The neurobiology of memory based predictions

Howard Eichenbaum* and Norbert J. Fortin

Center for Memory and Brain, Boston University, 2 Cummington Street, Boston, MA 02215, USA

Recent findings indicate that, in humans, the hippocampal memory system is involved in the capacity to imagine the future as well as remember the past. Other studies have suggested that animals may also have the capacity to recall the past and plan for the future. Here, we will consider data that bridge between these sets of findings by assessing the role of the hippocampus in memory and prediction in rats. We will argue that animals have the capacity for recollection and that the hippocampus plays a central and selective role in binding information in the service of recollective memory. Then we will consider examples of transitive inference, a paradigm that requires the integration of overlapping memories and flexible use of the resulting relational memory networks for generating predictions in novel situations. Our data show that animals have the capacity for transitive inference and that the hippocampus plays a central role in the ability to predict outcomes of events that have not yet occurred.

Keywords: episodic memory; rats; hippocampus; recollection; relational memory; transitive inference

1. INTRODUCTION

There are two general areas of considerable research activity aimed at understanding how we consciously make predictions about future events, one of which seeks to identify the brain structures that underlie imagining the future in humans and the other on the evolution of predicting future events through studies on animals.

In the research on humans, several clinical case studies have noted that patients with severe amnesia due to brain damage are impaired not only in remembering past experiences but also in describing personal future events (Talland 1965; Tulving 1985). Consistent with these findings, Dudai & Carruthers (2005) reminded us that the ancient philosophers viewed memory for the past and imagining the future as intimately linked. Considerable recent interest in the link between remembering the past and imagining the future was stimulated by a report that amnesic patients with primary damage to the hippocampus cannot imagine new experiences, and are particularly deficient in generating a spatial context into which elements of a possible future experience can be bound (Hassabis et al. 2007). Complementary evidence from functional imaging studies showed that a largely overlapping brain network, including the hippocampus, is involved in remembering personal past experiences and in imagining future events (Addis et al. 2007; Schacter & Addis 2007).

Why are the same brain areas that support laying down memories the same ones used in conceiving the future? The answer, many think, can be found in Bartlett’s (1932) early studies on remembering. By contrast to some more modern views of memory as an accurate and detailed record of past events, Bartlett introduced us to the reconstructive nature of recollection, by which we piece together memories of past events from a patchwork of salient details of specific experiences remembered within a framework of our knowledge about how the world works. This characterization of memory for past experiences also bears great similarity with how one might characterize the imagining of future events. Precisely along these lines, Schacter & Addis (2007) proposed the constructive episodic simulation hypothesis that explains ‘memory for the future’ in terms of a system that flexibly recombines details from past events (episodic memories) within a plausible scenario (semantic knowledge). They suggested that the hippocampus might play a specific role in recombining stored information into future event simulations, based on evidence that this structure performs a similar sort of relational processing for information contained in past events (Eichenbaum & Cohen 2001). Building on this framework, Bar (2007) proposed that much of the brain’s default activity involves using analogies and associations to make predictions that pre-sensitize brain representations involved in perception and thinking.

The other area of research that has focused on imagining the future, concerns whether or not animals have the capacities to recall the past and predict the future, with implications for how this capacity evolved in humans. Some of the most creative work on this question has explored the capacities of scrub-jays to remember particular foods that were cached at a particular time and place, capturing Tulving’s (1983) characterization of episodic memory as containing a combination of ‘what’, ‘where’, and ‘when’ prior experiences occurred (Clayton & Dickinson 1998). This line of research has also recognized a close connection between remembering the past and imagining the future (Clayton et al. 2003; Suddendorf 2006).
Thus, more recent studies have extended the capacity of scrub jays to cache food in ways that reflect future planning (Raby et al. 2007). Whether these abilities truly contain conscious awareness of past or future experiences has been challenged (Suddendorf & Busby 2003; Suddendorf & Corballis 2007). Nevertheless, there are now several examples of situations where different species can remember what happened when and where, and behave intuitively as if thinking ahead even beyond their immediate needs (e.g. De Kort et al. 2005; Zentall 2005; Mulcahy & Call 2006; Correa et al. 2007).

Whether or not the evidence justifies the conclusion that animals recall the past and imagine the future remains an area of controversy and has generated considerable effort towards the development of objective criteria for these capacities, rather than limiting the definitions of these abilities to subjective aspects of mental experience that can only be expressed verbally. In particular, Clayton and her colleagues highlighted three specific features of episodic recall and future planning that are shared by animals: (i) that the contents of the memory contain what, when and where information, (ii) that the what, where, and when information is closely integrated into a single representation for which remembering one kind of information automatically retrieves the other features, and (iii) that recovered information can be deployed flexibly to behave adaptively across many situations, including predictions about future events (Clayton et al. 2003).

