What–Where–When Memory in Pigeons

Shannon I. Skov-Rackette, Noam Y. Miller, and Sara J. Shettleworth
University of Toronto

The authors report a novel approach to testing episodic-like memory for single events. Pigeons were trained in separate sessions to match the identity of a sample on a touch screen, to match its location, and to report on the length of the retention interval. When these 3 tasks were mixed randomly within sessions, birds were more than 80% correct on each task. However, performance on 2 different tests in succession after each sample was not consistent with an integrated memory for sample location, time, and identity. Experiment 2 tested binding of location and identity memories in 2 different ways. The results were again consistent with independent feature memories. Implications for tests of episodic-like memory are discussed.

Keywords: matching to sample, memory, episodic-like memory, binding, pigeon

Do animals remember specific episodes in their personal past as such? Since the first report of episodic-like memory in scrub jays (Clayton & Dickinson, 1998), researchers studying a variety of species and tasks have sought to answer this question (see reviews in Hampton & Schwartz, 2004; Roberts, 2002, 2005). An obstacle to progress in their search is that nearly all evidence of episodic memory in humans is obtained verbally: People report a subjective sense of traveling back in time to a specific past experience. Such autonoetic consciousness is an essential feature of episodic memory by some recent definitions (Tulving, 2002). Many researchers studying nonverbal species have therefore followed Clayton and Dickinson (1998) in adopting the earlier definition (Tulving, 1972) of episodic memory as memory for a personal experience of what occurred, where, and when. This aspect of episodic memory, or what–where–when (W-W-W) memory, potentially can be demonstrated in nonverbal species. Because it is intrinsically impossible for such demonstrations to include the subject's report of its subjective state while demonstrating that it has the memory, W-W-W memory is referred to as *episodic-like* (Clayton & Dickinson, 1998).

The pioneering study of Clayton and Dickinson (1998; for a review, see de Kort, Dickinson, & Clayton, 2005) demonstrated W-W-W memory in a bird by showing that Western scrub jays (*Aphelocoma californica*) could remember what kind of food items they had cached, where they had cached those items, and how long ago they had cached them. A bird might cache peanuts in one set of distinctive sites and its preferred food of wax worms in other sites. Later, birds had the opportunity to retrieve both kinds of items. Birds that had been taught to expect worms to decay (i.e., become unpalatable) over time searched most in sites where worms had been cached at short retention intervals (RIs) and searched most in peanut sites at long ones. In contrast, birds that had not been taught to expect worms to decay always searched preferentially for worms. It is important to note that, because episodic memory is a kind of long-term memory (Hampton & Schwartz, 2004), the RIs in this study spanned periods of hours or days, and they were arranged so that time of day could not selectively enhance retrieval of one type of item.

The test for the scrub jays, like those it has inspired with other species, requires more than memory for a single episode. To express its W-W-W memory, the animal must encode two episodes, each involving encounters with two foods, and additionally learn a rule about how the value of at least one of those foods changes over time. This complexity might have contributed to the failure of some of the analogous tests with rats and monkeys, which used comparatively long RIs (e.g., Bird, Roberts, Abroms, Kit, & Crupi, 2003; Hampton, Hampstead, & Murray, 2005). However, under some conditions, rats do show evidence of W-W-W memory (Babb & Crystal, 2005, 2006). Rats can also learn a discrimination on the basis of odor, place, and relative time of events in a sequence of odor–place pairings (Ergorul & Eichenbaum, 2004). This has also been claimed to be an analogue of episodic memory, although RIs in this study were only a minute or 2 at most.

Tests of W-W-W memory for a single episode appear to have been more successful than those requiring memory for two such episodes. For example, Eacott and Norman (2004) pointed out that the when aspect of an episodic memory is not memory for a specific past time as much as for temporal and other aspects of the context in which an event occurred. Accordingly, they demonstrated that rats remember the visuospatial context in which they encountered a novel object up to an hour ago. Zentall (2005a; Zentall, Clement, Bhatt, & Allen, 2001) emphasized that, unlike

Shannon I. Skov-Rackette, Noam Y. Miller, and Sara J. Shettleworth, Department of Psychology, University of Toronto, Toronto, Ontario, Canada.

Correspondence concerning this article should be addressed to Sara J. Shettleworth, Department of Psychology, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada. E-mail: shettle@psych.utoronto.ca

most laboratory tests of animal memory, typical tests of episodic memory are unexpected, tapping the spontaneous encoding of interesting experiences that goes on all the time. Such spontaneous encoding was captured in Eacott and Norman's (2004) study in their measurement of exploratory behavior. Zentall et al. (2001) tested it in pigeons by training birds to "report" whether they had just pecked in one context and then showing that the birds used their reporting response appropriately when the pecking to be reported on was induced in a new way.

In the present article, we describe a novel nonverbal test of W-W-W memory for a single episode. It is based on the premise that if an animal can reliably answer any of the three relevant questions—What was it? Where was it? and When was it?—when it does not know which of those questions will be asked, the animal must have encoded all three features of the relevant episode. An episode in the present experiments consisted of the appearance of one of two colored shapes in one of eight locations on a video monitor. Birds were trained to match to sample, to match to location, and to report whether the RI was 2 s or 6 s. Therefore, like some researchers mentioned earlier, we tested only very short-term memory. However, in standard models of memory processing (e.g., Wagner, 1981), events do not enter into long-term memory without first being processed in short-term memory. Therefore, if one wants to test the generality of W-W-W or episodic-like memory across species, a good way to start may be to study how the identity, place, and time of single events are represented in short-term memory. Indeed, Gallistel (1990) claimed that "temporal and spatial coordinates are an obligatory part of every record kept by an animal brain" (p. 525). However, most of the evidence he cited for this claim came from studies of perception in humans. Hasher and Zacks (1979) made a similar claim based primarily on studies of memory in humans.

As Gallistel's (1990) statement suggests, there is more to episodic-like memory than memory for what occurred, where, and when. These features must be bound together in some way into an integrated representation of a single event. In addition, because episodic memory in humans is declarative, as opposed to procedural, animals should be able to use episodic-like memories flexibly, redeploying the behavior controlled by such memories appropriately when conditions change. The scrub jays studied by Clayton and Dickinson (1998; Clayton, Yu, & Dickinson, 2001) could use their W-W-W memories for food caches flexibly (see de Kort et al., 2005). One study (Babb & Crystal, 2006) has also demonstrated flexible use of W-W-W memory by rats. Clayton et al. (2001) have provided evidence for integration of the three features of W-W-W memory in scrub jays and suggested that one feature, either what or where in their studies, binds the other two together in memory. Each of the experiments reported in this article therefore includes one or more tests designed to discover whether memories for what, where, and when were integrated or independent.

Experiment 1

Experiment 1 began with separate, interleaved sessions of three different matching tasks (see Figure 1). In all three tasks, the sample was a red disk or a green triangle in one of eight locations around the periphery of a touch screen, and it occurred either 2 s or 6 s before the test, making 32 possible samples. In the test phase

Figure 1. Sequence of events during a trial in Experiment 1, showing the three types of tests that could occur. Samples as well as comparison stimuli on where tests appeared in any of eight locations around the periphery of the screen.

of the what task, the red disk and green triangle appeared in the center of the screen, and birds were reinforced for matching to sample. In the test phase of the where task, gray squares appeared in two of the peripheral locations, and birds were reinforced for matching to location. In the test phase of the when task, a yellow star and a blue paw shape appeared in the center of the screen; pecks to one shape were correct after the 2-s RI, and pecks to the other were correct after the 6-s RI. The birds were trained to a criterion with complete sessions of each task, one session of each task every 3 days, and then the tasks were mixed within sessions, starting with blocks of 16 successive trials per task. Block size was gradually reduced until finally the test presented on one trial did not predict the test on the next.

The final phase was designed to test whether location, time, and identity were encoded independently or bound together in memory by presenting two different tests in succession on occasional probe trials. If memories for the features were independent, the probability of responding correctly on the second test should not be related to the probability of responding correctly on the first test of the same trial. In contrast, bound W-W-W memory should result in a significant degree of dependence between first and second choice accuracy. However, accuracy on second tests would be expected to be lower overall than accuracy on first tests for several reasons. For example, the first test might be a source of interference, the RI was longer before the second test than before the first, second tests were uncommon, and the end of the first test of a trial usually signaled the intertrial interval (ITI) and the end of a need to remember the most recent sample. We minimized this latter effect by signaling the ITI throughout training by changing the background color of the touch screen. To minimize reduction in performance as a result of increasing the RI between the sample and the second test of a probe trial and to ensure that these tests were all preceded by a similarly increased RI, we did not present reinforcers after the first test of a probe trial, even though the birds were usually correct on those tests. It is unlikely that this treatment resulted in differential generalization decrement following correct

and incorrect first test responses because pigeons, unlike monkeys, do not show evidence of memory awareness (i.e., metamemory) in matching to sample (see Hampton & Schwartz, 2004). Thus, although the absence of reward on the first test of a probe trial might contribute to reduced accuracy on the second test, such an effect should be the same whether or not the choice on the first test was correct.

