Cognitive Memory: Cellular and Network Machineries and Their Top-Down Control

Yasushi Miyashita

A brain-wide distributed network orchestrates cognitive memorizing and remembering of explicit memory (i.e., memory of facts and events). The network was initially identified in humans and is being systematically investigated in molecular/genetic, single-unit, lesion, and imaging studies in animals. The types of memory identified in humans are extended into animals as episodic-like (event) memory or semantic-like (fact) memory. The unique configurational association between environmental stimuli and behavioral context, which is likely the basis of episodic-like memory, depends on neural circuits in the medial temporal lobe, whereas memory traces representing repeated associations, which is likely the basis of semantic-like memory, are consolidated in the domain-specific regions in the temporal cortex. These regions are reactivated during remembering and contribute to the contents of a memory. Two types of retrieval signal reach the cortical representations. One runs from the frontal cortex for active (or effortful) retrieval (top-down signal), and the other spreads backward from the medial temporal lobe for automatic retrieval. By sending the top-down signal to the temporal cortex, frontal regions manipulate and organize to-be-remembered information, devise strategies for retrieval, and also monitor the outcome, with dissociated frontal regions making functionally separate contributions. The challenge is to understand the hierarchical interactions between these multiple cortical areas, not only with a correlational analysis but also with an interventional study demonstrating the causal necessity and the direction of the causality.

Since the pioneering observations on patient H.M., who developed a severe and selective deficit in the formation of explicit (or declarative) memory after a bilateral resection of the medial temporal lobe (i.e., the hippocampus and nearby regions), subsequent studies of patients have located the source of various types of impairment in explicit memory in many brain areas (1). Notably, although patients with localized frontal lobe lesions do not have an amnesia typically observed in patients with medial temporal lobe lesions, they do exhibit impairments in memory of temporal context or temporal order, memory of the source of facts or events, or metamemory (i.e., knowledge about one’s memory capabilities and about strategies that can aid memory) (2–4).

The identified brain-wide distributed network, called here the cognitive memory system, is composed of three major subsystems, namely, the medial temporal lobe, the temporal cortex, and the frontal cortex (Fig. 1). Although the ultimate storage sites for explicit memories appear to be in the cortex [but see (5, 6) for another strong position], the medial temporal lobe plays a critical enabling role necessary for storage to take place. Domain-specific cortical regions in the temporal lobes are reactivated during remembering and contribute to the contents of a memory. The reactivation process is mediated by various signals, such as the top-down signal from the prefrontal cortex or the backward signal from the limbic cortex. Frontal regions mediate the strategic attempts for retrieval and encoding and also monitor its outcome, with the dissociated frontal regions making functionally separate contributions. This large-scale cognitive network was initially identified in humans by using neuropsychological and functional imaging. However, molecular, cellular, and network components of this cognitive system have been systematically dissected by recent technical advancements, particularly in animal studies. These include cell type–restricted gene manipulations in mice, a combination of molecular biology and single-unit recording in monkeys, and a sophisticated scan design of event-related functional magnetic resonance imaging (fMRI) in humans. This review aims to integrate some recent discoveries on cellular and network machineries at multiple levels of complexity, which will help us to understand how this brain-wide network orchestrates cognitive memorizing and remembering.

Semantic-Like Memory and Episodic-Like Memory in Nonhuman Species

It is widely held that there are multiple types of memory that are mapped onto distinct anatomical circuits in the brain. Although various taxonomic frameworks for different types of memory have been proposed, most of them share a common generic form (Fig. 2) (1, 7–8): A cardinal distinction lies between short-term (working) memory and long-term memory, and long-term memory is further divided into explicit (or declarative) and implicit (or nondeclarative) memory. Explicit memory is often further divided into semantic (fact) memory and episodic (event) memory: The former consists of facts about the world, such as the capital of the United States of America or that a horse has four legs and usually does not have wings (Fig. 1), whereas the latter was originally characterized as conscious recollection of specific events from one’s personal past in humans (7). For investigations of cellular and neural-network mechanisms, some of these original definitions of memories provide difficult obstacles to studies in nonhuman animals. Regarding semantic cognition, there are lines of evidence that nonhuman animals do not segment the world categorically into objects and that, although they lack verbal expression, they can display through their behavior that they “know” what these and other types of objects are (9, 10). Direct physiological evidence is also available regarding neuronal correlates of object category representation in the cortex (11, 12). Although previous approaches failed to capture some important semantic features of human memory, recent progress encourages us to investigate semantic-like memory, particularly in nonhuman primates (9, 13). Also, some recent studies modeled episodic memory in animals as memory of “when, where, and what” event had occurred (14), or memory of an animal’s own behavior (15). A detailed review of whether such episodic-like memory in animals meets the strict criteria required for human episodic memory (10, 16, 17) exceeds the scope of this article, and the following sections will focus on neurobiological bases of such episodic- or semantic-like memory.

