

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Benjamin Michael Basile

Date

Development of a Recall Memory Test
for Rhesus Monkeys (*Macaca mulatta*)

By

Benjamin M Basile
B.A. Carleton College, 2002

Psychology

Robert Hampton, Ph.D.
Adviser

Stephan Hamann, Ph.D.
Committee Member

Jocelyne Bachevalier, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

Date

Development of a Recall Memory Test for Rhesus Monkeys (*Macaca mulatta*)

By

Benjamin M. Basile
B.A., Carleton College, 2002

Advisor: Robert Hampton, Ph.D.

An Abstract of
A thesis submitted to the Faculty of the Graduate
School of Emory University in partial fulfillment
of the requirements for the degree of
Master of Arts
in Psychology
2009

Abstract

Development of a Recall Memory Test for Rhesus Monkeys (*Macaca mulatta*)
By Benjamin M. Basile

Human memory can be assessed with two types of tests: 1) recognition tests, in which subjects judge whether a current stimulus occurred previously, and 2) recall tests, in which subjects reproduce a previous stimulus in its absence. In contrast, there is a lack of comparable recall tests in nonhuman animals. Here, we describe the development of a touchscreen-based, shape reproduction task for rhesus monkeys that is similar to recall tests in humans. At study, monkeys saw a simple shape composed of two colored boxes located on a 5x5 grid. At test, one of the two boxes appeared in a new grid location. Monkeys earned food for reproducing the studied stimulus by touching the appropriate grid location for the second box. Performance was significantly above chance levels. In comparison with a recognition test matched for chance rate, monkeys showed similar accuracy and response latency for the two tests. This contrasts with human recall tests, which are normally harder than recognition tests. Possible reasons for this discrepancy are addressed and a new recognition test is proposed.

Development of a Recall Memory Test for Rhesus Monkeys (*Macaca mulatta*)

By

Benjamin M. Basile
B.A., Carleton College, 2002

Advisor: Robert Hampton, Ph.D.

A thesis submitted to the Faculty of the Graduate
School of Emory University in partial fulfillment
of the requirements for the degree of
Master of Arts
in Psychology
2009

Contents

Introduction	1
<i>Familiarity and recollection</i>	2
<i>Recall in nonhumans</i>	7
Recall training	9
<i>Method</i>	9
<i>Subjects</i>	9
<i>Apparatus</i>	9
<i>Procedure</i>	10
<i>Accuracy training</i>	10
<i>Recall training phase 1</i>	11
<i>Recall training phase 2</i>	11
<i>Recall training phase 3</i>	12
<i>Recall training phase 4</i>	12
<i>Recall training phase 5</i>	13
<i>Recall training phase 6</i>	13
<i>Data analysis</i>	14
<i>Results</i>	14
Comparison with recognition	15
<i>Method</i>	16
<i>Subjects and apparatus</i>	16
<i>Procedure</i>	16
<i>Recognition training with clipart</i>	16
<i>Recognition training with box shapes</i>	17
<i>Comparison with recall test</i>	17
<i>Data analysis</i>	19
<i>Results</i>	19
Discussion	23
<i>Group findings</i>	23
<i>Limitations and future directions</i>	24
<i>Individual findings</i>	25
<i>Conclusion</i>	26
References	27

Figures and Tables

Figure 1.	Image of the testing environment and computer touchscreen	10
Figure 2.	Schematic progression of a recall trial	14
Figure 3.	Individual performance on the last phase of recall training	15
Figure 4.	Schematic progression of a recognition trial	19
Figure 5.	Group and individual accuracy on recall and recognition	20
Table 1.	Chi square and p values for individual monkeys	21
Figure 6.	Group and individual response latencies on recall and recognition	21
Figure 7.	Group and individual forgetting rates across multiple delay lengths	22
Table 2.	Chi square and p values for individual monkeys at different delay lengths	22

Development of a Recall Memory Test for Rhesus Monkeys (*Macaca mulatta*)

Imagine that you are at a conference and you suddenly spot a familiar face. The person seems familiar, but you cannot recollect any details of meeting them. This familiar-face phenomenon illustrates an apparent dissociation in human memory between the systems used for recognizing whether you have previously experienced something and the systems used for recalling detailed events.

The underlying processes illustrated by the above example are often examined in the lab using tests of recognition and recall. In recognition tests, subjects indicate whether a stimulus has been previously experienced. In one type of recognition test, delayed match-to-sample, subjects might be shown an object, given a delay, and then be presented with that same object among distracter objects. The subjects show recognition by indicating which object they had previously seen (for example, see Postman et al., 1948; Reed et al., 1997). In recall tests, subjects verbally or visually reproduce a previously experienced stimulus. Subjects taking the Rey-Osterrieth Complex Figure Task, a nonverbal recall test, are shown a figure consisting of an abstract line drawing and asked to copy it. After a delay, subjects are given a blank piece of paper and a pencil. The subjects show recall by reproducing (drawing) the figure they had previously seen (for example, see Fastenau et al., 1999; Vargha-Khadem et al., 1997). The key procedural difference between recognition and recall tests is that the studied item is re-presented during a recognition test, but not during a recall test.

Familiarity and recollection

People perform differently on tests of recall than they do on tests of recognition. Psychologists have theorized this is because subjects use two different types of memory: familiarity and recollection (Kelley & Jacoby, 2000; Yonelinas, 2002). Familiarity refers to a judgment of novelty or recency. Items that have been experienced recently or repeatedly will be judged as more familiar than items that have not been experienced, have been experienced long ago, or have only been experienced briefly. In contrast, recollection is a more detailed retrieval of information about or associated with an experience. For example, you may recollect where an item was seen, who showed it to you, or what other items were nearby at the time. Because previously experienced items are re-presented to the subject in recognition tests, these tests can be solved either through recollecting the event of seeing the object or through a judgment of which object is most familiar. In contrast, because no items are present to trigger a familiarity response in recall tests, they can only be solved by recollection. In the Rey-Osterrieth Complex Figure Task discussed above, subjects cannot use familiarity to successfully reproduce the figure because it is not re-presented during test. Thus, recognition tests can be solved using either familiarity or recollection, whereas recall tests require recollection.

There is good evidence to support theoretical distinction between recollection and familiarity. Here, I will cover one of the more convincing findings, then two findings relevant to the current study, and finally briefly touch on three other lines of evidence. First, some amnesic patients, who have damage limited primarily to the hippocampus, show spared performance on recognition tests despite impaired performance on recall

tests. For example, patient Jon can recognize pictures he has previously seen and performs as well as normal control subjects (Baddeley et al., 2001), however, he cannot reproduce a previously seen shape from memory (Vargha-Khadem et al., 1997). Jon's poor performance is not the result of motor or perceptual problems, as he could copy the shape when it was present. In addition, Jon's impairment on recall tests encompasses both verbal and visual stimuli, suggesting that his condition is a failure of a general retrieval process and is not domain specific.

