Long-term associative memory capacity in man

Joel L. Voss
University of Illinois at Urbana-Champaign, Urbana, Illinois

Long-term episodic memory depends on the retention of associative information, such as the relationship between a friend's face and his name, a home and its neighborhood, and a mint and its odor. How many of these associative links can be stored and recalled? Many have considered memory's capacity, but relevant data are scant. Some previous experiments on humans have assessed the retention of thousands of visual impressions using forced-choice picture recognition and have concluded that there are virtually no constraints on how much information can be successfully retained. However, no previous experiments on humans have investigated the capacity of associative memory. I describe the first relevant data, which I obtained by systematically probing my own capacity during 58,560 memory trials for picture–response associations (approximately 1 year of testing). Estimated capacity was on the order of several thousand associations, and this and other indicators of memory function were remarkably similar to those obtained for baboons (*Papio papio*) under comparable circumstances. These findings, along with other data, suggest conservation of long-term memory mechanisms and effectiveness in humans relative to nonhuman primates, despite at least 20 million years of divergent evolution and vastly different behavioral and cognitive repertoires. The present findings also indicate that the associative processes that support our ability to remember episodes are limited in capacity relative to processes that support picture recognition.

How much information can be stored in memory? This question has daunted philosophers and scientists for over a thousand years (Dudai, 1997). No answer is currently available, partly because there are scant data on which to base capacity estimates or to constrain models of capacity (Landauer, 1986; Treves & Rolls, 1994).

Valuable evidence has been obtained by examining memory for large sets of pictures studied *en masse* (Brady, Konkle, Alvarez, & Oliva, 2008; Standing, 1973; Standing, Conezio, & Haber, 1970). In these studies, individuals attempted to memorize many briefly presented images (e.g., up to 10,000 photographs in Standing, 1973), and were later tested on their ability to discriminate studied images from novel images when they saw pairs comprising one of each type of image presented side by side in a forced-choice-format test. Discrimination was remarkably accurate in these studies, indicating that the brain's ability to serve as a storehouse for item-specific visual information is virtually limitless. However, long-term episodic memory also involves associative traces (Aggleton & Brown, 2006; Eichenbaum & Cohen, 2001; Lavenex & Amaral, 2000; Norman & O'Reilly, 2003; Tulving, 1983), and other approaches are necessary for determining capacity for associative information.

Data relevant to associative capacity in humans have been unavailable due to practical constraints on their collection. I therefore systematically probed my own capacity during 58,560 memory trials for picture–response associations distributed over approximately 1 year of testing. I used a modification of a paradigm originally designed to tax long-term memory in baboons (*Papio papio*) and pigeons (*Columba livia*) (Cook, Levison, Gillett, & Blaisdell, 2005; Fagot & Cook, 2006), so that information regarding the evolutionary trajectory of memory capacity could also be obtained. This paradigm is advantageous for making comparisons across species because associative traces comprise arbitrary pairings of visual stimuli with one of two behavioral response options, thus minimizing confounding influences of factors, such as conceptual/semantic learning strategies, which have previously undermined comparisons of humans and nonhuman animals (Ekstrand, Wallace, & Underwood, 1966).

METHOD

Visual stimuli comprised 4,980 unique color digital photographs of common matter (objects, people, buildings, animals, etc.) obtained from stock collections. All were resized to subtend a horizontal rectangle of approximately 6° × 12° of visual angle and were displayed at central fixation on a computer monitor.

The task involved learning an arbitrary association between each picture and one of two buttonpress responses on a keyboard (left shift key/right shift key). Once a set of picture–response associations was learned to criterion (described below), the set moved into an ever-increasing set of learned associations. Testing sessions provided opportunities for learning new picture–response associations and assessed retention for associations in the ever-increasing set of learned associations. It was thus possible to study the ongoing

J. L. Voss, joelvoss@illinois.edu

Figure 1. Schematic diagram of the experimental design. Each testing session (shown schematically on the right) comprised 120 trials of two types (shown schematically on the left). Half of the trials concerned picture–response associations for the 30-item learning set (two presentations each), and half were probe trials from the set of previously learned associations. Learning set items became part of the pool of previously learned associations once criterion was met. Trials during the testing session provided feedback-based learning of associations between each picture and one of two arbitrarily assigned button responses. Ongoing measures of retention and learning were thus provided for a large set of picture–response associations.

dynamics of new learning and of retention as the amount of material held in memory continuously increased. Figure 1 provides a schematic diagram of the overall experimental design.