Method

Subjects

Subjects were six 3–6-year-old White King pigeons (*Columba livia*) obtained from Palmetto Pigeon Plant (Sumter, SC). All had previous experience pecking touch screens in a different set of chambers in visual discrimination and memory experiments. Between experiments, the birds had lived communally in a large aviary. During the experiment, birds were individually housed in wire mesh cages (38 cm × 48 cm × 55 cm) on a 14:10 light–dark cycle with free access to water and grit. They were maintained at 85% ± 2% of their ad-lib weight by controlled feeding of mixed grain as necessary following experimental sessions.

Apparatus

The operant testing chambers measured 37 cm wide × 35 cm deep × 33 cm high. The walls were constructed of clear Plexiglas with a metal bar floor and with a Samsung 152T LCD 15-in. (38-cm) monitor composing most of the wall opposite the entry door to the chamber. The bottom of the 44-cm-high monitor screen was 3.5 cm above the floor of the chamber. The screen was covered by a thin sheet of clear, nonreflecting Plexiglas, over which was mounted a CarrollTouch 15-in. (38-cm) touch frame (Model d87587-000, EloTouch, Menlo Park, CA). Food pellets were dispensed by a MED Associates (St. Albans, VT) ENV-203-20 feeder into a 4-cm-wide × 14-cm-long × 1.5-cm-deep Plexiglas tray lined with black sandpaper that sat on the floor of the chamber with one edge at the side of the chamber. A 6-W light mounted above the clear ceiling of the chamber over the food tray provided illumination when food pellets were dispensed. During experimental sessions, white noise was broadcast into the room containing the operant chambers, and the room was dark. Control equipment was located in an adjoining room.

Procedure

Stimuli. The screen was divided into an invisible 3 × 3 grid of equal-sized rectangles in which stimuli could be centered. The stimuli were a blue cross (plus sign), a green equilateral triangle, a red disk, a white *X*, a gray square, a blue paw shape, and a yellow sunburst (see Figure 1). Each stimulus was approximately 2 × 2 cm at its widest extent. A peck to a given stimulus was registered when it was located within the 4-cm square centered on the image.

Pretraining. Each bird experienced two to four 50-trial sessions of autoshaping with a white disk centered on the screen. Once a bird was pecking on the majority of trials, it had two to three 162-trial sessions of continuous reinforcement for pecking each of the seven stimuli to be used in the experiment. These were presented in random order with a variable ITI averaging 15 s. Each stimulus could appear in any of the nine locations defined in the previous section. A stimulus remained on the screen until the pigeon pecked it once. At this stage, each reinforcer consisted of eight 20-mg Pigeon Food Pellets (Research Diets, New Brunswick, NJ) accompanied by illumination of the light above the feeder for 7 s.

Match-to-sample training. Immediately following pretraining, birds began entire sessions of W-W-W training. All trials began with darkening of the monitor screen and presentation of the blue cross in the center of the screen; the cross remained visible until pecked. One peck to the cross resulted in presentation of the red circle or the green triangle in one of the eight locations around the periphery of the screen. The first peck to the circle or triangle began a 3-s timer, at the end of which the first response advanced the bird to a 2-s or 6-s RI. Thus, each sample was presented for a minimum of 3 s. During the RI, the screen was dark. At the end of the RI, the white *X* appeared in the center of the screen, and a single peck to it started a what, where, or when test (see Figure 1). During a what test, both the red circle and the green triangle appeared in the center of the screen, 7 cm apart center to center, and one peck to the stimulus that matched the sample was reinforced. Left–right positions of the two stimuli changed randomly from trial to trial. On a where test, two gray squares appeared, one in the same location where the sample had appeared on that trial, and one in another of the eight peripheral locations on the screen. One peck to the square that matched the sample's location was reinforced. On a when test, the paw shape and the sunburst appeared in the center of the screen, 6 cm apart center to center, with left–right positions changed randomly from trial to trial. If the programmed RI had been 2 s, choice of the blue paw was reinforced; choice of the yellow sunburst was reinforced after the 6-s RI. Initially, each correct response was reinforced with eight pellets, as in pretraining; after 21 sessions, reinforcer size was reduced to six pellets and feeder light duration to 6 s. In all three types of trials, a peck to the incorrect stimulus led directly to the 15-s ITI, during which the screen was gray to distinguish the ITI from the RI. Sessions consisted of 160 trials and were generally run every day. Within each successive block of 32 trials, each combination of the 2 shapes × 8 locations × 2 RIs occurred once, in a different random order in each block.

Acquisition. Each acquisition session consisted entirely of trials of one type, what, when, or where. For the first 30 sessions, what, when, and where sessions occurred in a fixed order, so that each type occurred once every 3 days. Training on entire sessions then continued as necessary until the bird met a criterion of 80% correct on both RIs in all three session types for 2 successive sessions of each type. To allow additional training on more difficult tasks while attempting to avoid overtraining on the tasks on which the bird was already above criterion, whenever a bird had met criterion on a particular type of session, we had it skip the next scheduled session of that task. As long as accuracy on a high-performance task remained above 80%, that task occurred once every 5 sessions rather than once every 3 sessions. Once a bird met criterion on two of the tasks, sessions on the remaining task were given every 2nd day, with the two other tasks each given once every 4 days. After 80 sessions, criterion was reduced to 75% for 2 birds that had not achieved the 80% criterion; criterion was further reduced to 70% for 1 bird that had not met the reduced criterion after 90 sessions. Once a bird met its individual criterion, that bird began blocks training. The same individual criteria continued to hold in the next phase.

Blocks training. In this phase, sessions consisted of 144 trials, 48 of each type. Trials were presented in blocks of a single type, with a different type in each successive block. The order of blocks of different trial types was randomized so that no task predicted the next task to appear (e.g., a block of where tasks was equally likely to be followed by a block of when tasks or a block of what tasks). Block size was initially 16, then 8, 4, and 2. Birds advanced to the next smaller block size when they achieved their individual criterion proportion correct in each of the three trial types at each RI across four sessions. Across successive blocks of two sessions, each possible sample event occurred equally often in each of the three types of trials. Once a bird met the individual criterion set at the end of acquisition with tasks in blocks of 2 trials, it moved to random testing.

Random testing. For random testing, what, when, and where trials were mixed randomly, with the constraint that two trials of each type occurred in each successive block of 6 trials. Also, within each successive block of 96 trials, each sample event occurred once in each task (2 samples × 8 locations × 2 RIs × 3 tasks). As in blocks training, sessions were 144 trials long. Thus, across successive blocks of 2 sessions, each combination of sample event and test type occurred three times. Following the completion of 10 random sessions with the original sample presentation

time of 3 s, the programmed sample presentation time was reduced to 1.5 s for 10 more sessions.

Two-test phase. Immediately following random testing, sample presentation time was increased to 3 s, and sessions were increased to 168 trials. On 24 of these trials, the birds received two successive tests. These two-test trials were exactly like normal trials except that the white *X* reappeared immediately after the bird pecked one of the alternatives in the first test, whether the bird's choice was correct or not. One peck to the *X* led to a second test, always of a different kind from the first. Reinforcement was delivered or not on the basis of whether birds chose correctly on this second test. Each successive block of 7 trials consisted of 6 normal trials, randomized and counterbalanced as before, and 1 two-test trial at a randomly determined position within the block. The six possible sequences of two tests (e.g., what–where, what–when) were randomized such that each sequence appeared four times in each session. Birds had 10 two-test sessions with the 3-s presentation time, followed by 10 sessions with presentation time set at 1 s.

Analysis. Proportions of correct responses during acquisition were analyzed with parametric statistics. Analysis of performance in the two-test phase is described in the *Two-Test Phase* subsection of the *Results* section. Throughout, effects were considered significant if *p* < .05.

Results

Acquisition

By the end of the first 10 sessions of acquisition, mean performance on all tasks was close to 80% correct at both RIs (see Figure 2). A Session × Task × RI analysis of variance (ANOVA) on Sessions 9 and 10 of each task showed no effect of task or RI, $F(2, 10) = 1.26$, and $F(1, 5) = 0.09$, respectively. There was a significant effect of session, $F(1, 5) = 6.66$, indicating that the birds were still improving. There was also a significant interaction between task and RI, $F(2, 10) = 7.59$. All other effects were nonsignificant: Session × RI, $F(1, 5) = 1.12$; Session × Task, $F(2, 10) = 1.67$; Session × Task × RI, $F(2, 10) = 0.45$. We ran separate ANOVAs for each task to tease apart the interaction effects. These showed a significant effect of RI on the what and where tasks, $Fs(1, 5) = 8.08$ and 11.04, respectively, reflecting the fact that matching identity or location was less accurate after the longer RI. We did not expect an RI effect on the when task because birds were reporting on the length of the just-completed RI. There was also a significant effect of session in the when task, $F(1, 5) = 225.10$. No other effects were significant; what task: day, $F(1, 5) = 2.02$, Day × RI, $F(1, 5) = 0.07$; where task: day, $F(1, 5) = 0.22$, Day × RI, $F(1, 5) = 0.93$; when task: RI, $F(1, 5) = 4.41$, Day × RI, $F(1, 5) = 0.65$.

