Published in final edited form as:
*Trends Cogn Sci*. 2019 June ; 23(6): 510–524. doi:10.1016/j.tics.2019.03.006.

# Testing models of human declarative memory at the single-neuron level

**Ueli Rutishauser**[1,2,3,4]

[1]Department of Neurosurgery, Cedars-Sinai Medical Center, Los Angeles, CA

[2]Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA

[3]Center for Neural Science and Medicine, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA

[4]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA

## Abstract

Deciphering the mechanisms of declarative memory is a major goal of neuroscience. While much theoretical progress has been made, it has proven difficult to experimentally verify key predictions of some foundational models of memory. Recently, single-neuron recordings in human patients have started to provide direct experimental verification of some theories, including mnemonic evidence accumulation, balance-of-evidence for confidence judgments, sparse coding, contextual reinstatement, and the VTA-Hippocampus loop model. Here, we summarize the cell types that have been described in the medial temporal lobe and posterior parietal cortex, discuss their properties, and reflect on how these findings inform theoretical work. This body of work exemplifies the scientific power of a synergistic combination of modelling and human single-neuron recordings to advance cognitive neuroscience.

### Keywords

Declarative Memory; human single-neuron; hippocampus; dopamine; memory retrieval; medial temporal lobe

## Towards a Circuit-Level Understanding of Human Declarative Memory

**Declarative memories** (see Glossary) of events and facts are central to human behavior and are part of what defines each of our individual identities [1]. Consequently, deciphering the mechanisms that allow humans to form, maintain, and retrieve declarative memories is a major topic in cognitive neuroscience. Brain mapping and lesion techniques have started to reveal the brain networks that are involved in declarative memory formation, maintenance, consolidation, and retrieval [1]. This work has highlighted the distinct roles of different brain

Contact Author: Ueli Rutishauser, ueli.rutishauser@cshs.org.

areas, such as the hippocampus and other parts of the medial temporal lobe, cortical areas such as the posterior parietal cortex, and the basal ganglia in declarative memory. Theoretical work, on the other hand, has proposed both abstract cognitive and circuit-level mechanisms on how different types of neurons within these areas interact, and how their response changes as a function of learning to encode new memories (‘**engrams**’) [2–10]. Many such models make predictions that are testable using behavioral or non-invasive neuroimaging-based experiments, allowing for close interaction between theory and experiment that has been highly fruitful. Other model predictions, however, require either higher spatial and/or temporal resolution or high quality single-trial measurements to be tested directly. Due to this, it is challenging to test such predictions with non-invasive techniques, which have relatively low temporal and spatial resolution and typically require across-trial averaging. Invasive brain mapping techniques such as **electrocorticography (ECoG) and intracranial electroencephalography (iEEG)** performed in patients have significantly higher resolution and signal quality, an advantage that has already led to significant advances in our understanding of memory (see for example [11–14]). Nevertheless, there remain a considerable number of model predictions that are best tested at the level of individual neurons.

Invasive recordings in animals have provided an invaluable body of detailed knowledge of the circuits involved in memory at the systems, circuit, cellular, and molecular level. But, if our goal is to understand human memory, this animal work has to be complemented by human experiments at similar levels of resolution. To achieve this, invasive recordings in humans are critical for three reasons. First, findings from animals have to be validated in humans to build a bridge between humans and animal model systems and thereby establish the validity and limitations of the model system. Second, aspects of memory that are uniquely developed in humans or which cannot be practically studied in animals have to be investigated in humans. Third, some aspects of cognitive models of human memory can best be validated and refined using data recorded at the scale of individual neurons or small groups thereof. Here, we review what has been discovered from single-neuron recordings in humans (see Box 1) and illustrate how such recordings have been able to bridge between theory and experimentation and thereby provide new mechanistic insight into human memory.

## Single-Trial Learning: Assessing the Hippocampus VTA/SN Loop Model

Learning theory is largely focused on repetitive learning that requires many repetitions for a robust memory to be formed. In contrast, consider the kind of learning that allows a human being to recognize an image with high accuracy that was previously only seen once as part of a sequence of 10’000 images, each shown once for 1s [15]. How the declarative memory system achieves such rapid and high-capacity learning is a key open question.

A model that has motivated much research on the underlying mechanisms is the Hippocampus- **Ventral Tegmental Area (VTA)/Substantia nigra (SN)** loop model [2, 3, 16]. This model proposes that the rapid encoding of novel stimuli is facilitated by transient dopamine (DA) release that is triggered by novel stimuli (also see Box 2). Anatomically, the loop model (Fig. 1A) hypothesizes that, first, circuits within the hippocampus detect that a stimulus is novel. It has long been known that the hippocampus, along with upstream areas such as perirhinal cortex [17, 18], is crucial for detecting novelty [19]. This novelty signal, in turn, is thought to activate DA neurons, which then release DA in the hippocampus. DA release strengthens synaptic plasticity and enables late long-term potentiation (LTP) [3]. Among the predictions of the loop model are: First, subsets of DA neurons should be novelty sensitive, i.e. they should be activated by stimuli which have not been seen before. Secondly, this novelty signal should be conditional on a novelty signal being present in the hippocampus. Thirdly, the novelty signal should appear first in the hippocampus, followed later by DA neurons. Forth, following novelty-dependent activation, transient increases of DA occur in the hippocampus and these increases strengthen late LTP. Fifth, the model makes specific hypothesis on the anatomical pathways by which hippocampal novelty signals reach the SNc/VTA.

Single-neuron recordings have started to provide experimental support in humans for a subset of the above predictions. There are two groups of relevant findings: novelty responses in the hippocampus and other parts of the **medial temporal lobe (MTL)**, and novelty responses of DA neurons. In the MTL, a subset of neurons preferentially increase their firing rate only the first time a stimulus is presented [20, 21]. Comparing the response of such novelty-signaling neurons between the first and second time a stimulus is presented reveals a strong difference, supporting the notion that their response is different after a single exposure. This is in contrast to habituation, which leads to a gradual reduction in response strength of some human MTL neurons over many trials [22]. Habituation develops over tens of trials and responses recovers if other stimuli are shown intermittently. In contrast, novelty-neurons in the MTL do not respond again to the same stimulus even after long periods of time [20].