Before the experiment commenced, each picture was permanently assigned at random to either a left- or right-button response. Learning progressed in sets of 30 picture–response associations (the *learning set*). These 30 items appeared during each testing session until performance surpassed criterion (85% correct overall). After criterion was met, the learning set moved into the ever-increasing set of previously learned associations. The correct response for half of the items in each learning set was "left," so that guessing would yield 50% accuracy.

Each testing session, generally administered three times per day, comprised 120 trials. On each trial, a picture was displayed centrally until a left or right button response was registered. Correct responses cued the next trial. Incorrect responses cued a time-out (1,000 msec with the screen colored red) and feedback (3,500 msec). Feedback consisted of the same picture presented again, to which the correct response was required. Feedback responses were excluded from analyses. Thus, the association between a picture and one of two buttonpress responses was learned via feedback.

The 30 items in the learning set were shown on 60 of the 120 trials during a testing session (two presentations each). The learning set was repeated across consecutive testing sessions until 85% accuracy was achieved for the first presentation of a picture during the session. The other 60 trials of the testing session were probe tri-

als comprising pictures randomly selected from the ever-increasing set of previously learned associations. Trial order was randomized, with at least 10 intervening trials between consecutive presentations of each learning set picture.

Testing sessions were separated by at least 4 h, and the final session for a learning set was separated from the penultimate session by at least 1 night of sleep on all but six occasions. Testing sessions were completed during approximately 1 year. Analyses of the data were avoided during their collection, except as minimally required to ensure that responses were properly logged and accuracy properly computed. Furthermore, images were not viewed at any time other than during testing sessions.

The procedure described above is a modification of the paradigm used by Fagot and Cook (2006) to probe associative capacity in baboons and pigeons. The only significant modification made to accommodate a human subject is that keyboard buttons were used to register behavioral responses, as opposed to a manual device for baboons or screenpecking for pigeons. All other parameters were taken from the paradigm for nonhuman animals.

A subsidiary test of forced-choice picture recognition was conducted after all associative learning and testing were complete. The format was similar to that used in previous studies of forced-choice picture recognition (described above). Testing involved a set of 2,100 unique photographs of a format identical to those used in the main experiment, but which did not appear in the main experiment. For encoding, 2,000 images were presented for 2,000 msec each, with a
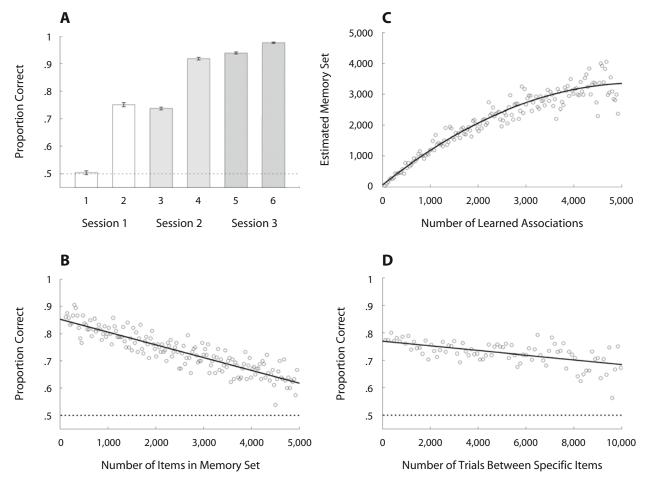
Figure 2. Long-term associative memory. (A) Mean accuracy is shown for 30-item learning sets for six learning trials (three testing sessions). Error bars indicate *SEM*s, and the dotted line indicates chance. Each value (but not the first) was significantly above chance ($p < .001$). (B) Accuracy for previously learned associations tested on probe trials is plotted as a function of the number of items in the memory set of previously learned associations. Each point indicates mean accuracy for an average of 180 probe trials and was above chance (indicated by the dotted line; $p < .05$). The best-fitting regression line is provided ($R^2 = .81$, $F = 676.2$, $p < .001$). (C) Estimated number of items stored in memory, as a function of the number of items in the pool of previously learned associations. The best-fitting quadratic is provided ($R^2 = .69$, $F = 352.0$, $p < .001$). The asymptote of the best-fitting quadratic was 3,370, and the mean of the last 20 estimates provided a similar maximum value of 3,315. (D) Accuracy, as a function of the number of intervening trials between two successive probe trials for the same item from the set of previously learned associations. The best-fitting regression line is provided ($R^2 = .34$, $F = 41.9$, $p < .001$). Longer lags were not assessed, due to the small *N*. The dotted line indicates chance.

