MARR'S THEORY OF THE HIPPOCAMPUS AS A SIMPLE MEMORY: DECADES OF SUBSEQUENT RESEARCH SUGGEST IT IS NOT THAT SIMPLE

Suzanna Becker

Abstract

Marr's 1971 simple memory theory of archicortex set the stage for over four decades of subsequent research on the medial temporal lobe memory system. Marr’s ideas contributed fundamentally to current computational models of the hippocampus. In particular, this theory led to the widely held view of the hippocampus as a temporary memory store that accurately memorizes events by creating orthogonalized representations (pattern separation), while using its associative pathways to retrieve the original event from a partial cue. Although Marr's model has garnered considerable empirical support, subsequent research also calls into question some of its key assumptions. Future extensions to the model must address these data. There have also been several game-changing neuroscientific discoveries, such as neurogenesis in the adult hippocampus and the fate of remote memories in densely amnesic patients; these findings require us to rethink the notion of the hippocampus as a simple memory system. Even the long-held view that the hippocampus performs a fundamentally different computation than the neocortex may be incorrect. Converging evidence suggests instead that the hippocampus is at the apex of a multi-level cortical hierarchy that encodes progressively more abstract information at each level, and is as involved in predictive perceptual coding as it is in memory.

Keywords

hierarchical representations, imagery, memorization, neurogenesis, pattern completion, pattern separation, prediction, recall, sparse coding, temporal coding,
Glossary

Archicortex: This is phylogenetically the oldest part of the brain, and includes the hippocampus and dentate gyrus.

CA: cornu ammonis, including the CA1, CA2, CA3 and CA4 hippocampal sub-regions

MTL: medial temporal lobe, including the hippocampus proper (dentate gyrus and CA regions), entorhinal, perirhinal and parahippocampal cortices

Pattern separation: A property of sparse distributed neural coding, such that the sets of output neurons activated by different input patterns have very little overlap, even when the input patterns are highly similar. This is desirable if the organism must respond differentially to very similar inputs.

Pattern completion: the process of recalling the entirety of a stored event when cued with a subset of the elements of an event.

Marr’s simple memory theory of archicortex

In 1971, Marr published his highly influential theory of the hippocampus. By making several assumptions about the hippocampal circuit, Marr derived constraints on the type of neural computations performed by this structure. Although today the entire model could feasibly be simulated on modern day computers, in Marr’s time, the model could only be evaluated as a theoretical exercise. Even so, Marr’s ideas have been tremendously influential, and permeate virtually every modern theory of hippocampal function. Given the influence Marr has had over current thinking, it is worth looking back at Marr’s key ideas, which aspects of the theory have
been supported by subsequent empirical investigations, which have not, and how the field should move forward in light of more recent discoveries.

**Key assumptions**

Marr’s hippocampal theory was derived out of necessity for the proper functioning of his neocortical model. Marr argued that the neocortex necessarily required an associative memory that could rapidly memorize a series of events, and retrieve an event quite accurately when cued with a subset of its elements. This memorizing device would relieve the neocortex of the burden of having to rapidly learn associations amongst different categorical inputs, as it was simultaneously attempting to learn the categories. Later, some of this associative knowledge could be transferred to the neocortex in the form of new categories, once it had been determined which aspects of the event were most critical to be retained. Importantly, Marr assumed that the hippocampus formed a temporary store, recording about one day’s worth of memories. At the end of each day, probably during sleep, some or all aspects of those memories would be transferred to neocortex.

According to Marr, the basic requirements of the hippocampal memory system were to store approximately $10^5$ complex events, the number of events that would be encountered at a rate of about one per second in a single day. Further, when cued with a partial version of an event, an associative retrieval process should recover the entirety of the original event. Associative retrieval had to be completed within the hippocampus, before the pattern was returned to the neocortex in its original form. Therefore, Marr argued that the job of the hippocampus was to form what he called “simple representations”. Rather than performing classification of the input, a simple representation should retain sufficient information to allow the original input event to be reconstructed from a sub-event. In Marr’s framework, a simple representation acts as a template
for each event, and involves relatively few cells. Marr argued that the lower the activity level, the more accurately could different events be encoded. These simple representations were contrasted with the type of representations learned by neocortical pyramidal cells. In the neocortex, but not in the hippocampus, the goal of learning was to discover relevant sets of co-occurring features and perform classification.

Given assumptions about the number of available coding cells and synaptic connections available in neocortical versus hippocampal neurons, Marr calculated the activity level within the hippocampus that would achieve the required capacity of $10^5$ memorized events. This resulted in an estimated sparseness level of about .001. It was assumed that the inhibitory interneurons performed both subtractive and divisive inhibition to keep activity levels normalized to within this target level of activation.