DA neurons in the VTA/SN are intermingled with other cell types, making it difficult to measure their activity with non-invasive methods. With extracellular recordings, however, it is possible to differentiate between different cell types due to differences in action potential waveforms and spike train statistics. This way, individual DA neurons in the human SN can be observed in patients undergoing **deep brain stimulation (DBS)** implantation for treatment of movement disorders, typically Parkinson’s disease (PD) [23–25] (see Fig. 1D). Note that while PD patients have dopamine cell loss, this loss typically begins in the ventral tier of the SNc, leaving the neurons in the dorsal tier of the SNc intact [26]. It is the dorsal tier that projects to the hippocampus, ventral striatum and cortex [26] and which is typically recorded in humans, making it suitable to investigate the DA system in human PD patients. Such experiments have revealed two sets of DA neuron responses with respect to novelty: one that increases its firing rate to novel stimuli, and one that does so for familiar stimuli [24, 27] (Fig. 1B–C shows the novelty type). Human DA neurons also signal aspects of rewards [23], but whether the same or different DA neurons respond to both reward and novelty remains unknown. The response properties of the novelty subtype of DA neurons are compatible with the loop model: their response diminishes after a single learning trial, the extent of response change between the first and second presentation is predictive of behaviorally assessed memory strength, and the latency of this novelty response is later than that of MTL novelty neurons [24]. Interestingly, this work also revealed a familiarity subtype of DA neurons, a type of response not predicted by the loop model. These neurons increased their firing rate only for previously seen stimuli, a response which might have a role in strengthening or consolidating existing memories. Together, this data illustrates the extent to which the loop model has guided ongoing work and the richness of insight that can be obtained by such model-guided experimental work in humans.

## Representations of memory content: Assessing sparse and distributed coding models

What are the features that are used to encode and retrieve a declarative memory? Declarative memory has two subsystems: **semantic memory and episodic memor**y, which represent knowledge of concept/facts, and details of autobiographical events, respectively [1, 28]. A theoretical question that has motivated much work is the level of granularity at which aspects of these two kinds of memory are represented. There are two opposing views [8–10, 29, 30]: on the one hand are models that indicate that a sparse and highly selective representation is optimal for fast learning [8, 9]. On the other extreme are fully distributed coding views [31, 32], in which only the pattern of activity across large groups of neurons can differentiate between different concepts and experiences. Under this second view, listening to a single neuron in the MTL would not be useful to make a high-level semantic decision such as ‘am I viewing an animal’. In between the two extremes are models that propose a sparse but distributed form of coding; under this view, a given neuron has high response sparsity (it only responds to a small subset of all possible stimuli) but there are many neurons that respond to the same subset of stimuli [8, 10, 33].

A second, orthogonal, question that is motived by the semantic vs. episodic memory subsystem view is whether the neurons encoding concepts/facts also encode aspects of episodic memory (such as memory strength, where, when) or whether the substrate for these two aspects of declarative memory are separate. Lesion studies indicate that they can be separate in the case of remote memories: whereas patients with hippocampal lesions are unable to form both episodic and semantic memories, they can retrieve remote but not recent semantic memories [34].

Compatible with the theoretical prediction of sparse but distributed coding, single-neuron recordings in humans reveal that concepts and facts are encoded in a sparse and invariant manner by single MTL neurons [35–39]. The response of these neurons is highly selective (sparse) but at the same time invariant: a given neuron only responds to a small subset of all tested stimuli, but the subset it responds to is closely related (i.e. all images of animals). At the same time, the code is distributed because there are many neurons with the same or highly similar selectivity [38]. Such neurons have become known as ‘category’ or ‘concept’ cells (Fig. 2A–C shows examples), respectively. Alternatively, the same two types of neurons are also commonly referred to as gnostic units and grandmother cells, respectively [40]. Here, were refer to this group of neurons as **‘visually selective’ (VS) cells**. VS cells are reproducible across tasks, recording techniques, and laboratories [35–39, 41–43] (see [41] for a review) and their responses are abstract and multimodal. For example, a cell tuned for a particular individual responds to a variety of images showing the person, the written name of that person as well as to auditory input saying the individual's name [44] (Fig. 2B). In addition to sensory input, VS cells can be activated by thought (including free recall and maintenance in working memory) [39, 45–48]. Also, for identical visual input, the response of VS cells varies as a function of whether the stimulus entered conscious awareness [42, 49, 50] and if so is indicative of subjective decisions made about the stimulus [51, 52]. These properties of VS cells are also properties of declarative memories, supporting the view that VS cells are part of the representation of such memories. VS cell responses are plastic: they are more likely to represent concepts of personal relevance [53, 54] and can change their tuning by associative pairing [55]. Together, this data motivates the hypothesis that VS cells in the human MTL represent sparse, abstract, and selective features of declarative memories, a view compatible with the sparse but distributed coding model. In contrast, this data is not compatible with fully distributed coding models. One hypothesis motivated by this data is that VS cells represent semantic memories, and thereby constitute an integral part of an engram [56].

A second type of response in the human MTL that has been characterized are **memory selective (MS) cells**. The activity of MS cells correlates with aspects of episodic memory and their activity exhibits single-trial plasticity [20] because their response is conditional on whether a stimulus has been seen before. Two types of MS cells (Fig. 2D) have been described [20, 21, 57, 58]: one that increases its firing only for novel stimuli, and one that increases its firing only for previously seen stimuli (similar cells exist in macaque perirhinal cortex [17]). The magnitude of MS cell activity in response to a previously seen stimulus is indicative of memory strength, with stronger memories resulting in larger changes both during **recognition memory** [58] and spatial cued recall tasks [59]. The firing rate of MS, but not VS, cells is indicative of the subjective confidence of recognition memory decisions. This constitutes a remarkably trial-by-trial correlation with the declarative aspect of episodic memories with only one type of cell [58]. Jointly, this indicates that MS cells represent aspects of episodic memory. Similar to VS cells, the code formed by MS cells is sparse and highly invariant, with the large majority of MS cells not also qualifying as VS cells [58]. Notably, MS cells respond later than VS cells, with an average response latency that is ~180ms longer relative to that of VS cells (a delay attributed to the theta oscillation by hippocampal models [5]). Together, this data shows that VS and MS cells in the human MTL together form an orthogonal code in both feature space and time. This result supports a sequential model of declarative memory, in which high-level sensory categorization first leads to reactivation of related concepts in semantic memory, followed by recognition whether the currently experienced stimulus has been seen before or not, and if so recall of related attributes and reinstatement of context.