Although, as just indicated, birds achieved criterion in all three tasks after similar numbers of sessions, performance during early acquisition in the when task differed markedly from that in the other two tasks. Most birds began when training 100% correct at one RI and 0% correct at the other, a fact reflected in the error bars for when tasks in Figure 2. The birds' strong idiosyncratic biases toward pecking the blue or the yellow comparison stimulus were gradually overcome with training. Apparently, no such biases affected performance on what and where tasks, perhaps because the birds were required to peck both colors and all eight locations around the touch screen in the sample phase of trials. Early acquisition in those two tasks followed a similar course (see Figure 2). However, location matching (where) was initially acquired faster than identity matching (what), as shown by a Session ×

Task × RI ANOVA on percentage correct on what versus where tasks over the first 10 days. This showed significant effects of session, task, and RI, $F(9, 45) = 17.94$, $F(1, 5) = 16.04$, and $F(1, 5) = 48.65$, respectively. Performance was better at the 2-s than at the 6-s RI. None of the interactions were significant: Session × Task, $F(9, 45) = 2.69$; Session × RI, $F(9, 45) = 1.24$; Task × RI, $F(1, 5) = 1.15$; Session × Task × RI, $F(9, 45) = 0.68$.

Blocks Training

When the three types of trials were first combined in single sessions, in blocks of 16, performance on what and where trials showed little change, but performance on when trials dropped precipitously (see Figure 3). A Phase (criterion vs. blocks) × Task × RI ANOVA compared the total percentage correct for each task and RI (i.e., total correct choices per total trials completed) in the last 2 acquisition sessions (320 trials) of each type with the total percentage correct for that trial type in the first 4 sessions (192 trials per type) in blocks of 16. There were significant main effects of phase, $F(1, 5) = 41.01$, and task, $F(2, 10) = 10.71$, and a significant Phase × Task interaction, $F(2, 10) = 76.60$. There was also a significant Task × RI interaction, $F(2, 10) = 7.81$. Separate ANOVAs run on each of the three tasks showed a significant effect of phase in the when task, $F(1, 5) = 107.23$, and a significant effect of RI in the where task, $F(1, 5) = 22.53$. All other effects were not significant; what task: all $Fs < 1.0$; where task: phase, $F(1, 5) = 6.01$, Phase × RI, $F(1, 5) = 0.66$; when task: RI, $F(1, 5) = 5.79$, Phase × RI, $F(1, 5) = 2.34$.

Recovering criterion levels of performance in when trials took up to 51 sessions plus 11 entire sessions of when trials. The latter were discontinued when it appeared that they were not improving performance. The birds spent a median of 35.5 (range = 7–51) sessions on blocks of 16. Four of the pigeons did not recover their criterion and were allowed to progress to the next stages at criteria of 75% (1 bird) or 70%. Training progressed more quickly after this point. Median numbers of sessions with blocks of 8, 4, and 2 were, respectively, 5 (range = 4–22), 4 (range = 4–17), and 5 (range = 4–6)—that is, close to the minimum of 4 sessions per block size. Exclusive of entire when sessions, the median total number of sessions of blocks training was 59 (range = 29–79).

Why did performance drop only on when trials once the tasks were mixed within sessions? One possibility is that, at the beginning of a block of when trials, the birds had difficulty switching attention from the physical features of the sample (i.e., its identity or location) to the duration of the RI. If that were the case, performance should be especially low at the beginning of a block of when trials and improve within the block. Such effects are seen in experiments on visual search, or search image formation, in pigeons when birds must switch attention between different features of target stimuli (Blough & Blough, 1997). For example, Reid and Shettleworth (1992) gave pigeons a task in which color or shape could be used to find grains on a speckled background. Performance fell immediately after a switch from the color to the shape task but recovered within about 10 trials.

To test for such an effect in the present data, we averaged performance over each 4 successive trials of the same type in the first four sessions, with the three trial types presented in blocks of 16. Thus, each block of 16 trials was divided into four sections (Trials 1–4, 5–8, 9–12, and 13–16). We analyzed performance in
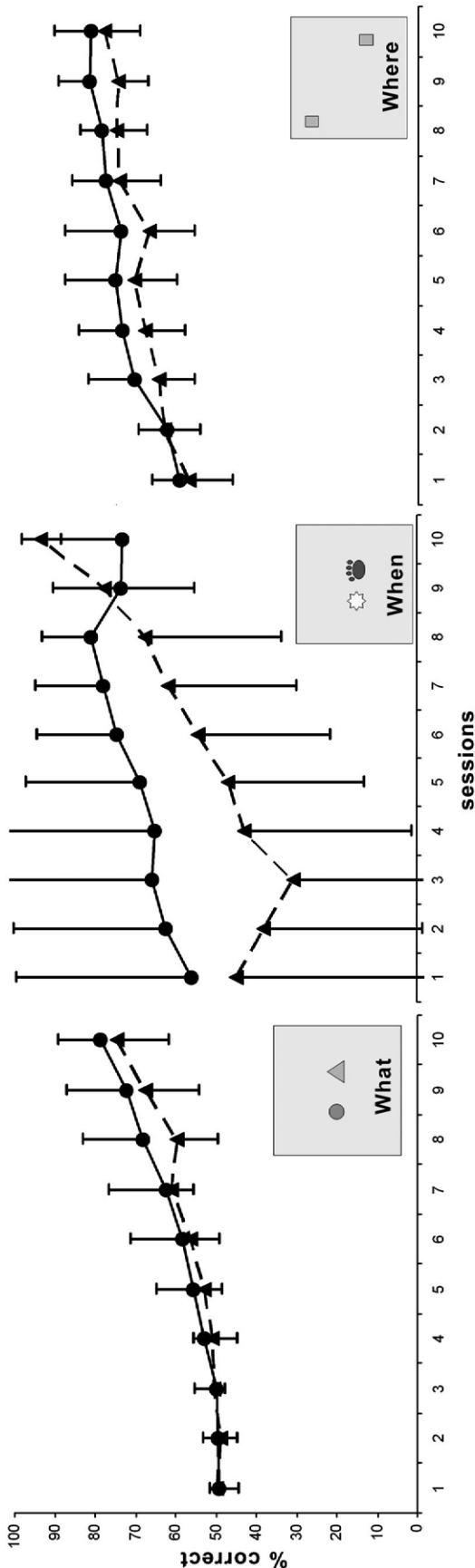
all three types of trials in this way to test whether such an attentional effect, if it occurred, was specific to when trials. A Task $\times$ Section ANOVA showed a significant effect of task, $F(2, 10) = 32.30$, but no effect of section, $F(3, 15) = 0.33$. There was also a significant Task $\times$ Section interaction, $F(6, 30) = 3.48$. However, separate ANOVAs showed no significant effect of section on what, where, or when trials, Fs$(3, 15) = 2.70, 2.63,$ and 0.81, respectively. In summary, analysis of performance within each block of 16 trials of the same kind provided no evidence of improvement across the block, such as would be consistent with a gradual switching of attention from one feature of the sample to another.

Random Testing

Performance with randomized presentation of what, where, and when trials was uniformly excellent with both 3-s and 1.5-s sample exposure times (see Figure 4). Matching accuracy was now well above the 70% or 75% criterion most birds needed to reach this stage. Performance did not improve significantly across sessions with the same exposure time, as shown by a Blocks of Sessions $\times$ Trial Type $\times$ RI ANOVA comparing overall percentages correct in the first five versus the last five sessions on each exposure time. Therefore, we combined all sessions for each exposure time in separate Task $\times$ RI ANOVAs. At the 3-s sample exposure time, there was a significant effect of task, $F(2, 10) = 5.03$, and of RI, $F(1, 5) = 11.62$, but no interaction, $F(2, 10) = 2.69$. At the 1.5-s sample exposure time, there was a significant effect of RI, $F(1, 5) = 15.07$, but not of task, $F(2, 10) = 1.27$. There was also a significant interaction, $F(2, 10) = 5.06$. Separate ANOVAs run on each task showed a significant effect of RI for the what and where tasks at both exposure times—what tasks: 3 s, $F(1, 5) = 39.30$, 1.5 s, $F(1, 5) = 8.58$; where tasks: 3 s, $F(1, 5) = 18.02$, 1.5 s, $F(1, 5) = 17.76$—but no effect of RI on the when task at either exposure time, Fs$(1, 5) < 1.0$.

Two-Test Phase

Performance on regular trials in the two-test phase remained high and similar to that in the preceding phase. On the 24 trials per session with two successive tests following the same sample, the birds averaged 87.4% correct on the first test and 79.5% correct on the second test over all trial types and sessions when the sample time was 3 s. With the 1-s sample time, these scores were, respectively, 89.4% and 78.0% correct (see Table 1). Paired-sample t tests on overall proportions correct showed that performance at both exposure times was significantly worse on the second test than on the first: 3 s, $t(5) = 5.99$; 1 s, $t(5) = 9.72$. For reasons discussed in the introduction to this experiment, reduced accuracy on the second test was to be expected. Clearly, however, performance on second tests was still well above chance, making it feasible to test for any dependency between first and second test performance.