1,000-msec interstimulus interval. Images were presented once each, and learning was divided across four consecutive sessions (500 images each), with a 5-min break between each session. Forced-choice recognition was tested 12 h after encoding (including 1 night of sleep) by pairing 100 images selected at random from the set of 2,000 (old) with 100 images that were not studied (new). One old/new pair was presented during each test trial for 3,000 msec, with the old item appearing on the left side of the screen for 50% of the trials and on the right side of the screen for 50% of the trials. A left- or right-button response was used to select the position of the old item.

RESULTS

I learned sets of picture–response associations until a criterion was met, and I was repeatedly tested using pictures from the ever-increasing set of previously learned associations (Figure 1). Six feedback-learning trials (distributed across three testing sessions) were generally re-

quired for each association in order to reach criterion for a learning set of 30 associations (Figure 2A). A small minority of learning sets required two testing sessions (2%) or four testing sessions (3%). The average acquisition rate was approximately four times faster than the average rate for baboons (Fagot & Cook, 2006).

Figure 2B shows accuracy for probe trials from the set of previously learned associations, plotted as a function of set size, and indicates that a set size of approximately 7,500 would lead to chance performance. This is approximately three times lower than for baboons (Fagot & Cook, 2006). Note that this number does not represent an estimate of memory capacity, given that it is unclear how many associations would have to be stored in order to support the observed level of performance (see below).

Two estimates of my memory capacity were remarkably similar to those obtained for baboons and were vastly

lower than estimates obtained using forced-choice picture recognition paradigms in humans. These estimates aimed to determine how many picture–response associations were retained in memory by considering performance as a mixture of memory-guided responding and guessing. The first estimate derived from an item analysis aimed at determining, for the entire set of learned associations, how many were retained with sufficient fidelity to support moderately above-chance performance (>65%) for the entirety of testing. That is, members of the set of previously learned associations were tested on multiple probe trials throughout the year of testing, and I computed the mean accuracy for each picture–response association for all occasions that it appeared during the experiment (not including learning trials) and tallied those with a mean accuracy greater than 65%. This was 4,105 associations, or 82.4% of the total set size. The average for baboons was 83.6% (Fagot & Cook, 2006).

The second estimate was obtained by fitting a conservative, all-or-none, high-threshold model to performance on probe trials (Fagot & Cook, 2006; Standing, 1973). This model incorporated the assumption that performance was an additive combination of responses based on memory (probability correct = 1) and of responses based on guessing (probability correct = .5). The size of the memory set was estimated by deriving the best-fitting probability using these two trial types. One estimate was obtained for each 30-item increment in association-set size (180 test trials, on average). This model indicated that at least 3,370 picture–response associations had been retained in my memory (Figure 2C). The average estimate for baboons was 3,433, which was approximately four times the estimated capacity for pigeons (Fagot & Cook, 2006).

