Object, spatial, and temporal memory: A behavioral analysis of visual scenes using a what, where, and when paradigm

Stephanie J. Babb and Ruth M. Johnson
Tulving (1972) originally distinguished between episodic memory (the encoding and retrieval from memory of unique, personal past experiences) and other types of memory. This early distinction focused on the difference between personally experienced events and knowledge of general facts about the world (i.e., semantic memory). In particular, Tulving (1972) proposed that “Episodic memory receives and stores information about temporally dated phases or events, and temporal-spatial relations among these events” (p. 385). Therefore, episodic memory is defined by what happened, and when and where the event occurred.

Although knowledge of what, where, and when remains an important component of episodic memory (Nyberg, McIntosh, Cabeza, Habib, Houle, & Tulving, 1996), it is not the only feature of episodic memory. In particular, Tulving and colleagues (e.g., Tulving, 1983; Tulving & Kim, 2007; Tulving & Markowitsch, 1998) have focused on three additional features of episodic memory: the ability to recognize subjective time, autonoetic consciousness, and knowledge of a “self”, all of which are deemed necessary for mental time travel (i.e., the ability to subjectively re-experience an event). According to that perspective, the additional features distinguish between recalling a personal past experience and remembering an impersonal fact. However, the behavioral criteria for episodic memory remain important test measures, especially in non-human studies (Babb & Crystal, 2005, 2006a, 2006b; Clayton & Dickinson, 1998).

Clayton and colleagues (Clayton, Bussey, & Dickinson, 2003; Clayton & Dickinson, 1998; Clayton & Griffiths, 2002; Clayton, Griffiths, Emery, & Dickinson, 2001) have argued that behavioral studies of episodic memory should focus on knowledge of what, where, and when as criteria for episodic-like memory in animals, instead of using subjective states of
consciousness, which are difficult, if not impossible, to test in non-human, non-verbal animal subjects.

In a series of experiments with rats, Babb and Crystal (2005, 2006a, 2006b) used Clayton's paradigm (Clayton & Dickinson, 1998) to present evidence that rats have detailed knowledge of what, where, and when, and therefore provided evidence of episodic-like memory. These studies were the first published evidence of episodic-like memory in mammals using Clayton's behavioral criteria of what, where, and when, and have been replicated by other researchers (Naqshbandi, Feeney, McKenzie, & Roberts, 2007).

Most studies of episodic memory focus on an integrated component consisting of all three what-where-when features. However, to understand the brain mechanisms responsible for each component of episodic memory, they must also be studied independently (Hoffman, Beran, & Washburn; 2009; Skov-Rackette, Miller, & Shettleworth; Ungerleider & Haxby, 1994). Skov-Rackette and colleagues (2006) have shown that pigeons can perform well on tasks involving separate what, where, and when components, but there was no evidence that these components were bound together. Hoffman et al. (2009) found that rhesus monkeys can integrate the three components, but show different levels of performance for what, where, and when.

Nonverbal, behavioral tests of episodic memory have been adapted for humans. Holland and Smulders (2011) required that participants hide four different coin types (what) in distinct locations (where) on two different occasions (when). Subjects reported using mental time travel to complete the task, and the authors argue that this demonstrates the use of episodic memory during the task.

Hayes and colleagues (Hayes, Ryan, Schyner, & Nadel, 2004) separately tested memory for what, where, and when in human subjects using visual images taken from a recently-viewed videotape of furnished homes while subjects were scanned using functional Magnetic Resonance Imaging (fMRI). However, the Hayes et al. study used one test image that had been previously seen in the videotape, and one test image that had not been seen; therefore, the subjects’ answers could have been based on familiarity instead of actual recollections. Recognition memory utilizes two independent mechanisms: episodic recollection of a specific event and a sense of familiarity of a previously experienced stimulus (Sauvage, Fortin, Owen, Yonelinas, & Eichenbaum, 2008).