Roles of Local Neuronal Circuits Within the Hippocampus

The medial temporal lobe, particularly the hippocampus, is a major site of multimodal convergence. It contains neurons that are sensitive to the configuration of many
environmental stimuli as well as to the behavioral context in which events occur (5, 18–20). The hippocampus is considered to be critical to the formation of long-term explicit memories, which depends, most theories assume, on this type of configurational information and also on unique hippocampal neural circuits (5, 19, 20).

Anatomically, the hippocampus can be divided into several major subfields, namely, the dentate gyrus (DG), CA3, CA1, and subiculum, and each subfield has a distinct anatomical organization (21). Theoretical studies have proposed a complementary but computationally distinct mnemonic role for each of the hippocampal subfields. Some predictions were experimentally tested and confirmed. For example, the N-methyl-D-aspartate (NMDA) receptor, a type of glutamate receptor, has long been known to be critical to the formation of spatial memory in rats and mice (22), and some theories ascribed it to the NMDA receptors of the Schaffer collateral synapses on CA1 pyramidal cells (20, 23). Cell type–restricted gene manipulations in mice, in which the NMDA receptor gene was ablated specifically in CA1 pyramidal cells, demonstrated that CA1 is indeed a major site involved in the storage of spatial reference memory (24, 25).

However, some hypotheses on different aspects of the hippocampal circuit have not been tested. CA3 pyramidal cells are massively interconnected by recurrent collaterals (more than 10,000 synapses per pyramidal cell in the rat). This recurrent network in CA3 may be critical to various hippocampus-dependent memory functions, particularly “pattern completion” (26). The configuration of environmental stimuli and their behavioral context in daily life are unique and rarely repeated exactly, and thus the input patterns would be able to reactivate only a part of a stored memory and would be unable to activate the whole pattern unless a special computational process recovers the whole from a part. This “the-whole-from-a-part” process is called pattern completion, and the recurrent collateral network with modifiable synapses in CA3 was suggested to perform this computation. This conjecture was recently tested empirically in a genetically engineered mouse strain (called CA3-NR1 KO) where the NMDA receptor gene was ablated specifically in CA3 pyramidal cells of adult mice (27). Plasticity at CA3 recurrent collateral synapses was abolished in these mutants, whereas the plasticity at DG mossy fiber-CA3 synapses remained intact because the latter do not depend on the NMDA receptors. The mutant mice were normal in the acquisition and retrieval of spatial memory with repeated learning, tested with use of the Morris water maze. However, when memory of the hidden platform location was tested after removal of three of the four major extramaze cues, the mutants exhibited a deficit of memory retrieval. The lack of robustness to cue removal of the animal’s spatial behavior was also mirrored in the responses of “place cells,” which links NMDA CA3 receptors to behavior (5, 18). Theoretically, pattern completion can contribute more to wide computational processes in episodic-like memory than shown above (5, 19, 20, 23). One key feature, though not a sufficient criterion, of episodic-like memory is that it is acquired rapidly in a single trial and involves trial-specific information. A delayed matching-to-place (DMP) version of the Morris water maze task was used as a single-trial learning model, and an intrahippocampal infusion of the NMDA antagonist, D(-)-2 amino-5-phosphonovaleric acid (D-AP5), impaired the DMP task (28). This deficit in the DMP task can also be induced by NMDA receptor ablation restricted in CA3 pyramidal cells (29). In another example of an episodic-like memory test, rats encoded paired associates (flavors of food and their spatial locations) and recalled one item when cued by the other (30). When pairings of a particular food and its location were never repeated, ensuring unique “what-where” attributes that are other key features of episodic-like memory, intrahippocampal infusion of D-AP5 impaired memory encoding but not memory recall. Infusion of an AMPA receptor antagonist, CNQX, impaired both encoding and recall. In contrast, when paired associates were trained repeatedly over 8 weeks, the blockade of hippocampal AMPA receptors did not affect their recall. This indicates the differential roles of hippocampal and extrahippocampal neural circuits for nonrepeated and repeated learning.

