Neural models assist in characterizing the processes carried out by cortical and hippocampal memory circuits. Recent models of memory have addressed issues including recognition and recall dynamics, sequences of activity as the unit of storage, and consolidation of intermediate-term episodic memory into long-term memory.

Current Opinion in Neurobiology 1999, 9:184–188
http://biomednet.com/elecref/0959438800000184
© Elsevier Science Ltd ISSN 0959-4388

Abbreviations
AMPA \(\alpha\)-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid
EEG electroencephalogram
GABA \(\gamma\)-aminobutyric acid
NMDA \(N\)-methyl\(\alpha\)-aspartate

Introduction

Recent evidence suggests that several different neuronal substrates underlie different types of memory behavior (see reviews in [1–3]), including autobiographical memory for specific events in the environment (often referred to as ‘episodic memory’) and the memory for general information about the world (referred to as ‘semantic memory’). In the famous case of patient HM, surgical removal of the anterior hippocampus and adjacent structures bilaterally caused an inability to form new memories of events. Patient HM retains, however, the memory necessary to converse or perform computations, as long as he is not interrupted. In addition, he retains memories of his life before the surgery. Thus, his loss appears specific to intermediate-term episodic memory.

Understanding the properties of these and other types of memory requires understanding the intrinsic properties of neurons and their complex, circuit-level interactions. Here, we will first review neural models of recognition and recall in memory tasks. We will then discuss models that use sequences of activity patterns as the unit of memory storage, as well as models that address consolidation of long-term episodic memory and semantic memory.

Modeling recognition and recall

Recent models have incorporated directly data from human cognitive memory tasks using representations based on extensive physiological and anatomical knowledge of the hippocampal formation [4•,5•]. Basic components of one model [5•] are summarized in Figure 1. These hippocampal models use many features discussed in earlier models of the hippocampus [6–8]. Individual stored items such as words are represented as patterns of active and inactive neurons in the entorhinal cortex, which provides input to the hippocampus and which, in turn, receives converging input from a broad range of neocortical structures. Activity spreads from this input layer into subregions of the hippocampus, including a structure with extensive excitatory recurrent connections — region CA3. Strengthening of the recurrent synapses connecting active neurons within region CA3 provides a mechanism for associating different components of each stored pattern. The experimental phenomenon of long-term potentiation provides support for this mechanism of synaptic modification (see [9] for a review). Associative storage in region CA3 has been included in most models of episodic memory function, which differ primarily in the details of learning rules and activation rules. These models allow behavioral features of memory to be related to specific neural substrates.

Most humans are familiar with the difference in effort required for recognizing a name versus recalling a name. Recall is more sensitive than recognition to injections of the acetylcholine receptor blocker scopolamine before encoding, and the cellular basis of this sensitivity has been analyzed in a network model [5•]. In this model, neurons in entorhinal cortex represent the subject’s memory for experimental context — such as the testing room and apparatus — while separate neuronal populations in this area represent individual items — such as words on a list. During encoding, activity spreads through the dentate gyrus into region CA3, where connections are strengthened between the neurons representing context, between the neurons representing individual words, and between the context and word neurons. The experimentally demonstrated effects of scopolamine were represented by reducing the rate of synaptic modification, reducing the depolarization of neurons, and increasing feedback excitation.

Simulation of scopolamine effects impaired subsequent recall but not recognition. During free recall, a subject is asked ‘What words were on the list?’ In the model, this is simulated with activation of entorhinal context units, which activate the context representation in region CA3. This context representation then sequentially activates the representations of individual words via strengthened connections (competition between words prevents simultaneous recall). During recognition, subjects are presented with individual words and asked which ones they recognize. In the model, neurons representing individual words are activated. If spread of activity across strengthened connections is sufficient to evoke activity in the context neurons, then the item is counted as correctly recognized. Because the context is present more frequently than the
words during encoding, it has stronger excitatory feedback and is easier to activate, allowing recognition to persist even when slower synaptic modification during encoding prevents effective recall. This model suggests specific parameters of memory function that should be affected by specific drugs [5•]. Drug effects on conditioning phenomena in rats have also been modeled [10], but space does not allow us to review neural models of conditioning (see [11]).