If an incoming pattern of activity from the neocortex had a number of active units above some threshold level, the pattern would be treated as novel and would be stored in the associative memory. Otherwise, a cued recall process would be triggered. Pattern completion had to take place within the hippocampal memory system, and should be complete by the time the event was projected back to the neocortex, so as to require as few neocortical synapses as possible. Thus, the return projections to neocortex were presumed to be used only for indexing the original event, rather than performing an additional associative memory function.

Having worked out the broad computational constraints of an associative memory, Marr then fleshed it out as a simple neural circuit model of the hippocampus. He first considered a two-layered structure, consisting of an input layer and a second layer that combined the functions of sparse coding and associative retrieval into a single stage. This earlier model was rejected in favor of a three-layered structure that separated the functions of coding and associative retrieval
into successive stages. Marr argued that storage and recall must be handled separately. He suggested that if an input pattern consists of a small subset (as few as 1/3) of the features of a previously stored event, recall will be initiated; otherwise the event will be stored as a new event. Thus, the switch between storage and recall was governed by whether the number of active features in the input was greater than some threshold.

The main elements of Marr’s simple memory circuit are shown in Figure 1.

![Figure 1](image)

*Figure 1. The three-layer circuit Marr favored for his simple memory theory of the hippocampus. Blue arrows: feedforward excitatory pathways. Green arrows: recurrent and feedback excitatory pathways. Black lines with filled circle endings: inhibitory pathways.*

Marr suggested neuroanatomical correlates for each of the 3 layers shown in Figure 1. He associated the input layer P1 with the neocortex, the middle layer P2 with the entorhinal cortex and presubiculum, and the output layer P3 with the dentate gyrus (DG) and CA (cornu ammonis) pyramidal regions (the latter combining the CA3 and CA1 sub-fields). Interestingly, Marr argued on the one hand that the DG shared many architectural features of the CA pyramidal regions and should be considered to perform the same basic operations. On the other hand, he acknowledged
the unique aspects of the DG: the lack of recurrent excitatory collateral connections seen in CA3, and the very sparse but large mossy fiber connections projecting from DG to CA3, suggesting to Marr that the DG code should be even sparser than that of the CA regions.

**Empirical support**

The three-layer circuit proposed by Marr has been re-interpreted by subsequent modellers as mapping nicely onto known properties of the trisynaptic circuit in the hippocampus. The cortical input in most subsequent hippocampal models is assumed to be the entorhinal cortex. A second stage of sparse coding could be achieved by the granule cells of the dentate gyrus (although in Marr’s model the P2 and P3 layers were actually smaller and less sparse than the cortical input layer P1), while the associative retrieval function could be performed plausibly by CA3 pyramidal cells via their dense set of recurrent collaterals.

Sparse coding in the dentate gyrus, as predicted by many models that build upon Marr’s, if not by Marr himself, is supported by converging evidence from a wide range of experimental methods. Unit recordings of ongoing neuronal activity and immediate early gene (IEG) markers of recent neuronal activity in the DG in rodents consistently indicate that only 2-4% of dentate granule cells are active in any given environment (Jung and McNaughton, 1993; Gothard et al., 2001; Chawla et al., 2005; Leutgeb et al, 2007; Marrone et al., 2010; Satvat et al., 2011). Moreover, distinct subsets of dentate granule cells respond to similar contexts when small changes are made to the environment (Leutgeb et al, 2007). Remarkably, IEG labelling revealed that even a subtle change in task demands, while holding the spatio-temporal context constant, resulted in distinct populations of DG cells being activated in the two conditions (Satvat et al, 2011). Evidence from human neuroimaging with fMRI indicates that the DG/CA3 area shows greater differences in response to patterns that are highly similar to previously seen items,
relative to other hippocampal regions that show less difference in response (Bakker et al., 2008). While these data support the idea that the dentate gyrus generates sparse less overlapping codes, the question remains, do these codes have anything to do with event memories? Liu et al. (2012) addressed this question by using optogenetic techniques to label a population of neurons that were involved in encoding a contextual fear memory, and then optically re-activating the same population of neurons. The light stimulation induced increased freezing in the mice, indicating that re-activation of the same neurons induced fear memory recall. Taken together, these data are consistent with the idea that the dentate gyrus performs pattern separation, generating distinctive neural codes for event memories, even when the original events are highly overlapping.

