The relative frequency with which each sample type appeared was 0.33. Likewise, the relative frequency of each of the three exemplars of each sample type was 0.33. Each type of sample was presented 25 times during each session in a quasirandom order.

The peck requirement for the sample was three pecks to the sample within 5 sec of its onset. Failure to satisfy this requirement terminated the trial and initiated an ITI of 10 sec (range = 5–30 sec). If the peck requirement was satisfied, the sample was terminated immediately and two comparison items were presented in the bottom row of the stimulus display: the sample item (S1) and one other stimulus from the category from which S1 had been selected. The positions of the comparison items were varied randomly from trial to trial on three response locations (5, 6, and 7). A response to the item displayed as the sample terminated the trial with a reinforcer. A response to the incorrect item resulted in a 10-sec timeout (TO) during which the houselight was extinguished. The TO also delayed the onset of the next trial by 10 sec. A correction procedure repeated trials on which an incorrect response occurred and allowed any trial to be repeated as many as 10 times. A new trial was programmed if a correct response did not occur within 10 trials of the original error. Trials on which a subject responded correctly were followed by an ITI the mean value of which was 20 sec (range = 5–30 sec).

Phase 1B. MTS with three comparison stimuli displayed in four positions. S1 was presented as previously, in Position 2. Training parameters were the same as those in Phase 1A except that the number of comparison stimuli presented on each trial was increased from two to three, and the number of positions on which the comparison stimuli were presented was increased from three to four. The positions used were 5, 6, 7, and 8.

Phase 1C. MTS with comparison stimuli displayed in all eight possible positions. During this phase of training, the three comparison stimuli were programmed to appear in Positions 1–8. All other aspects of the training procedure were identical to those used during Phase 1B.

Phase 1D. MTS with distractors. Distractors from the two categories not represented by the sample were added in two steps. One

Table 1
Sequence of MTS and DMTSS Training Procedures in Experiments 1 and 2

Phase	Number of Comparison Stimuli	Number of Positions in Display	Number of Distractors	Comparison Display Continuous	Number of Trials per Session
A. Matching to Sample					
1A	2	3	0		75
1B	3	4	0		75
1C	3	8	0		75
1D	3	8	6		75
B. Delayed Matching to 2 Items					
2A	3	3	0	No	96
2B	3	3	0	Yes	96
2C	3	8	6	Yes	144
C. Delayed Matching to 3 Items					
3	3	3	0	Yes	81

Note—MTS, matching-to-sample; DMTSS, delayed matching-to-successive-samples.

distractor from each of the unrepresented categories was presented during the first step. If, for example, the sample was a color, the choices were the three colors, one shape, and one line orientation. After a subject satisfied the accuracy criterion with two distractors, the remaining four distractors were displayed. The presentation of all possible distractors was staggered across successive trials because a display of three comparison stimuli and six distractors from the unrepresented categories would require nine response locations. If, as in the previous example, the sample was a color, the choices on some trials would consist of three colors, three shapes, and two orientations; on others, three colors, two shapes, and three orientations. The relative frequency of each type of distractor was essentially the same during each session. All other aspects of the training procedure were identical to those used during Phase 1B.

Phase 2. Two-Item DMTSS Training

During the second phase of training, each subject was trained to respond to two items in the order in which they were displayed. Items were again presented in Position 2. The first item of the sample (S1) was presented for 4 sec; the second (S2) for 2 sec. Pecks to the sample items were recorded but had no programmed effect. Each sample item was followed by a 50-msec dark period. During this and all subsequent phases of training, sample items were presented for the full period for which they had been programmed. The rationale for displaying the sample items for the entire period was to maximize the likelihood that the subject would attend to sample items for the programmed duration of the trial. In a pilot experiment, subjects pecked each item throughout the interval during which it was displayed on virtually every trial.

Phase 2A. Initial training on multiple samples. Three comparison stimuli were displayed in Positions 5–7 following the offset of S2. Trials were terminated with the delivery of a reinforcer following the execution of a correct sequence (S1→S2) or with a TO following the occurrence of an error. Trials were also terminated if the sequence S1→S2 was not executed within 8 sec of the onset of the comparison display. During Phase 2A only, a response to S1 during the recall phase of the trial resulted in its offset. When S1 was removed, the subject was left with a choice of the two remaining comparison stimuli.

Phase 2B. Comparison Stimulus S1 presented continuously. With one exception, all training parameters were the same as those in effect during Phase 2A. During Phase 2B, the comparison stimuli corresponding to Samples S1 and S2 were presented continuously until the end of the trial.

Phase 2C. Comparison stimuli displayed in all positions along with distractor stimuli. Distractor stimuli were added in two phases, as in Phase 1D. By the end of Phase 2C, eight comparison stimuli were displayed after a subject responded to S2: three comparison stimuli from the S1 category, three comparison stimuli from the S2 category, and two from the remaining category. As in Phase 1D, it was not possible to present all of the comparison stimuli and distractors on an eight-position display. Accordingly, only two distractors from the unrepresented category could be presented on each trial. During each session, the relative frequency of each type of distractor was approximately equal.

Phase 3. Three-Item DMTSS Training

Sample items were again displayed in Position 2. Each sample consisted of three items: a color (red, green, or blue), a white achromatic shape presented on a black background (triangle, circle, or square), or a pair of black parallel lines on a white background shown in one of three orientations (−45°, 90°, and +45°). The three different types of item that could be presented in a sample on each trial [color (C), shapes (S), and line orientation (O)] gave rise to six trial types: C→O→S, C→S→O, S→O→C, S→C→O, O→S→C, and O→C→S. Since there were three exemplars of each item type, there were nine versions of each trial type. Accordingly, there were 162 physically different sequences that could be presented as a sam-

ple. Each sample type was presented once over the course of two successive sessions in a quasirandom sequence. With the following exceptions, all training parameters were the same as those in effect during Phase 2A. S1 was presented for 8 sec, S2 for 4 sec, and S3 for 2 sec. Three comparison stimuli were displayed in Positions 5–7 following the offset of S3. The display was presented continuously until the end of the trial. The subject's task was to respond to the items presented during the recall phase of the trial in the order in which they were presented as sample items.