Other patients with damage restricted to the hippocampus show a similar pattern of spared familiarity memory and damaged recollection memory. Patient Y.R. performs as well as controls when asked to choose which of 4 highly-similar options she has seen before (Holdstock et al., 2002), however, she performs significantly worse than controls when the drawings are absent and she is asked to recall them by name. Like Jon, her deficit is not due to trouble processing the drawings themselves, as she had previously been able to name the drawings and make a "man-made/natural" judgment about each when it was present. Y.R. has been extensively tested; she performed statistically as well as controls on 39 out of 43 recognition tests. In contrast, she is significantly impaired on 32 out of 34 recall tests (Mayes et al., 2002). Jon's and Y.R.'s spared performance when the stimulus was re-presented and impaired performance when it was not, suggests they can accurately judge the familiarity of a currently-perceived stimulus, but not recollect that stimulus in its absence. As their brain damage is limited primarily to the hippocampus, some researchers have suggested that the hippocampus supports recollection, but not familiarity (Aggleton & Brown, 1999; Yonelinas, 2002). However, this neurological dissociation is controversial, as the majority of amnesic patients show

deficits in both recall and recognition tests (Haist et al., 1992; Manns et al., 2003; Wais et al., 2006).

A second type of evidence that recall and recognition assess different types of memory is that recognition tests are usually easier than recall tests. This is proposed to be because subjects can use two memory systems to solve recognition tests, but only one memory system to solve recall tests. One of the most reliable findings in memory research is that subjects can typically recognize more items than they can recall (Brown, 1975). For example, when subjects memorized a list of 48 nonsense syllables, they were later able to recognize an average of 25.9 syllables using a four choice test and recall an average of 8.1 syllables (Postman et al., 1948). Equating these two tests is often difficult though, because the success rate expected by chance is often known for recognition tests, but not for recall tests. One way to circumvent this problem is to compare success on recognition tests that can be readily solved on the basis of a familiarity strategy with recognition tests that are difficult to solve using familiarity. Subjects are faster and more accurate at recognizing individual words (where some stimuli are familiar and some are novel) than they are at recognizing word-word associations (where all items are familiar; Hockley, 1992; but see Mayes et al., 2004 for evidence that some amnesic patients perform well on some tests of association). This effect is reliable for both forced-choice and yes-no recognition tests. It is difficult to know why recognition is easier than recall, even under matched conditions. However, because the most salient procedural difference is the presence of the studied stimulus at test (and thus the availability of making a familiarity judgment), it is likely that superior performance on recognition tests is due to the ability to use familiarity in addition to recollection.

A third line of evidence that recall and recognition assess different types of memory is that performance on recognition tests declines rapidly over short delay periods if the test can be solved more easily with familiarity judgments than by recollection of stimulus details. When subjects were given recognition tests for individual words, their performance declined as a function of the number of trials between study and test (Hockley, 1992). In contrast, when they were tested for recognition of word-word associations using the same stimuli, their performance remained stable (at least over the short to intermediate delays that were tested). The decrease in item recognition, but not association recognition, was reliable for both forced-choice and yes-no recognition tests. This difference in the forgetting rates for the two tests suggests the presence of two memory systems with differential contributions to each test. Additionally, because familiarity judgments are a more effective strategy for recognizing individual items than for recognizing associations, the faster decline in item recognition provides evidence that familiarity decays faster than recollection (see Yonelinas & Levy, 2002 for similar findings comparing recognition of words and recall of the word's color).

Other manipulations also have differential effects, which vary based on how easily the particular test can be solved using familiarity. Accurate familiarity judgments can be made more quickly than accurate recollection judgments (Hintzman et al., 1998). Additionally, aging appears to affect performance on recall tests more than on recognition tests (Craik & McDowd, 1987). Finally, ROC curves (plots of yes-no recognition judgments at different response criteria) show different patterns based on whether the test can be solved by a familiarity judgment. Normal human subjects show a curvilinear and asymmetrical ROC curve during normal recognition tests (Yonelinas et

al., 1998). Amnesic patients, who show primary deficits in recollecting information, show a curvilinear and symmetrical curve. Normal subjects doing a paired-associate recognition test, in which all items are equally familiar, show a linear and asymmetrical pattern (Yonelinas, 1997). Thus, curvilinear and symmetrical ROC curves are diagnostic of populations that are likely making familiarity judgments, linear and asymmetrical patterns are diagnostic of tests that require recollection of information, and curvilinear and asymmetrical curves are diagnostic of tests that can be solved using either a familiarity judgment or recollection of information.

The converging evidence from tests of amnesic patients, test difficulty, forgetting rates, processing speed, aging, and ROC curves suggests that multiple memory systems contribute to normal memory retrieval in humans. However, the details of how these systems interact and what their neural substrates might be are still controversial (Eichenbaum et al., 2007; Squire et al., 2007). Investigation of the similarities and differences in performance on recall and recognition tests in nonhuman animals would greatly inform current thinking about the theoretical construct of familiarity and recollection. Additionally, studies comparing humans and nonhumans across similar tests could reveal qualitative and quantitative differences in how different memory systems have been preserved across evolutionary history. Finally, identification of valid recall tests in animals would provide a useful models of factors that preferentially affect recall, such as amnesia (Vargha-Khadem et al., 1997) or aging (Craik et al., 1987).

Recall in nonhumans

Investigating the possible dissociation between familiarity and recollection in animals has been problematic (see Aggleton et al., 1999 and ; Shettleworth, 1998 ch. 6 for a brief discussion). This is due, in part, to the lack of recall tests for nonhumans, especially recall tests that have comparable recognition tests in nonhumans and comparable recall tests in humans. This contrasts with the wide availability of recognition tests that have been implemented across a variety of species (e.g., Murray & Mishkin, 1998; Nakagawa et al., 2004; Zhang et al., 2005).

There is evidence that animals possess recollective memories. One approach has been to examine ROC curves for rats' memory of a sequence of odors at short delays, long delays, and after radiofrequency lesions of the hippocampus (Fortin et al., 2004). Normal rats showed curvilinear and asymmetric ROC curves similar to those of normal humans. Rats with hippocampal lesions showed a curvilinear and symmetrical curve like that of human amnesics. The normal rats showed a linear and asymmetrical curve when tested with a long sample-test delay, results which parallel those of human subjects performing the paired-associate recognition test. Thus, rats and humans show similar ROC curves on recognition tests that are designed to stress either familiarity or recollection. However, methods like these are controversial, as it is unclear whether these manipulations can provide good estimates of one process in the absence of the other (Squire et al., 2007).