Marr’s conjecture that associative retrieval takes place in the CA3 region is also supported by a broad range of empirical data. For example, knockout mice lacking NMDA receptors in CA3 show normal acquisition and normal place fields in the Morris water maze task, but a loss of spatial selectivity in both areas CA3 and CA1 after visual cue removal (Nakazawa et al., 2002). Moreover, human volumetric MR imaging shows that the size of area CA3, but not of other hippocampal sub-regions, predicts how much retrieval confusion people experience between previously stored episodic memories (Chadwick et al., 2014). If CA3 neurons are performing pattern completion, many models predict that attractor dynamics should be observed. Thus, when presented with an ambiguous pattern that is similar to two or more stored memories, the nearest stored pattern should be retrieved, rather than a blend of two or more patterns. To some degree, this prediction has been born out. When CA3 place cell recordings were made in familiar circular and square activity boxes, distinct CA3 cell ensembles fired within each box (Wills et al., 2005). On the other hand, when the box was gradually morphed from being circular to being square, or vice versa, many (but not all) place cells abruptly remapped at some point in the transition from
circular to square, broadly consistent with the idea that the CA3 recurrent collaterals are used for associative retrieval of the nearest stored pattern.

Whereas Marr’s assumptions about coding and associative retrieval are broadly supported by a wealth of data, his assumption about the temporary nature of the hippocampal memory trace is more controversial. Evidence in support of this assumption comes mainly from rodent studies where hippocampal lesions were made at varying intervals post-learning. For some (but not all) hippocampal-dependent tasks, lesions made one day post learning cause severe memory deficits, while hippocampal lesions made weeks after learning do not (for a review, see e.g. Nadel and Moscovitch, 1997). This is consistent with Marr’s assumption that memories are temporarily stored in the hippocampus, but are eventually consolidated elsewhere. We return to this issue later in the chapter where we consider evidence against the assumption that the hippocampus acts as a temporary memory store.

Regardless of whether memories in the hippocampus remain hippocampally dependent or can be consolidated elsewhere and become hippocampal-independent, there is now ample evidence for Marr’s conjecture that memories for recently experienced events are re-played during sleep. For example, when rats navigated through a maze, sequential patterns of hippocampal place cell activation were recorded; later, similar patterns of sequential activation were recorded from the rat hippocampi during slow wave sleep (Wilson and McNaughton, 1994). Moreover, simultaneous recordings in the hippocampus and parietal cortex indicated that the hippocampal-neocortical circuits were jointly re-replaying recent patterns of activation during sleep (Qin et al., 1997). Interestingly, such reactivation patterns have been recorded in all regions of the hippocampus including the dentate gyrus (Shen et al., 1998), suggesting that the entire hippocampal circuit may be involved in associative retrieval. This finding necessitates a re-
evaluation of Marr’s assumption that the hippocampus strictly separates the functions of sparse coding and associative recall into separate hippocampal regions, but still fits within the spirit of his original theory of archicortex.

**Data that challenge Marr’s assumptions**

While there has been considerable empirical support for Marr’s model, as discussed above, some of its underlying assumptions are overly simplifying. See also Willshaw, Dayan and Morris’s (2015) recent commentary on Marr’s model. By relaxing these assumptions and incorporating additional details of the hippocampal circuitry, subsequent theorists have been able to extend the model to account for a broader range of data.

**Three-layered circuit structure**

Marr’s simple memory theory of the hippocampus was strongly constrained by the anatomical data available at the time. While little was known about the activity levels of each hippocampal sub-region, gross connectivity patterns and morphology of different cell types had been studied across many mammalian species. However, even relative to what was known at the time, the model was highly simplified in terms of its layering and connectivity. The architecture of the rat hippocampus, based on more recent data, is shown in Figure 2.
Figure 2. The neural circuitry of the hippocampus. This figure was created by Martin Pyka and Stefanie Bothe. It has not been modified from its original form and is publicly available at https://github.com/MartinPyka/NeuroSVG under the terms of the Creative Commons Attribution 4.0 International License http://creativecommons.org/licenses/by/4.0/

While simplification is intrinsic to modelling, the modeller aims to retain the critical features of the thing being modelled, so as to provide insight into the underlying mechanisms. The risk is always that the modeller may omit some critical aspects of the circuitry that are essential for understanding the functions of the circuit being modelled.

Marr first considered a simpler two-layer model but rejected this on the grounds that it could not achieve adequate storage capacity, and therefore he adopted the three-layer model.

Interestingly, however, using Marr’s computational constraints on memory capacity, sparseness,
input layer size and connectivity, Willshaw and Buckingham (1990) demonstrated in computer simulations that the two-layer and three-layer models described by Marr achieved similar performance. Why then does the hippocampus require the extra layers?