Results and Discussion

After completing preliminary training on MTS, all subjects mastered the two-item DMTSS task in Phase 2. However, after 90 sessions of training on the subsequent three-item DMTSS task (7,290 trials), none of the subjects responded at a greater than chance level of accuracy. Performance on the two-item DMTSS task compared favorably with that reported in other studies (Devine et al., 1979; MacDonald, 1993). This was true with respect to the amount of training needed to satisfy the accuracy criterion, the relative difficulty of different types of samples, and the relative frequency of different types of errors. Below we summarize the salient features of performances leading up to the (unsuccessful) attempt to obtain recall on the three-item DMTSS task.

Table 2 shows the mean number of trials needed to satisfy the accuracy criterion during each phase of MTS and two-item DMTSS training. The total of those means was 33,475 trials. At that point in training, all subjects had responded correctly on at least 75% of the trials during two successive sessions on *all* types of sample. During MTS training, for example, a subject might correctly complete 75% of the trials on which a color sample was presented during the 24th and 25th sessions, 75% of the trials on which color and shape samples were presented during the 36th and 37th sessions, and 75% of the trials during which color, shape, and line orientation were presented during the 46th and 47th sessions. In this instance, the subject fulfilled the accuracy criterion in the 47th session. For each phase of training, the two sessions in which the subject responded correctly on 75% of the trials on which each type of sample was presented will be referred to as the "criterial sessions."

In a study of two-item DMTSS performance with pigeons (MacDonald, 1993), 5 of the 8 subjects learned to respond at accuracy levels that exceeded chance perfor-

Table 2
Mean Number of Trials Needed to Satisfy Accuracy Criterion

Phase	Experiment 1	Experiment 2			
		Experimental Group	Control Group		
			1	2	3
1A	1,650	1,170	1,659	2,316	2,494
1B	356	408	384	572	759
1C	319	458	684	394	497
1D	2,888	2,798	3,975	4,425	3,450
2A	3,782	2,390	3,864	4,992	4,260
2B	8,640	7,392	10,332	10,140	9,312
2C	15,840	13,003	17,820	14,815	15,355
Σ	33,475	27,619	38,718	37,654	36,127

mance after 14,688 trials. On average, those subjects responded correctly on 66.2% of the trials. In MacDonald's study (1993), the first item correctly pecked in recall (S2) was removed from the display following a correct peck. This reduced the number of comparison stimuli from three to two. In the present study, S1 was turned off following a correct peck only during Phase 2A. Finally, unlike MacDonald's subjects, the pigeons of this study were confronted with distractors from categories other than those from which the sample items had been selected (Phase 2C). In view of these differences, the appropriate point of comparison between the results of this and MacDonald's study is the end of training on Phase 2A. The mean number of trials needed to complete Phase 2A was 8,995, approximately 5,000 trials fewer than that needed by MacDonald's pigeons.

The performance of the pigeons in the present experiment also compares favorably with that of rhesus monkeys. The subjects in the Devine et al. (1979) experiment needed more than 10,000 trials to master a two-item DMTSS task in which three items were presented as comparison stimuli: the two sample items and a single distractor. As was true of MacDonald's (1993) pigeons, Devine et al.'s monkeys also had the benefit of a reduction in the number of choices following a correct response to S1 (from three to two). Devine et al.'s monkeys responded correctly on more than 90% of the trials.

MTS Training (Phases 1A–1D)

A reliable difference in the difficulty of the different types of sample was observed during each phase of MTS training. Colors were least difficult, shapes of intermediate difficulty, and line orientations most difficult. Difficulty was measured by the number of trials needed to respond correctly on 75% of the trials on which a particular type of sample was displayed during two successive sessions. The relevant data are shown in Table 3A.

Table 3A also shows that the first and the last phases of MTS training were the most difficult. The relative difficulty of those phases reflects the time needed to master the MTS task at the start of training (Phase 1A) and to accommodate to the addition of distractors (Phase 1D). During the two intermediate phases (1B and 1C), the only changes in procedure were increases in the number of response positions on which the choice items were displayed. A two-way analysis of variance (ANOVA) of the number of trials needed to satisfy the accuracy criterion at each phase of training (sample type × phase) revealed a significant effect of sample type $[F(2,14) = 38.47, p < .05]$, a significant effect of phase $[F(3,21) = 149.7, p < .05]$, and a significant interaction between sample type and phase $[F(6,42) = 9.5, p < .05]$.[1] Post hoc Tukey analyses showed that the difficulty of each sample type differed, that the difficulty of Phases 1A and 1D differed from each other and all other phases, and that all interactions between sample type and phase were significant.

The relative difficulty of matching each type of sample was also evident in the accuracy of responding during the criterial sessions, as shown in Table 3B. At the end of training on Phases 1B and 1D, accuracy of responding was higher to color samples than to shape or line-orientation samples. These results are similar to those obtained with monkeys (Devine et al., 1979). On a two-item DMTSS task, monkeys performed more accurately with color than with shape samples. A two-way ANOVA on the accuracy of responding during the criterial sessions (sample type × phase) revealed a significant effect of sample type $[F(2,14) = 4.8, p < .05]$, but not of phase $[F(3,21) = 2.9, p > .05]$. Post hoc Tukey analyses showed that accuracy of responding to color samples was higher than it was to orientation samples. The interaction between sample type and phase was significant $[F(6,42) = 3.9, p < .05]$.

Two-Item DMTSS (Phases 2A–2C)

The three phases of two-item DMTSS training were progressively more difficult. During Phase 2A, the comparison item corresponding to S1 was removed from the display of comparison stimuli following a correct response. That reduced from three to two the number of items from which the subject had to select S2. During Phases 2B and 2C, all of the comparison stimuli were displayed until the end of each trial. During Phase 2C, distractors were added. The mean number of trials to criterion is shown at the bottom of the second column of Table 2.