Another example of an approach aimed at investigating recall in animals is a series of studies using Panzee, a lexigram-trained chimpanzee (Menzel, 1999). After seeing food hidden outside, Panzee would later enlist the help of care staff, point to the

tunnel that led outside, and touch the correct symbol for the hidden food on her lexigram board. Crucially, Panzee requested the foods from inside, without being able to respond to the familiarity of the baited location, and sometimes after delays of hours or days. Further experiments with Panzee have shown that she can correctly indicate the location of the food even when the baiting was observed from different vantages or even over a television monitor (Menzel, 2005). It seems that Panzee was freely recalling what food she had seen and where it was hidden. However, these tests do not provide a clear way to compare Panzee's recall performance with matched recognition performance. It is also not easy to compare her recall performance to the large literature on human recall performance, which typically requires reproduction of words or drawings. Additionally, the performance of one highly-trained chimpanzee will not provide tools for understanding the neurobiology underlying recall. "Panzee did not literally draw a lexigram or a map of the forest" (Menzel, 2005, p. 214), but her performance suggests that she might be able to if she could draw. If an animal could draw or reproduce a previously seen stimulus, it would provide a powerful test to compare nonhuman recall to human recall tests such as the Rey-Osterrieth complex figure test.

The current study describes the development of a recall test for monkeys. This test was designed to be comparable to human recall tests that involve reproduction of a figure. Monkeys saw shapes composed of two colored boxes located on a 5x5 grid on a computer touchscreen. At test, one of the two boxes appeared in a new location on the grid. Monkeys earned a food reward for reproducing the studied shape by touching the appropriate grid location for the second box. To establish the characteristics of this recall test, we also compared it to a recognition test (delayed match to sample), which was

matched for chance rate, stimulus characteristics, and effort. Based on the finding in humans that recall tests are harder than recognition tests (Brown, 1975; Postman et al., 1948), we hypothesized that monkeys would be more accurate on the recognition test compared to the recall test. In addition, we examined the forgetting rates of the two tests, based on the previous research suggesting faster forgetting rates for tests that rely primarily on familiarity judgments (Hockley, 1992; Yonelinas et al., 2002). However, the current investigation used a relatively small stimulus set, all of which were very familiar to the monkeys by the time recall and recognition were compared. It was therefore unclear how useful familiarity judgments would be in solving the recognition test. Consequently, we made no specific predictions about the forgetting rates.

RECALL TRAINING

Method

Subjects

Subjects were six adult male rhesus macaques (*Macaca mulatta*). Monkeys were pair housed on a 12:12 light-dark cycle with onset at 7 am. All animals were fed a full ration of food, with four animals fed after daily testing and two fed ad libitum, depending on individual motivation. Water was available ad libitum.

Apparatus

All tests were presented using automated computer testing rigs mounted on the monkeys' home cage (Figure 1). Each rig contained a 15" 3M ChassisTouch 250 touch screen (3M, Saint Paul, MN), food dispensers for flavored grain pellets and miniature chocolate candies (Med Associates, St. Albans, VT), and generic stereo speakers.

Stimulus presentation and data collection were controlled by Presentation behavioral testing software (Neurobehavioral Systems, Albany, CA).

Procedure

Accuracy Training

Reproducing a shape on a touchscreen requires manual dexterity. Stimuli were composed of small, adjacent boxes (148 pixels X 148 pixels). In previous training with a

separate group of monkeys, we found that many rhesus tended to touch the screen with broad strokes of their hand or by biting/licking. To ensure that monkeys had the manual dexterity required to accurately touch individual boxes, monkeys underwent accuracy training and continued to receive accuracy trials before each session of recall training.

Monkeys initiated accuracy trials by touching a green “start box” (100 pixels X 100 pixels) located at the bottom of the screen. They were then presented with a 5 X 5 response grid composed of 25 individual boxes. The grid and boxes were the same size as the response grid for the recall test. A white cross (20 pixels X 20 pixels) appeared in the center of one of the boxes. The outer 15 pixels inside each box were a “dead zone”.

Touches to the dead zone produced a low-pitched tone signaling an invalid response.

Touches to the center area of any box produced a high-pitched sound signaling a valid response. Two valid responses to any box activated that box, turning it red (all responses in subsequent phases and experiments are also FR2). Monkeys earned food rewards by



Figure 1. Two of the portable rigs used to test the monkeys. Front center: the touchscreen side of a rig. Rear left: a monkey in his home cage. Rear right: the back of a rig as it hangs on the cage.

activating the target box containing the cross. Correct responses produced a distinct reward sound and incorrect responses produced a distinct error sound (“woo-hoo” and “d’oh”, respectively). Monkeys were required to score 100% correct for one complete session of 200 trials before proceeding to the recall training. Monkeys continued to receive 20 trials of accuracy training prior to each recall training session.

Recall Training Phase 1

To train up the reproduction rule, we started by having monkeys reproduce shapes in a single stationary location and gradually moved to shapes that could appear in any location at study and move to any other location at test. We first trained monkeys to reproduce a red stimulus box adjacent to a stationary, centered blue anchor box. Trials were initiated by touching the green “start box”. The monkey was then presented with the 5 X 5 response grid with the center box colored blue and one of the eight adjacent boxes colored red. Touching the blue box extinguished the red box and produced white crosses in all boxes except the center, indicating that they were now responsive. The correct response was to touch the previous location of the red box (which would activate that box, turning it red) and the incorrect response was to touch any other box in the grid. All touches were responsive at FR 2. Monkeys moved to the next phase of testing after completing a single session at 60% correct.

Recall Training Phase 2

All procedures remained the same as in Recall Training Phase 1, with the exception that the blue anchor box could appear in any of the nine central locations. This allowed for 72 possible stimulus combinations (9 possible locations for the blue anchor box X 8 possible locations for the adjacent red box). Session length was doubled to 144

trials in order to prevent monkeys from learning that response locations could be exhausted (i.e., that a correct grid location would not be correct again). Criterion was 60% correct in one session.

Recall Training Phase 3

To prepare monkey for tests with varying delay intervals between study and test, we first accustomed them to a short 1-second delay. During the delay, only the empty grid was presented and the screen was unresponsive. To discourage non-mnemonic strategies, such as repeatedly touching the correct location during the delay, the delay interval timer reset if the monkey touched the screen during the delay. After the delay, the blue anchor box and white crosses appeared, allowing the monkey to make a response. Criterion was the same as in previous phases. If the monkey's performance dropped to chance levels for two consecutive sessions, the delay was reduced back to 0-seconds (same conditions as Phase 2). The delay was then increased in half-second increments as the monkey met criterion at each delay.