The present study adapted the behavioral criteria established by Clayton and Dickinson (1998) to also independently test what, where, and when memory in humans. Participants in this study viewed unique visual scenes of the interiors of furnished homes in the Houston, Texas area. The scenes each contained multiple objects in various arrangements, including images of completely furnished rooms. In two experiments, participants were shown the images on a computer screen in a blocked or mixed design, and were then separately tested for memory for the object that was missing from a scene (What condition), the correct spatial arrangement of the objects in the scene (Where condition), and the temporal order of the scenes (When condition). We predicted that participants would be able to successfully complete each condition significantly above chance and therefore show memory for what, where, and when. We also predicted that the mixed design would result in increased accuracy and faster reaction times. The establishment of a non-verbal, human model of episodic memory based on Clayton’s animal model would validate the animal model of episodic memory and therefore provide an important model for future functional, pharmacological, and neuroanatomical studies.
Experiment 1

Experiment 1 was designed to measure participants’ accuracy and reaction time when recalling information about visual scenes; specifically, the objects viewed in those scenes, the spatial arrangements of those scenes, and the temporal order of the scenes. Experiment 1 consisted of 3 blocks (What, Where, and When conditions) of 20 visual scenes each. After viewing the images, participants were then tested on each condition.

Method

Participants. Thirty-two participants (7 males/25 females; 18-47 years of age with an average age of 23; 4 left-handed/28 right-handed) were recruited from the Psychology research subject pool at the University of Houston-Downtown (UHD). Before the experiment began, participants gave their consent to be in the study. This experiment was approved by the Committee for the Protection of Human Subjects.

Apparatus. This experiment was conducted using PC computers in the UHD psychology data laboratory. The experimental software was custom written using E-Prime (Psychology Software Tools, PA) and the data analyzed using SPSS 17.0. The visual images were taken at a resolution of 640 x 480 pixels using a digital camera set on a tripod stand to ensure that the images remained identical when an item was moved to a new spatial location or removed from the scene.

Design. In Experiment 1, participants were given three randomly presented blocks of 20 trials each. Each block of trials was a different condition (What, Where, and When) in which participants were shown 20 different scenes for 3 seconds each (see Figure 1). While viewing each block of photographs, participants were unaware which test condition corresponded with the images. In the What condition test trials, participants were given 20 trials where they were shown one of the previously viewed images, but an object was removed from the scene (see Figure 2). Subjects were asked to recall the object that was removed from the original visual scene. After viewing the object-removed scene for 3 seconds, participants were then shown the target object and another object they had seen in a different scene. Participants were instructed that they needed to make a response within 5 seconds or the computer program would automatically move to the next trial. Subjects pressed the ‘1’ key to indicate the object on the left was the object that was removed or the ‘2’ key to indicate the object on the right was the target object. A correct response in the What condition was determined by choosing the target object that belonged in the visual scene that subjects had just viewed. The Where condition was identical to the What condition, except the 20 test trials consisted of the original scene on one side of the monitor and a scene in which the target object was moved to a different spatial location on the other side (see Figure 3). Participants were asked to recall which of the images was the original visual scene viewed in the previous set of images. Subjects again pressed the ‘1’ key to indicate the image on the left was the original image or the ‘2’ key to indicate the image on the right was the original image. A correct response in the Where condition was determined by choosing the visual scene with the correct spatial arrangement. In the When condition test trials, subjects were shown two of the previously viewed scenes and asked to judge which scene they had observed earlier in the sequence of pictures (see Figure 4). Participants pressed the ‘1’ key to indicate the image
on the left was the scene that was shown earlier in the set or the ‘2’ key to indicate the image on the right was the scene that was shown first. A correct response in the When condition was determined by choosing the visual scene that was shown earlier during the original set of images. This experiment had three blocks of 20 trials each for a total of 60 trials. Subjects’ accuracy rates and reaction times (RT) were recorded on all trials.

Figure 1.

Visual scenes of the interiors of furnished homes were displayed in blocks of 20 images. Each image was shown for 3 s in Experiment 1 and 4 s in Experiment 2.

Figure 2. Example of the What condition testing object memory.