Here, we will consider how our recent work offers an opportunity to bridge the research in humans and animals. Both lines of study have highlighted a strong connection between remembering the past and imagining the future. The studies on humans have discovered that the hippocampal system is central to both remembering the past and imagining the future and have suggested that the relational processing supported by this system may play the same role in both kinds of cognition. The studies on animals have suggested that what, when and where memories are integrated representations and that these memories can be deployed creatively to solve new problems. Our own research programme has been aimed at understanding the neural circuitry that supports recalling past experiences (Eichenbaum et al. 1999, 2007; Eichenbaum 2000, 2004; Fortin 2008). However, when viewed in the light of these recent findings, our work can also help to answer the following questions: (i) do animals have the capacities for recollection and prediction? and (ii) what are the fundamental information processing functions of the hippocampus that support both remembering the past and imagining the future? Below, we will first consider these questions with regard to recalling the past, then address predicting the future.

2. DO ANIMALS HAVE THE CAPACITY TO RECALL THE PAST, AND IS THIS CAPACITY SUPPORTED BY THE HIPPOCAMPUS?

The distinctive nature of episodic recall or recollection, is perhaps the best illustrated by William James’ (1890) prescient characterization of recollection (what he called ‘memory’): ‘What memory goes with is...a very complex representation, that of the fact to be recalled plus its associates...known in one integral pulse of consciousness... and demanding probably a vastly more intricate brain-process than that on which any simple sensorial image depends’. Thus, James’ (1890) early writings remarkably captured both the contents of our present day view of recollection (fact plus associates) and its retrieval dynamics (integral pulse of consciousness), and contrasted recollection with something like familiarity (a ‘sensorial image’). Although, initially rooted in introspection, the study of recollection has also been approached experimentally in recent years, by directly comparing the contributions of recollection and of familiarity to recognition memory (Yonelinas 2002). In a recognition memory task, subjects can recognize a recently presented stimulus in two ways: they can recollect the specific experience (e.g. ‘I remember seeing the word ‘cat’ because it made me think of my childhood cat’), or they may simply have a sense that the particular item is familiar (e.g. ‘I do not distinctly remember seeing cat, but I’m confident it was on the list’). The two processes are fundamentally distinct in terms of their retrieval dynamics. Recollection involves the ability to recall a specific experience along with its spatial and temporal context and is best characterized by a threshold process, whereas, familiarity is based on the strength of a perceptual match of the current memory cue to previously experienced stimuli and is viewed as an incremental process (i.e. there are degrees of familiarity).

In studies on humans, the use of receiver operating characteristic (ROC) analyses of recognition memory performance is one of the most compelling methods to explore the differences in retrieval dynamics between recollection and familiarity. In a typical experiment, subjects initially study a list of words then distinguish re-presentations of those words as ‘old’ from additional words as ‘new’. The resulting ROC analysis plots ‘hits’, correct identifications of old items, against ‘false alarms’, incorrect identifications of new items as old, across a range of response-bias levels typically measured by confidence ratings. The resulting ROC curve typically involves an asymmetric function characterized by an above-zero threshold of recognition at the most conservative criterion (zero false alarm rate) and thereafter a curvilinear performance function (Yonelinas 2001; see figure 1a). The positive y-intercept is viewed as an index of recollection, whereas the degree of curvature reflects familiarity as typical of a signal-detection process (figure 1b,c); Parks & Yonelinas 2007, but see Wixted 2007). A body of experiments indicates that the human hippocampus is differentially involved in recollection and not familiarity (Eichenbaum et al. 2007, but see Squire et al. 2007).

Are animals also capable of recollection? Given the controversy about whether animals have these capacities (Suddendorf & Busby 2003; Suddendorf & Corballis 2007), it behoves us to provide evidence for the validity of such studies on animals. To answer this question, we adapted the ROC approach used in humans and examined the retrieval dynamics of recognition memory in rats. We used a recognition
task that exploits rats’ superb memory capacities with odours, and varied their response biases by manipulation of reward payoffs and response effort (Fortin et al. 2004). Similar to the findings on humans, we found that the ROC curve of intact rats was asymmetric (figure 1d), containing both a recollection component (above-zero y-intercept) and a strong familiarity (curvilinear) component. Furthermore, following selective hippocampal damage, the recollection component of the ROC function was lost, sparing the familiarity component, i.e. the ROC function became fully symmetrical and curvilinear (figure 1e). Importantly, simply reducing the strength of memory by extending the retention period resulted in a different pattern characterized primarily by loss of the familiarity signal, indicating that hippocampal damage does not just reduce the strength of memories (figure 1f). These findings indicate that, according to measures of retrieval dynamics defined in studies on humans, rats have the capacity for recollection and normally employ both recollection and familiarity to make recognition judgements. Also, as in humans, the rat hippocampus plays a critical and selective role in this ability to recall the past.