We tested whether performance on the second test of a probe trial depended on performance on the first test in the following way. For each bird at each sample exposure time, we cast the total frequencies of the four possible sequences of correct or incorrect choices on first and second tests into a 2 $\times$ 2 table. The marginal totals gave the overall proportions of correct first and second test

Figure 2. Mean (plus or minus standard deviation, denoted by the error bars) percentages of correct choices for the first 10 sessions of each kind in Experiment 1.
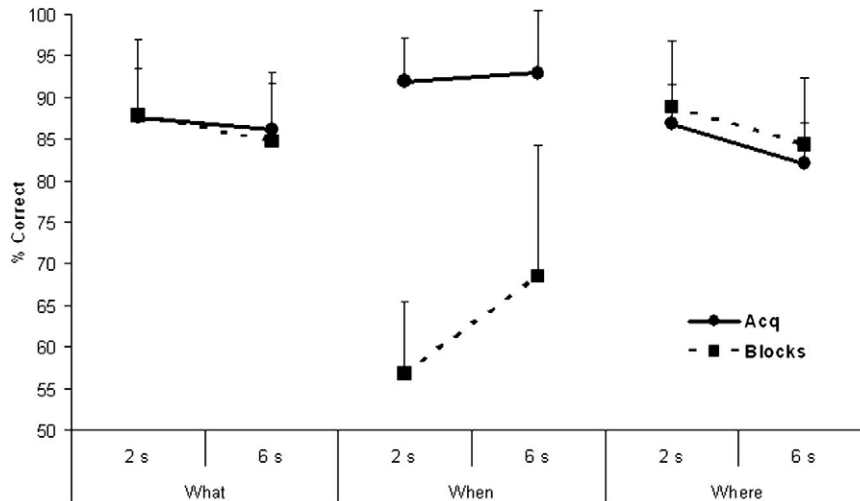
Figure 3. Mean (plus or minus standard deviation, denoted by the error bars) percentages of correct choices at the 2-s and 6-s retention intervals in the last two sessions of acquisition (Acq) and the first four sessions of the blocks phase in Experiment 1.

choices, analyses of which we have reported. If what, when, and where memories are independent, a bird with the overall proportions correct reported for the group should be correct on both Test 1 and Test 2 on 69.5% (i.e., 87.4% × 79.5%) of probe trials with a 3-s sample. Similarly, it should be incorrect on both Test 1 and Test 2 on 2.6% (12.6% × 20.5%) of those trials. If memories for what, when, and where are bound, numbers of correct choices on both tests and of incorrect choices on both tests should exceed these expectations. Frequencies of each sequence of correct and incorrect choices expected if choices on the two tests were independent were compared with the observed frequencies via chi-square. The data and results of the chi-square tests for the 6 individual birds and the two exposure times are displayed in Table 1. As can be seen, observed frequencies were remarkably close to those predicted from independence. Values of $\chi^2(1, 216 \leq N \leq 360)$ were not greater than 1.04 ($ps > .30$).

Figure 4. Mean (plus or minus standard deviation, denoted by the error bars) percentages of correct choices at the 2-s and 6-s retention intervals in each type of trial during the random phase of Experiment 1.

Equivalently but perhaps more intuitively, if what, when, and where memories are an integrated unit, then performance should be better on Test 2 if the bird was correct rather than incorrect on Test 1. However, the data displayed in Table 1 imply that no such difference existed. For example, with 3-s samples, birds were correct on Test 2 following a correct choice on Test 1 on 69.4% of all trials, on average. Because they were correct on Test 1 on 87.4% of trials, this means that the probability of being correct on Test 2 given a correct choice on Test 1 was .794 (or 69.4/87.4). Similar computations show that the conditional probability of a correct choice on Test 2 given an incorrect choice on Test 1 was .0801 (10.1/12.7), on average.

These overall analyses could obscure effects in particular test sequences. For example, perhaps only location and identity are bound in memory. Because each possible sequence of tests occurred only 40 times per bird at each sample exposure time, we combined data across the two exposure times for each of the six different sequences of test types. However, because errors on Test 1 occurred on only 10%–15% of trials, performance on Test 2 following an error was still based on limited data. Table 2 summarizes the outcomes of these analyses as the group mean observed and expected choice proportions and the maximum values of $\chi^2(1, 76 \leq N \leq 96)$ for each of the six sequences of test types. Just as with the overall analyses, the frequencies did not differ from those expected if what, where, and when memories were independent. One may note that the lowest average performance on Test 2 occurred on trials with when as the second test, which perhaps reflects difficulty in reporting the sample–Test 1 interval correctly when extra events had intervened between the sample and the test.
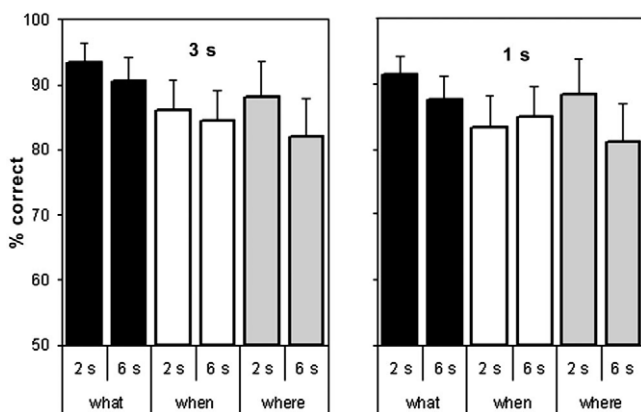
Discussion

This experiment was designed to determine whether pigeons encode the location, identity, and time of appearance of a single item by training them to match each of these three features sepa-

Table 1
Percentages Correct on Tests 1 and 2, Proportions of Trials With Correct and Incorrect Choices, and Chi-Square Values for the Two-Test Phase of Experiment 1

Bird	Correct		T1+, T2+		T1+, T2−		T1−, T2+		T1−, T2−		$\chi^2(1)$
	T1	T2	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	
					3-s samples (% of all trials)						
32	86.7	80.8	69.6	70.1	17.1	16.6	11.3	10.8	2.1	2.6	0.23
38	88.6	77.5	68.1	68.7	20.6	19.9	9.4	8.8	1.9	2.6	0.78
51	84.6	80.0	67.5	67.7	17.1	16.9	12.5	12.3	2.9	3.1	0.03
53	88.5	76.0	67.7	67.3	20.8	21.2	8.3	8.7	3.1	2.8	0.22
57	85.8	80.0	69.6	68.7	16.3	17.2	10.4	11.3	3.8	2.8	1.04
58	90.0	82.9	74.2	74.6	15.8	15.4	8.8	8.3	1.3	1.7	0.40
M	87.4	79.5	69.4	69.5	17.9	17.9	10.1	10.0	2.5	2.6	
					1-s samples (% of all trials)						
32	89.6	75.0	67.1	67.2	22.5	22.4	7.9	7.8	2.5	2.6	0.01
38	89.8	76.9	69.4	69.0	20.4	20.8	7.4	7.8	2.8	2.4	0.23
51	86.3	79.6	68.8	68.6	17.5	17.6	10.8	10.9	2.9	2.8	0.01
53	92.1	79.6	73.2	73.4	19.0	18.8	6.5	6.3	1.4	1.6	0.08
57	88.9	79.6	71.3	70.8	17.6	18.1	8.3	8.9	2.8	2.3	0.36
58	89.6	77.1	69.2	69.1	20.4	20.5	7.9	8.0	2.5	2.4	0.02
M	89.4	78.0	69.8	69.7	19.6	19.7	8.2	8.3	2.5	2.3	

Note. T1 = Test 1; T2 = Test 2; + = correct; − = incorrect; Obs = observed; Exp = expected.

rately and then mixing the tasks together randomly. The birds matched correctly on 80% or more of trials (see Figure 4) even when they could not predict from one trial to the next which test would occur. This means that all three features must have been encoded on at least 52% ($0.8 \times 0.8 \times 0.8$) of the trials, but they might have been encoded independently rather than as an integrated whole.

Two lines of evidence from training and testing suggest that encoding was independent. Most important are the results from the tests of memory for two features of the same sample. Performance on the second test was unrelated to performance on the first test. Some dependence might have been expected because variations in attention to the sample from trial to trial should affect performance on both tests in the same way. However, the fact that the birds were required to peck to start the trial, to peck the sample to move to the RI, and to peck the center of the screen to start a test might

have ensured that they were always attending quite well to the samples as well as to the tests.