What aspect of the highly sparse and abstract responses of VS and MS cells in the human MTL are computed locally and which are inherited from upstream areas such as perirhinal or inferotemporal cortex? One important insight on this question comes from analysis of response latencies of individual neurons relative to stimulus onset. The response latencies of VS cells (which respond earlier than the MS cells) in the human amygdala, hippocampus, and entorhinal cortex are ~300–400ms [35, 58, 60], a delay that is considerably longer than in higher visual areas and which would be too slow for many perceptual processes [61] (but see [62] for an argument that the MTL in addition also has a critical role in visual perception). Also, the latencies of VS cells are inversely proportional to selectivity, with more selective responses occurring later [35]. Intriguingly, VS cells recorded in parahippocampal cortex are less selective and respond about 100ms earlier than those in entorhinal cortex, hippocampus, and amygdala, indicating that as information propagates through the MTL, responses become more and more selective and sparse. In macaques, perirhinal cortex neurons and the BOLD-fMRI signal differentiate familiar from novel stimuli, suggesting that they encode mnemonic aspects of the stimuli shown rather than their physical attributes [17, 63]. Also, selective perirhinal lesions impair recognition memory [64], indicating that the perirhinal cortex might be the first anatomical area downstream to high level sensory areas in which responses become contingent on experience [65]. However, its exact contribution to declarative memory remains unclear [18, 66, 67]. To our knowledge, no single-neuron responses have been performed in human perirhinal cortex, leaving their latency, novelty-selectivity, and sensitivity to explicit declaration of memories unknown. Such recordings will be critical to determine what specific transformations in both selectivity and latency signals undergo as they propagate from perirhinal cortex into the hippocampal system.

## Retrieving existing Memories: Assessing the Reinstatement of Temporal Context Model

What allows the episodic memory system to differentiate similar memories that happened at different points of time and to selectively retrieve temporal clusters of memories? While temporal clustering of retrieval is ubiquitous [68, 69], the aspect of memory search that give rise to this effect remain poorly understood. A class of theoretical models for studying these questions are temporal context models (TCM) [70–73]. TCMs propose that neural activity in a subset of cells drifts as time progresses, leading to identical stimuli perceived at different times being accompanied by different neuronal states. This contextual information is then combined with sensory information and encoded into memory. Within this framework, successful retrieval of a memory results in re-instatement of the neuronal state at the time of encoding (“contextual reinstatement”). As a result, a ‘jump back in time’ or ‘mental time travel’ occurs, which is a defining feature of episodic memory [74]. As part of reinstatement, the context serves as a cue to then retrieve other associated attributes of a memory, a process leading to the subjective impression of ‘recollecting’ [74]. Similarly, the reinstated context serves as a cue to recall other nearby memories, thereby explaining effects of temporal encoding order on recall [70–72]. A second prediction of TCM is that strong memories, which result in recollection, are accompanied by reinstatement, whereas weaker memories are not. This is because the former is associated with a ‘jump back in time’ [75], a feature of reinstatement sensitive to MTL damage [76]. Intracranial recordings have revealed direct evidence for these predictions of TCM.

Neural activity in the MTL drifts slowly over both short (seconds)- and long (minutes) timescales at both the single-neuron [77, 78] and field potential level [79, 80]. This effect is visible as auto-correlation of activity at the single-neuron level over minutes [77, 78]. This temporal drift was not restricted to neurons which do not respond to sensory inputs because it is also visible in neurons that are visually tuned [77]. This data shows that items which are seen at similar periods of time will be accompanied by more similar neural activity, thereby resulting in a more similar memory representation.

The neural state at encoding is reinstated, that is the patterns of activity during encoding and retrieval of the same item are similar. Reinstatement is visible at the single-neuron level in several areas of the temporal lobe, including the hippocampus, amygdala, and middle temporal gyrus (MTG) [77, 78, 81] (Fig. 2E–G). What remains poorly understood is what information is reinstated – is it specifically temporal context or rather other internal or external features? One notable exception is spatially tuned cells, which reveal specific reinstatement of spatial location [82], indicating that reinstatement can be specific. Remarkably, reinstatement is sufficiently powerful to be visible at the aggregate field potential [79, 80] and **blood-oxygen level dependent (BOLD)** functional magnetic resonance imaging (fMRI) [83] level. Reinstatement of activity has been seen both during recall (cued [80, 81] and free [79] recall) as well as recognition memory [77, 78]. In the former, such reinstatement occurs 0–2s before a subject indicates recall of an item [80, 81], whereas in the later reinstatement occurs following stimulus onset. Third, the extent of reinstatement is related to the quality of memory retrieval: it is predictive of whether cued recall occurs correctly or incorrectly [81] (Fig. 2E) and whether an item is recognized with high or low confidence [77].

In the case of recognition memory, the same items are shown during both encoding and recognition, making it necessary to disambiguate between activity representing stimulus-specific sensory input from reinstated activity. However, the reinstated context changes only gradually, and should thereby still be similar to different encoded items which were shown closely (in time) before or after the recognized item. Indeed, both forward-and backward contiguity effects are visible at the single-neuron level [77] in recognition memory (Fig. 2F–G). Uniquely, in recognition memory tasks, recency effects can be differentiated from contiguity effects [57]. Using this approach, it has been shown that contiguity effects (due to reinstatement) are present at the single-neuron level in the absence of recency effects [77]. While forward-and backward contiguity effects remain to be shown during cued or free recall at the single-neuron level, field potential studies show robust reinstatement during recall as well [79, 80]. Together, there is thus evidence at the single-neuron level for several predictions of TCMs, making them excellent candidates for continued theoretical study. Of note, TCM models in addition also make predictions that relate to the process and experience of recollection and memory search during free recall. While important, here our focus is only on the aspects of TCM that have been studied in the context of recognition memory.