3. DO ANIMALS USE INTEGRATED WHAT, WHEN AND WHERE REPRESENTATIONS TO REMEMBER UNIQUE EVENTS, AND ARE THESE REPRESENTATIONS SUPPORTED BY THE HIPPOCAMPUS?

The previous experiment investigated recollection in terms of its distinct retrieval dynamics, but did not explore the content of what is recalled as an animal remembers past experiences. In an attempt to shed light on whether rats integrate what, when and where information, suggested as defining criteria for recollection by James (1890) and Clayton et al. (2003), we trained rats on a task that required them to remember when and where each of a list of odours (what) had recently been experienced (Ergorul & Eichenbaum 2004). Rats began each trial by serially sampling each of the four odours located along different walls of a square arena (figure 2a). Then what, when and where memory was subsequently tested in a choice between two of the stimuli, randomly selected from the four presented items, in which the animal was rewarded for selecting the earlier presented item. Because each trial involved a novel sequence of odours and locations, and because the animals did not know in advance which two odours would be tested on any particular trial, they had to remember each odour and when and where it was experienced on that unique trial. Normal rats performed well on selecting which odour had appeared in a particular location at a particular time (figure 2b). Additional measures indicated that performance was not based simply on memory for location, such that animals were less accurate in their initial approach to a stimulus before they could smell the odour than the final choice following sampling of the odour at that location. Also, on probe trials, when the odour was left out of the stimulus, choice performance fell to chance, indicating rats relied on the appropriate odour to make the final choice. These observations indicate that rats normally use integrated representations of which odour was presented in each place to make the judgement about when the stimuli
were sampled. Furthermore, selective damage to the hippocampus reduced choice performance to chance, and rats with hippocampal damage tended to initially approach the wrong stimulus location (figure 2c). These and other (e.g. Babb & Crystal 2005) findings indicate that normal rats use an integrated representation of what happened where and when. Furthermore, what, where and when memory depends on the hippocampus and, in the absence of a hippocampal representation, rats are influenced by other brain systems to repeat the most recently rewarded spatial response.

What is the nature of the neural representation in the hippocampus that supports integrated what, when and where memory? In order to address the what and where components of recognition memory, we recorded the activity patterns of hippocampal principal neurons in rats performing an odour recognition task in which the stimuli were presented in any of several locations (figure 3a; Wood et al. 1999). We observed that during the period when the rats sampled the odours, some hippocampal cells fired differentially in association with specific odours (what), other cells were activated when the rat sampled any odour at a particular location (where; figure 3b). In addition, yet other cells fired in association with whether the odour matched or differed from that of the previous trial, regardless of odour identity or location, indicating additional representation of the recognition status of each odour. Finally,
the largest subset of hippocampal neurons reflected the integration of these event features by firing differentially in association with a particular combination of the odour, the location where it was sampled, and whether the odour was recognized as the same. These findings indicate that hippocampal neurons, both individually and as a population, represent the critical memory stimulus along with its location and its meaningful context. In an additional recent study, we also found that, when rats encode a sequence of odours, hippocampal neurons carry information not only about the odours and where they were sampled, but also a gradually changing ensemble code that represents the temporal context of each odour sampling event that predict subsequent memory accuracy (Manns et al. 2007; see also Fortin et al. 2002). The combined findings from our recording studies complement the data from the experiments that examined the effects of hippocampal damage, and show that neuronal ensembles within the hippocampus encode integrated what, when and where representations of unique experiences.