Recall Training Phase 4

To prepare monkeys for trials in which the sample and test phases of trials occurred in different locations, we required the monkeys to touch the blue anchor box after the delay period and before reproducing the red box. This drew the monkeys' attention to the new location of the blue anchor box. After the delay, only the blue anchor box re-appeared; the remaining grid locations were empty and unresponsive. Touching the blue anchor box produced the white crosses in the other grid locations, signaling that they were now responsive. Criterion was the same as previous phases.

Recall Training Phase 5

We expected that monkeys would have a hard time transitioning to shapes that moved between study and test. Accordingly, we intermixed trials on which the shape moved with trials on which it remained stationary. This was done to give the monkeys experience with moving trials while maintaining a high enough proportion of stationary trials to ensure continued participation. On half of the trials, the blue anchor box reappeared in a new grid location during test. The new anchor location could be any of the eight remaining center boxes of the grid. Moving trials were assigned pseudorandomly, with two moving trials in each block of four trials. Monkeys moved on to the next phase after having had 4 sessions of experience with the intermixed trials.

Recall Training Phase 6

Phase 6 was similar to Phase 5, with the exception that the blue anchor box moved between study and test on all trials. To prevent responses to the same absolute location used during the study phase, we now presented the white crosses only in the eight grid locations surrounding the blue anchor box. We also constrained the movement of the anchor box so that the surrounding response boxes at test did not overlap with the absolute location of the red box during study. Criterion was 60% correct in one complete session of 144 trials. See Figure 2 for a schematic of the final recall test.

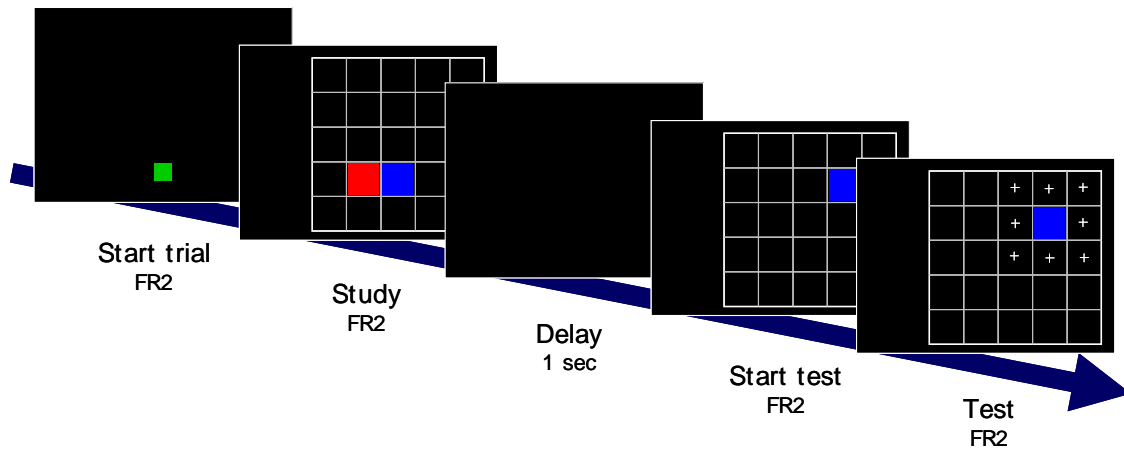


Figure 2. Schematic progression of a recall trial. Monkeys started the trial by touching the green “start box”, saw the sample in a random location, and touched the blue box to indicate they had seen the sample. After a delay, the blue box appeared in a new location, the monkeys touched it to bring up the possible response locations, and then could reproduce the studies shape by touching the appropriate grid location for the red box.

Data Analysis

We conducted individual analyses on each monkey using a binomial test to compare the frequency of correct responses with levels expected by chance. For all analyses, we use 12.5% correct as our chance level, the monkeys had to choose one of the eight grid locations adjacent to the blue anchor box. All statistical values were considered significant at the $p = .05$ level.

Results

In accuracy training, monkeys performed over 90% correct after an average of 2.33 sessions ($SD = 1.86$) and reached criterion of 100% correct after an average of 11.83 sessions ($SD = 10.59$). In continued accuracy training, all monkeys have maintained performance of 90% or greater.

No monkey met criterion at Recall Training Phase 6. However, all monkeys performed above chance levels from the first session (see Figure 3; each monkey $p < .01$,

one-tailed binomial test). As some monkeys had been on this stage for almost ten thousand trials without improvement, we decided to compare the recall test to a matched recognition test using the current recall performance. Because monkeys had different amounts of training at this phase, we looked at the last four sessions completed by each monkey to determine their current performance. All monkeys were still above chance levels (see Figure 3; all monkeys $p < .01$, one-tailed binomial test).

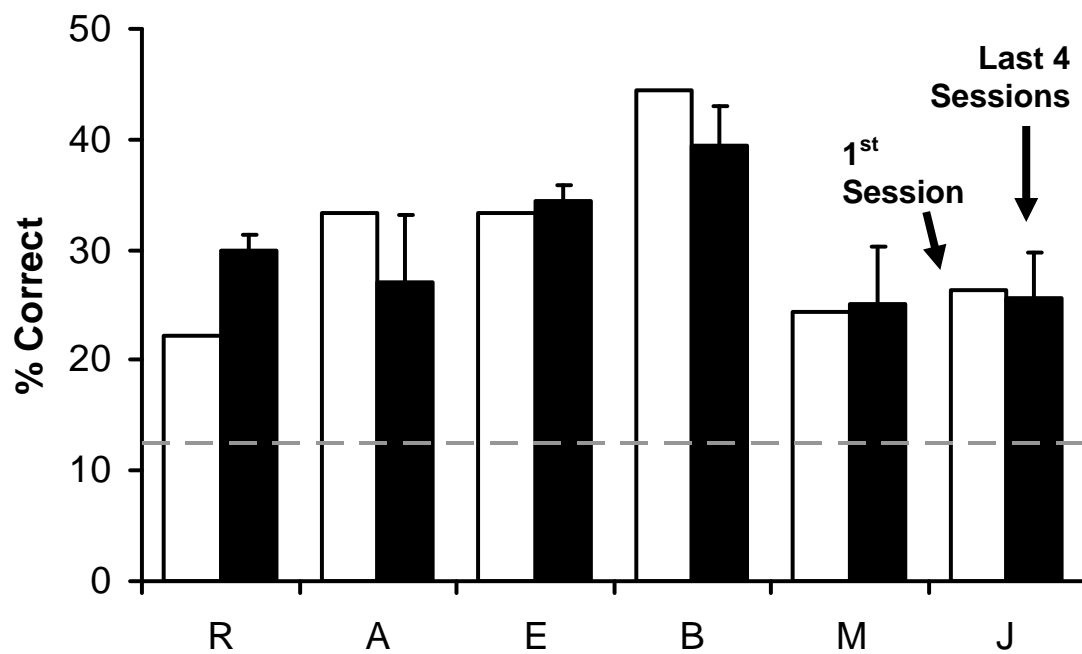


Figure 3. Individual performance on the last phase of recall training for the first session (white) and the last four sessions (black; +SD). Dashed line indicates performance expected by chance. Letters on the x-axis represent individual monkeys.