Two other indicators of memory function were highly similar to those obtained in both baboons and pigeons. The forgetting rate was determined by examining performance for two successive probe trials for a picture–response association as a function of the number of intervening trials, or lag (Figure 2D). The slope of the best-fitting regression line was .0001. Values for baboons and pigeons were .0001 and .0002, respectively, indicating similar forgetting rates (Fagot & Cook, 2006). The second similarity is that response times provided evidence for parallel memory search rather than serial memory search, in that they were faster during the second half of the experiment than during the first, despite the need to search a larger memory set (1,208 msec and 1,176 msec, respectively, for correct responses, $p < .001$; 1,458 msec and 1,289 msec, respectively, for incorrect responses, $p < .001$). Moreover, responses to incorrect items were slower than to correct items for both halves ($ps < .001$), potentially indicating additional, unsuccessful retrieval effort. The same patterns in response time effects were identified in both baboons and pigeons (Fagot & Cook, 2006).

DISCUSSION

My estimates suggest that the efficacy of associative memory degrades substantially with a load on the order of several thousand traces. Why is this number so much lower than estimates approaching the infinite obtained in studies of forced-choice picture recognition (Brady et al., 2008; Standing, 1973; Standing et al., 1970)? One possibility is that my memory is abnormally poor. However, it is unreasonable to assume that my estimates are low due to below-average general intellectual ability, given that I recently scored above average (>100) on the WAIS–III standardized test for IQ. To provide evidence that my memory capabilities were not responsible for different estimates than those obtained via picture recognition, I participated in an additional study on forced-choice picture recognition for 2,000 images studied *en masse*, using methods similar to those in the aforementioned studies. I achieved 91% accuracy, indicating that my ability to perform this task is similar to that of subjects in prior experiments. Therefore, the sources of divergence for my estimates and previous estimates likely include differences in the memory mechanisms probed by tests of associative recollection versus object recognition.

One could argue that my estimates are unrepresentative of all humans or are otherwise biased. However, fundamental insights into memory have been obtained previously through rigorous self-study (Ebbinghaus, 1885/1964). Small sample sizes and self-study (see, e.g., Blakemore & Sutton, 1969) and single-case research methods have provided many enduring insights (Newell & Simon, 1972; Skinner, 1956). Moreover, individual differences in capacity for baboons ($N = 2$) and pigeons ($N = 2$) were negligible (Fagot & Cook, 2006), and there is no reason to assume greater variability in humans.

Nevertheless, because of the limitations imposed by studying only 1 subject, the present results should be taken as an approximation of capacity for one type of associative trace on the order of several thousand. If capacity were orders of magnitude higher, such as tens of thousands of associations or practically infinite—as was suggested by Fagot and Cook (2006) and is likely true for forced-choice picture recognition (Standing, 1973)—then finding a limitation of merely several thousand in 1 individual with normal intellectual capabilities would be highly improbable.

Great care was taken in designing the paradigm to ensure that the memory processes supporting human performance would be as similar as possible to the processes supporting performance for baboons and pigeons. The paradigm has already been successfully employed in cross-species comparisons between baboons and pigeons (Fagot & Cook, 2006). Moreover, the paradigms used for probing my capacity and that of baboons and pigeons were virtually identical on many parameters, such as the mean retention interval, the number of learning opportunities, the temporal distribution of learning opportunities across testing sessions, and the method of reinforcing correct responses. In addition, memory was tested in all species over a prolonged period, which is advantageous for making comparisons across species. This is because nonhuman animals generally require extensive training on behavioral tasks, and thus could uncover learning strate-

gies not available to humans if testing in humans were conducted during only a small number of sessions (as discussed in Voss & Paller, 2009b).

It is important to note that the fundamental unit of memory was the arbitrary association between a picture and one of two button responses, and this feature is advantageous for making comparisons across species. Individual pictures that comprised elements of the associations were undoubtedly of different behavioral/conceptual relevance for baboons and pigeons than for me, and this could have caused the level of noise against which memory judgments were made for the pictures themselves to differ across species. That is, individual pictures could have been more or less distinctive for me than for the other species. If this distinctiveness was able to cue recollection of the association between the picture and the button response, semantics could have influenced my associative memory to a greater extent than that for pigeons and baboons. However, the association between a picture and a response was arbitrary and thus the chirality of conceptual features could not have been systematically related to the correct behavioral responses. It is, therefore, unlikely that retention of associations was based on conceptual cues available to me but unavailable to nonhuman animals. Likewise, salient perceptual features were available to all species tested, and many previous findings suggest that all species tested should have been incredibly adept at recognizing individual pictures. Distinctiveness for each item should, therefore, have been high for all species (Fagot, 2000; Wasserman & Zentall, 2006). Picture recognition could have cued associative memory in all species; but because the associative information was unrelated to the meaning of each picture, it is unlikely that semantics provided substantial additional benefits to me that were not provided by accurate picture recognition in all species tested.