Participants were asked “Which object was removed?” and then given 5 trials. The What trials consisted of viewing a previously viewed image with one object removed for 4 s, followed by a choice between two previously seen plausible objects.
In Experiment 1, subjects were required to respond within 5 seconds or the program automatically advanced to the next trial. If participants failed to respond on a trial, that trial was excluded from further analysis. A total of 173 trials were excluded from the experiment (9.0% total). In addition, six of the 32 participants failed to respond on more than 50% of trials in one of the three blocks. Given that this happened at least once across all three conditions and each of the six subjects responded to almost all trials in the other two conditions, the subjects’ missing data was replaced with the mean for that condition, instead of excluding all six participants’ data. However, when all six subjects’ data were excluded, the means for each condition varied by approximately 1% or less. Therefore, we attempted to keep as much of the data as possible and not exclude the subjects from further analyses.

Accuracy rates for Experiment 1 are shown in Figure 5. In the What condition, accuracy was 71.5%, compared with 59.6% in the Where condition, and 59.3% in the When condition.
condition. As predicted, accuracy rates in all three conditions were significantly greater than chance (50% detection) (all \( t \)'s ≥ 3.5, all \( p \)'s ≤ .001). An ANOVA was conducted on the accuracy data across the conditions (What, Where, and When). Overall, the accuracy rates across the three conditions were significantly different from one another, \( F(2,62)= 8.286, p = .001 \). Participants were significantly more accurate in the What condition compared to the Where and When conditions, \( t(31)= 3.501, p = .001 \), two-tailed and \( t(31)= 4.063, p < .001 \), two-tailed, respectively. However, the Where and When conditions were not significantly different, \( t(31)= 0.070, p = .945 \), two-tailed. Serial position effects were also analyzed across the three blocked conditions in Experiment 1, but significant differences were not found.

Figure 5. Object, Spatial, and Temporal Memory

Experiment 1 and 2 results demonstrating object, spatial, and temporal memory for visual scenes. Bar graphs indicate accuracy (shown as percent correct) and solid lines represent reaction time (shown in ms). Accuracy rates were significantly above chance in all conditions across both experiments (all \( p \)'s < .001). Participants were significantly faster and more accurate in the What condition compared to the Where and When conditions in Experiment 1 and 2 (\( p \)'s < .05). Furthermore, participants were significantly faster and more accurate in Experiment 2 compared to Experiment 1 (\( p \)'s < .001).

Reaction times (RTs) for Experiment 1 are also shown in Figure 5. In the What, Where, and When conditions, participants RTs were 1748, 2899, and 2309 ms, respectively. An ANOVA was conducted on the reaction times across the conditions. Overall, the RTs across the three blocks were significantly different from one another, \( F(2,62)= 86.614, p < .001 \). Furthermore, participants were significantly faster in the What condition compared to the Where and When conditions, \( t(31)= 15.211, p < .001 \), two-tailed and, \( t(31)= 6.076, p < .001 \), two-tailed, respectively. Subjects were also faster in the When condition compared to the Where condition, \( t(31)= 6.326, p < .001 \), two-tailed.
Experiment 2

Experiment 2 was designed to increase participants’ accuracy rates. We hypothesized that shorter test blocks and longer viewing time would result in greater recall; therefore, we presented the participants with a mixed design and an extra second of viewing time. A Control condition was also included to access participants’ baseline levels for visual search.

Method

Participants. Thirty-eight participants (7 males/31 females; 18-58 years of age with an average age of 26; 3 left-handed/35 right-handed) were recruited from the Psychology research subject pool at UHD. This experiment was approved by the Committee for the Protection of Human Subjects, and subjects gave their informed consent before the experiment began.

Apparatus. The apparatus was identical to Experiment 1.