COMPARISON WITH RECOGNITION

Although monkeys were performing significantly above chance levels, their performance was still low. To explore the properties of our recall test, we compared it to a recognition test that was matched for chance rate, stimuli, and required responses. If our

recall test is similar to human recall tests, then it should also be more difficult than a recognition test (Brown, 1975; Postman et al., 1948). If, however, the recall test is difficult because of other aspects of the testing setup (e.g., stimuli are too similar, delay is too long, etc), then performance on the recall and recognition tests should be equivalent.

Method

Subjects and Apparatus

Subjects and apparatus were the same as those used in the recall training.

Procedure

Recognition Training with Clipart Images

Subjects were trained to match to sample using a set of 1000 visually distinct, session unique, color clipart images (296x296 pixels). We used a large image set to facilitate the learning of a general rule rather than of stimulus-specific rules (Wright et al., 2003). Monkeys initiated trials in the same way as during recall training, by touching the green “start box” at the bottom center of the screen. At study, one clipart image appeared in the center of the screen. After touching it, the monkey advanced to a delay interval of 1 second. Prior to test, a blue box (148 x 148 pixels) appeared in the center of the screen and the monkey had to touch it to advance to the test. Presentation of this blue box equated the recognition test and the recall test for the amount of time and the number of responses required to complete a trial (compare Figures 2 and 4). At test, the studied image and three distractor images appeared in the four corners of the screen. The location of the studied image varied pseudorandomly such that it appeared twice in each corner of the screen during a block of eight trials. If the monkey touched the studied image, he received a food reward accompanied by a “correct” sound and proceeded to a ten-second

intertrial interval. If the monkey touched a distractor, he received an “incorrect” sound and proceeded to a twelve-second intertrial interval. Criterion was 60% correct in a single session of 250 trials.

Next, we gradually moved monkeys to a set-size of eight images to equate the stimulus set size with that used in the recall. Monkeys received one 250-trial session at each of the following set sizes: 1000, 500, 100, 20, and 8.

Recognition Training with Box Shapes

Each shape used in the recognition test was designed to be visually identical to one of the two-box configurations used in the study phase of the recall test. Just as in the recall test, each was composed of a red box and a blue box that shared either a full side or a corner, resulting in eight possible shapes.

We introduced the box shapes by intermixing them with the trained clipart stimuli, such that each stimulus type occurred twice in each block of four trials. To avoid biasing monkeys’ performance in favor of the recognition test, we did not require them to meet an accuracy criterion with the box shapes. Instead, we gave them four sessions of experience with intermixed stimuli. As location and stimulus type could not both be balanced within a 250-trial session, sessions were shortened to 200 trials, 100 of each stimulus type.

Comparison with Recall Test

As all monkeys had more experience with the recall test than with the recognition test, we attempted to equate the two tests by bringing both through a series of common criteria. For each test, we started with one distractor at test and increased the number of distractors by one, up to a maximum of four. For the recognition test, this meant that the

studied shape was presented with one, two, or three distractor shapes. For the recall test, the blue anchor box was presented with two, three, or four possible adjacent choices -- indicated by the white cross -- one correct and one, two, or three incorrect. Limiting the possible responses on the recall test to four allowed us to equate the recall and the recognition tests for the rate of accuracy that would be expected by chance (25%). Monkeys had to perform better than expected by chance (one-tailed binomial test: 61%, 44%, and 32% for 1, 2, and 3 distractors, respectively) before additional distractors were added. For the recognition test, all stimuli were box shapes and the number of trials per session was reduced to 144, to equate it with the recall test sessions. Monkeys received one session of each type per day. The order of the two tests alternated between days.

After meeting criterion with three distractors (4 choices total), the monkeys received ten sessions of each test type to ensure that they had significant experience with both. We then tested each monkey's performance for stability by testing them for six more sessions and comparing the first block of two sessions against the third block of two sessions. If performance differed between the two blocks, as measured by a binomial test, the monkeys were given additional blocks of two sessions until performance stabilized.

Finally, we ran eleven sessions of each test type. During these sessions, we varied the delay interval between study and test to assess the monkeys' forgetting rates for both recall and recognition tests. Monkeys received the trained delay length of 1 second on half the trials, with the delay lengths of 2, 4, 8, 16, 32, and 64 seconds mixed pseudorandomly among the other half of the trials. As the addition of varied delays was novel, we discarded the first comparison session to avoid an initial performance decrease. Thus, for each test type, our final comparison consisted of 720 trials at the trained delay

and 120 trials at each of the other delays. See Figure 4 for a schematic of the final recognition test.

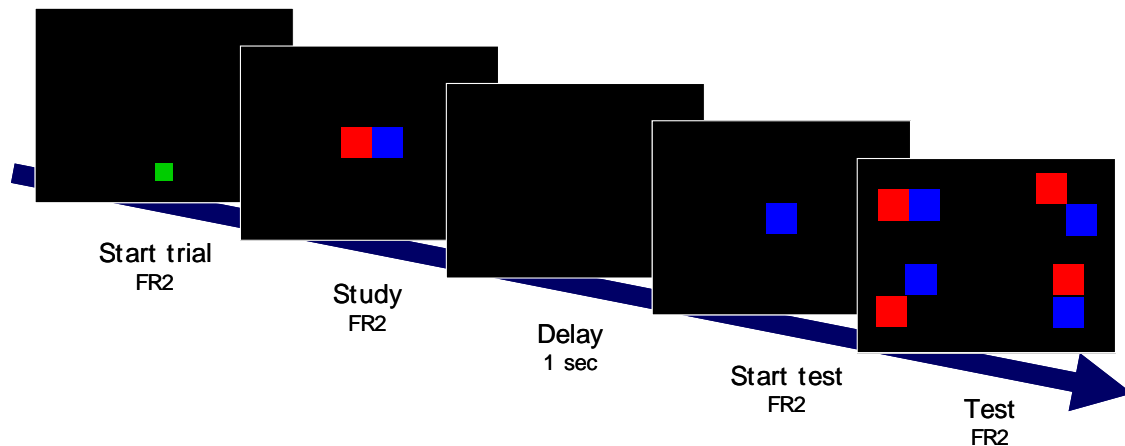


Figure 4. Schematic progression of a recognition test. Monkeys started the trial by touching the green “start box”, saw the sample and touched it to indicate they had seen it. After a delay, the blue box appeared in the center of the screen and the monkeys touched it to bring up the possible choices. During test, the studied stimulus and three distractors appeared in the four corners. During the comparison with the recall test, the trained delay of 1 second varied in length.