There was little evidence of systematic variation in the number of trials needed to achieve the 75% level of accuracy with respect to sample type during Phase 2A. During Phases 2B and 2C, however, samples consisting of achromatic items (S→O and O→S) were more difficult than samples on which a color was presented (C→S, S→C, C→O, O→C). The relevant data are shown in Figure 1. A two-way ANOVA of the mean number of trials to criterion during Phases 2A–C (phase × sample type) revealed a significant effect of phase $[F(2,14) = 553.04, p < .05]$, and a significant effect of sample type $[F(5,35) = 42.0, p < .05]$. There was also a significant interaction between phase and sample type $[F(10,70) = 7.1, p < .05]$. Post hoc Tukey comparisons showed that significantly more trials were needed to satisfy the accuracy criterion when the samples contained only achromatic items than when the samples contained a color item. The relative difficulty of achromatic

Table 3
Matching to Sample in Experiment 1

Phase	Color	Shape	Orientation
A. Mean Number of Trials Needed to Satisfy Accuracy Criterion			
1A	1,078	1,453	1,650
1B	188	281	356
1C	159	234	319
1D	1,669	2,334	2,888
B. Percent Mean Accuracy During Criterial Sessions			
1A	79.5	79.1	79.6
1B	82.1	80.8	78.4
1C	81.4	82.4	83.6
1D	86.0	79.5	76.0

Note—Extensive analyses failed to reveal any effect on performance of the specific sample displayed during MTS and DMTSS training—for example, $color_1$, $color_2$, and $color_3$ for color samples on trials on which the color was a sample. Likewise, for the different shapes and orientations.
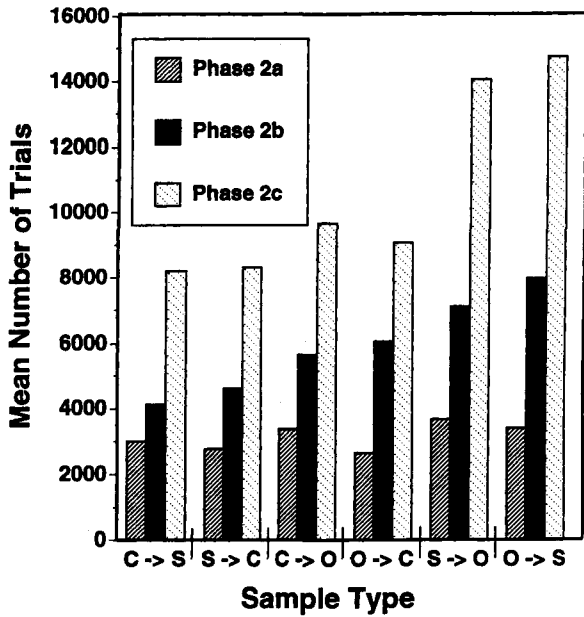
Figure 1. Number of trials needed to satisfy the accuracy criterion for each sample type during Phases 2A–2C of Experiment 1 (C, color; S, shape; O, line orientation). The number of trials needed to satisfy the accuracy criterion for each *phase* was greater than that shown in this figure (see Table 2) because it was necessary to satisfy the accuracy criterion for *all* sample types to complete each phase.

samples became more pronounced during Phase 2C. During Phase 2C, subjects required significantly more trials to achieve an accuracy level of 75% correct for two successive sessions for S→O and O→S samples than for any of the other four sample types (C→S, S→C, O→C, or C→O).

The relative difficulty of samples containing only achromatic items and samples containing at least one color was evident in two other aspects of performance: the ranks of the number of trials needed to satisfy the 75% accuracy criterion for each type of sample during each phase of training and the accuracy of responding to each of the sample types during the criterial sessions of Phase 2C. On average, the order of difficulty (from easiest to hardest) was C→S, S→C, O→C, C→O, S→O, and O→S. Kendall's coefficient of concordance was .82 ($p < .05$). The second column of Table 4 shows the mean percentage of correctly completed trials during the two criterial sessions for each sample type. A one-way ANOVA of accuracy of responding yielded a significant effect of sample type [$F(5,35) = 11.2, p < .05$]. Post hoc Tukey comparisons showed that accuracy of responding on S→O and O→S samples was significantly lower than it was on C→S, S→C, and C→O samples (but not O→C samples) during the criterial sessions.

Errors

For all sample types, the vast majority of errors occurred when subjects made their first response during the recall phase of each trial. The relevant data from Phase 2C are shown in the third and fourth columns of Table 4. A two-way ANOVA of the number of errors (position × sample

type) yielded a significant effect of the position of errors [$F(1,7) = 186.0, p < .05$] and a significant effect of sample type [$F(5,35) = 11.1, p < .05$]. The interaction of position and sample type was not significant [$F(5,35) = 2.5, p > .05$].

Four types of error could occur as the first response during the recall phase of each trial: a response to (1) S2 (an order error), (2) one of the items from the set from which S2 was selected (an S2 category error), (3) an item from the category from which S1 had been selected (an S1 category error), or (4) a distractor. By a substantial margin, order errors were the most frequent type of error. The relevant data are shown in Figure 2. The relative frequency of S1 and S2 category errors, and of distractor errors, is shown collectively as "other" errors because of their low values. A two-way ANOVA of the relative frequency of each type of error during the two criterial sessions of Phase 2C (error type × sample type) yielded significant differences between types of errors [$F(3,21) = 41.5, p < .05$] but not differences between sample types [$F(5,35) = 1.2, p > .05$]. There was, however, a significant interaction between error type and sample type [$F(15,105) = 2.0, p < .05$]. Post hoc Tukey comparisons showed that order errors occurred significantly more frequently than each of the other types of error and that S1 category errors were significantly more frequent than S2 category errors. As we will discuss in our analysis of the results of Experiment 2, this pattern of error provides evidence of a recency effect.

Following a correct response to S1, a subject could make three types of error: a response to (1) one of the items from the S1 set (an S1 category error), (2) an incorrect item from the S2 set (an S2 category error), or (3) a distractor. Very few errors were made in the second position during the two criterial sessions of Phase 2C (on average, less than 1.2 per session). Of those errors, 47% were S1 category errors, 25% were S2 category errors, and the remaining 28% were distractor errors. A two-way ANOVA of the relative frequency of each type of error (error type × sample type) did not reveal any significant effects.