## Converting Memories into Decisions: Assessing the Mnemonic Evidence Accumulation Model

Many of the decisions we make in daily life depend on previous experience. Take, for example, deciding whether a person you see on the bus is the same person you met last night at a party. This decision relies on memory retrieval, the integration of different kinds of information, and meta-cognitive processes to assess certainty. A model that encapsulates one point of view of how the nervous system makes decisions is the drift diffusion model (DDM) [84]. The DDM proposes that **leaky integrators** [85] accumulate evidence in favor of all possible choices and that the action for a particular choice is initiated once the total accumulated evidence for a choice exceeds a threshold (Fig. 3A). While the DDM was originally developed for memory-based decisions [86], much of the experimental work so far has focused on perceptual decisions. In particular, individual neurons have been found in macaques [87–89] and rodents [90] whose firing rate reflects integrated sensory evidence and predicts choices [84]. This discovery has led to an unprecedented mechanistic understanding of perceptual decision making. It remains unknown whether a similar mechanism is at work for declarative memory-based decisions and if so, whether the same or different neurons support this process.

Based on prior work on perceptual decision making in macaques [87–89], connectivity analysis for areas that receive hippocampal output, lesion studies, and neuroimaging, one candidate area in humans that has emerged for where mnemonic evidence-integrating neurons might be located is the left posterior parietal cortex (PPC) [91, 92] (Fig. 3B; see [91–93] for detailed reviews). In recognition memory, different parts of PPC exhibit differences in BOLD signal activation between new vs. old, recollected vs. recognized, or high vs. low confidence items [91, 92]. Scalp EEG source localization similarly indicates that PPC is the source of an ERP that differentiates between new and old items [94]. PPC lesions largely do not reduce recognition accuracy (but see [95]), but rather manifest as impoverished autobiographical recall, reduced likelihood of recollection, and reduced confidence. Thus, while information is present, PPC lesions lead to an inability to properly access this information for further processing – a kind of ‘memory neglect’ [93]. While it remains unclear what specifically the PPC contributes to memory retrieval, a hypothesis is that a key contribution is the accumulation of mnemonic evidence [91, 96].

Invasive recordings from human PPC [97–99] provide insight into this prediction. ECoG recordings from left PPC [97] reveal a striking functional heterogeneity when comparing response patterns of high-gamma band power (HGP) between different parts of the PPC (Fig. 3B). In the intraparietal sulcus (IPS), HGP was higher for old compared to new items, whereas in the superior parietal lobule (SPL), HGP power was higher for new compared to old items. The time course of the signal relative to stimulus onset also varied between the two areas: HGP first differentiated between new and old stimuli in SPL (200–300ms), with signals in IPS differentiating only later (300–700ms). Aligned to button-press, signals in IPS differentiated between old and new stimuli up to ~200ms before button press, but not later. In light of the mnemonic integration framework, this data suggests that neurons in IPS integrate evidence for old, but not new, stimuli, and once a threshold is reached a motor action is initiated.

Single-neuron recordings in the IPS [100] of two human subjects participating in a brain-machine interface clinical trial provide further evidence for a role of the PPC in memory retrieval [98]. Within a small 4×4 mm patch of PPC, two types of neurons with signals relevant for memory retrieval were found: memory-selective (MS) and confidence-selective (CS) neurons (Fig. 3D). There were two types of MS neurons: one that increased its firing rate for familiar stimuli, and one that increased its firing rate for novel stimuli (below referred to as the preferred stimulus). This firing rate increase was graded with memory strength as measured by the reported confidence (Fig. 3C–D). Notably, this modulation by confidence was restricted to the preferred stimulus of the MS cell, with the activity during non-preferred trials not modulated. CS cells, on the other hand, increased their firing rate either for high-or low confidence retrieval decisions regardless of whether the stimulus has been seen before. In contrast to recordings during the same task in the MTL [58], there was no evidence in PPC for neurons carrying information about stimulus identity. Rather, both CS and MS neurons signaled non-stimulus specific mnemonic information. Errors trial analysis further revealed that these neurons carried choice signals (Fig. 3E). The recorded neurons started to differentiate between the two choices well before the motor response but significantly later than the latency at which the MS signal is available in the MTL [58, 97, 99, 100]. This result reveals a single-neuron candidate for mnemonic integrators in human cortex.

Although neurons in PPC are remarkably heterogeneous [98], BOLD-fMRI and ECoG studies [97] show that novelty, memory strength, or recollection is indicated by the average activity across large subareas of PPC [99]. This raises the important question of how heterogenous signals at the single-neuron level can give rise to such differences on a larger scale. One limitation is that the recordings discussed here were performed at the border between SPL and the supramarginal gyrus (SMG), making it possible that a transition area exists between the two. A critical open question is how declarative-memory based information is accessed by neurons in PPC. One putative mechanism is that PPC neurons transiently coordinate their activity with hippocampal theta oscillations when integrating hippocampal information but testing this prediction will require simultaneous recordings. While such recordings have not yet been performed, it is known that during autobiographical recall, theta-oscillations transiently synchronize between PPC and MTL [101].