Data Analysis

We arcsine transformed all proportions and log transformed all latencies prior to analysis (Howell, 2002). All tests were considered significant at the $p = .05$ level.

Results

For training on the recognition test using clipart images, monkeys reached criterion in an average of 6.2 sessions ($SD = 2.1$) and a paired t-test showed that accuracy did not decrease as monkeys moved from a set size of 1000 to a set size of 8 (81.1% and 83.5%, respectively; $t_{(5)} = -0.6$, $p = .57$). When the box shapes were first introduced, intermixed with the clipart shapes, no monkey transferred performance to the box shapes at level above chance (chance = 25%, mean accuracy = 25.7%).

Comparing accuracy at the trained delay, a paired t-test revealed no significant difference between the recall and recognition tests ($t_{(5)} = -0.23, p = .83$; see Figure 5). However, there was a significant differences in the monkeys' response latencies on correct trials, with correct responses on recognition trials taking longer than correct responses on recall trails ($t_{(5)} = -4.65, p < .01$; see Figure 6). Finally, a repeated measures ANOVA of accuracy at different delay intervals revealed a main effect of delay, but no main effect of test type and no interaction between test and delay (delay: $F_{(6,30)} = 35.27, p < .01$; test: $F_{(1,5)} = 0.56, p = .49$; text x delay: $F_{(6,30)} = 1.21, p = .33$; see Figure 7).

As there were no group differences in accuracy, we conducted exploratory analyses of each monkey's performance. Chi square analysis of accuracy data showed that two monkeys performed better on the recognition test, two performed equally well on both tests, and two performed better on the recall test (see Figure 5 and Table 1).

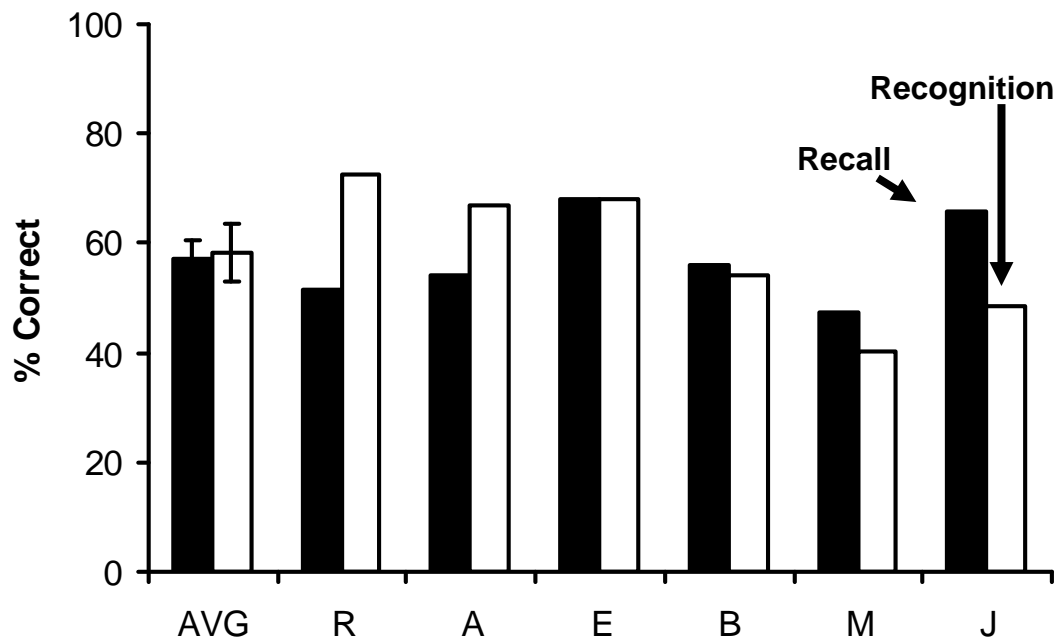


Figure 5. Individual performance at the trained delay of 1 second on the recall test (black) and the recognition test (white). AVG = group data (\pm SEM) and letters represent individual monkeys.

Table 1 Chi-square and *p* values

Monkey	R	A	E	B	M	J
χ^2 (df=1)	69.01	24.06	0.01	0.55	7.63	45.74
<i>p</i>	< .01	< .01	.91	.46	< .01	< .01

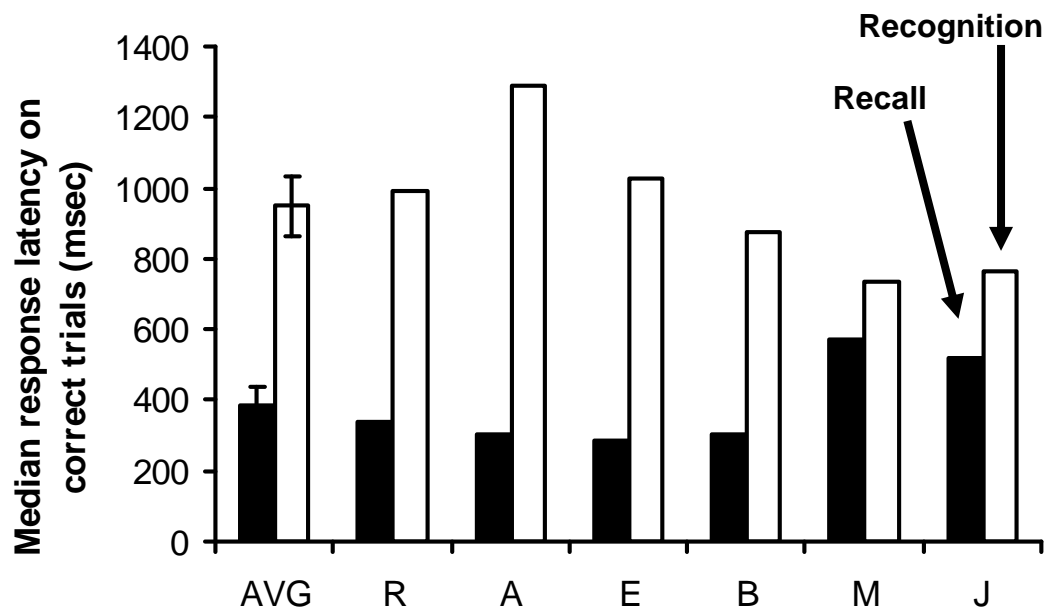


Figure 6. Median response latency on correct trials at the trained delay of 1 second on the recall test (black) and the recognition test (white). AVG = group data (\pm SEM) and letters represent individual monkeys.