Design. Experiment 2 was similar to Experiment 1 with two major changes: a mixed-condition design and an additional Control condition. In this experiment, participants were given four sets of photographs, and each block of 20 visual scenes contained 5 images of each of the four conditions (What, Where, When, and Control). After viewing 20 visual scenes for 4 seconds each, participants were randomly given 5 trials of each of the four conditions (for a total of 20 trials). Before each of the 5 test trials, participants were notified which condition they were about to view (i.e. What, Where, When, or Control) and asked a simple question to help them understand how to respond. In the What condition, subjects were asked “What object was removed?” and were then shown one of the previously viewed scenes with one object removed for 4 seconds and asked to chose whether the object shown on the right or left was the removed target object. Like Experiment 1, subjects pressed the ‘1’ key to indicate the image on the left or the ‘2’ key to indicate the image on the right. A correct response in the What condition was determined by choosing the target object that belonged in the visual scene that subjects had just viewed. In the Where condition, participants were asked “Which image did you see before?” and then chose between the original scene and a scene in which the target object was moved to a different spatial location. A correct response in the Where condition was determined by choosing the visual scene with the correct spatial arrangement. In the When condition, subjects were asked “Which image did you see first?” and then chose between two previously viewed scenes. A correct response in the When condition was determined by choosing the visual scene that was shown earlier during the original set of images. In the Control trials, subjects saw two identical images of a previously viewed scene; however, one image had the letter X overlaid on top of it and the other had the letter O (see Figure 6). Participants were asked “Which image has the letter X in it?” and chose the image with the letter X overlaid on the image. This experiment had four blocks of 20 trials each for a total of 80 trials. Subjects’ accuracy and RTs were recorded on all trials.
Results and Discussion

20 In Experiment 2, subjects were required to respond within 4 seconds or the program automatically advanced to the next trial. If participants failed to respond on a trial, that trial was excluded from further analysis. A total of 94 trials were excluded from the experiment (3.1% total). Accuracy rates for Experiment 2 are shown in Figure 5. In the What condition, accuracy was 76.8%, compared with 69.1% in the Where condition, and 65.5% in the When condition. Accuracy in the Control condition was 99.4%. Accuracy rates in all four conditions were significantly greater than chance (50% detection) (all t’s > 6.5, all p’s < .001). An ANOVA was conducted on the accuracy data across the conditions (What, Where, When, and Control). Overall, the accuracy rates across the four conditions were significantly different from one another, F(3,111)= 52.074, p < .001. Furthermore, participants were significantly more accurate in the What condition compared to the Where and When conditions, t(37)= 2.296, p = .027, two-tailed and t(37)= 3.273, p = .002, two-tailed, respectively. However, the Where and When conditions were not significantly different, t(37)= 1.030, p = .310, two-tailed. As expected, participants were significantly more accurate in the Control condition compared to the other three conditions (all t’s > 9.0, all p’s < .001).

21 Reaction times (RTs) for Experiment 2 are also shown in Figure 5. In the What, Where, When, and Control conditions, participants’ RTs were 1550, 2377, 1975, and 762 ms, respectively. An ANOVA was conducted on the RTs across the conditions. Overall, the RTs across the four conditions were significantly different from one another, F(3,111)= 588.362, p < .001. Furthermore, participants were significantly faster in the What condition compared to the Where and When conditions, t(37)= 36.844, p < .001, two-tailed and t(37)= 8.393, p < .001, two-tailed, respectively. Subjects were also faster in the When condition compared to the Where condition, t(37)= 8.366, p < .001, two-tailed.

22 Between Experiment Analyses. Experiment 2 was designed to increase participants’ accuracy. Therefore, a 2 x 3 mixed-factor ANOVA was performed on the accuracy rates with experiment (1 and 2) as the between-subjects factor and condition (What, Where, and When) as the within-subjects factor. A main effect of condition was found, F(2,136)= 13.261, p < .001, as well as a main effect of experiment, F(1,68)= 4921.345, p < .001;
however, the interaction between condition and experiment was not significant, \( F(2,136)=.405, p=.668 \). As shown in Figure 5, the mixed design in Experiment 2 increased participants’ accuracy across the What, Where, and When conditions. A 2 x 3 mixed-factor ANOVA was also conducted on the RT data with experiment (1 and 2) as the between-subjects factor and condition (What, Where, and When) as the within-subjects factor. A main effect of condition was found, \( F(2,136)=228.569, p<.001 \), as well as a main effect of experiment, \( F(1,68)=2595.334, p<.001 \). The interaction between condition and experiment was also significant, \( F(2,136)=6.208, p=.003 \). Participants responded more quickly in Experiment 2 compared to Experiment 1, especially in the Where condition which likely accounted for the significant interaction in reaction times (see Figure 5).

**General Discussion**

23 This study adapted Clayton and Dickinson’s (1998) behavioral model for testing episodic-like memory in animals to examine what, where, and when memory in humans using a nonverbal task. In two experiments, participants demonstrated memory for objects, spatial arrangements, and temporal order in visually presented scenes. Memory for objects is analogous to scrub jays’ recall of what type of food was cached; spatial configuration is analogous to where the food was cached in the trays; and the temporal order of the scenes is comparable to the birds’ remembrance of when, or how long ago, food types were cached.