## Meta-Cognitive Confidence Judgments: Assessing the Balance of Evidence Model

The assessment of confidence is a hallmark of declarative memory. Theoretical work has advanced several potential ways by which **confidence judgments** might be made [102–105]. One class of models is the balance of evidence model (BEM) [105], which is an extension of the evidence accumulation model [106] (Fig. 3F). In the case of two possible choices (i.e. old or new), the BEM consists of two integrators that each accumulate evidence. The “balance of evidence” is the absolute difference between the two accumulated evidence values and is proportional to the confidence. Importantly, the two decisions (the actual choice and the confidence) are made at the same time using the same mechanisms. This contrasts with other models, in which the two decisions are made sequentially [107]. The BEM makes specific predictions about the underlying neural correlates (Fig. 3F). These include: there are separate neurons that integrate evidence only for old or new stimuli, the neurons that supply the evidence as well as the neurons that integrate the evidence are modulated by underlying memory strength for only their preferred stimuli (i.e. new or old), and the RT, accuracy, and confidence is proportional to the balance of evidence. The latter prediction implies that confidence is influenced by the accumulated evidence for both the winning and the losing choice.

Support for some of the BEM’s predictions have been found in both macaques [87] as well as humans [58, 98] at the single-neuron level. First, during declarative memory-based decisions, putative neurons that integrate evidence in PPC [98] or that supply input to the integration process in MTL [58] are modulated by evidence strength as reported by subjective confidence (Fig. 3G–H). This modulation is only apparent for the preferred, but not the non-preferred stimulus. MS cells in the MTL and memory-choice cells in the PPC thus represent signals at the input and output stages of decision making that correlate with confidence as predicted by the BEM model. Second, estimating the balance of evidence in individual trials from the firing rate of MS neurons reveals that the balance of evidence correlates well with RT, accuracy, and declared confidence. Together, this data thus reveals a remarkable predictive power of the theoretical quantity of balance of evidence [58], indicating that the BEM model has good explanatory power and can bridge experiment and theory. Note that these experiments make excellent use of the unique ability of humans to declare their confidence in a decision, thereby allowing a direct (albeit subjective) assessment of memory strength.

## Concluding remarks

In this review, we summarized what invasive recordings at the single-neuron level in humans have revealed about the mechanism of human declarative memory. Our emphasis was on illustrating the power of this approach by demonstrating how it has enabled direct testing of predictions made by five models of relevance for declarative memory: the hippocampus/VTA loop model, the sparse coding model, the contextual reinstatement model, the evidence accumulation model, and the balance of evidence model (see Table 1 for a summary). While there are of course many other important models, here we focus on this subset because together they provide a good perspective of the power of the overall approach. In each case, the combination of human behavior and simultaneous single-neuron recordings has revealed critical new insights into different aspects of human memory. Jointly, this data now allows us to synthesize a view of the processing elements and the information flow among these processing elements in the human brain during memory encoding and retrieval (Fig. 4, Key Figure) and to tentatively map aspect of this circuitry to specific kinds of neurons in particular brain areas (of course this abstract model is far from being an actual implementable circuit). While here the focus was specifically on declarative memory, a similar combination of model-driven analysis of intracranial recordings has started to provide essential new insights into other human cognitive processes. While still in its early stage, this powerful approach has a bright future and we anticipate much exciting future work of this kind that will push ahead our understanding of human cognition in ways not possible with other experimental approaches (see Outstanding Questions).

## Acknowledgements

I thank all members of the Rutishauser Laboratory, Adam Mamelak, Ralph Adolphs, Tyson Aflalo, Richard Andersen, Doris Tsao, and Wolfram Schultz for discussion. The work described here was supported by NIH

(R01MH110831, U01NS103792, P50MH094258), the McKnight Endowment for Neurosciences, and the NSF (BCS-1554105).

## Glossary

**Blood oxygen level-dependent signal (BOLD)**

measure indirect measure of neural activity acquired using fMRI.

**Confidence judgment**

A subjective assessment of the likelihood that a given decision was correct or not, typically assessed on a 3 to 10-point scale ranging from “guessing” to “very confident”.

**Contiguity effect**

Temporal clustering of retrieval at the behavioral (the item retrieved next is most likely the one studied right before or after the previously retrieved item) and neuronal (reinstated context is most similar to that present close in time to the retrieved item) level.

**Declarative memory**

Memories that can be brought into conscious awareness and that can be described and assessed verbally. Includes memories for past events (episodic) and facts (semantic memory).

**Deep Brain Stimulation (DBS)**

A clinical treatment for movement disorders that is also sometimes used to treat psychiatric disorders. Applies high-frequency extracellular stimulation, which is thought to inhibit neural activity in the target area.

**Dopamine (DA)**

A neuromodulator important for memory (see Box 2).

**Electrocorticogram (ECoG)**

refers to recordings performed with subdural strip or grid electrodes that do not penetrate the brain.

**Engram**

The physical substrate of a specific memory.

**Episodic memory**

Memories of personally experienced events.

**Intracranial electroencephalography (iEEG)**

We use the term iEEG to refer to recordings with low-impedance macro electrodes along the shaft of depth electrodes deep inside the brain (see Box 1).

**Leaky integrator**

An integrator with a decay rate such that inputs that rely further back in time have less influence on the current value than more recent inputs

**Memory selective cells (MS cells)**

cells that respond differently as a function of whether a stimulus is novel (never seen before) or familiar (seen before at least once).

**Recognition memory**

The ability to identify a previously seen stimulus as familiar.

**Semantic memory**

Memories of concepts and facts about the world.

**Substantia nigra (SN)**

A part of the midbrain that contains cell bodies of dopamine neurons.

**Ventral tegmental area (VTA)**

A part of the midbrain that contains cell bodies of dopamine neurons.

**Visually selective cells (VS cells)**

Summary term that includes all variants of category-and concept cells.