The model of recognition described above used one representation of recognition processes; however, humans may use two different processes in recognition: a process of rich recollection, in which the details of the specific item are recalled in a specific context; and a general sense of familiarity based on component features of an item [12]. The rich recollection process may depend on the hippocampus, whereas the sense of familiarity may depend on other cortical structures, including parahippocampal structures. Experimental data relating to these differences have been addressed in a recent paper [4•] describing a model of the dynamics of recollection in a simulation of the hippocampal formation. The activation of representations in region CA3 of this model depends on a conjunction of cues in the entorhinal cortex. This requirement for a specific conjunction of features prevents recollection from being induced by items containing only a portion of the features of a particular memory. For example, the model was trained on patterns representing word pairs such as ‘window–reason’ and ‘car–oyster’ and then tested on re-paired lures such as ‘window–oyster’. The model was able to recollect studied word pairs and reject many re-paired lures by retrieving the correct pair.

The simulation also demonstrated the experimentally observed property that recollection-based recognition decreases with the number of words on a list [12]. This contrasts with the increase in false alarms due to the more vague familiarity-based recognition. This familiarity-based recognition may take place outside the hippocampus and contribute to many aspects of memory function. Lesions of the hippocampus alone cause significant impairments in episodic memory, including recognition memory [13], but research indicates that lesions including structures adjacent to the hippocampus cause stronger overall memory impairments [14]. Both of the models described here build on previous hippocampal models to demonstrate how specific biological processes could underlie components of observed phenomena in human memory experiments.

**Memories stored as sequences of neural activity patterns**

Many early models represented memories with a single spatial pattern distributed across a population of neurons. During retrieval, activity induced by a memory cue caused activity to evolve toward one of these single spatial patterns — an ‘attractor’ state. This final attractor state would persist indefinitely and was not linked directly to other patterns. Because of this single stable final state, these networks are referred to as ‘fixed-point attractor’ systems.

Fixed-point attractors can be generated in networks with extensive excitatory feedback connections, and could therefore exist in region CA3 of the hippocampus or in
neocortical structures. One danger of such strong excitatory feedback connections is the possibility that activity can increase exponentially within the network. In early models, units were prevented from firing at rates higher than a particular maximum value, and memory states commonly involved firing at these high rates [15]. More realistic networks obtained attractor dynamics with lower rates of firing by balancing feedback excitation with different types of inhibition, including shunting inhibition [16], subtractive inhibition [5•,17,18] or normalization of total activity [19]. Though fixed-point attractors in models can persist indefinitely, it is likely that neural circuits only come under the influence of individual attractors for brief periods. In very detailed biological models with spiking neurons, attractors require larger numbers of units to be stable, but can be obtained by using very specific point-to-point inhibitory connectivity [20] or saturating synapses [21].

Memories can also be represented as sequences of activity patterns within a network. In this framework, each pattern of activity in a population of neurons such as region CA3 is associated with a different subsequent pattern during encoding. During retrieval, presentation of an early pattern then elicits a chain of different subsequent patterns in the network that can be repeated in a limit cycle. Sequences provide a simple means of representing inter-item associations in memory tasks, as well as pathways through the environment. Recently, models of sequence storage in region CA3 of the hippocampus [22–24,25•] have been used to address behavioral tasks, including the transitivity phenomenon of ‘theta-phase precession’ [27,28]. This experimental phenomenon is suggestive of sequence storage within the hippocampus. As a rat runs along a continuous track, individual neurons (‘place cells’) in its hippocampus will fire as the rat traverses a location specific to that cell (the ‘place field’). The firing of these cells has been compared with the phase of a high-amplitude oscillation in the hippocampal EEG called the ‘theta rhythm’. As the rat enters the place field associated with a particular place cell, the cell will fire late in the theta cycle. As the rat crosses and leaves the place field, the place cell will fire earlier and earlier, suggesting that the cell was initially the end of a sequence being read out in the hippocampus, and as the rat crosses the field, the cell becomes an earlier component of the sequence.

Several models of theta phase precession have been published. These models all involve a read-out of sequences across time, but two of them [29,30] involve slow read-out of sequences across the full cycle of the theta rhythm (which has a period of about 200 ms). In a model by Tsodyks, Skaggs, Sejnowski and McNaughton [29], this slow read-out is obtained with very weak excitatory connections. In a model by Jensen and Lisman [30], this slow read-out is obtained with the slow dynamics of the NMDA receptor. In contrast, more rapid read-out with AMPA receptor kinetics is used in another model [24]. In this latter model, the theta phase precession is obtained by read-out of sequences to different lengths during different phases of the theta cycle, due to phasic changes in the regulation of synaptic strength by activation of GABAB receptors. This phenomena could enhance retrieval of weak sequences despite stronger prepotent sequences [31,32]. Finally, still other models have proposed that the theta phase precession does not result from sequence read-out, but from a precession attributable to theta oscillations running at different frequencies in the soma versus the dendrites of pyramidal cells [33]. Several of these hypotheses can be tested with pharmacological investigation of the phenomenon of theta phase precession.