The associative task described here was limited in the sense that only one type of information was pertinent: the arbitrary association between a picture and one of two button responses (however, this was an ideal feature with respect to the goal of making evolutionary comparisons, as indicated above). In contrast, many of the associations encountered every day are conceptually meaningful (e.g., a pie in a bakery, a child at a playground). This factor likely exerts a strong influence on associative capacity, given the strong influences of conceptual meaning on most memory measures. Additional evidence will be necessary in order to determine how capacity differs for conceptually meaningful versus meaningless associations.

An outstanding issue is that my capacity for retention of arbitrary picture–response associations was approximately reached, and yet I could presumably retain other forms of novel associative information. Indeed, I did not notice any degradation in my everyday uses of memory during the experiment. Likewise, I entered the experiment with a vast storehouse of well-learned arbitrary associations, such as the meanings of words. To what extent is memory capacity domain specific? Does the storage of large amounts of one kind of novel associative information (such as a picture and a button response) interfere with the ability to store other kinds of novel associative information (such as the meaning of new vocabulary words)? Answers to these questions will be of great importance for understanding the cognitive and neural constraints on memory capacity and will require substantial additional efforts.

The present findings underscore distinctions between the memory processes critical for forced-choice picture recognition versus associative memory. Forced-choice tests of picture recognition can rely on item-specific visual information, and accurate performance in forced-choice tests for pictures has been found with minimal response confidence (Tulving, 1981) and when confidence is entirely absent (Voss, Baym, & Paller, 2008; Voss & Paller, 2009a). In contrast, associative memory generally involves consistency between accuracy and confidence (Yonelinas, 2001). Furthermore, patients with amnestic insults and marked deficits in associative recollection have exhibited relatively spared forced-choice picture recognition (Holdstock, Mayes, Gong, Roberts, & Kapur, 2005; Westerberg et al., 2006), indicating that forced-choice picture recognition need not involve associative recollection. These distinctions emphasize two important facets of memory capacity: (1) the brain's ability to serve as a storehouse for vast amounts of basic sensory information, and (2) its relatively limited ability to build arbitrary associations. Both memory types should be treated in future considerations of memory capacity. The former function may be crucial for abstract reasoning and thought (as reviewed in Brady et al., 2008), whereas the latter function has been emphasized in contemporary theories of long-term episodic memory (Aggleton & Brown, 2006; Eichenbaum & Cohen, 2001; Lavenex & Amaral, 2000; Norman & O'Reilly, 2003), and the relative fallibility of associative memory may explain the memory failures and distortions characteristic of episodic memory (Schacter, 2001).

The present findings are relevant to our understanding of the evolution of cognition in primates. Despite at least approximately 20 million years of divergent evolution (Goodman et al., 1998) and marked differences in behavioral and cognitive repertoires, my associative memory capacity was no greater than a baboon's. Furthermore, other indicators of memory function (forgetting rate and parallel memory search) were highly similar to those of baboons.

These results (if they can be generalized to other humans) suggest that the basic mechanisms for associative memory storage and retrieval have changed little throughout primate evolution. Indeed, the architecture of brain structures necessary for associative memory is highly conserved in primates (Burwell & Agster, 2008; Lavenex & Amaral, 2000; Squire & Zola-Morgan, 1991). These findings are problematic for suggestions that relatively gradual expansion in the long-term memory capabilities of mammals have supported relatively gradual developments in complex cognition (Fagot & Cook, 2006). An alternative possibility is that the ability to understand symbolic representations (and to communicate them with language) followed its own evolutionary trajectory (Pinker & Bloom, 1990). Relatively discontinuous changes (Eldredge & Gould, 1972; Gould & Eldredge, 1977), such

as rapid increases in cortical surface area that occurred approximately 2 million years ago (Boyd & Silk, 2006; Deacon, 1997), could have spawned novel mental faculties rather than reorganized fundamental processes, such as long-term associative storage and retrieval.