These performance differences partially extended to the different delay lengths. Monkeys R and A, who had performed better on the recognition test, also performed significantly better on the recognition test at delays of 2, 4, and 8 seconds and 16, 32, and 64 seconds, respectively. Of the two monkeys that showed equal performance on the two tests, one performed significantly better on the recognition test at a delay of 4 seconds. Of the two monkeys that performed better at the recall test, only one was significantly better at any of the other delays (monkey J, 2-second delay; see Figure 7 and Table 2).

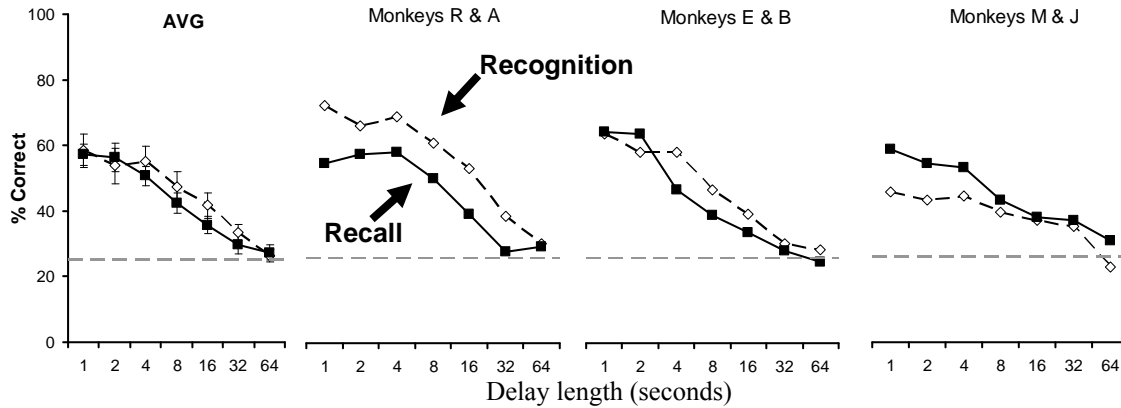


Figure 7. Mean performance on the recall test (black line) and recognition test (dashed line) at different delay lengths. Group means are presented on the left (\pm SEM) and monkeys that performed similarly at the trained delay are grouped together in the right three panels. The horizontal dashed line indicates performance expected by chance.

Table 2 Chi-square and *p* values by Delay Length

Monkey		R	A	E	B	M	J
2 sec	χ^2 (df=1)	6.34	0.02	1.31	0.42	1.1	5.4
	<i>p</i>	< .01	.90	.25	.52	.30	.02
4 sec	χ^2 (df=1)	5.8	0.84	1.71	4.34	3.31	0.6
	<i>p</i>	.02	.36	.19	.04	.07	.44
8 sec	χ^2 (df=1)	4.29	1.36	2.4	0.69	0.02	0.84
	<i>p</i>	.04	.24	.12	.41	.90	.36
16 sec	χ^2 (df=1)	0.42	12.4	1.71	0.19	0.44	0.17
	<i>p</i>	.52	< .01	.19	.67	.51	.68
32 sec	χ^2 (df=1)	0.02	11.62	0.47	0	0.17	0.89
	<i>p</i>	.89	< .01	.50	1.00	.68	.35
64 sec	χ^2 (df=1)	3.4	4.66	0.2	0.78	0.82	3.5
	<i>p</i>	.07	.03	.66	.38	.37	.06

Discussion

Group findings

In the recall test, all monkeys learned to reproduce the two-box shapes and continued to do so at above-chance levels throughout testing. This required that they remember the location of the red box relative to the blue box. As the shape moved between study and test, they could not solve the task by relying on the recent familiarity of studied grid locations. Additionally, as they only reproduced the red box at test and not at study, it is unlikely that they could solve the task by repeating a motor response.

In the comparison between recall and recognition, we found no evidence at the group level to support our hypothesis that recall would be harder than recognition. Accuracy data were similar for both tests. This contrasts with findings in humans that performance on recall tests is typically lower than performance on recognition tests (Brown, 1975; Postman et al., 1948). Additionally, there was no indication that recognition performance had a steeper forgetting curve than recall performance, as might be predicted from some human studies (Hockley, 1992; Yonelinas et al., 2002). While we had no specific predictions about this outcome, the results are not surprising. Based on the previous studies, we should see a difference in forgetting rates if one test is best solved by making familiarity judgments and the other requires recollection of additional stimulus attributes. All shapes were very familiar; there are only eight possible configurations and they were used in both tests. Although more recently seen shapes are presumably more familiar than the distractors, it is unclear whether this small familiarity difference would be useful enough that the monkeys would primarily use familiarity judgments to solve the recognition test.

One consistent difference between the recall and recognition tests was that correct recognition choices took more than twice as long as correct recall choices. This was most likely due to the number of items on screen during test. During the recall test, only one partial shape was present (a single blue box) close to the center of the screen, whereas during the recognition test, four shapes were present (each composed of a blue and red box) near the corners of the screen. Thus, responses on the recognition tests likely took longer because the monkeys had to visually scan more information spread across a larger area. The observed difference in response latencies likely represents an artifact of the testing setup rather than a genuine difference in processing time.

Limitations and future directions

Although we attempted to equate the two tests, it is likely that they were not matched for the amount of memory interference. In the recognition test, the target was presented along with three distractors (other shape configurations). It is possible that the current perception of the distractor shapes interfered with the monkey's memory of the studied shape. As no other shapes were presented during the choice phase of the recall trials, the recall test had less interference than the recognition test. This likely resulted in artificially low scores on the recognition test, possibly helping to explain why monkeys did not show better recognition performance, as predicted.

The presence of distractors in the recognition test may also explain the difference in the response latency data. Compared to the recall test, decisions in the recognition test might have required more time because the monkeys had to visually search a larger portion of the screen. In the recognition test, the four options were located in the corners

of the screen, whereas in the recall test, the four options were clustered around the blue anchor box. If the two tasks were equated for the amount of screen area that was necessary for the monkey to search, it is likely that the response latencies for the recognition test would be lower than was currently observed.

To equate the two tests for interference and visual search time, we are currently comparing recall performance to performance on a yes-no recognition test. In the yes-no recognition test, one shape will appear at test along with a distinctive “non-match” symbol. The monkey can earn food for touching the shape if it is the same shape as during study or by touching the “non-match” symbol if the shape is different. This will eliminate the interference caused by the distractors in the match-to-sample recognition test. Additionally, as only one shape is present at test, the monkey will not have to visually search the entire screen, thereby better equating the time required at test. The yes-no recognition test will likely provide a better-equated comparison between recall and recognition tests.