4. CAN RATS PREDICT THE FUTURE AND DOES THIS ABILITY INVOLVE THE HIPPOCAMPUS? The above considerations have focused on whether animals have the capacity for recollection and on the contribution of the hippocampus to this ability. Our observations indicate that animals do have the ability for recollecting past experiences, as tested by objective measures of retrieval dynamics and memory content. Also, our observations indicate that the hippocampus plays a critical role in recollective memory and that it does so through representations of items (what) in their spatial (where) and temporal (when) context. Does the same hippocampal memory processing that underlies recollection also support predicting the future? Our position in addressing this question begins with the premise that imagining the future depends on much of the same neural machinery that is needed to remember the past, and involves what Schacter & Addis (2007) called constructive episodic simulation. According to this view, the hippocampus participates in creating a memory for the future by assisting in the adaptive and flexible recombination of information from past events within a plausible scenario. We extend this notion here, proposing that constructive episodic simulation requires the participation of the hippocampus in the initial learning of multiple episodic and semantic memories that overlap in information content, and the consequent ability to integrate them into a network that links memories by their common elements and thereby represents relationships among memories. When called upon to imagine the future, hippocampal processing constitutes a 'surfing' of the network, recovering past events and episodes that can be applied in a variety of ways. In particular, such a relational memory network can support the capacity to make inferences that predict relationships among items that have never been experienced together. By this view, relational processing by the hippocampus provides a mechanism by which information from our vast array of episodic and semantic memories can be accessed flexibly to reconstruct a past experience and to imagine the outcomes of events that have not yet occurred. Importantly, the capacity for predicting events that have never occurred is viewed as qualitatively different from predictions about events that have also previously occurred, for example, predicting that a reward will be given for repeating a behavioural response that has previously resulted in reward.

We have explored the acquisition of multiple overlapping memories and their integration into relational networks that support flexible, inferential and predictive judgements on novel problems. One study examined the ability of rats to learn a set of arbitrary associations between pairs of odours, wherein sets of pairs shared a common element (Bunsey & Eichenbaum 1996; figure 4a). In this study, animals were trained on two sets of overlapping odour-paired associates (A-B and B-C, or X-Y and Y-Z). On each trial, they were presented with the first element of one paired associate (A or X) then presented with a choice between the second elements of both paired associates (B versus Y); a reward was given for selecting the correctly assigned associate (B given A, or Y given X). Thus subjects were required to learn the associations A-B and X-Y. Subsequently, subjects were trained on the second set of pairs wherein the former second element became the first element of the new pairings (B-C, or Y-Z). Following success in learning all four associations, we tested whether the representations of the overlapping paired associations had been interleaved into relational memory networks (A-B-C and X-Y-Z). If the relational network exists, and if the animal has the ability to employ these networks to make novel predictions, then they should be able to make the associative inferences A-C and X-Z. Note that rats should be able to make these judgements as predictions, that is, guesses about associations on which they had never been trained. To test this capacity, we presented rats with one of the first elements from the initial pairs (A or X) and then tested them with the second elements from the second pairs (C versus Z). On these probe trials, no rewards were given and we measured their preference for the inferential choice. Normal rats learned the paired associates and showed strong transitivity in the form of the ability to make the correct predictions in the probe tests (figure 4b). Rats with selective hippocampal lesions also learned the odour pairs over several trials but were severely impaired in the probes, showing no evidence of transitivity and prediction.

In another experiment, we extended the number of associations that rats had to integrate into a relational memory network that had an organized structure (Dusek & Eichenbaum 1997; figure 5a). To accomplish this, rats learned a hierarchical series of overlapping odour choice judgements (A>B, B>C, C>D, D>E; where ‘.’ means ‘should be chosen over’), then they were probed on the relationship between indirectly related items (e.g. B>D). Thus, in this problem, rats had to learn a series of distinct but overlapping pairings, e.g. choose A over B, choose B over C, etc. then integrate them by their common elements to form a hierarchical relational memory network (A>B>C>D>E). The critical test involved probing whether rats
Figure 4. Associative inference in rats. (a(i)(ii)) Rats initially learned two sets of overlapping paired associates (A-B, X-Y, B-C, X-Y and Y-Z). Subsequently, rats were probed with novel pairings (A versus C, A versus Z or X versus C, X versus Z), which can only be solved by associative inference (A-B-C or X-Y-Z). (b(i)(ii)) Both groups of animals successfully learned the two sets of overlapping paired associates. However, only control animals correctly predicted the association in the novel pairings, by showing a preference for the transitive items. These findings indicate that an intact hippocampus is necessary for interleaving the representations of the overlapping paired associates into a relation memory network that supports associative inference. White bars, control; grey bars, hippocampus lesioned.