Subsequent to Marr’s 1971 paper, other modellers elaborated on Marr’s theory by incorporating additional physiological and anatomical features of the hippocampus. The direct versus indirect pathways through the hippocampal circuit have been hypothesized to play distinct roles in memory storage versus recall. The CA3 and CA1 regions receive both direct input from the EC via the performant path, and indirect input through the trisynaptic circuit via the DG. The DG granule cells project to the CA3 field via mossy fiber synapses, which are few in number but are among the largest in the brain, such that only a few Mossy fiber synapses may be sufficient to activate a CA3 pyramidal cell (Brown and Johnston, 1983). It has therefore been suggested that these terminals act as “detonator synapses,” so that during encoding, a sparse pattern of activation in the DG mandatorily causes a postsynaptic CA3 cell to fire (McNaughton and Morris, 1987; Treves and Rolls, 1992). On the other hand, during retrieval, the CA3 recurrent collaterals and CA3-to-CA1 Shaffer collaterals may dominate in driving CA3 and CA1 cells to perform associative recall (Treves and Rolls, 1992; Hasselmo and Schnell, 1994; Hasselmo et al, 1996). The switch between storage and retrieval dynamics may be controlled by levels of the neuromodulator Acetylcholine (Hasselmo et al., 1995; Hasselmo, 1999; Hasselmo and McGaughy, 2004). For a more detailed discussion of this, see the chapter by Hasselmo in this book. There also exist back-projections from the CA3 to the DG that have both excitatory and inhibitory effects (Scharfman, 2007). These connections have been ignored by virtually all modellers from Marr’s time to the present, but could also have important implications for neural coding and/or retrieval.
Time and memory

Marr side-stepped the thorny issue of time in memory, but acknowledged that events are not typically isolated at specific time points. Instead, events continuously unfold over time. The importance of the hippocampus for temporal associative memory has long been supported by results from lesion studies. For example, rodents with CA1 lesions are impaired at forming associations between items separated by a 10 second delay (Kesner et al., 2005) and on tasks that required judgements of temporal order (Farovik et al., 2009). Similarly, human patients with MTL amnesia are impaired on memory for temporal order (Downes et al., 2002).

Consideration of how the brain represents time raises many questions, for example: How are temporally discontiguous events integrated into a single episode? How are different episodes parsed into separate events? While answers to these questions remain elusive, some progress has been made.

One way in which the hippocampus may contribute to temporal coding is by representing a gradually evolving temporal context signal (Manns et al., 2007), as predicted by the Temporal Context Model (TCM) of memory (Howard and Kahana, 2002; Howard et al., 2005). Such a mechanism could allow the hippocampus to create a representational bridge that spans across time delays. Further, the hippocampus may be important for forming predictions of future events, a topic that we return to in the final section of this chapter. Data from human neuroimaging studies supports this hypothesis (for a review, see Davachi and Dubrow, 2015). For example, when participants read narratives that contained event boundaries, activation in the MTL, prefrontal cortex and caudate ramped up as successive words of a sentence were read, but dropped suddenly at event boundaries (Ezzyat and Davachi, 2011). If the hippocampus is
forming a temporal predictive model of its inputs, then a sudden violation in its prediction could trigger a reset of its temporal context and the formation of a new episodic trace.

**Game-changing discoveries**

The data reviewed above challenge some of the core assumptions of Marr’s hippocampal theory. However, these challenges do not necessarily indicate that the theory is fatally flawed and should be rejected. Others have extended and further advanced Marr’s model to address the finer nuances of recall and retrieval dynamics and temporal coding. On the other hand, there have been several fundamental discoveries in the last few decades that have been truly “game changers” for many hippocampal theorists.

**Memory versus perception**

A central tenet of Marr’s theories was the sharp divide between the functions of the hippocampal system / archicortex and neocortex: “Archicortex is essentially *memorizing cortex*, in the sense that a given area of archicortex is likely to contain one or more layers of a simple memory… Neocortex, on the other hand, although undoubtedly used a great deal for simple associational storage, can probably be regarded as *classifying cortex*” (Marr, 1971). Broadly speaking, this divide between memorizing and classifying cortex is supported by data from individuals with damage to the medial temporal lobe (MTL) versus the neocortex. A hallmark feature of MTL amnesia is a deficit in memory for complex associations. Such deficits were first documented in great detail for the famous patient HM (Scoville and Milner, 1957). HM, after undergoing bilateral hippocampal transections to treat his intractable epilepsy, was severely impaired in his ability to form new memories for people, places and events. This type of memory is known as autobiographical or episodic memory. Although HM’s memory for events that happened in the
distant past seemed to be intact, he also exhibited retrograde amnesia for the events in the
months leading up to this surgery. Similar patterns of deficits have since been observed in many
other MTL patients. On the other hand, patients with neocortical damage that spares the MTL
appear to have intact episodic memory, but impaired domain-specific knowledge or abilities; the
specific domain of the knowledge impairment, e.g., perceptual, procedural, emotional, or
semantic, depends on the locus of the damage (for a review, see Squire, 2004). The wide range
of specific deficits associated with damage to different areas of the brain has led to the
broadening of the dual memory systems view into a “multiple memory systems” perspective
(Squire, 2004).