---

**Fig. 1.** A schematic drawing of the brain-wide network discussed in this article. Specific functional localizations are assigned on the basis of the observations on semantic-like memory in humans and monkeys (other types of memory are supported in different brain structures). The associative representations of long-term memory in the temporal cortex can be reactivated by either of two types of memory retrieval process; active retrieval process is supported by the signal that runs from the frontal cortex (top-down signal), whereas automatic retrieval signal is generated within the network of the temporal lobe and spreads backward. The cortical representations of experience or knowledge are schematically drawn as a semantic network (57). Neurobiologically they are likely mediated by pair-coding memory neurons that are created by a structural reorganization of neural circuits. 

---
A similar conclusion was obtained by single-unit recording of the activity of hippocampal neurons in monkeys while they learned a new association between a scene and an eye movement direction in several successive trials (31). Hippocampal neurons changed their firing rate and stimulus selectivity during the location-scene learning, suggesting the involvement of these neurons in the initial formation of new associative memories.

These observations provide experimental supports for the theoretical considerations in the previous section. First, the unique configurational association between environmental stimuli and behavioral context, which is likely the basis of episodic-like memory, depends on distinct hippocampal neural circuits. Second, memory traces representing repeated associations are consolidated in neural circuits outside the hippocampus. With repetition, only the gist of the configurational information is reactivated and would be multiply represented among various contextual information that accompanied the events, eventually being associated with the subject’s general knowledge as semantic-like memory (8, 13). There is a debate on whether the hippocampus is necessary for retrieval of episodic memory per se (5–8), whereas it is agreed that other types of memory (e.g., semantic memory) are formed over time that do not depend on the hippocampal system for retrieval, though they may require it for acquisition. Third, as conjectured many times (5, 7), human semantic and episodic memory may have evolved from the spatial knowledge and learning of the ancestors of humans. Studies on spatial memory have been providing many important works concerning the role of the hippocampus, which link molecular/genetic, single-unit, lesion, and functional imaging studies across various species. Many fine reviews of those studies are available and should be consulted (5, 32, 33).

Cortical Representation of Associative Memory: Temporal Association Mechanism

Where, outside the hippocampus, are memory traces representing such repeated associations located, and how are they organized? Several lines of evidence suggest that they are organized in the higher order association cortex for semantic-like memory (13, 34–36) as well as in subcortical structures, such as the basal ganglia, for some nondeclarative memory (35, 37). The neuronal correlates of associative long-term memory were first reported in the monkey inferior temporal (IT) cortex (38, 39). In order to investigate the neurobiological basis of semantic-like memory, monkeys were trained to perform the pair-association memory task, a well-known neuropsychological test that is widely used for assessment of dysfunction of the medial temporal lobe system in humans. Single neurons in the IT cortex were found to create linkages between representations of operationally associated but physically unrelated visual stimuli. These associations turned out to be formed by temporally correlated activity in the network, as also confirmed by later studies (40, 41). This work opened the door to neurobiological investigations of the cortical semantic-like memory network by reducing a complex network into its elementary associative links between two objects and then by seeking molecular and morphological machineries underlying such elementary associative links (13, 36).

Physiological mapping and comparison of memory representations in different temporal cortical regions [for anatomical definitions, see Fig. 3A and (42, 43)] revealed that the percentage of memory neurons (the “pair-coding” neurons) that encode both paired associates was significantly higher in the limbic cortex (area 36, 33%) than in the adjoining neocortex (area TE, 4.9%; 44). The functional architecture was also different between TE and area 36: Pair-coding neurons were found to form a local cluster of about 1 mm in area 36 but were distributed more sparsely in TE. Thus, although neurons in both areas acquire stimulus selectivity through learning, the association between representations of visual paired associates proceeds forward within the IT cortex in multiple steps.

Molecular and morphological bases of neural circuit reorganization. It has long been hypothesized that the memory engrams of declarative knowledge in the cortex develop with the structural reorganization of neural circuits (34–36). This reorganization of neural circuits would be accomplished through a cellular program of gene expression leading to increased protein synthesis and then to an alteration of synaptic connections (45). This hypothetical framework has been primarily investigated in invertebrates and lower mammals.