Three-Item DMTSS (Phase 3)

During three-item DMTSS training, the probability of responding correctly by chance was 8.3%. During 90 sessions of training on Phase 3 (7,290 trials), the highest level of accuracy during any session was 10.3% correct (as averaged over the six basic sample types). A one-way ANOVA

Table 4
Performance on Phase 2C of Experiment 1, Criterial Sessions

Sample Type	Percentage Correct	% of Total Errors to S1	% of Total Errors to S2*
C→S	85.6	78.4	21.6
S→C	88.0	83.0	17.0
C→O	84.5	63.2	36.8
O→C	80.0	80.5	19.5
S→O	74.5	71.2	28.8
O→S	73.5	73.5	26.5

Note—C, Color; S, Shape; O, Orientation.   *Since an error could only occur to S1 or S2, the sum of the last two columns must be 100%.
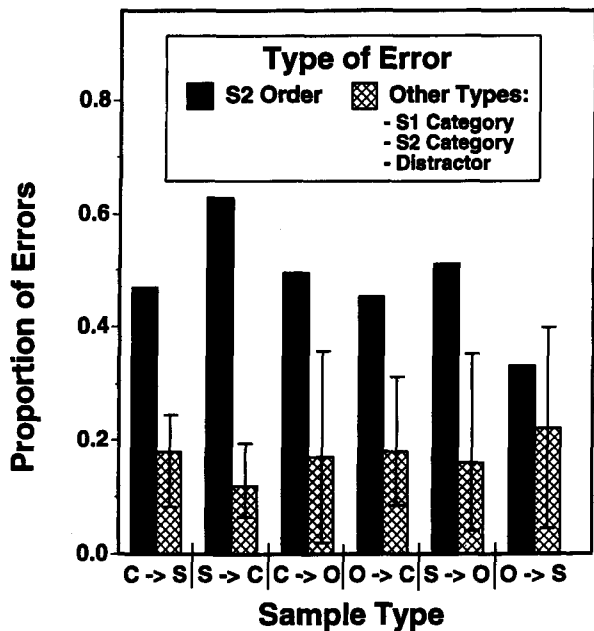
Figure 2. The proportion of S2 order errors and the mean proportion of other types of errors (S1 category, S2 category, and distractor) that could occur in the first position of the sequence during the criterial sessions of Phase 2C of Experiment 1. The vertical bars show the range of the mean proportion of "other" errors. The proportion of S2 order errors, the mean of the other types of errors, and the values of the range of other types of errors sum to 1.0. See text for additional details.

of the percentage of correct trials for each sample type was not significant [$F(5,35) = 4.5, p > .05$]. Experiment 1 was terminated after 90 sessions on Phase 3 since there was no evidence of improvement from the start of three-item DMTSS training.

During three-item DMTSS training, it was possible to make four types of errors: three forward and one backward. At the first position, a subject could skip forward to S2 or S3. At the second position, a subject could skip forward to S3. At the last position, it could make a backward error by responding to S1.[2] Regardless of sample type, the vast majority of errors (89.2% in the last two sessions) occurred at the first position and, of those, 77.5% were two-skip errors to S3. The remaining errors in the first position were one-skip errors to S2. On the few occasions on which subjects responded correctly to S1, 84.4% of subsequent errors were forward errors to S3, and the remainder were backward errors to S1 following a correct response to S2.

EXPERIMENT 2

In Experiment 1, pigeons were able to recall the order in which two successive samples were presented at an accuracy level of 75%. That ability did not generalize to three-item samples. After 90 sessions of training on the three-item DMTSS task, none of the pigeons exceeded the chance level of accuracy during the recall phase of each trial. One reason for that failure might be decay in the strength of

representations of the first and the second items of a three-item sample. The amount of time during which the sample was presented more than doubled following the shift from two- to three-item samples (from 6 to 14 sec). On a two-item DMTSS task, the comparison stimuli appeared 2 sec after the offset of S1. On a three-item DMTSS task, the interval between the offset of S1 and the appearance of the comparison stimuli increased to 6 sec.

A study in which pigeons learned to *produce* a particular sequence before training on sequence discrimination (Terrace, 1986) suggests a way to improve a pigeon's recall of arbitrary sequences. Pigeons that were trained to produce the three-item sequence A→B→C acquired a discrimination between A→B→C and non–A→B→C sequences more rapidly than subjects that hadn't learned to produce the sequence A→B→C. Control groups showed that the facilitation of sequence discrimination by prior training on sequence production could not be attributed to familiarity with Items A, B, and C.

In Experiment 2, a similar two-phase training procedure was used to evaluate the influence of production training on the recall of sequences presented in a DMTSS task. Prior to training on a three-item DMTSS task, each experimental subject was trained to produce one of three lists that was selected from the set of 162 possible three-item samples that were used during three-item DMTSS training.

Learning to produce a three-item list could facilitate performance on the DMTSS task for at least two reasons. First, subjects are required to produce a three-item sequence during the recall phase of the DMTSS task. Second, when executing a three-item sequence, a pigeon must keep track of its position in a representation of those items. Accordingly, the ability to represent a three-item sample prior to training on a three-item DMTSS task might help pigeons to maintain item and order information about sample items during the interval between the offset of S1 and the appearance of the comparison items. It is possible, however, that training on sequence production could facilitate sequence recall for reasons that have nothing to do with the ability to produce the simultaneous chain A→B→C. That possibility was evaluated by training three control groups to perform nonserial tasks in which they were exposed to A, B, and C for the same amount of time as the experimental subjects. It is conceivable, for example, that recognition of the sequence A→B→C could be facilitated by the experience of responding to Items A, B, and C even when they are presented successively.

Consider first a subject who is shown Items A, B, and C successively, each at a different response location. In this "pseudoproduction" procedure, subjects are always rewarded for responding to individually presented items in the sequence A→B→C and are never rewarded for responding to other sequences of those items. Control Group 1 assessed the effect of such training on subsequent recall of three-item sequences. Each subject was paired with a subject from the experimental group. On each pseudoproduction trial, the sequence of items and the occurrence of reinforcement were determined by the performance of the production (experimental) subject.

One possible limitation of the pseudoproduction procedure is the absence of an explicit contingency that requires subjects to discriminate A, B, and C. That limitation was addressed in a second control group that was trained to discriminate A, B, and C in a nonsequential task. Finally, a third control group was trained first on the nonsequential task and then on the pseudoproduction paradigm. The nonsequential task ensured that subjects were able to discriminate A, B, and C from the very start of pseudoproduction training.

Method

Subjects and Apparatus

The subjects were 32 experimentally naive male White Carneaux pigeons maintained under the same conditions as those described in Experiment 1. Subjects were randomly divided into four groups (N = 8): an experimental group and three control groups. The apparatus was identical to that used in Experiment 1.