In contrast to the trend to divide the cortex into ever more fine-grained memory systems, there
has been an opposing trend toward a more unifying perspective. In particular, the boundary
between perceptual and memory systems has been called into question. A core issue in this
debate is whether the MTL memory system is merely retrieving and remembering information,
or is also involved in the ongoing classification and perception of stimuli. The divide between
perceptual versus memory systems is broadly analogous to Marr’s distinction between the
classifying versus memorizing functions of neocortex and allocortex respectively. Marr argued
that the sole job of the hippocampus was memorization, for the purpose of subsequent re-coding
and storage in neocortex. According to this view, the hippocampus should not be required for
ongoing perceptual decisions. However, it appears that under some circumstances it is.

A prime example is the coding of space. In rodents, hippocampal “place cells” reflect the
animal’s current location in space and are critical for navigation (e.g. O’Keefe, 1976). O’Keefe
and Nadel (1978) argued that this collection of neurons forms the basis of a cognitive map and
provides the rat’s internal allocentric representation of the environment. The same appears to be
true in humans; many hippocampal neurons recorded during a virtual navigation task reflected the person’s current location in the virtual town (Ekstrom et al, 2003).

The importance of the MTL for representing perceptual information also extends to non-spatial perceptual judgements. A growing body of evidence from both lesion and neuroimaging studies points to a central role for the hippocampus in representing higher order perceptual features and making complex perceptual decisions (for a review, see Lee, Yeung and Barense, 2012). For example, patients with selective hippocampal lesions were impaired at deciding which of a pair of altered scenes was closer to a previously viewed target scene; importantly, the same deficit was observed when the two test scenes were viewed simultaneously with the target scene, eliminating the memory component from the task (Lee et al, 2005). Moreover, healthy individuals exhibited increased activation in the hippocampus and perirhinal cortex when discriminating meaningful faces and objects relative to novel stimuli, even when those faces and objects were not subsequently well remembered (Barense et al, 2011). These findings point to a key role for the hippocampus and surrounding MTL structures in perceptual matching and classification. Such a role is difficult to reconcile with Marr’s view that the hippocampus is purely for memorization and not classification.

**Recall versus imagination and future thinking**

Marr’s view of the hippocampus was that it acts as a memorization device. This implies that the main goal of memory retrieval is to reconstruct the original event as precisely as possible. This too has been called into question. Instead, a large body of evidence suggests that the hippocampal “memory system” is as much involved in imagining and predicting the future as it is in remembering the past. Patients with MTL damage not only show autobiographical memory retrieval deficits; they also have great difficulty imagining future scenarios (e.g. Tulving, 1985;
Hassabis et al, 2007; Rosenbaum et al, 2009; Andelman et al, 2010). The future scenarios that they attempt to imagine, like the past events that they attempt to remember, are notably lacking in episodic detail. These data imply that the same neural systems subserving recollection of past events are engaged in imagining future events.

Neuroimaging studies in healthy individuals lend further support to this notion. When asked to imagine a future event, people activate the same neural structures, including the hippocampus, parahippocampal regions, retrosplenial cortex, and posterior parietal areas, as during autobiographical memory retrieval tasks (e.g. Okuda et al, 2003; Addis et al, 2007; Szpunar et al, 2007). Moreover, those amnesic patients who do have the ability to construct fictitious or real future events may rely on residual functions in these same brain regions (Mullally et al, 2012).

The fact that the hippocampal system is involved in imagery and predicting the future is inconsistent with the notion that its chief function is exact memorization. Instead, it may play a central role in planning for the future and selecting the most appropriate response option in the current context. This makes sense from an adaptive perspective. There is obvious utility in being able to recognize the similarities between one’s present circumstances and past contexts that have been experienced, in order to make the best future choices. However, the current context is never precisely identical to a previously experienced event. Thus, one needs the capacity to flexibly retrieve one or more relevant contexts in order to predict a future scenario that may involve a novel combination of one or more past events.