The hypothesis that pathways through the environment are stored as sequences of place cell activity gives rise to another prediction, that as a pathway becomes familiar, the place field should expand and move backward along the path [34,35•]. This prediction has recently been confirmed experimentally [36••]. New models have proposed that representations of neural space involve learning of multiple different pathways within that space, which can then be effectively integrated in a flexible, relational structure [25•,35•,37•]. Many models start with an array of simulated place cells that encode the environment, and then modify the connections between these cells to store potential pathways toward specific goals [35•,38•]. This can be viewed as instantiating the assumption that space is the dominant parameter for neuronal response, although it is possible that place cells arise from a generic sparse conjunctive coding scheme [39]. One recent model assumes that generic two-dimensional maps of space are precoded in region CA3, and learning of a new environment involves modification of excitatory input to individual maps, rather than modification of recurrent connections within a map [40•]. In contrast to models starting with a spatial map, models that start with learning individual sequences can draw on a range of features in each sequence, building a response to task elements beyond just the spatial layout [25•]. Most models use a particular goal to influence the activity of other neurons in order to direct network activity toward a particular location [25•,35•,37•,38•]. This remains an important issue as experiments have not demonstrated ‘goal cells’ in any particular structure [41]. Functional models of navigation provide an important means of interpreting available physiological evidence from this important experimental paradigm.

**Consolidation**

Lesions of the hippocampal formation do not appear to impair pre-existing semantic memory, but there is some loss of episodic memories from the time before the lesion; however, more recently stored information appears to be affected more strongly — a phenomenon
termed ‘temporally graded retrograde amnesia’ (reviewed in [42]). This suggests that the hippocampus mediates the gradual formation of neocortical memory representations. Models of the formation of semantic memory demonstrate that gradual, interleaved learning of new episodic information with existing semantic representations is essential to prevent distortion of previously stored semantic representations [42]. Thus, the hippocampus may provide a temporary store for associations that then gradually modify neocortical representations.

The potential effect of the loss of hippocampal training (i.e. hippocampal effects) on semantic memory has been investigated in studies of children with perinatal damage to the hippocampus [43–45]. These subjects show a profound impairment of episodic memory, and as might be expected from the model, their development of semantic memory requires extensive training over a longer period than in normal children — the external world must take the place of an internal mechanism for interleaved learning. Temporally graded retrograde amnesia does not arise consistently in all behavioral tasks [46,47], but in the case of human subjects, temporally graded retrograde amnesia appears to be particularly prominent for patients with damage selective to the hippocampus proper [48].

Neural models of neocortex do not have sophisticated representations of semantic memory. Thus, models of consolidation usually incorporate different time courses for memory formation in the two structures, but do not directly address the problems of differences in the nature of representation in the two structures. A number of models have explicitly addressed the two-stage process of memory formation [49–51]. These models usually assume slower synaptic modification in neocortical structures than in hippocampus. This allows the initial formation of attractor states in the hippocampus, but not the neocortex. Then, during a period in which no input from the external world is presented, distributed activity in the hippocampus activates the attractor states. The spread of activity from these attractor states reactivates components of the association in neocortex, allowing the gradual strengthening of representations in the neocortex.

Potential physiological mechanisms for a two-stage model of memory formation have been proposed [52]. The initial encoding in hippocampus has been proposed to take place during theta rhythm oscillations, and the subsequent transfer to neocortex during sharp waves in quiet waking and slow-wave sleep [52]. The dramatic decrease in acetylcholine levels during slow-wave sleep could contribute to these different dynamic states, as it will greatly enhance the strength of excitatory feedback in the hippocampus [18]. However, two-stage memory models have not demonstrated that the temporal dynamics of sharp wave initiation in region CA3 are such that coded information could be effectively transferred without serious distortions.

Conclusions

Neural simulations demonstrate that specific properties of memory function can be linked to dynamic properties of cortical networks. This modeling will allow increased use of electrophysiological and anatomical data in developing theoretical accounts for memory behavior. At this point, several different models can often account for much of the same data, but further empirical work will increase the constraints, and the models are strongly influencing the course of ongoing experimental investigations.

Acknowledgements

The authors acknowledge support from National Institutes of Health grants MH-47566 to JL McClelland and MH-52732 to ME Hasselmo and National Science Foundation grant IIB 9723947 to ME Hasselmo.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:
• of special interest
•• of outstanding interest