Procedure

The pecking response for all subjects was conditioned by the same autoshaping procedure as that used in Experiment 1. After four successive trials on which a peck occurred, 40 trials were presented on which food was contingent on pecking. Contingent key-peck training occurred daily until each subject responded on at least 36 of the 40 programmed presentations of A and B during a daily session. No subject required more than two sessions to satisfy this criterion. Except for occasional apparatus failures, each bird was trained at the same hour, 7 days a week, in the experimental apparatus to which it was assigned during all phases of training.

Production training for experimental group. Each experimental subject was initially trained to produce one of three lists that were selected from the 162 possible sample sequences: [blue→circle→ −45°], [90°→red→square] or [triangle→ +45°→green]. List production was trained in successive phases on the simultaneous chains A→B and A→B→C (Straub & Terrace, 1981). The items that constituted those two- and three-item sequences were displayed simultaneously, each item appearing on one of the eight response locations of the experimental chamber. On A→B trials, Items A and B were presented simultaneously for a maximum of 10 sec. On subsequent A→B→C trials, Items A, B, and C were presented simultaneously for a maximum of 15 sec. Failure to respond within the prescribed time produced a 15-sec TO. During training on both A→B and A→B→C sequences, trials were separated by an ITI whose value was determined by a 15-sec variable interval (VI) schedule. A new trial was signaled by diminishing the intensity of the houselight for 0.5 sec 3 sec prior to the onset of trial.

A new configuration of list items was generated on each trial. On A→B trials, two locations were illuminated, one with Item 1 (e.g., blue, 90°, or triangle), the other with Item 2. Item 2 was selected from the two categories not represented by Item 1. On an A→B trial, for example, A might be projected onto Response Location 3, and B onto Response Location 5. On A→B→C trials, the configuration of list items was similarly varied. All told, there were 56 possible two-item configurations and 336 three-item configurations. The probabilities of generating a particular A→B and A→B→C configuration on any given trial during two- and three-item training were, respectively, 1/56 and 1/336. Each session was terminated after 80 trials.

Reinforcement occurred if, and only if, the pigeon responded to the items displayed on the video monitor in the sequence A→B during the first phase of production training and in the sequence A→B→C during the second phase of production training. Repeat pecks to a particular item were considered correct. Thus, in addition to the sequences A→B and A→B→C, sequences such as A→A→B, A→A→A→A→B→B→C, and so on, were also rewarded. No differential feedback followed *correct* responses until the end of the se-

quence. An incorrect response (a forward or backward error or a peck to a dark key) terminated the trial and produced a 15-sec TO. Training on A→B and A→B→C sequences continued until a subject satisfied a criterion of two successive sessions in which it responded correctly on at least 75% of the trials.

Pseudoproduction training for Control Group 1. Each subject in Control Group 1 was matched to 1 of the experimental subjects. For control subjects, the events on each trial were programmed to match those of the corresponding experimental subjects during training to execute the list A→B→C. This ensured that matched pairs of subjects from Control Group 1 and the experimental group experienced (1) the same sequences of items, (2) the same differential reinforcement with respect to each sequence, and (3) the same distribution of reinforcers during each session. For control subjects, however, Items A and B and Items A, B, and C were presented *successively* in the order in which those items were pecked by the experimental subject during A→B and A→B→C training, respectively. Consider, for example, a trial in which an experimental subject responded in the sequence A→B→C when Items A, B, and C were presented simultaneously on Response Locations 3, 5, and 1. For the corresponding pseudoproduction subject, the sequence of items was A on Location 3, followed by B on Location 5, and finally, C on Location 1.

The programmed duration of each item was 5 sec. The first peck to each item terminated that item and resulted in the immediate appearance of the next scheduled item. The pseudoproduction subject was reinforced at the end of a given trial if, and only if, the production subject to which it was matched was reinforced on the corresponding trial. On trials on which the experimental subject did not perform correctly, items were presented to the control subject in the order in which the experimental subject responded to them. If, for example, Items A and B were presented in Locations 7 and 1, respectively, and the experimental subject responded to Location 1 (a forward error), the control subject was only shown Item B (on Location 1). Similarly, if Items A, B, and C were presented on Locations 8, 1, and 5, respectively, and the experimental subject responded in the sequence 8→1→8 (a backward error), the pseudoproduction subject was shown Item A on Location 8, then Item B on Location 1 and, finally, Item A on Location 8.

As was true on production trials, a peck to an "incorrect" item on the corresponding pseudoproduction trial resulted in a TO. If a pseudoproduction subject did not respond to an item within 5 sec, that item was terminated and the next scheduled item was presented. Failure to respond to each of the items presented on a trial resulted in the omission of a reinforcer that was programmed for that trial. The number of A→B and A→B→C training sessions provided for each pseudoproduction subject was determined by the number of sessions that the corresponding production subject needed to satisfy the accuracy criterion in each instance.

Discrimination training for Control Group 2. The preliminary training of each of the 8 subjects of Control Group 2 on the Items A, B, and C was identical to that given to subjects of Control Group 1 (pseudoproduction training) except that during autoshaping, A, B, and C were presented (rather than just A and B).

Three adjacent response locations were illuminated on each trial throughout discrimination training for Control Group 2. The same item (A, B, or C) was displayed in each location. A C-A-B rule was used to define the correct responses to each type of configuration. On trials in which A appeared, the first response to the middle location produced reinforcement. On trials in which B appeared, pecks to the right-hand location resulted in reinforcement. Pecks to the left-hand location were reinforced on trials in which C was presented. A C-A-B rule [as opposed to a left-right (A-B-C) or a right-left (C-B-A) rule] was used to avoid any possibility of establishing a spatial representation of the Items A, B, and C that might facilitate their production in the sequence A→B→C.

The location of each triplet of discriminative stimuli was varied randomly to minimize bias to any particular response location. Triplets

of identical items were presented on all sets of adjacent response locations: 1, 2, and 3; 2, 3, and 4; 5, 6, and 7; and 6, 7, and 8. The probability of A, B, or C appearing on a particular triplet of response locations was .33 on each noncorrection trial. The probability that a particular triplet of adjacent response locations would be used on each trial was .25. On average, each of the response locations was associated with the same number of reinforcers, as were each of the items.

Pecks to an incorrect response location resulted in a 15-sec TO. Failure to respond within 5 sec was also considered an error and resulted in a 15-sec TO. In order to discourage position habits, a correction procedure was in effect: Whenever a subject pecked an incorrect location, the same items were repeated on the next trial on the same response locations. This procedure was followed until the subject responded correctly. Discrimination training continued for each subject until it responded correctly to at least 75% of the initial presentations of each triplet during two successive sessions (noncorrection trials).