Recently, this hypothesis has been tested with the pair-association task for semantic-like memory in a series of molecular biological studies carried out in monkeys. Up-regulation of mRNAs encoding proteins, thought to be involved in structural reorganization, occurs during the formation of the pair-association memory in a narrow cortical area where pair-coding neurons are physiologically located (Fig. 3, B to D) (46). In these studies, reverse transcription polymerase chain reaction mRNA quantitation was combined with an experimental strategy using split-brain monkeys, in which the anterior commissure and the entire extent of the corpus callosum were transected. This preparation enables the comparison of mRNA expressions in both hemispheres (the pair-association hemisphere and the control hemisphere) within the same monkey, thereby eliminating genetic and cognitive variations between individuals. mRNA encoding the gene of brain-derived neurotrophic factor (BDNF) and an immediate-early gene, zif268, were found to increase in area 36. The spatial distribution of mRNAs was also visualized with the use of in situ hybridization. BDNF mRNA–positive cells and ZIF mRNA–positive cells accumulated as a patchy cluster in area 36, extending for at least 0.4 mm along the anterior-posterior axis.

BDNF is considered to mediate activity-dependent synaptic plasticity, even in mature nervous systems, and the BDNF Val66met
polymorphism was demonstrated to affect activity-dependent secretion of BDNF and human episodic memory (47). Because zif268 encodes a transcription factor (a protein that binds to DNA and controls a transcription of other genes), its expression may trigger a cascade of gene activation that leads to cellular events underlying the neuronal circuit reorganization. This hypothesis was tested by a morphological approach (48). In monkeys trained in a pair-association task, neurons selective to learned pictures formed a focal patch in area 36 (“hot spot”). Three types of retrograde tracer were injected into area 36. One tracer was injected into the hot spot. Two different tracers were injected into two regions adjacent but outside the hot spot. Then, the distribution of retrogradely labeled neurons and electrophysiologically recorded neurons was compared in TE. Picture-selective neurons in TE projected less divergently and more specifically to the hot spot than other neurons, suggesting that, after visual learning, axonal arbors originating from learning-related neurons are pruned to connect specifically to the patch in area 36 whereas those from other neurons retain their divergence. An interesting conjecture is that this learning-induced axonal pruning outside the hot spot is related to the enhanced local extension of axonal arbors within the hot spot (48), leading to the reorganization of local networks in the hot spot that is detected electrophysiologically as a change in neuronal stimulus selectivity, that is, the emergence of pairing-coding neurons.

**Activation of Memory Representations: Active Versus Automatic Retrieval**

The associative long-term memory stored in the temporal cortex can be retrieved by either of two types of memory retrieval processes: one occurs when we need no effort to recall and the other when we have to strive toward a successful recall. We refer to the former as automatic retrieval and to the latter as active, or effortful, retrieval (13). The concept of active retrieval stems from the controlled processing in cognitive theories (49, 50), which was characterized to be capacity limited and operating when the task cannot be accomplished through automated stimulus-response mapping. I now examine evidence supporting the hypothesis that active retrieval and automatic retrieval are supported by retrieval signals generated within the temporal lobe network and by signals that run from the frontal cortex to the temporal cortex, respectively (13).

**Automatic retrieval signal: backward spread of memory signal in the temporal lobe.** The cognitive theory of semantic network postulates retrieval of an item as an activation of a corresponding node in the network (51). The neural correlate of such a node activation was first reported in a pair-association task (39). The response was referred to as a pair-recall response. Then, by using a modified pair-association task (PA with a color switch task), Naya et al. showed that this pair-recall response indeed corresponds to the recall of the target in the subject’s mind, because IT neurons started firing immediately after a color switch that signaled the necessity and timing of memory retrieval during a delay period (52). IT neurons also stopped firing immediately after another color switch that signaled the retrieval of other memorized items. Recently, propagation of the pair-recall activity in the temporal lobe has been investigated (53). The onset of the pair-recall activity was much earlier and the activity developed more rapidly in area 36 neurons than in TE neurons. The median retrieval time was over 300 ms longer in TE than in area 36. Therefore, memory retrieval signals appeared first in the limbic cortex (area 36), after which neocortical (TE) neurons were gradually recruited to represent the sought target. Thus, the mnemonic information that was extracted from long-term storage spreads backward from the limbic cortex to the neocortex in the temporal lobe.

**Top-down signaling appears when active retrieval is required.** A clinical case study highlights active retrieval in humans and provides a clue to an experimental model with which active retrieval can be investigated (54). An epileptic patient who had undergone posterior callosotomy (i.e., partial disconnection of the commissural fibers connecting the left and right cerebral hemispheres) was presented a word in his left visual field. He could not read the name of
 primitive retinotopically mapped visual features (64).