24 Across both experiments, the data showed that performance was highest and RT was fastest for the What condition compared to the Where and When conditions (all \( p’s < .05 \)). In the What condition, participants likely responded significantly faster to the test items because they were more confident in their answers, which was reflected in higher accuracy rates. Both humans and animals are generally less accurate at reporting when an event took place, because temporal memory is less salient than object and spatial memory (Bird, Roberts, Abroms, Kit, & Crupi, 2003; Roberts & Roberts, 2002). However, in this study, spatial memory was not significantly more accurate than temporal memory, most likely due to the difficulty of the task.

25 Participants were significantly faster and more accurate in Experiment 2 when compared to Experiment 1 (all \( p’s < .001 \)). Therefore, participants seemed to benefit from shorter, mixed test blocks, and perhaps from the extra second of viewing time for the images. Because subjects could have used recognition memory, or familiarity, in the When condition, there is a possibility that the participants were using trace strength (Craik & Tulving, 1975) to solve the question of temporal order; however, because significant serial position effects were not found, it does not seem likely that the participants were using this strategy. The present study attempted to reduce the possibility of using familiarity as a strategy by only using images that had previously been viewed. Nevertheless, in the Where condition, participants could have chosen the correct spatial arrangement based on familiarity; however, because both test scenes were identical except for the location of one object, it would have been difficult to choose the correct scene based on familiarity alone. Clayton and colleagues (Clayton & Dickinson, 1998, 1999a, 1999b; Clayton, Yu, & Dickinson, 2001, 2003) have also ruled out use of relative familiarity as a strategy by modifying the caching paradigm to remove the differential relative familiarity of the tray. Their studies have shown that food-storing birds can remember what type of food they cached, where they cached it, and when they cached it (Clayton & Dickinson, 1998,
If characterizing episodic memory in animals is based on the information encoded rather than the subjective experience that accompanies the memory, then there is evidence that scrub jays possess episodic-like memory.

In the present study, participants were shown photographs without knowing which specific features they would be tested on later. Therefore, they had to simultaneously encode object, spatial, and temporal information about each slide. Results showed that participants were above chance in all trial types; however, performance on the three components differed from one another, and therefore could possibly be stored separately. Participants were significantly more accurate in the What condition, which is similar to Hoffman et al.’s (2009) findings, in which monkeys were more accurate in the What condition when all 3 memory components were integrated. This discrepancy could be due to a participant strategy of focusing more on objects located within the scenes rather than on the entire scenes themselves. Also, on the What trials, participants only had to recall the object that belonged in the scene; on the Where trials, participants had to recall the prior location of an object, which also requires some remembrance that the object belonged in that particular scene; on the When trials, participants had to recall the sequential order of the entire scenes, plus they had probably memorized some object information within the scenes. In other words, spatial and temporal trials probably also incorporated some object recognition as part of the task. Hayes and colleagues (2004) conducted a similar study, and also found that object trials were easier to recall than spatial trials, which were easier to recall than temporal-order trials.

This study used the behavioral model of episodic memory to separately test each what-where-when component, and therefore provided a nonverbal, behavioral model for future studies of the cortical mechanisms responsible for encoding and recalling these types of memory. These types of functional studies would validate Clayton and colleagues’ what-where-when model of episodic memory in animals. Validation of the mammalian behavioral model of episodic memory (Babb & Crystal, 2005) is necessary in order to explore the neural, anatomical, and pharmacological mechanisms underlying human episodic memory diseases and disorders.

Currently, there is some functional data for the brain regions responsible for memory in terms of what, where, and when. The ventral and dorsal pathways (Ungerleider & Haxby, 1994) likely contribute to the cortical mechanisms underlying episodic recall for object and spatial memories, respectively. The lateral occipital complex is also important in object recognition and recall in humans (Grill-Spector, Kourtzi, & Kanwisher, 2001).