Discrimination training followed by pseudoproduction training for Control Group 3. The first two phases of training for the 8 subjects of Control Group 3 were identical to those provided for the subjects of Group 2: autoshaping of the pecking response followed by the paradigm that trained subjects to discriminate A, B, and C. Each subject of Group 3 was then matched to a particular subject from the experimental group and given pseudoproduction training identical to that given to the subjects of Control Group 1.

After completing the training described above, all subjects were trained on the MTS and the two- and three-item DMTSS procedures used in Experiment 1.

Results and Discussion

Initial Training

The range of the number of sessions needed by the experimental subjects to satisfy the accuracy criterion for producing a three-item list was 22–63 (1,760–5,040 trials). This range was comparable to that reported in previous experiments (Straub & Terrace, 1981; Terrace, 1987; Terrace & Chen, 1991a, 1991b). All of the control subjects completed their prescribed tasks without any difficulty in a manner similar to that observed by Terrace (1986).

Mastery of Three-Item DMTSS Task

Each experimental subject mastered the three-item DMTSS task. On average, it took 74 sessions (5,994 trials) to satisfy the accuracy criterion: correct recall of at least 75% of the 162 physically different three-item sequences that could be presented as a sample during two successive sessions. The range of the number of sessions needed to satisfy the accuracy criterion was 59–87. After 90 sessions of training (7,290 trials), none of the control groups achieved an accuracy level greater than 9.2% correct on the three-item DMTSS task. That level of accuracy did not differ significantly from the level predicted by chance (8.3%).

During all but two phases of training preceding the three-item DMTSS task, experimental subjects satisfied the accuracy criterion more rapidly than did control subjects and the subjects of Experiment 1. The relevant data are shown in Table 2. During Phases 1B and 1C (in which the fewest number of errors occurred), the pattern of errors made by the different groups did not differ systematically. A two-way ANOVA of the average number of trials to criterion (group × phase) produced a significant effect of group

$[F(4,28) = 38.2, p < .05]$, a significant effect of phase $[F(6,42) = 10.97, p < .05]$, and a significant interaction $[F(6,42) = 4.3, p < .05]$. Post hoc Tukey comparisons showed that the performance of the experimental group trained in Experiment 2 differed significantly from that of the control groups and the group trained in Experiment 1. The control groups differed from the group trained in Experiment 1, but not from each other.

The data shown in Table 2 raise two questions. What advantages did the experimental subjects of Experiment 2 have over the subjects in Experiment 1? Second, why did the control subjects of Experiment 2 need more MTS and two-item DMTSS training than the subjects in Experiment 1? The most likely explanation of the advantage of the experimental subjects during Phases 2A–2C is the ability to represent the first item of each sample. The disadvantage of the control groups is more difficult to explain. The large number of errors they made during Phase 1A might have resulted from interference from the tasks that were trained prior to MTS training. One would expect, however, that interference from earlier tasks would have dissipated during the later phases of DMTSS training. The fact that it did not is puzzling.

As can be seen in Table 5, accuracy of responding by the experimental group to each of the six types of samples was uniformly high during the two criterial sessions. In contrast to the data shown in Table 4, fewer systematic difficulties were observed with respect to exemplars of shape (S) and orientation (O) in the case of three-item samples. A one-way ANOVA showed a significant effect of sample type $[F(5,35) = 2.9, p < .05]$. Post hoc Tukey comparisons were significant in only one instance. Performance was poorer following O→C→S samples than following S→ O→C samples. The difficulties caused by S and O items in Experiment 1 appear to have been dissipated by one or both of the following factors: the longer durations of the sample sequence in Phase 3 and the combination of exemplars of S and O with those of C.

Recency Effect

The results of Experiments 1 and 2 provide evidence of a strong recency effect. During the recall phase of each trial, premature responses to the last item of the sample accounted for the vast majority of errors. The relevant data are shown in Figure 2 and Table 6.

The tendency to respond first to the last item of the sample was true for all sample types. As can be seen in Fig-

Table 5
Mean Accuracy of the Experimental Group in Experiment 2 on the Criterial Sessions of Phase 3

Sample Type	Percentage Correct
C→S→O	77.9
C→O→S	76.6
S→O→C	81.0
S→C→O	78.9
O→S→C	74.8
O→C→S	73.9

Note—C, color; S, shape; O, orientation.

ure 2, for example, the relative frequency of forward errors to S2 exceeded the relative frequency of all other errors combined by a substantial margin during the criterial sessions of Phase 2C. The range of the relative frequency of forward errors to S2 was .33–.63.

The recency effect was just as pronounced when subjects were making few errors as when they were making many. This can be seen by comparing the first, second, and fifth rows of Table 6, which illustrates the types of errors that occurred in Phase 3. The first row shows the relative frequency of errors for each group in Experiments 1 and 2. The range is from 18.4% (experimental group of Experiment 2) to 93.3% (Control Group 3 of Experiment 2). The second row shows the relative frequency of errors at the first position of the recall sequence; the fifth row shows the proportion of position errors that were responses to the last item of the sample (S3). Regardless of group or phase of training, responding first to S3 was the most likely error.

For example, consider the last two sessions of three-item DMTSS training in Experiment 1 (the 89th and 90th sessions). The initial response during the recall phase of each trial accounted for 89.2% of the errors. Of these errors, 77.5% were recency errors, that is, incorrect responses to S3. Of the errors that followed the occurrence of a correct response to S1 (9.1% of all errors), 84.4% were recency errors, that is, incorrect responses to S3. This pattern of forward errors to the last item of the sample can also be seen in the data of the group that made the fewest errors. During their criterial sessions, 18.4% of the responses of the experimental group of Experiment 2 were errors. Of those, 85.1% occurred at the first position; 78.2% of first position errors were responses to S3.