Cognitive Control and Top-Down Signaling

What types of cognitive process in the frontal lobes are sending the top-down signals? Patients with localized frontal lobe lesions are impaired in such tasks that tap into, for example, memory of temporal order (recency memory), source memory, or metamemory (1–4, 66, 67). There are many functional imaging literatures on prefrontal contributions to controlled memory retrieval, but relatively small number of reports directly investigated the above types of tasks [for reviews, see (58–61)]. A flavor of recent attempts to break down such complex cognitive functions into elementary processes can be seen in an example of a feeling-of-knowing (FOK) task (68, 69). FOK is a subjective sense of knowing an item or a word before recalling it and is a well-established tool for assessing the metamemory system (Fig. 4, top) (70). Event-related fMRI revealed multiple frontal regions that showed stronger activity when the subjects had a greater FOK, including the bilateral inferior frontal gyri (BA 47), left middle frontal gyrus (MFG) (BA 46/9), frontopolar area (BA 10), and anterior cingulate/supplementary motor areas (BA 32/24/6) (Fig. 4, bottom). In human neuroimaging, the identification and dissociation of distinct frontal regions have been extensively pushed forward by using simpler, controlled memory-retrieval tasks, such as an episodic recognition task or a source retrieval task (58–61). Cognitive and neuropsychological theories suggest several controlled processes common for these controlled retrieval tasks and the FOK task (49, 50, 70, 71). The retrieval-cue-specification process systematically analyzes possible semantic relationships between the retrieval cue and the known characteristics of the potential targets. If self-generated cues trigger necessary semantic knowledge or explicit source recollection unique to the targets, then an appropriate memory judgment can be made. The recollection monitoring process evaluates the products of memory retrieval with respect to their relevance to the demands. In verbal or verbalizable tasks, phonological maintenance and rehearsal processes are recruited. However, the specific functional roles of the identified...
frontal regions related to controlled memory retrieval are currently still under debate (58–61, 72), though many researchers suggest that the anterior portion of the left MFG near BA10 is related to high-level retrieval strategy (61). At present, a similar fMRI study of these complex cognitive tasks in monkeys looks to be simply a dream. But I believe that a powerful approach to test the suggested functional roles of the frontal regions will be provided in monkey experiments by the combination of monkey fMRI (73, 74) and reversible cortical inactivation with a local drug injection.

Conclusions

Neuroimaging studies in humans identified a large brain-wide network of cognitive memory. A subsystem in the parietal cortex (58–60, 75) was not discussed in this article in spite of its importance, because there are few animal data for possible cellular and network machineries that substantiate suggested parietal functions. Initial neuroimaging studies on memory, particularly positron emission tomography and fMRI studies, often emphasized only the activation of the frontal regions without detectable activation in the posterior cortices, which apparently disagreed with neuropsychological literature [for historical review, see (1)]. However, a consensus appears to be emerging. Recent event-related fMRI successfully detected activation in the posterior cortices in the controlled retrieval tasks (62–65) and spatial navigation tasks (33). The frontal regions are critically involved in manipulating and organizing to-be-remembered information and in devising strategies and monitoring for retrieval, although the frontal regions themselves may not be intimately involved in the binding of information into long-term memory (1–4, 58–61). This consensus was promoted by cellular- and network-level observations in mice, rats, and monkeys. The neuronal representations of mnemonic contents in the medial temporal lobe and the posterior cortices, as well as the interacting signals between and within these structures, have been analyzed by genetic and single-unit recording approaches.

The existence of top-down signaling from the frontal cortex to the temporal cortex was directly demonstrated in monkeys (55, 56), and neuronal activities were observed in the monkey frontal cortex in relation to various cognitive memory tasks (3, 57, 76, 77). However, we still lack basic knowledge on how these frontal neurons are incorporated into network machineries that send out the top-down signal and/or support the cognitive processes identified by neuropsychology and neuroimaging in humans (1–4, 76). The challenge is to clarify hierarchical interactions or couplings between multiple cortical areas as initially demonstrated by an effective connectivity analysis or a correlation analysis with fMRI in humans (78, 79). Obviously, the time resolution of these hemodynamic analyses is low when compared with cell-level signaling. Moreover, because these analyses are correlational, only an interventional study will settle the causal relation and the direction of the causality. Thus, the necessary technical breakthrough would be, I believe, to construct an animal model of source memory or metamemory and to apply an interventional approach such as a reversible inactivation with a genetic engineering or a local drug injection. When our knowledge on these cortical interactions advances, we may be able to understand the neurobiological basis of our metacognition.

References and Notes

27. K. Nakazawa et al., Science 297, 211 (2002); published online 30 May 2002 (10.1126/science.1077959).