## References cited

1. Squire LR, Stark CE, and Clark RE (2004). The medial temporal lobe. Annu Rev Neurosci 27, 279–306. [PubMed: 15217334]
2. Lisman JE, and Grace AA (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron 46, 703–713. [PubMed: 15924857]
3. Lisman J, Grace AA, and Duzel E (2011). A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. Trends Neurosci 34, 536–547. [PubMed: 21851992]
4. Howard MW (2018). Memory as Perception of the Past: Compressed Time inMind and Brain. Trends Cogn Sci 22, 124–136. [PubMed: 29389352]
5. Hasselmo ME, Bodelon C, and Wyble BP (2002). A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. Neural Comput 14, 793–817. [PubMed: 11936962]
6. Wixted JT (2007). Dual-process theory and signal-detection theory of recognition memory. Psychol Rev 114, 152–176. [PubMed: 17227185]
7. Kepecs A, and Mainen ZF (2012). A computational framework for the study of confidence in humans and animals. Philos Trans R Soc Lond B Biol Sci 367, 1322–1337. [PubMed: 22492750]
8. Marr D (1971). Simple memory: a theory for archicortex. Philos Trans R Soc Lond B Biol Sci 262, 23–81. [PubMed: 4399412]
9. Mcclelland JL, Mcnaughton BL, and Oreilly RC (1995). Why There Are Complementary Learning-Systems in the Hippocampus and Neocortex - Insights from the Successes and Failures of Connectionist Models of Learning and Memory. Psychological Review 102, 419–457. [PubMed: 7624455]
10. Norman KA, and O'Reilly RC (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol Rev 110, 611–646. [PubMed: 14599236]
11. Solomon EA, Kragel JE, Sperling MR, Sharan A, Worrell G, Kucewicz M, Inman CS, Lega B, Davis KA, Stein JM, et al. (2017). Widespread theta synchrony and high-frequency desynchronization underlies enhanced cognition. Nat Commun 8, 1704. [PubMed: 29167419]
12. Kahana MJ, Sekuler R, Caplan JB, Kirschen M, and Madsen JR (1999). Human theta oscillations exhibit task dependence during virtual maze navigation. Nature 399, 781–784. [PubMed: 10391243]

13. Raghavachari S, Kahana MJ, Rizzuto DS, Caplan JB, Kirschen MP, Bourgeois B, Madsen JR, and Lisman JE (2001). Gating of human theta oscillations by a working memory task. J Neurosci 21, 3175–3183. [PubMed: 11312302]

14. Johnson EL, and Knight RT (2015). Intracranial recordings and human memory. Curr Opin Neurobiol 31, 18–25. [PubMed: 25113154]

15. Standing L, Conezio J, and Haber RN (1970). Perception and Memory for Pictures - Single-Trial Learning of 2500 Visual Stimuli. Psychonomic Science 19, 73–74.

16. Lisman JE, and Otmakhova NA (2001). Storage, recall, and novelty detection of sequences by the hippocampus: elaborating on the SOCRATIC model to account for normal and aberrant effects of dopamine. Hippocampus 11, 551–568. [PubMed: 11732708]

17. Tamura K, Takeda M, Setsuie R, Tsubota T, Hirabayashi T, Miyamoto K, and Miyashita Y (2017). Conversion of object identity to object-general semantic value in the primate temporal cortex. Science 357, 687–692. [PubMed: 28818944]

18. Xiang JZ, and Brown MW (1998). Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. Neuropharmacology 37, 657–676. [PubMed: 9705004]

19. Knight R (1996). Contribution of human hippocampal region to novelty detection. Nature 383, 256–259. [PubMed: 8805701]

20. Rutishauser U, Mamelak AN, and Schuman EM (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. Neuron 49, 805–813. [PubMed: 16543129]

21. Viskontas IV, Knowlton BJ, Steinmetz PN, and Fried I (2006). Differences in mnemonic processing by neurons in the human hippocampus and parahippocampal regions. J Cognitive Neurosci 18, 1654–1662.

22. Pedreira C, Mormann F, Kraskov A, Cerf M, Fried I, Koch C, and Quiroga RQ (2010). Responses of human medial temporal lobe neurons are modulated by stimulus repetition. J Neurophysiol 103, 97–107. [PubMed: 19864436]

23. Zaghloul KA, Blanco JA, Weidemann CT, McGill K, Jaggi JL, Baltuch GH, and Kahana MJ (2009). Human substantia nigra neurons encode unexpected financial rewards. Science 323, 1496–1499. [PubMed: 19286561]

24. Kaminski J, Mamelak AN, Birch K, Mosher CP, Tagliati M, and Rutishauser U (2018). Novelty-Sensitive Dopaminergic Neurons in the Human Substantia Nigra Predict Success of Declarative Memory Formation. Curr Biol 28, 1333–1343 e1334. [PubMed: 29657115]

25. Ramayya AG, Zaghloul KA, Weidemann CT, Baltuch GH, and Kahana MJ (2014). Electrophysiological evidence for functionally distinct neuronal populations in the human substantia nigra. Front Hum Neurosci 8, 655. [PubMed: 25249957]

26. Fearnley JM, and Lees AJ (1991). Ageing and Parkinson’s disease: substantia nigra regional selectivity. Brain 114 (Pt 5), 2283–2301. [PubMed: 1933245]

27. Mikell CB, Sheehy JP, Youngerman BE, McGovern RA, Wojtasiewicz TJ, Chan AK, Pullman SL, Yu Q, Goodman RR, Schevon CA, et al. (2014). Features and timing of the response of single neurons to novelty in the substantia nigra. Brain Res 1542, 79–84. [PubMed: 24161826]

28. Squire LR (2004). Memory systems of the brain: a brief history and current perspective. Neurobiol Learn Mem 82, 171–177. [PubMed: 15464402]

29. Rolls ET, and Treves A (2011). The neuronal encoding of information in the brain. Prog Neurobiol 95, 448–490. [PubMed: 21907758]

30. Wixted JT, Squire LR, Jang Y, Papesh MH, Goldinger SD, Kuhn JR, Smith KA, Treiman DM, and Steinmetz PN (2014). Sparse and distributed coding of episodic memory in neurons of the human hippocampus. Proc Natl Acad Sci U S A 111, 9621–9626. [PubMed: 24979802]

31. Rogers TT, and McClelland JL (2014). Parallel Distributed Processing at 25: further explorations in the microstructure of cognition. Cogn Sci 38, 1024–1077. [PubMed: 25087578]

32. Rumelhart DE, McClelland JL, and PDP Research Group (1986). Parallel distributed processing : explorations in the microstructure of cognition, (Cambridge, Mass.: MIT Press).