Few if any computational models have addressed the complex operations implied above, namely, using past remembered episodes to imagine and plan for the future. The “BBB” model (Byrne and Becker, 2004; Byrne, Becker and Buress, 2007) proposed a parietal-frontal-hippocampal neural circuit by which a single remembered event could be transformed from an allocentric
representation in long-term memory into an egocentric mental image. Moreover, this mental
image could be translated and rotated, to generate imagined navigation sequences. Such a model,
with the ability to generate a mental image of one’s remembered past, could form the basis for a
planning and spatial navigation system. While the BBB model is highly simplified, it sets the
stage for future model developments that may explain how we are able to imagine never before
experienced future scenarios as novel combinations of past memories.

A temporary memory system?

Marr’s justification for a “simple memory system” residing in the hippocampus was grounded in
memory capacity calculations. He argued that the pyramidal cells of the neocortex required all of
their available synaptic connections to learn the appropriate combinations of input features for
classification. These pyramidal cells did not have enough inter-connectivity to also support
associative learning. Therefore, a memorizing device was required that could quickly lay down a
trace of each event encountered throughout the day, without the need to re-code or classify the
information. This system could itself run into capacity issues, however, if it were not a temporary
memory store. Thus, Marr further assumed that the hippocampus is only temporarily involved in
memory formation. Once the neocortex has had time to work out which aspects of a memory are
relevant and commit them to long-term storage, the contents of the hippocampal memory trace
could safely be discarded. This view of memories temporarily being laid down in the
hippocampus and gradually being consolidated in neocortex has come to be known as the
“systems consolidation” hypothesis.

There is support for the systems consolidation hypothesis from studies of both non-human
animals with hippocampal lesions and humans with MTL damage. For example, rats who were
given hippocampal lesions one day after a contextual fear conditioning paradigm lost the
contextual specificity of their conditioned fear response; on the other hand, when rats were lesioned 28 days after the conditioning, their contextual fear memory was intact (Kim and Fanselow, 1992). Moreover, when rats were given tone-shock pairings in one context, further tone-shock pairings 50 days later in a second context, followed by hippocampal lesions one day later, their remote memory for the initial contextual fear association was intact whereas their memory for the more recent contextual fear event was impaired (Anagnostaras et al, 1999).

These data suggest a time-limited role for the hippocampus in contextual fear memory, after which time the memory is consolidated elsewhere (presumably in neocortex) in a more permanent form. Further, the rodents’ retrograde amnesia was graded, with remote memories being relatively spared. Similar findings have been reported in humans. Patient HM, who had extensive bilateral hippocampal damage, showed a graded retrograde amnesia, with remote memories for events that occurred long ago being spared relative to memories for more recent events that occurred just prior to HM’s surgery (Scoville and Milner, 1957). McClelland, McNaughton and O’Reilly (1995) fleshed out and extended Marr’s ideas about the complementary roles of the hippocampus versus the neocortex. Their highly influential model laid out why it is computationally advantageous to have both fast and slow learning systems, with the former hippocampal system being optimal for learning of exceptional or novel information, and the latter neocortical system being optimal for gradual learning of general statistical regularities. A single system that attempted to do both would suffer from “catastrophic interference”, with new knowledge over-writing old knowledge.

Although the cases in the literature of graded retrograde amnesia after MTL damage seem to support the systems consolidation hypothesis, when one looks more closely at the data, there are problems with this interpretation. A meta-analysis of many such studies indicates that the
temporal extent of the retrograde memory loss can vary widely from one patient to the next, spanning months to years to many decades (Nadel and Moscovitch, 1997). There are even patients who exhibit a flat gradient, with virtually no ability to recall any specific life events, in spite of intact knowledge of skills, perceptual and semantic information (Nadel and Moscovitch, 1997). Moreover, many non-human animal studies of retrograde amnesia also show a flat gradient, with equal deficits when hippocampal lesions are made 1 day versus several weeks after the learning (Nadel and Moscovitch, 1997). Such findings of retrograde amnesia spanning weeks to years do not accord with Marr’s view of the hippocampus. Marr’s calculations of the capacity, sparsity and connectivity of the hippocampus were based on the assumption that the hippocampus would need to store approximately one day’s worth of memories. He envisioned a consolidation process that would take on the order of days, not weeks and certainly not years. Thus, a retrograde memory loss spanning many decades is not consistent with Marr’s notion of the hippocampus as a temporary memory store. Moreover, as Nadel and Moscovitch (1997) aptly put it, “It is difficult to conceive of an adaptive basis for a consolidation process that is almost as long as the average human lifespan throughout much of history”. While not all types of memory were equally affected by MTL lesions, memories for personal autobiographical events were universally and devastatingly affected. Therefore, they concluded that autobiographical episodic memory must always be dependent upon the integrity of the hippocampus.