Similar recency effects have been observed in studies in which dolphins and sea lions were trained by a procedure that resembles the DMTSS paradigm (Herman, Richards, & Wolz, 1984; Schusterman & Gisiner, 1988). Both species were trained to perform sequences of specific actions (e.g., fetch, tail-touch, under, over, etc.) involving one or more objects (pipe, ball, ring, disc, cone, etc.). The procedure used in those studies differed in two important respects from that used to train pigeons and monkeys to recall sequences by the DMTSS procedure. Responding was rein-

forced after the first item was recalled and the items were three-dimensional objects to which the subjects swam. Those differences notwithstanding, most of the errors occurred when subjects made their first response during the recall phase of each trial (Herman et al., 1984, Table 11; Schusterman & Gisiner, 1988, Table 9).

GENERAL DISCUSSION

A pigeon's ability to correctly recall a three-item sample raises a variety of issues concerning the development of serial skills in animals and, indirectly, in human infants. Most important is the need to learn to produce a three-item list before the ability to recall three items can manifest itself. What cognitive factors were crucial to the success of the experimental subjects of Experiment 2 and to the failures of the subjects of Experiment 1 and the control subjects of Experiment 2? Our answers will have to be speculative, given that the only other studies on DMTSS in animals used two-item samples (Devine et al., 1979; Devine & Jones, 1975; MacDonald, 1993).

The samples used during three-item DMTSS training were selected from an exhaustive set of 162 three-item sequences. Why would a pigeon's ability to recall three-item samples the constituents of which varied from trial to trial be facilitated by learning to produce just *one* of those sequences prior to training on the DMTSS task? An analysis of the cognitive skills needed to execute a three-item list provides at least a partial answer to that question. While executing a three-item list, the pigeon is not given any external cues regarding its position on the list. As a consequence, the pigeon must track its position within a representation of the sequence as it moves from one item to the next. The list-learning experience of the subjects of Experiment 1 and of the control groups of Experiment 2 was limited to the two-item sequences they were required to produce at the end of each DMTSS trial. Those sequences could be executed by a "find Item 1 and respond to the other item by default" strategy. For this same reason, the representational demands of a two-item DMTSS task are much simpler than those imposed by a three-item list. A pigeon can "recall" a two-item list when shown a two-item

Table 6
Percent Types of Errors Occurring in Phase 3 (Three-Item DMTSS) in Last Two Sessions of Training

| | | Experiment 2 | | | |
| | | Experimental Group | Control Groups | | |
	Experiment 1		1	2	3
Proportion of all responses	90.9	18.4	92.3	91.7	93.3
Proportion of all errors					
Position 1	89.2*	85.1	88	89	89.1
Position 2	9.1	13.8	10.3	9.2	9.0
Position 3	1.7	1.2	2.0	1.9	1.9
Forward errors					
→C	77.5†	78.2	74.7	76.4	74.2
A→C	84.4‡	87.8	83.7	82.9	82.9

*Percentage of all errors. †Percentage of Position 1 errors. ‡Percentage of Position 2 errors.

display by identifying the first item, and then responding to the second item by default. That strategy would also work when a two-item sample is followed by three comparison stimuli. The pigeon need only recall the first item and recognize which of the remaining items on the display was included in the sample. On a three-item list, a default rule will not disambiguate the second and third items.

Subjects in Experiment 1 failed to master a three-item DMTSS task because they were unable to represent a three-item sequence. Studies of list *production* have identified two important features of a pigeon's representation of a three-item list: the unidirectional nature of a list and the salience of the start and end items. The high frequency of forward (as opposed to backward) errors is evidence that pigeons represent a list in a forward (as opposed to a backward) manner (Terrace, 1993). The reliable occurrence of a serial position effect is evidence of the distinctiveness of the first and the last items (Straub & Terrace, 1981). Two-item sequences can be produced correctly without relying on either of these features. Thus, there is no reason to assume that a pigeon who can produce a two-item sequence during the recall phase of a two-item DMTSS trial would grasp the complexities of a longer list.

Pigeons that learned to produce a three-item list not only mastered the three-item DMTSS task but, unlike the subjects of Experiment 1 and the control groups of Experiment 2, they showed positive transfer during two-item DMTSS training. It seems likely that the positive transfer during two-item DMTSS training reflects a "search for a salient start item" strategy that subjects of the experimental group acquired when learning to produce a three-item list.

The crucial role of production training in establishing a pigeon's ability to recall three-item sequences is an instructive example of the value of investigating the ontogeny of serially organized behavior in animals. Animals lack not only language but also the sophisticated serial abilities that investigators of human memory take for granted in their subjects. Those deficiencies, however, can benefit investigators because they reduce the number of variables that can contribute to the acquisition of serially organized behavior. In this respect, subjects of experiments on animal memory are comparable to human infants. In each instance, the subject lacks language and formal experience with serial tasks at the start of the experiment.

Practical issues constrain how much can be learned about the development of serial memory from experiments on infants since it is difficult to obtain systematic data from a particular infant for more than a few sessions. Thus, an important advantage of studying the development of serial memory in animals is the opportunity to observe particular subjects over as many trials as is practical from the experimenter's point of view. In this study, that was almost 40,000 trials.

The potential of longitudinal studies of serial learning has been confirmed by other experiments with pigeons and monkeys. Experiment 2 of this study was modeled after one that showed that the acquisition of a sequence discrimination was facilitated by prior training on sequence production (Terrace, 1986). Studies of serial learning by monkeys show that they become progressively more efficient at list learning as they master successive lists (Chen et al., in press; Swartz et al., 1991). During the course of learning successive lists, they also learn to encode knowledge of the ordinal position of list items (Chen et al., 1996). The paradigms used in each of these experiments afford opportunities for observing the emergence of a complex serial skill from one or more simpler skills. As such, they demonstrate the feasibility of a novel approach to the study of "higher order" serial processes, one that eliminates the uncontrolled influences of language and other serial skills.

REFERENCES

BROWN, P. L., & JENKINS, H. M. (1968). Auto-shaping of the pigeon's keypeck. *Journal of Experimental Analysis of Behavior*, 11, 1-8.

CAPALDI, E. J., MILLER, D. J., ALPTEKIN, S., & BARRY, K. (1990). Organized responding in instrumental learning: Chunks and superchunks. *Learning & Motivation*, 21, 415-433.

CHEN, S., SWARTZ, K., & TERRACE, H. S. (1996). Knowledge of the ordinal position of list items by rhesus monkeys. Manuscript submitted for publication.

CHEN, S., SWARTZ, K., & TERRACE, H. S. (in press). Serial learning by rhesus monkeys: II. The emergence of a trial and error strategy during mastery of successive lists. *Psychological Science*.