REFERENCES

Aggleton, J. P., & Brown, M. W. (2006). Interleaving brain systems for episodic and recognition memory. *Trends in Cognitive Sciences*, **10**, 455-463.

Blakemore, C., & Sutton, P. (1969). Size adaptation: A new after-effect. *Science*, **166**, 245-247.

Boyd, R., & Silk, J. B. (2006). *How humans evolved*. New York: Norton.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, **105**, 14325-14329.

Burwell, R. D., & Agster, K. L. (2008). Anatomy of the hippocampus and the declarative memory system. In H. E. Eichenbaum (Ed.), *Systems and neuroscience* (pp. 47-66). Oxford: Elsevier.

Cook, R. G., Levison, D. G., Gillett, S. R., & Blaisdell, A. P. (2005). Capacity and limits of associative memory in pigeons. *Psychonomic Bulletin & Review*, **12**, 350-358.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York: Norton.

Dudai, Y. (1997). How big is human memory, or, on being just useful enough. *Learning & Memory*, **3**, 341-365.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover. (Original work published 1885)

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. New York: Oxford University Press.

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. (1966). A frequency theory of verbal-discrimination learning. *Psychological Review*, **73**, 566-578.

Eldredge, N., & Gould, S. J. (1972). Punctuated equilibria: An alternative to phyletic gradualism. In T. J. Schopf (Ed.), *Models in paleobiology* (pp. 82-115). New York: Freeman.

Fagot, J. (2000). *Picture perception in animals*. East Sussex, England: Psychology Press.

Fagot, J., & Cook, R. G. (2006). Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proceedings of the National Academy of Sciences*, **103**, 17564-17567.

Goodman, M., Porter, C. A., Czelusniak, J., Page, S. L., Schneider, H., Shoshani, J., et al. (1998). Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Molecular Phylogenetics & Evolution*, **9**, 585-598.

Gould, S. J., & Eldredge, N. (1977). Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology*, **3**, 115-151.

Holdstock, J. S., Mayes, A. R., Gong, Q. Y., Roberts, N., & Kapur, N. (2005). Item recognition is less impaired than recall and associative recognition in a patient with selective hippocampal damage. *Hippocampus*, **15**, 203-215.

Landauer, T. K. (1986). How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science*, **10**, 477-493.

Lavenex, P., & Amaral, D. G. (2000). Hippocampal–neocortical interactions: A hierarchy of associativity. *Hippocampus*, **10**, 420-430.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, **110**, 611-646.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral & Brain Sciences*, **13**, 707-784.

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers*. New York: Basic Books.

Skinner, B. F. (1956). A case history in scientific method. *American Psychologist*, **11**, 221-233.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, **253**, 1380-1386.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, **25**, 207-222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, **19**, 73-74.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, **4**, 374-391.

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning & Verbal Behavior*, **20**, 479-496.

Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon.

Voss, J. L., Baym, C. L., & Paller, K. A. (2008). Accurate forced-choice recognition without awareness of memory retrieval. *Learning & Memory*, **15**, 454-459.

Voss, J. L., & Paller, K. A. (2009a). An electrophysiological signature of unconscious recognition memory. *Nature Neuroscience*, **12**, 349-355.

Voss, J. L., & Paller, K. A. (2009b). Recognition without awareness in humans and its implications for animal models of episodic memory. *Communicative & Integrative Biology*, **2**, 1-2.

Wasserman, E. A., & Zentall, T. R. (2006). *Comparative cognition: Experimental explorations of animal intelligence*. New York: Oxford University Press.

Westerberg, C. E., Paller, K. A., Weintraub, S., Mesulam, M.-M., Holdstock, J. S., Mayes, A. R., & Reber, P. J. (2006). When memory does not fail: Familiarity-based recognition in mild cognitive impairment and Alzheimer's disease. *Neuropsychology*, **20**, 193-205.

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General*, **130**, 361-379.