We have implicitly assumed that if the second test were the same as the first, we would not have found independence. This experiment did not include probe trials in which the same type of test was given twice, however, primarily to minimize the frequency of unrewarded first choices entailed by probe trials. In the one relevant study we could find in the literature, Roitblat (1980, Experiment 1) gave pigeons two successive tests after a single sample in a three-alternative color-matching task. However, the implications of his data for the present question are unclear because second tests were given only after incorrect choices on the first test. Moreover, they involved only the sample and the one distractor not chosen on the first test—that is, the birds chose among three alternatives on the first test but between only two on the second test. Consistent with the assumption that performance

Table 2
Mean Percentages Correct on Tests 1 and 2, Mean Proportion of Trials With Correct and Incorrect Choices, and Range of Chi-Square Values for the Two-Test Phase of Experiment 1

Test sequence	% correct ± *SD*		T1+		T1−		$\chi^2(1) \leq$
	T1	T2	T2+	T2−	T2+	T2−	
When–what	86.2 ± 4.4	86.8 ± 6.4	74.8	11.5	12.1	1.7	0.91
Where–what	87.7 ± 5.7	87.6 ± 5.2	76.3	11.4	11.3	1.0	1.01
When–where	88.1 ± 3.8	80.2 ± 2.3	70.9	17.2	9.3	2.6	1.86
What–where	92.1 ± 2.7	79.3 ± 4.7	72.6	19.5	6.7	1.2	0.72
Where–when	85.9 ± 4.1	66.2 ± 5.5	58.1	27.8	8.1	6.0	2.69
What–when	89.7 ± 2.8	72.4 ± 3.4	64.6	25.1	7.8	2.6	2.27

Note. T1 = Test 1; T2 = Test 2; + = correct choice on a test; − = incorrect choice.

on the first test of a given feature would predict performance on a second test of the same feature, at short sample durations, most birds performed at or only slightly above chance on second tests. However, Roitblat (1980, Experiments 1 and 3) also found that as first test performance improved with longer sample durations, some birds' performance on second tests after incorrect choices on the first test rose above chance. This finding is consistent with some independence between two tests of the same sample feature, but without comparable data from trials on which the bird chose correctly on the first test, the implications for the present study are unclear.

In addition to the results of the two-test phase, evidence consistent with independent memories for what, when, and where in the present study comes from the beginning of blocks training. When the three tasks were first combined, performance on when tasks fell dramatically, whereas performance on the other two tasks was unaffected. Such a dissociation would not be expected if what, when, and where memories were bound. A possible account of the selective decline in when performance is that the when task involved symbolic matching, whereas the what and where tasks involved literal matching. However, it is hard to see how this difference can explain why the when task was the only task to suffer when the tasks were first mixed within sessions. The drop in accuracy on the when task could reflect difficulty in shifting attention from the physical features of the sample, which must be encoded for successful matching of location or identity, to some cue associated with the RI, presumably either time or strength of memory for the sample. However, the analysis of performance within the 16-trial blocks shows that if this were the case, shifting attention to the appropriate cue would take more than 16 trials. Another possibility is that it was not equally easy to report on all three features of the sample. This was evidently the case with respect to the what versus where tasks. Both in this experiment and in Experiment 2, the where (location matching) task was acquired more quickly than the what (identity matching) task, perhaps because there were eight possible location samples and only two identity samples, resulting in less intertrial interference for the where task. In the long run, however, birds were clearly able to perform all tasks very well indeed (Figure 4).

In summary, it is not clear why the when task was affected differently from the what and where tasks when all three were first mixed within sessions, but the effect is consistent with the results of the two-test phase in suggesting that, in this task, the pigeons formed independent memories for location, time, and identity. Experiment 2 was designed to test for independent versus integrated memories for a sample's features in other ways.

Experiment 2

What, where, and when memories might have appeared to be independent in the tests at the end of Experiment 1 because the birds had first been trained extensively with the three tasks individually. In Experiment 2, the spontaneous binding of features of an event, as hypothesized by Gallistel (1990; see the introduction to this article), was predicted to affect performance from the very beginning of training. We focused on what and where tasks, which made our tests analogous to some tests of binding in human memory (Chalfonte & Johnson, 1996; Kohler, Moscovitch, & Melo, 2001). For example, Chalfonte and Johnson had participants

study a 7 × 7 array of locations, 30 of which were occupied by drawings of common objects. The authors tested memory for object location alone by presenting the same array with Xs in 10 old and 10 new locations. They tested memory for object identity alone by presenting a 4 × 5 array of objects, 10 of which were old and 10 of which were new. They tested bound memory for object and location by presenting 20 of the objects seen in the study phase, 10 of them in new locations in the 7 × 7 array. In each kind of test, participants were asked to identify the items that were the same in the tested features as those seen in the study phase. Chalfonte and Johnson (1996) and Kohler et al. (2001) primarily varied conditions at encoding within each of their experiments by telling people what to study. In contrast, we kept conditions constant in the study phase but varied the type of test.

The samples in the present experiment consisted of one of two differently colored shapes in one of eight locations, as in Experiment 1, but there was only one RI. There were initially four types of tests, as diagrammed in Figure 5. The unbound what and unbound where tests were the same as the what and where tests in Experiment 1: The rewarded alternative in the test phase retained only the feature being tested. In contrast, in bound what tests, the sample and comparison shapes appeared on either side of the location where the sample had been earlier in the trial, and in bound where tests, the sample and comparison were both identical in shape and color to the sample for that trial. If the birds had formed an integrated memory for the sample, the untested feature presented in bound tests should help retrieve memory of the tested feature. Moreover, accurate performance should be facilitated in bound tests by the fact that the correct alternative was identical to the sample (or very nearly so) in both location and identity. Chalfonte and Johnson (1996) suggested that "when information is bound together, individual feature information may be less available" (p. 413), and this notion, too, seems to predict better performance in our bound than in our unbound tests. Contrary to this hypothesis, one might suggest that the presence of the untested feature in bound tests would increase errors because the distractor matched the sample in one respect. In any case, as we show, there was actually no consistent effect of bound versus unbound testing in either direction.

Therefore, we included a further test of whether what and where memories are bound or independent by introducing trials with four alternatives in the final phase of the experiment (see Figure 5, lower right). One alternative matched the sample in both features, one matched it in location only, one matched it in identity only, and one did not match it at all. We compared the pattern of choices on these tests with that expected if what and where memories were independent, using the same statistical approach as for the two-test probes of Experiment 1.

As in Experiment 1, training began with entire sessions of each type of trial, with each type of session occurring equally often. Then trial types were mixed within sessions in blocks of gradually decreasing size, and finally they were mixed randomly. Performance in all four tasks was close to ceiling by the time the birds reached the random stage, so we decreased exposure time and increased RI in an attempt to reveal potential differences among the four types of tests by lowering overall accuracy. This same lowered level of performance was maintained when the four-alternative tests were added, because the distribution of errors on these tests was of major interest.
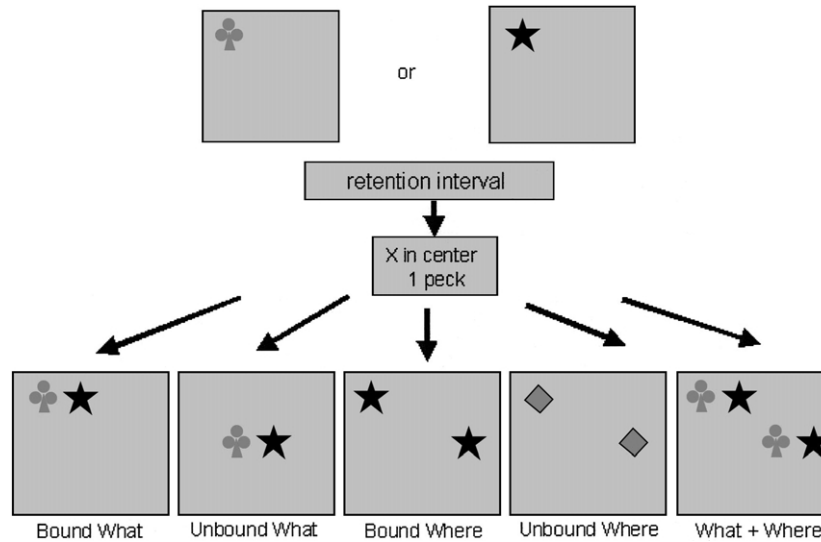
Figure 5. Sequence of events during a trial in Experiment 2, showing the five types of tests that occurred in the final phase of the experiment. The Bound Where test is an example for a trial in which the star was the sample.

Method

Subjects

Subjects were six 3–6-year-old White King pigeons (*Columba livia*) similar to those in Experiment 1, maintained in the same way. All of them had previous experience in a location-matching task with the same apparatus and stimuli used in Experiment 1.

Procedure

Apparatus. This was the same as the apparatus used in Experiment 1.

Stimuli. The stimuli were a light blue cross (plus sign), a yellow club shape (see Figure 5), a dark blue star, a white *X*, and a pink diamond. Each stimulus was 2 cm square at its widest extent. The invisible grid for placing stimuli on the screen and the method of detecting pecks at a stimulus were the same as in Experiment 1.

Pretraining. As all subjects had prior experience in the same apparatus used in the present experiment, minimal pretraining was necessary. Prior to beginning the match-to-sample training, each bird had two sessions consisting of 112 continuous reinforcement trials, programmed like those in Experiment 1, in which they were rewarded with eight pellets for pecking each of the stimuli that were to be used in the match-to-sample phase.