33. Kreiman G (2017). A null model for cortical representations with grandmothers galore. Lang Cogn Neurosci 32, 274–285. [PubMed: 29204455]

34. Manns JR, Hopkins RO, and Squire LR (2003). Semantic memory and the human hippocampus. Neuron 38, 127–133. [PubMed: 12691670]

35. Mormann F, Kornblith S, Quiroga RQ, Kraskov A, Cerf M, Fried I, and Koch C (2008). Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. Journal of Neuroscience 28, 8865–8872. [PubMed: 18768680]

36. Kreiman G, Koch C, and Fried I (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. Nat Neurosci 3, 946–953. [PubMed: 10966627]

37. Quiroga RQ, Reddy L, Kreiman G, Koch C, and Fried I (2005). Invariant visual representation by single neurons in the human brain. Nature 435, 1102–1107. [PubMed: 15973409]

38. Waydo S, Kraskov A, Quian Quiroga R, Fried I, and Koch C (2006). Sparse representation in the human medial temporal lobe. J Neurosci 26, 10232–10234. [PubMed: 17021178]

39. Kaminski J, Sullivan S, Chung JM, Ross IB, Mamelak AN, and Rutishauser U (2017). Persistently active neurons in human medial frontal and medial temporal lobe support working memory. Nat Neurosci 20, 590–601. [PubMed: 28218914]

40. Coltheart M (2017). Grandmother cells and the distinction between local and distributed representation. Lang Cogn Neurosci 32, 350–358.

41. Quiroga RQ (2012). Concept cells: the building blocks of declarative memory functions. Nat Rev Neurosci 13, 587–597. [PubMed: 22760181]

42. Reber TP, Faber J, Niediek J, Bostrom J, Elger CE, and Mormann F (2017). Single-Neuron Correlates of Conscious Perception in the Human Medial Temporal Lobe. Curr Biol 27, 2991–2998 e2992. [PubMed: 28943091]

43. Reber TP, Bausch M, Mackay S, Bostrom J, Elger CE, and Mormann F (2019). Representation of Abstract Semantic Knowledge in Populations of Human Single Neurons in the Medial Temporal Lobe PLoS Biol in press.

44. Quian Quiroga R, Kraskov A, Koch C, and Fried I (2009). Explicit encoding of multimodal percepts by single neurons in the human brain. Curr Biol 19, 1308–1313. [PubMed: 19631538]

45. Cerf M, Thiruvengadam N, Mormann F, Kraskov A, Quiroga RQ, Koch C, and Fried I (2010). On-line, voluntary control of human temporal lobe neurons. Nature 467, 1104–1108. [PubMed: 20981100]

46. Gelbard-Sagiv H, Mukamel R, Harel M, Malach R, and Fried I (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. Science 322, 96–101. [PubMed: 18772395]

47. Kornblith S, Quian Quiroga R, Koch C, Fried I, and Mormann F (2017). Persistent Single-Neuron Activity during Working Memory in the Human Medial Temporal Lobe. Curr Biol 27, 1026–1032. [PubMed: 28318972]

48. Kreiman G, Koch C, and Fried I (2000). Imagery neurons in the human brain. Nature 408, 357–361. [PubMed: 11099042]

49. Kreiman G, Fried I, and Koch C (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. Proc Natl Acad Sci U S A 99, 8378–8383. [PubMed: 12034865]

50. Nir Y, Andrillon T, Marmelshtein A, Suthana N, Cirelli C, Tononi G, and Fried I (2017). Selective neuronal lapses precede human cognitive lapses following sleep deprivation. Nat Med 23, 1474–1480. [PubMed: 29106402]

51. Wang S, Yu R, Tyszka JM, Zhen S, Kovach C, Sun S, Huang Y, Hurlemann R, Ross IB, Chung JM, et al. (2017). The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. Nat Commun 8, 14821. [PubMed: 28429707]

52. Wang S, Tudusciuc O, Mamelak AN, Ross IB, Adolphs R, and Rutishauser U (2014). Neurons in the human amygdala selective for perceived emotion. Proc Natl Acad Sci U S A 111, E3110–3119. [PubMed: 24982200]

53. Viskontas IV, Quiroga RQ, and Fried I (2009). Human medial temporal lobe neurons respond preferentially to personally relevant images. Proc Natl Acad Sci U S A 106, 21329–21334. [PubMed: 19955441]

54. De Falco E, Ison MJ, Fried I, and Quian Quiroga R (2016). Long-term coding of personal and universal associations underlying the memory web in the human brain. Nat Commun 7, 13408. [PubMed: 27845773]

55. Ison MJ, Quian Quiroga R, and Fried I (2015). Rapid Encoding of New Memories by Individual Neurons in the Human Brain. Neuron 87, 220–230. [PubMed: 26139375]

56. Josselyn SA, Kohler S, and Frankland PW (2015). Finding the engram. Nat Rev Neurosci 16, 521–534. [PubMed: 26289572]

57. Faraut MCM, Carlson AA, Sullivan S, Tudusciuc O, Ross I, Reed CM, Chung JM, Mamelak AN, and Rutishauser U (2018). Dataset of human medial temporal lobe single neuron activity during declarative memory encoding and recognition. Sci Data 5, 180010. [PubMed: 29437158]

58. Rutishauser U, Ye S, Koroma M, Tudusciuc O, Ross IB, Chung JM, and Mamelak AN (2015). Representation of retrieval confidence by single neurons in the human medial temporal lobe. Nature Neuroscience 18, 1041–1050. [PubMed: 26053402]

59. Rutishauser U, Schuman EM, and Mamelak AN (2008). Activity of human hippocampal and amygdala neurons during retrieval of declarative memories. Proc Natl Acad Sci U S A 105, 329–334. [PubMed: 18162554]

60. Minxha J, Mosher C, Morrow JK, Mamelak AN, Adolphs R, Gothard KM, and Rutishauser U (2017). Fixations gate species–specific responses to free viewing of faces in the human and macaque amygdala. Cell Reports 18, 878–891. [PubMed: 28122239]

61. Kirchner H, and Thorpe SJ (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. Vision Res 46, 1762–1776. [PubMed: 16289663]