CLAUSON, H. D., IZATT, E. J., & SHIMP, C. P. (1985). An infrared system for the detection of a pigeon's pecks at alphanumeric characters on a TV screen: The dependency of letter detection on the predictability of one letter by another. *Journal of the Experimental Analysis of Behavior*, 43, 257-264.

CROWDER, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

D'AMATO, M. R., & COLOMBO, M. (1988). Representation of serial order in monkeys Cebus apella. *Journal of Experimental Psychology: Animal Behavior Processes*, 14, 131-139.

DEVINE, J. D., BURKE, M. W., & ROHACK, J. J. (1979). Stimulus similarity and order as factors in visual short-term memory in nonhuman primates. *Journal of Experimental Psychology: Animal Behavior Processes*, 5, 335-354.

DEVINE, J. V., & JONES, L. C. (1975). Matching to successive samples: A multiple-unit memory task with rhesus monkeys. *Behavior Research Methods & Instrumentation*, 7, 438-440.

FERSTER, C. B., & SKINNER, B. F. (1957). *Schedules of reinforcement*. Englewood Cliffs, NJ: Prentice-Hall.

GRAYSON, R. J., & WASSERMAN, E. A. (1979). Conditioning of two-response patterns of key pecking in pigeons. *Journal of the Experimental Analysis of Behavior*, 31, 23-29.

HAGGBLOOM, S. J., BIRMINGHAM, K. M., & SCRANTON, D. (1992). Hierarchical organization of series information by rats: Series chunks and list chunks. *Learning & Motivation*, 23, 183-199.

HERMAN, L. M., RICHARDS, D. G., & WOLZ, J. P. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition*, 16, 129-219.

JITSUMORI, M., & SUGIMOTO, S. (1982). Memory for two stimulus-response items in pigeons. *Journal of the Experimental Analysis of Behavior*, 38, 63-70.

MACDONALD, S. E. (1993). Delayed matching-to-successive-samples in pigeons: Short-term memory for item and order information. *Animal Learning & Behavior*, 21, 59-67.

MCGONIGLE, B., & NEAPOLITAN, N. (in press). Serial ordering skills in young children. *Quarterly Journal of Experimental Psychology*.

MURDOCK, B. B. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.

ROBERTS, W. A., & GRANT, D. S. (1976). Studies of short-term memory in the pigeon using the delayed matching to sample procedure. In D. L. Medin, W. A. Roberts, & R. T. Davis (Eds.), *Processes of animal memory* (pp. 79-112). Hillsdale, NJ: Erlbaum.

ROITBLAT, H. L., BEVER, T. G., HELWEG, D. A., & HARLEY, H. E. (1991). On-line choice and the representation of serially structured

stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, **17**, 55-67.

ROITBLAT, H. L., SCOPATZ, R. A., & BEVER, T. G. (1987). The hierarchical representation of three-item sequences. *Animal Learning & Behavior*, **15**, 179-192.

SCHUSTERMAN, R. J., & GISINER, R. (1988). Artificial language comprehension in dolphins and sea lions: The essential cognitive skills. *Psychological Record*, **38**, 311-348.

SHIMP, C. P. (1976a). Organization in memory and behavior. *Journal of Experimental Analysis of Behavior*, **26**, 113-130.

SHIMP, C. P. (1976b). Short-term memory in the pigeon: Relative recency. *Journal of the Experimental Analysis of Behavior*, **25**, 55-61.

SHIMP, C. P., & MOFFITT, M. (1974). Short-term memory in the pigeon: Stimulus-response associations. *Journal of the Experimental Analysis of Behavior*, **22**, 507-512.

SKINNER, B. F. (1957). The experimental analysis of behavior. *American Scientist*, **45**, 343-371.

STRAUB, R. O., & TERRACE, H. S. (1981). Generalization of serial learning in the pigeon. *Animal Learning & Behavior*, **9**, 454-468.

STROMER, R., & MACKAY, H. A. (1993). Human sequential behavior: Relations among stimuli, class formation, and derived sequences. *Psychological Record*, **43**, 107-131.

SWARTZ, K. B., CHEN, S., & TERRACE, H. S. (1991). Serial learning by rhesus monkeys. I: Acquisition and retention of multiple four-item lists. *Journal of Experimental Psychology: Animal Behavior Processes*, **17**, 396-410.

TERRACE, H. S. (1984). Simultaneous chaining: The problem it poses for traditional chaining theory. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behavior: Discrimination processes* (pp. 115-138). Cambridge, MA: Ballinger.

TERRACE, H. S. (1986). Positive transfer from sequence production to sequence discrimination in a non-verbal organism. *Journal of Experimental Psychology: Animal Behavior Processes*, **12**, 215-234.

TERRACE, H. S. (1987). Chunking by a pigeon in a serial learning task. *Nature*, **325**, 149-151.

TERRACE, H. S. (1993). The phylogeny and ontogeny of serial memory: List learning by pigeons and monkeys. *Psychological Science*, **4**, 162-169.

TERRACE, H. S., & CHEN, S. (1991a). Chunking during serial learning by a pigeon: II. Integrity of a chunk on a new list. *Journal of Experimental Psychology: Animal Behavior Processes*, **17**, 94-106.

TERRACE, H. S., & CHEN, S. (1991b). Chunking during serial learning by a pigeon: III. What are the necessary conditions for establishing a chunk? *Journal of Experimental Psychology: Animal Behavior Processes*, **17**, 107-118.

WASSERMAN, E. A., NELSON, K. R., & LAREW, M. B. (1980). Memory for sequences of stimuli and responses. *Journal of the Experimental Analysis of Behavior*, **34**, 49-60.

WEISMAN, R. G., & DODD, P. W. D. (1979). The study of associations: Methodology and basic phenomena. In A. Dickinson & R. A. Boakes (Eds.), *Mechanisms of learning and motivation: A memorial volume for Jerzy Konorski* (pp. 337-363). Hillsdale, NJ: Erlbaum.

WEISMAN, R. G., WASSERMAN, E. A., DODD, P. W., & LAREW, M. B. (1980). Representation and retention of two-event sequences in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **6**, 312-325.

NOTES

1. For all statistical tests, differences were considered significant if their probability of occurring by chance was less than .05.

2. It was not possible to make a backward error at S2, since repeat responses to each item were not considered errors.