Match-to-sample training. Following pretraining, birds began entire sessions of bound what, unbound what, bound where, and unbound where training. One of each occurred per 4-day block, in a fixed order, with one of the types testing memory for what alternating with one of those testing memory for where. Up to the end of the RI, the sequence of events in each trial was as in the training phase of Experiment 1, except that the sample shapes were the club and the star and the RI was always 2 s. Unbound what and unbound where trials proceeded in the same way as what and where trials, respectively, in Experiment 1, except that the stimuli in the test phase of unbound where trials were pink diamonds. For the test phase of a bound what trial, both the club and the star appeared 5.5 cm apart, center to center, centered in one of the eight peripheral locations on the screen, and a peck to the one that matched the sample was reinforced. Left–right positions of the two stimuli changed randomly from trial to trial. In a bound where test, two stimuli identical to the sample for the trial appeared, one where the

sample had appeared on that trial, and one in a different, randomly selected one of the eight peripheral locations on the screen. Sessions consisted of 144 trials and were generally run every day. Because of experimenter error, for the first one or two sessions per bird, reinforcement was only six pellets. Reinforcer size was then restored to eight pellets until the end of the first four sessions of each trial type, before being set at six pellets for the rest of the experiment.

After at least 24 sessions (6 days per type) and 80% correct responding on the last 2 days of each type, the pigeons were advanced to blocks training. All but 1 bird met the 80% criterion on both kinds of where sessions before reaching it on what sessions. These birds received extra bound and unbound what sessions, so that there were 2 of each of these sessions for every 1 bound or unbound where session. Training continued in this way until the criterion of 80% in the last 2 days of every task was met.

Blocks training. As in Experiment 1, the four trial types were intermixed within a session in blocks of gradually decreasing size, beginning with blocks of 16 and ending with blocks of 2. Each session consisted of 128 trials, 32 of each of the four types. Block order was randomized such that each type of trial was equally likely to follow any other type of trial. All possible combinations occurred once in each block of 64 trials. Block size was reduced for each bird once it achieved 80% correct on all four trial types for 4 days in a row. Thus, the minimum exposure to any block size was 4 days.

Random testing. In this phase, the four trial types were mixed randomly, with the constraint that in each block of 8 trials each trial type occurred twice. As in blocks training, sessions were 128 trials long, and, across successive blocks of two sessions, each possible sample event occurred equally often in each of the three types of trials. Following 6 random sessions with the original sample presentation time of 3 s and a 2-s RI, the sample presentation was reduced to 1 s, and the RI was increased on an individual bird basis until the lowest two of the scores on the four trial types were between 70% and 80% correct. The RI was increased first to 4 s, then to 8, 12, and 16 s if necessary. Following the first 3 days at a given RI, if the mean performance over the last 2 days on each of the worst two of the four trial types was between 70% and 80%, the same RI was maintained for another 3 days. If performance on all conditions was 80% or more, the RI was increased to the next longer length; if it was 70% or

below, the RI was returned to a lower value and maintained there. For purposes of analysis, the block of six sessions at the original 3-s exposure and 2-s RI were designated Random 1; the six sessions at the final RI (different for different birds) and 1-s exposure were designated that bird's Random 2.

Five-test phase. We added a final trial type to the previous four trial types by presenting a yellow club and a blue star 5.5 cm apart, center to center, around the center point of both the sample location and a randomly selected distractor location during the test phase of the trial. Thus, there were four stimuli on the screen, as depicted in Figure 5 (lower right). The left–right position of the paw and club was randomized independently at the sample and distractor locations. Reinforcement was given for pecking the alternative that matched the sample in both identity and location. Pigeons completed 20 sessions of 128 trials, in which 64 trials were of the new type and 64 trials were evenly divided among the four original types. Trials were randomized in such a way that in each block of 16 trials there were 8 trials of the new type and 2 trials of each of the four original types. All 16 possible sample events in the new trial type occurred once in each two blocks of 16 new plus old trials and, thus, four times in each session. Exposure time remained at 1 s, and each bird continued to experience the RI determined for it at the end of the random phase.

Results

Initial Acquisition

One bird had not achieved criterion performance on blocks of eight after 86 sessions of blocks training, whereas the other birds completed this stage in a median of 42 sessions (range = 34–53). Training was discontinued for this bird, and its data were excluded from all analyses. The remaining 5 birds required a median of 9 each of bound or unbound what sessions and 7 each of bound or unbound where sessions, or a total of 24–39 sessions to complete the acquisition phase.

Mean performance on the first six sessions of each type is depicted in Figure 6. Just as in Experiment 1, matching location (where) was acquired more rapidly than matching identity (what), but there was no indication of better performance in bound than in unbound tests. To the contrary, in the what task, performance was better in unbound than in bound tests, but in the where task, performance in the two kinds of tests overlapped throughout early
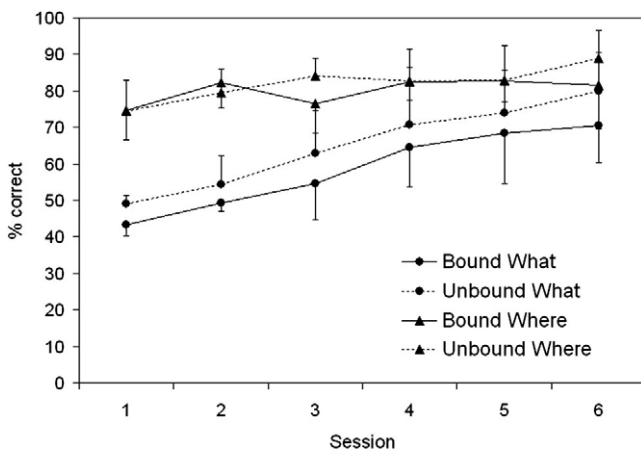

Figure 6. Mean (plus or minus standard deviation, denoted by the error bars) percentages of correct choices for the first six sessions of each kind in Experiment 2.

acquisition. This impression was confirmed by a Task × Bound–Unbound × Day ANOVA for the data from the first six sessions per type. There were significant effects of task, $F(1, 4) = 33.48$, bound–unbound, $F(1, 4) = 102.10$, and day, $F(5, 20) = 17.43$, and a significant Task × Bound–Unbound interaction, $F(1, 4) = 11.35$, Task × Day interaction, $F(5, 20) = 11.35$, and Day × Bound–Unbound, $F(5, 20) = 2.97$, interaction. The only nonsignificant interaction was Task × Bound–Unbound × Day, $F(5, 20) < 1$. We ran separate ANOVAs for each task to tease apart the interaction effects. For the what task, there were significant effects of both bound–unbound, $F(1, 4) = 60.00$, and day, $F(5, 20) = 19.95$, but no significant Bound–Unbound × Day interaction, $F(5, 20) < 1$. For the where task, there was a significant effect of day, $F(5, 20) = 4.80$, but not of bound–unbound, $F(1, 4) = 7.37$. There was also a significant Bound–Unbound × Day interaction, $F(5, 20) = 2.91$.

Blocks Training

Just as with the what and where tasks in Experiment 1, accuracy showed essentially no change when the four trial types were mixed within sessions. Birds spent a median of 17 (16–20) sessions in the blocks phase, close to the minimum required to progress through the four block sizes at above 80% correct.

Random Testing

By the time the birds reached the random testing phase, mean performance was near 90% on both what and where tasks and did not differ significantly between bound and unbound tests (Table 3). A Task × Bound–Unbound ANOVA showed no effect of task, $F(1, 4) < 1$, or of bound–unbound, $F(1, 4) = 4.58$, nor was there a significant Bound–Unbound × Task interaction, $F(1, 4) = 2.22$. Thus, the early slight superiority of unbound to bound trials for the what task disappeared with continued training.

The RI necessary to achieve the required lower level of performance with the 1-s sample exposure was 8 s for 3 of the birds, 4 s for another, and 12 s for the 5th. However, there was still no evidence of better performance on bound than on unbound tests (Table 3). A Task × Bound–Unbound ANOVA showed no significant effects of either task, $F(1, 4) < 1$, or bound–unbound, $F(1, 4) = 2.21$, but there was a significant Task × Bound–Unbound interaction, $F(1, 4) = 13.06$. Individual ANOVAs suggested that the source of this interaction was the difference between bound and unbound where tests (see Table 3), $F(1, 4) = 3.97$, $p = .117$ (other Fs < 1.0).

Four-Alternative Testing

Performance did not change noticeably across the 20 sessions that incorporated the new four-alternative tests. Group means for all scores on the second block of 10 sessions were within 1%–2% of scores in the first block. Therefore, all 20 sessions were combined for analysis. Mean performance on the four original types of trials is shown on the left of Figure 7. An overall Task × Bound–Unbound ANOVA for the regular trials for all 20 days showed no significant effect of either task, $F(1, 4) = 2.89$, or bound–unbound, $F(1, 4) = 1.97$. Unlike the case in the immediately

Table 3

Mean Percentages Plus Standard Deviation of Correct Responses in the Four Types of Trials in the Random Phases of Experiment 2

	What		Where	
RI	Bound	Unbound	Bound	Unbound
2 s	89.5 ± 3.7	93.7 ± 2.8	90.7 ± 5.5	90.7 ± 4.2
4–12 s	79.6 ± 6.5	81.1 ± 6.9	76.9 ± 2.8	81.4 ± 2.6

Note. RI = retention interval.

preceding stage, there was not a significant Task × Bound–Unbound interaction, $F(1, 4) < 1$.