Figure 5. Transitive inference in rats. (a) Rats learned a series of overlapping premise pairs, presented as odour choice judgements (A > B, B > C, C > D, D > E; where > means should be chosen over). The series of premise pairs could be integrated by their common elements to form a hierarchical relational memory network (A > B > C > D > E). The critical tests involved probing whether rats could predict the relationship between elements that had never been experienced together: the pure predictive probe B versus D, and the probe A versus E, which can be solved using the reward history of the items (A was always rewarded in the premise pairs, whereas E never was). (b) Both groups learned the premise pairs (only average of inner premise pairs is shown), but only control rats showed robust transitive inference in the ability to predict the correct item on the relational probe pair (B versus D). White bars, control; grey bars, hippocampus lesioned.
could predict the relationship between two elements that had never been experienced together. Since any probe that involved an end anchoring item (A or E) could be judged based on consistent experience with that item (A was always rewarded, E was never rewarded), the only pure predictive probe was B versus D, wherein these stimuli had never appeared together during training and had been equally associated with reward. Normal rats learned the series and showed robust transitive inference in the ability to predict the correct item on the B > D probe test (figure 5b). Rats with hippocampal damage also learned each of the initial premises but failed to show transitivity. The combined findings from these studies show that rats with hippocampal damage can learn even complex associations, such as those embodied in the odour paired-associates and conditional discriminations. But, without a hippocampus, they do not interleave the distinct experiences according to their overlapping elements to form a relational network that supports flexible and inferential memory expression to make the correct prediction. Studies on humans have similarly indicated a selective role for the hippocampus in supporting inferences from memory (Heckers et al. 2004; Preston et al. 2004).

There is also substantial complementary evidence from the analyses of firing patterns of hippocampal neurons that hippocampal neurons encode elements that are common among different experiences—these representations could provide links between distinct memories that allow the formation of relational memory networks (reviewed in Eichenbaum 2004). For example, in the Wood et al. (1999) study on odour recognition memory introduced above, some neurons fired association with the same odour across trials where the odour was sampled at different locations and with different meaning (figure 3b). Other cells fired during a particular phase of the approach towards any stimulus cup across trials that occurred at different places. Others fired differentially as the rat sampled at a particular location, across trials with different odours and meanings. Yet other cells fired differentially associated with the recognition status of the odour, across many trials with different odours and at different locations. The observation that hippocampal cells might link experiences by the encoding of common features has also been highlighted in recent studies on monkeys and humans (Kreiman et al. 2000; Hampson et al. 2004). This combination of findings across species provides compelling evidence that hippocampal networks represent common elements among the distinct episodes that could serve to link memories obtained into a relational network.

5. HOW DOES THE HIPPOCAMPUS SUPPORT REMEMBERING THE PAST AND IMAGING THE FUTURE?

Bar (2007) has suggested that a frequent off-line (default) function of the declarative memory system is to explore future possibilities through activation and recombination of memories. The findings reviewed here are entirely consistent with this notion, and offer evidence about the nature of hippocampal representations that can support this function. Data from other studies suggest how the relational representations of the hippocampus emerge in the circuitry of the medial temporal lobe and how the memory and predictive functions of the hippocampus are seamlessly connected (Eichenbaum et al. 2007). The medial temporal lobe is composed of the hippocampus and the immediately surrounding parahippocampal region. Anatomical studies of this region show a segregation of inputs through the parahippocampal region, such that representations of distinct items (e.g. people, objects, events) and information about the context in which those items are experienced (where and when), are processed in distinct areas (Suzuki & Amaral 1994; Burwell & Amaral 1998). What, when and where information converge on and are associated within the hippocampus. Subsequently, when an item is presented as a memory cue, the hippocampus may complete the full pattern and mediate a recovery of the contextual representations, as well as representations of associated items in the appropriate areas of the parahippocampal region (Eichenbaum et al. 2007). According to this view, the recovery of context and item associations constitutes the experience of recollection.

Here, we speculate that imagining the future is based on similar information processing. Representations of information that is common to multiple memories will probably generate the activation of representations of associated items and contextual information for multiple overlapping memories. Furthermore, this information is probably delivered to brain areas that are the recipients of medial temporal output that perform logical processing, including prefrontal areas, that can then assess the validity of relationships between information that is only indirectly related via the common associates. Consider, for example, the associative inference problem described above. Having been trained on A-B and B-C, when an animal is first asked, whether there is a relationship between A and C, hippocampal activations will generate the common associate B and send this information to prefrontal areas. The prefrontal system, then, may evaluate these associations and deduce the indirect association between A and C. Notably, within this conception, the hippocampus itself supplies recovered memories in the service of constructing a plausible future; the hippocampus does not itself generate future scenarios but leaves this constructive processing to cortical systems (for another view, see Lisman & Redish 2009).

We suggest this paradigm for cortical and hippocampal roles in information processing provides a basis for memory-based prediction that can be applied across many more complex problems in predicting the future. Although it is unlikely that animals are capable of the same elaborate mental simulation of the future as humans (Suddendorf & Busby 2003), we argue that the same medial temporal lobe memory processing that supports this capacity in humans is also present in animals, suggesting that the fundamental features of memory-based prediction extend to animals as well.

Preparation of this paper was support by grants from the National Science Foundation and National Institute of Mental Health.
REFERENCES