As an alternative to the systems consolidation theory, Nadel and Moscovitch (1997) put forward the Multiple Trace Theory (MTT). MTT builds on many of Marr’s ideas, but with some important differences. Consistent with Marr’s and related models, MTT postulates that the hippocampus rapidly lays down memory traces of events using sparse distributed codes.
However, the entire hippocampal-neocortical ensemble in MTT constitutes the memory trace for an episode. Moreover, each time a memory trace is reactivated, it is laid down with slightly different subsets of neurons participating in the ensemble. Thus an event that has been re-experienced or remembered often will be supported by many variable memory traces. This explains why certain remote memories – those that are highly salient and likely to have been remembered often – should be more robust to partial damage. Furthermore, the neocortical component of a memory trace of an event will be a generalized version, lacking the original episodic details that are encoded in the hippocampal component of the trace. Thus, in the absence of hippocampal input, a retrieved neocortical memory trace will be more schematized. This accords with studies of remote memory in MTL amnesics. For example, patient KC, who had extensive bilateral hippocampal damage, exhibited deficits in both remote autobiographical and spatial memories although he was able to recognize highly salient landmarks from remotely learned neighborhoods (Rosenbaum et al, 2000).

**Neurogenesis in the adult hippocampus**

If the hippocampus is not in fact a temporary memory store, and yet it maintains a very high level of plasticity, then a dilemma arises. How does the hippocampus overcome interference between successively stored memories? Sparse coding can help to reduce collisions between similar memories and increase pattern separation, as demonstrated in computer simulations by O’Reilly and McClelland (1984). Further, it is now well established that there is ongoing generation of new neurons throughout the lifespan in the dentate gyrus of the hippocampus. Adult hippocampal neurogenesis was first discovered in rodents in the 1960’s (Altman and Das, 1965) and has since been found in a wide range of mammalian species including humans (Eriksson et al., 1998; Knoth et al. 2013; Spalding et al., 2013).
The ongoing neurogenesis in the hippocampus has profound implications for neural coding. Whereas mature dentate granule cells fire very sparsely, as they are under tight control by inhibitory interneurons, the newly generated neurons have very different properties. When they reach about 3-4 weeks of age, they are able to fire action potentials, and yet they are much more plastic (Schmidt-Hieber et al, 2004) and more highly excitable (Mongiat et al, 2009). Over the next several weeks, the young neurons mature and become progressively more like adult DG granule cells. How would this varying population of dentate granule cells affect learning and memory? Simulations of a multi-level computational model of the hippocampus, including the entorhinal cortex, DG, CA3 and CA1 regions, suggest that the incorporation of neurogenesis in the dentate layer protects the hippocampus from interference (Becker, 2005). In Becker’s model, shown in figure 3, the increased levels of excitability and plasticity in the simulated younger population of neurons caused the younger neurons to be preferentially recruited into novel memory formation, thereby protecting older memories from interference. The incorporation of neurogenesis was particularly crucial when the model was challenged to store a set of highly overlapping memories. The predictions from Becker’s model that neurogenesis helps to mitigate against both proactive and retroactive interference have been tested and validated experimentally. Rodents with suppressed neurogenesis were less able to overcome the proactive interference induced when they were challenged to learn to discriminate olfactory odour pairs that overlapped with previously learned pairs (Luu et al., 2012). Additionally, rodents with reduced neurogenesis were more susceptible to the retroactive interference induced by a secondary learning task involving a similar visual discrimination; on subsequent probe trials, animals with reduced neurogenesis showed greater forgetting of the previously learned visual discrimination (Winocur et al., 2012).
Figure 3. Becker’s model of the hippocampus, including neurogenesis in the dentate gyrus (DG). Young neurons are shown in blue; a subset of these are active, shown as filled circles. The input layer to the model represents the entorhinal cortex (EC). Blue arrows: during encoding, activation flows from the input via the trisynaptic circuit through the DG, as well as via direct monosynaptic projections from the EC to the CA3 and CA1. Green arrows: during retrieval, activation flows in the reverse direction from the CA3 and CA1 back to the EC.