The group mean proportions of four-alternative test trials with a correct response and each type of error appear on the right side of Figure 7. Choices of the unrewarded alternative that matched the sample in one feature were clearly more frequent than choices of the alternative that was wrong in both location and identity. To test whether the results of the four-alternative tests are consistent with responding based on independent what and where memories, we cast each bird's choice frequencies on these trials into a 2 × 2 table representing what correct versus incorrect by where correct versus incorrect. For example, correct choices were counted as being correct on both what and where, and the marginal total number of correct what choices was the sum of those correct choices and choices of the distractor that matched the sample in identity only (the choices represented by the bar labeled *What* in the right-hand part of Figure 7). Each bird's data were subjected to a chi-square test of independence. Just as for the related tests in Experiment 1 (see Tables 1 and 2), none of the distributions were even close to being significantly different from that expected if responding was based on independent what and where memories. Individual birds' values of $\chi^2(1)$ were between 0.00 and 1.35 ($ps > .24$).

Discussion

The results of this experiment are consistent with those of Experiment 1 in indicating that pigeons can perform very well on a matching task in which the feature to be matched changes randomly from one trial to the next. This means that they must be encoding multiple features of the sample (in the present case, location and identity) on a substantial proportion of trials. However, this experiment provides no evidence that the memories for location and identity are spontaneously bound together. We suggested that if they were bound, performance on the bound what and bound where tasks should have exceeded performance on the comparable unbound tasks, but, across the experiment, there were no consistent effects of bound versus unbound testing. Indeed, early in acquisition, birds were slightly but significantly worse on bound than on unbound what tests, contrary to what we suggested bound memories for the sample features would predict. However, as we pointed out in the introduction to this experiment, there also might be reason to expect more errors on bound than on unbound tests, which would make this comparison perhaps less than conclusive (as it turned out to be). Therefore, the most important finding is that when four alternatives were present in the test phase of trials, the pattern of errors did not differ from that expected if the birds' choices were based on independent memories for location and identity.

In the studies of memory for object location and identity in humans (Chalfonte & Johnson, 1996; Kohler et al., 2001) that we described in the introduction to this experiment, which features the subjects were asked to attend to during the study phase of trials influenced what features they best recognized at test. This is taken to show that memory for location and identity are somewhat independent. However, some encoding processes, such as those going on during object naming, support memory for both location and identity. Kohler et al. (2001) concluded from such findings that "successful recovery of object location and object identity from episodic memory relies on an intricate interplay between
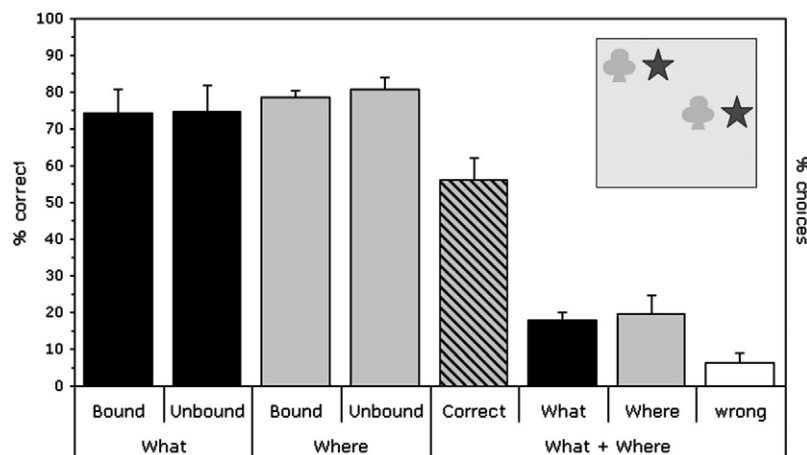

Figure 7. Experiment 2, final phase. Mean (plus or minus standard deviation, denoted by the error bars) percentages of correct choices in bound and unbound what and where trials (the four leftmost bars) and mean (plus or minus standard deviation, denoted by the error bars) percentages of four-alternative (what + where) trials with each kind of choice (the four rightmost bars). Scale is the same for all parts of the figure.

domain-specific and common encoding processes" (p. 958). The present findings are consistent with this conclusion in showing that location and identity of a sample are remembered independently. In that reducing exposure time and increasing the RI reduced performance uniformly on all types of trials (see Table 3), the results are also consistent with the suggestion that there are some common encoding processes.

There is an apparent conflict between our results and findings that humans can selectively encode location or identity. Testing the same feature for a block of trials can be taken as analogous to a manipulation of attention, as in studies of search image in pigeons (e.g., Reid & Shettleworth, 1992). However, the data analyzed from the blocks phase of Experiment 1 provide no evidence that reinforcing the pigeons for attending to a specific feature of the sample selectively improved memory for that feature in succeeding trials. This finding is, however, consistent with the results of studies on divided attention in pigeons in suggesting that they process location and identity independently and concurrently. Although pigeons' matching performance suffers when they must divide attention between the color and the orientation of a sample, no divided attention effect is seen when sample location is one of the features to be attended to, which suggests that location does not compete for attention with features that define the identity of the sample (see the review in Zentall, 2005b).

General Discussion

The experiments reported in this article were designed to test animals' W-W-W, or episodic-like, memory in a simpler way than in some previous studies reviewed in the introduction to this article. In Experiment 1, we presented pigeons with a single event—a particular colored shape in a particular location on a touch screen a particular number of seconds ago—and tested memory for its location, identity, or time of occurrence. The birds performed very well on all of these tests, even when they could not predict from one trial to the next which one would occur. Therefore, some proportion of the time, they were encoding what, when, and where all together. However, three different tests indicated that features were independent rather than bound or integrated in memory. In Experiment 1, the pattern of correct and incorrect choices on two successive tests of the same item was remarkably close to and not significantly different from that predicted if memories for what, when, and where were independent. During both acquisition and random testing in Experiment 2, birds did not do best on tests in which the binding of what and where was predicted to improve performance. If anything, there was a slight but not consistent tendency to perform better on unbound than on bound tests. Finally, the pattern of correct choices and errors in the final phase of Experiment 2 was no different than, and again very close to, that predicted by independent memories for location and identity.

Relation to Other Studies of Matching to Sample in Pigeons

For a valid test of episodic-like memory, subjects must be required to, in some sense, travel back in time to the episode in question, as opposed to producing a response that is habitual (Hampton & Schwartz, 2004). One way to ensure this is to use unique to-be-remembered episodes. Another is to ask the subject an unexpected question (Zentall, 2005a; Zentall et al., 2001). At first glance, our paradigm does not do either of these things. The to-be-remembered events were far from unique, because the pigeons were trained for thousands of trials, during which each possible combination of sample and test occurred many times. Moreover, although birds could not predict from one trial to the next what feature of the sample would be tested in the random phase of each experiment, the questions asked in the test were hardly unexpected in the sense promoted by Zentall (2005a) because there were only three possible kinds of questions.

Information about what pigeons actually learn when trained to match to sample provides some insights into what might have been going on in the present experiments. To begin with, pigeons apparently do not learn to match as such unless they are trained very extensively (see review in Shettleworth, 1998; Wright, Cook, & Rivera, 1988). For example, the fact that pigeons typically acquire symbolic matching as fast as literal matching indicates that they do not learn a matching concept but rather learn rules relating particular samples to particular comparisons (Zentall, Urcuioli, Jagielo, & Jackson-Smith, 1989). In addition, studies of pigeons' performance in more conventional matching tasks than the one used in this work indicate that the number of different samples or comparisons being used in the task and the mapping between them influences how samples are encoded. Coding tends to take the form that intuitively imposes the least cognitive demand. If each of two samples is mapped onto a single comparison—that is, in many-to-one matching—samples tend to be coded prospectively, in terms of the common correct choice on the test. Coding tends to be retrospective—that is, in terms of features of the sample itself— when two or more comparisons are associated with the same sample. This takes place in one-to-many matching, for example, when either a red shape or a vertical line is correct after the same sample.

Because the present tasks involved both more than two samples and more than two comparisons, it is not clear how we can extrapolate past findings to predict what kind of coding would be used. For example, the unbound what task can be seen as many-to-one matching, because the sample could appear in any of eight locations, whereas the test display was always a choice between the club and the star in the middle of the screen. The unbound where task also involved eight samples, but each was mapped to a unique correct choice in the test, so perhaps it was one-to-one matching. Finally, in random testing, three (Experiment 1) or four (Experiment 2) different kinds of test displays could appear after each sample, which ought to make it a test of one-to-many matching, except for the fact that different features of the (moveable and changeable) samples were mapped to the comparisons in different kinds of tests. Thus, the random phase included both common tests for different samples and different tests for common samples. The task is most easily described in human terms as matching the location, identity, and/or (in Experiment 1) recency of the sample, but, of course, this does not guarantee that the pigeons did eventually encode it that way.