Individual findings

The results from the individual analyses are promising. Two of the monkeys (R and A) were more accurate on the recognition test than the recall test across several different delay lengths. Additionally, a third monkey (E) was more accurate on recognition at a delay of four seconds. This is interesting because 1) they had much more experience with recall, 2) the recognition test likely had more memory interference, and 3) the recognition test required subjects to visually scan a larger area. However, we

should interpret these individual findings cautiously, as two monkeys (M and J) showed the opposite pattern and one (B) showed no differences.

Conclusion

We developed a novel shape reproduction test for monkeys, designed as an analogue of human recall tests. All monkeys learned to reproduce a simple two-box configuration at levels above chance. We then compared their performance to that on a match-to-sample recognition test, matched for chance rate. Unlike human recall tests, this recall test was not harder than the recognition test. This finding may be due to more memory interference and visual search time in the recognition test. Future studies using a yes-no recognition test may provide a better-equated comparison of recall and recognition in monkeys. Other manipulations, such as speeded retrieval tests, hippocampal lesions, or investigations of ROC curves may help to determine whether our recall test is functionally similar to recall tests in humans. If the present recall test in monkeys shows functional similarities to tests of recall in humans, it will provide a powerful tool to compare human and nonhuman memory, to assess the validity of the familiarity/recollection distinction, and to explore the neural correlates of recall performance.

References

- Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia and the hippocampal-anterior thalamic axis. *Behavioral And Brain Sciences*, 22(3), 425-444.
- Baddeley, A. D., Vargha-Khadem, F., & Mishkin, M. (2001). Preserved recognition in a case of developmental amnesia: Implications for the acquisition of semantic memory? *Journal of Cognitive Neuroscience*, 13(3), 357-369.
- Brown, J. (1975). *Recall and recognition*. London ; New York: Wiley.
- Craik, F. I. M., & McDowd, J. M. (1987). Age-Differences in Recall and Recognition. *Journal of Experimental Psychology-Learning Memory and Cognition*, 13(3), 474-479.
- Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, 30, 123-152.
- Fastenau, P. S., Denburg, N. L., & Hufford, B. J. (1999). Adult norms for the Rey-Osterrieth complex figure test and for supplemental recognition and matching trials from the extended complex figure test. *Clinical Neuropsychologist*, 13(1), 30-47.
- Fortin, N. J., Wright, S. P., & Eichenbaum, H. (2004). Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature*, 431(7005), 188-191.
- Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On The Relationship Between Recall And Recognition Memory. *Journal Of Experimental Psychology-Learning Memory And Cognition*, 18(4), 691-702.
- Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory & Cognition*, 26(3), 449-462.
- Hockley, W. E. (1992). Item Versus Associative Information - Further Comparisons of Forgetting Rates. *Journal of Experimental Psychology-Learning Memory and Cognition*, 18(6), 1321-1330.
- Holdstock, J. S., Mayes, A. R., Roberts, N., Cezayirli, E., Isaac, C. L., O'Reilly, R. C., et al. (2002). Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans? *Hippocampus*, 12(3), 341-351.
- Howell, D. C. (2002). *Statistical Methods for Psychology* (Fifth Edition ed.). Pacific Grove, CA: Duxbury.

- Kelley, C. M., & Jacoby, L. L. (2000). Recollection and Familiarity. In E. Tulving & C. I. M. Fergus (Eds.), *The Oxford handbook of memory* (pp. 215-228). Oxford ; New York: Oxford University Press.
- Manns, J. R., Hopkins, R. O., Reed, J. M., Kitchener, E. G., & Squire, L. R. (2003). Recognition memory and the human hippocampus. *Neuron*, 37(1), 171-180.
- Mayes, A. R., Holdstock, J. S., Isaac, C. L., Hunkin, N. M., & Roberts, N. (2002). Relative sparing of item recognition memory in a patient with adult-onset damage limited to the hippocampus. *Hippocampus*, 12(3), 325-340.
- Mayes, A. R., Holdstock, J. S., Isaac, C. L., Montaldi, D., Grigor, J., Gummer, A., et al. (2004). Associative recognition in a patient with selective hippocampal lesions and relatively normal item recognition. *Hippocampus*, 14(6), 763-784.
- Menzel, C. (2005). Progress in the Study of Chimpanzee Recall and Episodic Memory. *Terrace, Herbert S (Ed); Metcalfe, Janet (Ed), (2005). The missing link in cognition: Origins of self-reflective consciousness. (pp. 188-224). xx, NY, US: Oxford University Press.*
- Menzel, C. R. (1999). Unprompted recall and reporting of hidden objects by a chimpanzee (*Pan troglodytes*) after extended delays. *Journal of Comparative Psychology*, 113(4), 426-434.
- Murray, E. A., & Mishkin, M. (1998). Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. *Journal of Neuroscience*, 18(16), 6568-6582.
- Nakagawa, S., Etheredge, R. J. M., Foster, T. M., Sumpter, C. E., & Temple, W. (2004). The effects of changes in consequences on hens' performance in delayed-matching-to-sample tasks. *Behavioural Processes*, 67(3), 441-451.
- Postman, L., Jenkins, W. O., & Postman, D. L. (1948). An Experimental Comparison of Active Recall and Recognition. *American Journal of Psychology*, 61(4), 511-519.
- Reed, J. M., Hamann, S. B., Stefanacci, L., & Squire, L. R. (1997). When amnesic patients perform well on recognition memory tests. *Behavioral Neuroscience*, 111(6), 1163-1170.
- Shettleworth, S. J. (1998). *Cognition, Evolution, and Behavior*. New York Oxford University Press.
- Squire, L. R., Zola-Morgan, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience*, 8(11), 872-883.
- Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., VanPaesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, 277(5324), 376-380.

- Wais, P. E., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron*, *49*(3), 459-466.
- Wright, A. A., Rivera, J. J., Katz, J. S., & Bachevalier, J. (2003). Abstract-concept learning and list-memory processing by capuchin and rhesus monkeys. *J. Exper. Psychol.: Anim. Behav. Proc.*, *29*(3), 184-198.
- Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, *25*(6), 747-763.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441-517.
- Yonelinas, A. P., Kroll, N. E. A., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, *12*(3), 323-339.
- Yonelinas, A. P., & Levy, B. J. (2002). Dissociating familiarity from recollection in human recognition memory: Different rates of forgetting over short retention intervals. *Psychonomic Bulletin & Review*, *9*(3), 575-582.
- Zhang, S. W., Bock, F., Si, A., Tautz, J., & Srinivasan, M. V. (2005). Visual working memory in decision making by honey bees. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(14), 5250-5255.