Manipulations that suppress neurogenesis have also been shown to impact performance on a wide range of other hippocampal-dependent tasks, including distinguishing between similar contexts, environments, objects, and spatial locations (Saxe et al., 2006; Winocur et al., 2006; Warner-Schmidt, Madsen and Duman, 2008; Wojtowicz et al., 2008; Hernandez-Rabaza et
al., 2009; Kitamura et al., 2009; Ko et al., 2009; Creer et al., 2010; Guo et al., 2011; Sahay et al., 2011; Pan et al., 2012a; Kohman et al., 2012; Nakashiba et al., 2012). Neurogenesis has also been implicated in the long-term retention of memories (Snyder et al., 2005; Deng et al., 2009; Jessberger et al., 2009; Pan et al., 2012b, 2013) and in tasks that require forgetting or overcoming previously learned task demands in order to respond effectively to new ones (Saxe et al., 2007; Pan et al., 2012a). All of these neurogenesis-dependent memory tasks have a high interference component. In addition to mitigating interference, it has also been proposed that neurogenesis contributes to the formation of temporal associations (Aimone et al, 2006; Becker and Wojtowicz, 2007) and remote memories (Snyder et al., 2005; Déry, Goldstein and Becker, 2015), and to the clearance of old memories from the hippocampus (Chambers et al, 2004; Deisseroth et al, 2004; Feng et al, 2001; Weisz and Argibay, 2012; Frankland et al, 2013). Future theoretical developments are needed to reconcile the potential role of the hippocampus, including neurogenesis, in these various functions.

**Future of HC modelling: A hierarchy of memory systems**

In light of the empirical findings and further modelling developments in the 45 years since Marr’s published his model of the hippocampus, two things are clear. First, Marr’s ideas have had a tremendous impact on our thinking about the functions computed by the hippocampus, and have infiltrated virtually every subsequent model of hippocampal coding. Second, many of Marr’s core assumptions were incorrect. The hippocampus is not just a memorization device; it is also important for perception and classification of stimuli. The hippocampus is not a temporary memory store. At least when it comes to richly detailed episodic memories, the hippocampus may always be required to retrieve such memories. It is also not a static neural circuit, but rather, it supports ongoing neurogenesis throughout the lifespan. The range of
properties in younger versus more mature dentate granule cells may facilitate the encoding of novel information while preserving memories for stable properties of the environment. The hippocampus does not merely encode a static series of snapshot memories, but rather, it is integrally involved in encoding temporal sequences and temporal context. It is also very likely involved in predictive coding. This fits well with findings suggesting that the hippocampus is as important for imaging future scenarios as it is for remembering past events.

It is clear that a new kind of model of the hippocampus is called for. The basic functions of the hippocampus must be re-thought. The hippocampus operates integrally with the neocortex to encode, retrieve and predict information. As a first step in this direction, Kali and Dayan (2004) proposed a hierarchical model of the cortex, with the hippocampus at the top of the hierarchy. They proposed that replay of recently stored events served to maintain the correspondence between neocortical and hippocampal representations, as both may be evolving over time. Future developments of this sort of model could potentially account for a wide range of the data reviewed here.

References


Kempermann G (2010) Murine features of neurogenesis in the human hippocampus
across the lifespan from 0 to 100 years. PLoS one 5: e8809.

ablated hippocampal neurogenesis on the formation and extinction of contextual fear

Voluntary wheel running enhances contextual but not trace fear conditioning. Behav

(2005) Perceptual deficits in amnesia: challenging the medial temporal lobe ‘mnemonic’

Lee ACH, Yeung LK, Barense, MD (2012) The hippocampus and visual perception. Frontiers in
Human Neuroscience 91 DOI:10.3389/fnhum.2012.00091

Leutgeb S, Leutgeb JK, Moser MB, Moser EI (2007) Pattern separation in the dentate gyrus and

Optogenetic stimulation of a hippocampal engram activates fear memory recall. Nature
484:381–385.


Mongiat LA, Espósito MS, Lombardi G, Schinder, A.F. Reliable Activation of Immature Neurons in the Adult Hippocampus. PLOS One DOI: 10.1371/journal.pone.0005320


Pan Y-W, Chan GCK, Kuo CT, Storm DR, Xia Z (2012a) Inhibition of adult neurogenesis by inducible and targeted deletion of ERK5 mitogen-activated protein kinase specifically in adult neurogenic regions impairs contextual fear extinction and remote fear memory. J Neurosci 32: 6444-6455.

Pan Y-W, Storm DR, Xia Z (2012b) The maintenance of established remote contextual fear memory requires ERK5 MAP kinase and ongoing adult neurogenesis in the hippocampus. PLoS ONE 7(11): e50455.


Shen, JM, Kudrimoti, HS, McNaughton, BL, Barnes, CA (1998) Reactivation of neuronal ensembles in hippocampal dentate gyrus during sleep after spatial experience. J Sleep Research 7:6-16