Nevertheless, the validity of our paradigm as a test of W-W-W memory rests to some extent on the assumption that the birds would code the sample retrospectively as an object in a place and call on this representation to make their choices in the tests rather than learning the task as a set of associations between location +

identity samples and reinforced choices. Although pigeons can learn large sets of specific sample–choice associations (for a review, see Shettleworth, 1998), at least some of the present results argue against the birds having learned such associations in the present experiments. Consider a sample that might have appeared in Experiment 2—for example, the blue star in the upper left corner, as in Figure 5—and suppose for the moment that it was encoded as an integrated place plus object episode, or configural sample. This sample was followed on 25% of trials—that is, on unbound what trials—by the blue star and the yellow club in the center of the screen, and pecking the star was reinforced. On another 25% of trials, the unbound where trials, pecking a pink diamond in the upper left corner was reinforced in the test. On a full 50% of trials, the bound what and bound where trials, pecking a blue star in the upper left corner was reinforced. If the birds were encoding each of the 16 samples as a unique object plus location compound and learning a set of associations between these samples and the choices reinforced after each one, then they should have done equally well on bound what and bound where trials, because these trials required the same choice after a given sample. The birds should have done better on bound trials than on unbound trials, because each correct unbound choice occurred only half as often as the common correct bound choice. Instead, however, both where tasks were acquired more quickly than either what task, and bound versus unbound never made any difference in the predicted direction, consistent with our description of the tasks as separate tests of memory for location and for identity.

Implications for Tests of Episodic-Like Memory

The fact that pigeons remembered the what, where, and when of a single event very well in the present experiments might suggest that they would also perform well in a paradigm involving two or more events, analogous to the tests for scrub jays and rats described in the introduction to this article. In fact, before conducting the experiments reported in this article, we trained pigeons in several variations of such a paradigm, all of which followed procedures similar to those of the present experiments. In each of these variations, birds were exposed to a red and a green sample in succession, each in a different one of the eight peripheral locations on the touch screen. Then, following a 2-s or a 6-s RI, they were presented with two or more gray squares, including one in each of the places where a sample had been. The red sample's location was correct after the short RI, and the green sample's location was correct after the long one. To encourage the birds to learn the contingencies relating sample color, RI, and reinforcement, we displayed the samples again briefly once a location had been chosen.

This design is logically identical to the paradigm used by Clayton and Dickinson (e.g., 1998) for scrub jays. The display of gray squares plays the role of the sand-filled caching tray in which the jays chose the locations where worms had been cached after a short RI and the locations of peanut caches after a long one. Correct performance in such a paradigm depends, among other things, on the remembered cache (i.e., sample) locations retrieving a memory of cache identity and time since caching (i.e., time since exposure to the samples). This description makes clear why pigeons uniformly failed our two-item task. Regardless of RI and sample identity, in the test phase they generally chose the location

of the more recent sample. If pigeons encode identity and location independently, as indicated by the results of the present experiments, then presenting only location information in the test would not allow retrieval of the necessary information about the identity of the samples that had occupied those locations. It has been suggested that the episodic-like memory of scrub jays is an adaptation for retrieving stored foods that spoil at different rates (see de Kort et al., 2005). Because the present tests of W-W-W memory for pigeons involved very minimal "episodes" that differed in many ways from those experienced by Clayton and Dickinson's (1998) scrub jays, it is impossible to say whether the pigeons' failure to integrate the separable features of these episodes is attributable to their lack of such an adaptation, to some difference between the present tests and those given to the jays, or both.

Our findings suggest that other reported failures to demonstrate W-W-W memory in animals (e.g., Bird et al., 2003; Hampton et al., 2005) might also reflect not so much failures of memory for what occurred, where, and when but failures to bind or integrate independent memories for the three features in the way required by the tests (see also Thorpe, Bates, & Wilkie, 2003). Of course, encoding something with distinct spatial, identity, and temporal parameters as a unique event is not all that is required of episodic-like memory. For example, as Zentall (2005a) discussed, one might imagine training pigeons on a task similar to ours but restricted to the minimum number of samples (eight) necessary to require that time, place, and identity all be encoded to match each sample to a unique comparison. The memory involved in successful performance on such a task would not be considered episodic-like because the subject could solve the task by learning eight sample–comparison associations. Unlike our task, however, it would require simultaneous use of what, where, and when memories. By demonstrating excellent but apparently independent memories for what, where, and when, the present studies highlight the importance of integrating features in memory for success in tests of episodic-like memory and suggest that further analysis of the conditions necessary for integration could be a worthwhile topic for future research.

References

Babb, S. J., & Crystal, J. D. (2005). Discrimination of what, when, and where: Implications for episodic-like memory in rats. *Learning and Motivation, 36,* 177–189.

Babb, S. J., & Crystal, J. D. (2006). Discrimination of what, when, and where is not based on time of day. *Learning & Behavior, 34,* 124–130.

Bird, L. R., Roberts, W. A., Abroms, B., Kit, K. A., & Crupi, C. (2003). Spatial memory for food hidden by rats (*Rattus norvegicus*) on the radial maze: Studies of memory for where, what, and when. *Journal of Comparative Psychology, 117,* 176–187.

Blough, D. S., & Blough, P. M. (1997). Form perception and attention in pigeons. *Animal Learning & Behavior, 25,* 1–20.

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition, 24,* 403–416.

Clayton, N. S., & Dickinson, A. (1998, September). Episodic-like memory during cache recovery by scrub jays. *Nature, 395,* 272–274.

Clayton, N. S., Yu, K. S., & Dickinson, A. (2001). Scrub jays (*Aphelocoma coerulescens*) form integrated memories of the multiple features of caching episodes. *Journal of Experimental Psychology: Animal Behavior Processes, 27,* 17–29.

de Kort, S. R., Dickinson, A., & Clayton, N. S. (2005). Retrospective

cognition by food-caching western scrub-jays. *Learning and Motivation, 36,* 159–176.

Eacott, M. J., & Norman, G. (2004). Integrated memory for object, place, and context in rats: A possible model of episodic-like memory? *Journal of Neuroscience, 24,* 1948–1953.

Ergorul, C., & Eichenbaum, H. (2004). The hippocampus and memory for "what," "where," and "when." *Learning & Memory, 11,* 397–405.

Gallistel, C. R. (1990). *The organization of learning.* Cambridge, MA: MIT Press.

Hampton, R. R., Hampstead, B. M., & Murray, E. A. (2005). Rhesus monkeys *(Macaca mulatta)* demonstrate robust memory for what and where, but not when, in an open-field test of memory. *Learning & Memory, 36,* 245–259.

Hampton, R. R., & Schwartz, B. L. (2004). Episodic memory in nonhumans: What and where, is when? *Current Opinion in Neurobiology, 14,* 192–197.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 3,* 356–388.

Kohler, S., Moscovitch, M., & Melo, B. (2001). Episodic memory for object location versus episodic memory for object identity: Do they rely on distinct encoding processes? *Memory & Cognition, 29,* 948–959.

Reid, P. J., & Shettleworth, S. J. (1992). Detection of cryptic prey: Search image or search rate? *Journal of Experimental Psychology: Animal Behavior Processes, 18,* 273–286.

Roberts, W. A. (2002). Are animals stuck in time? *Psychological Bulletin, 128,* 473–489.

Roberts, W. A. (2005). Introduction. *Learning and Motivation, 36,* 107–109.

Roitblat, H. L. (1980). Codes and coding processes in pigeon short-term memory. *Animal Learning & Behavior, 8,* 341–351.

Shettleworth, S. J. (1998). *Cognition, evolution, and behavior.* New York: Oxford University Press.

Thorpe, C. M., Bates, M. E., & Wilkie, D. M. (2003). Rats have trouble associating all three parts of the time-place-event memory code. *Behavioural Processes, 63,* 95–110.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organisation of memory* (pp. 381–403). New York: Academic Press.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53,* 1–25.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale, NJ: Erlbaum.

Wright, A. A., Cook, R. G., & Rivera, J. J. (1988). Concept learning by pigeons: Matching-to-sample with trial-unique video picture stimuli. *Animal Learning & Behavior, 16,* 436–444.

Zentall, T. R. (2005a). Animals may not be stuck in time. *Learning and Motivation, 36,* 208–225.

Zentall, T. R. (2005b). Selective and divided attention in animals. *Behavioural Processes, 69,* 1–15.

Zentall, T. R., Clement, T. S., Bhatt, R. S., & Allen, J. (2001). Episodic-like memory in pigeons. *Psychonomic Bulletin & Review, 8,* 685–690.

Zentall, T. R., Urcuioli, P. J., Jagielo, J. A., & Jackson-Smith, P. (1989). Interaction of sample dimension and sample-comparison mapping on pigeons' performance of delayed conditional discriminations. *Animal Learning & Behavior, 17,* 172–178.