The hippocampus has been implicated in spatial memory tasks, and may provide a basis for the neural mechanisms underlying memory for the spatial arrangements of the items in the scenes (O’Keefe & Nadel, 1978; Hayes et al., 2004). The right hippocampus processes spatial locations, while the left hippocampus is involved in episodic memory (Burgess et al., 2002; Maguire, 2001). The left and right prefrontal cortices are implicated in encoding and retrieval of episodic memory, respectively (Tulving, 2002). Hayes and colleagues (2004) found activation in the frontal and medial temporal regions during episodic recall using a what, where, and when task with videotaped images.

The hippocampus also plays an important role in the acquisition of new memories; it incorporates temporal information from the frontal lobes, thus providing a basis for the neural mechanisms underlying memory for the temporal order of the visual scenes (Burgess, Maguire, & O’Keefe, 2002). Eichenbaum and Fortin (2003) argued that memory
for the sequential order of unique events provides a model for episodic memory, and that
the hippocampus is involved in remembering these sequences of events (Fortin, Agster, &
Eichenbaum, 2002). Furthermore, patient research has also found disruptions in
sequential order memory recall after frontal lobe lesions (Mangels, 1997).

In order to understand how episodic memory fails in patients with Alzheimer’s disease,
schizophrenia, autism, and other memory disorders, research must first help us to better
understand how normal memory works. Only by understanding how normal memory
works will it be possible to identify failures of memory, which aspects of memory have
failed, what treatments might be effective, and the degree to which treatments are
effective. Any complete understanding of human memory requires an understanding of
memorial processes in nonhuman species because verbal memory processes, such as
verbal coding and rehearsal, cannot be used by these species and cannot be employed as
hypothetical constructs to explain the memory results. Human memory may be different
from memory in other animals, possibly due to language or a more fully developed
neocortex, but these differences may be quantitative in nature rather than qualitative
(Shettleworth, 1993). Studies such as this one that are directly comparable to studies in
animals are the first step in understanding human memory diseases and disorders, and
can provide a foundation for investigating basic processes of episodic memory with
neuropsychological recordings, manipulations of brain area-specific neurotransmitters,
and drugs to improve degraded memory.

BIBLIOGRAPHY

for episodic-like memory. Learning and Motivation, 36, 177-189.

Babb, S. J., & Crystal, J. D. (2006 a). Discrimination of what, when, and where is not based on time
of day. Learning and Behavior, 34(2), 124-130.


by rats (Rattus norvegicus) on the radial maze: Studies of memory for what, where, and when.
Journal of Comparative Psychology, 117, 1-12.


Clayton, N. S., Bussey, T. J., & Dickinson, A. (2003). Can animals recall the past and plan for the


Clayton, N. S., & Dickinson, A. (1999). Scrub jays (Aphelocoma coerulescens) remember the relative time of caching as well as the location and content of their caches. Journal of Comparative Psychology, 113, 403-416.


ABSTRACTS

Tulving (1972) defined episodic memory as memory for what, where, and when. Clayton and Dickinson’s (1998) behavioral model for animals was adapted to examine what, where, and when memory in humans. Participants viewed unique visual scenes of furnished homes in both blocked and mixed designs. They were tested separately for memory of objects (What), spatial configurations (Where), and temporal order of the scenes (When), and participants’ accuracy and reaction times were examined for each condition. Performance was highest and reaction time was fastest for the What condition. Participants were also faster and more accurate in the mixed design experiment compared to the blocked design experiment. This study established a behavioral analysis of episodic memory in humans based upon Clayton and Dickinson’s (1998) animal model, which will provide a basis for functional episodic memory studies to separately characterize the cortical mechanisms for processing episodic memory using the object, spatial, and temporal task with humans.

INDEX

Keywords: episodic memory, visual scenes, visual memory, what-where-when memory

Mots-clés: mémoire épisodique, scènes visuelles, mémoire visuelle, le "quoi" le "où" et le "quand" de la mémoire

AUTHORS

STEPHANIE J. BABB
University of Houston-Downtown
Department of Social Sciences One Main Street Houston, TX 77002
BabbS@uhd.edu

RUTH M. JOHNSON
University of Houston-Downtown
Department of Social Sciences One Main Street Houston, TX 77002
JohnsonRu@uhd.edu