62. Murray EA, Bussey TJ, and Saksida LM (2007). Visual perception and memory: a new view of medial temporal lobe function in primates and rodents. Annu Rev Neurosci 30, 99–122. [PubMed: 17417938]

63. Landi SM, and Freiwald WA (2017). Two areas for familiar face recognition in the primate brain. Science 357, 591–595. [PubMed: 28798130]

64. Buffalo EA, Reber PJ, and Squire LR (1998). The human perirhinal cortex and recognition memory. Hippocampus 8, 330–339. [PubMed: 9744420]

65. Naya Y, Yoshida M, and Miyashita Y (2001). Backward spreading of memory-retrieval signal in the primate temporal cortex. Science 291, 661–664. [PubMed: 11158679]

66. Squire LR, Wixted JT, and Clark RE (2007). Recognition memory and the medial temporal lobe: a new perspective. Nat Rev Neurosci 8, 872–883. [PubMed: 17948032]

67. Brown MW, and Aggleton JP (2001). Recognition memory: what are the roles of the perirhinal cortex and hippocampus? Nat Rev Neurosci 2, 51–61. [PubMed: 11253359]

68. Miller JF, Lazarus EM, Polyn SM, and Kahana MJ (2013). Spatial clustering during memory search. J Exp Psychol Learn Mem Cogn 39, 773–781. [PubMed: 22905933]

69. Kahana MJ (2012). Foundations of human memory, (New York: Oxford University Press).

70. Polyn SM, Norman KA, and Kahana MJ (2009). A Context Maintenance and Retrieval Model of Organizational Processes in Free Recall. Psychological Review 116, 129–156. [PubMed: 19159151]

71. Howard MW, Fotedar MS, Datey AV, and Hasselmo ME (2005). The temporal context model in spatial navigation and relational learning: toward a common explanation of medial temporal lobe function across domains. Psychol Rev 112, 75–116. [PubMed: 15631589]

72. Howard MW, and Kahana MJ (2002). A distributed representation of temporal context. J Math Psychol 46, 269–299.

73. Sederberg PB, Howard MW, and Kahana MJ (2008). A context-based theory of recency and contiguity in free recall. Psychol Rev 115, 893–912. [PubMed: 18954208]

74. Tulving E (2002). Episodic memory: From mind to brain. Annual Review of Psychology 53, 1–25.

75. Yonelinas AP (2001). Components of episodic memory: the contribution of recollection and familiarity. Philos Trans R Soc Lond B 356, 1363–1374. [PubMed: 11571028]

76. Palombo DJ, Di Lascio JM, Howard MW, and Verfaellie M (2018). Medial Temporal Lobe Amnesia Is Associated with a Deficit in Recovering Temporal Context. J Cogn Neurosci, 1–13.

77. Folkerts S, Rutishauser U, and Howard MW (2018). Human Episodic Memory Retrieval Is Accompanied by a Neural Contiguity Effect. J Neurosci 38, 4200–4211. [PubMed: 29615486]

78. Howard MW, Viskontas IV, Shankar KH, and Fried I (2012). Ensembles of human MTL neurons “jump back in time” in response to a repeated stimulus. Hippocampus 22, 1833–1847. [PubMed: 22488671]

79. Manning JR, Polyn SM, Baltuch GH, Litt B, and Kahana MJ (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. Proc Natl Acad Sci U S A 108, 12893–12897. [PubMed: 21737744]

80. Yaffe RB, Kerr MS, Damera S, Sarma SV, Inati SK, and Zaghloul KA (2014). Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. Proc Natl Acad Sci U S A 111, 18727–18732. [PubMed: 25512550]

81. Jang AI, Wittig JH Jr., Inati SK, and Zaghloul KA (2017). Human Cortical Neurons in the Anterior Temporal Lobe Reinstate Spiking Activity during Verbal Memory Retrieval. Curr Biol 27, 1700–1705 e1705. [PubMed: 28552361]

82. Miller JF, Neufang M, Solway A, Brandt A, Trippel M, Mader I, Hefft S, Merkow M, Polyn SM, Jacobs J, et al. (2013). Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. Science 342, 1111–1114. [PubMed: 24288336]

83. Tompary A, Duncan K, and Davachi L (2016). High–resolution investigation of memory-specific reinstatement in the hippocampus and perirhinal cortex. Hippocampus 26, 995–1007. [PubMed: 26972485]

84. Gold JI, and Shadlen MN (2007). The neural basis of decision making. Annu Rev Neurosci 30, 535–574. [PubMed: 17600525]

85. Usher M, and McClelland JL (2001). The time course of perceptual choice: the leaky, competing accumulator model. Psychol Rev 108, 550–592. [PubMed: 11488378]

86. Ratcliff R, Smith PL, Brown SD, and McKoon G (2016). Diffusion Decision Model: Current Issues and History. Trends Cogn Sci 20, 260–281. [PubMed: 26952739]

87. Kiani R, and Shadlen MN (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. Science 324, 759–764. [PubMed: 19423820]

88. Shadlen MN, and Newsome WT (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 86, 1916–1936. [PubMed: 11600651]

89. Schall JD (2003). Neural correlates of decision processes: neural and mental chronometry. Curr Opin Neurobiol 13, 182–186. [PubMed: 12744971]

90. Hanks TD, Kopec CD, Brunton BW, Duan CA, Erlich JC, and Brody CD (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature 520, 220–223. [PubMed: 25600270]

91. Wagner AD, Shannon BJ, Kahn I, and Buckner RL (2005). Parietal lobe contributions to episodic memory retrieval. Trends Cogn Sci 9, 445–453. [PubMed: 16054861]

92. Sestieri C, Shulman GL, and Corbetta M (2017). The contribution of the human posterior parietal cortex to episodic memory. Nat Rev Neurosci 18, 183–192. [PubMed: 28209980]

93. Davis SW, Wing EA, and Cabeza R (2018). Chapter 27 - Contributions of the ventral parietal cortex to declarative memory In Handbook of Clinical Neurology, Volume 151, Vallar G and Coslett HB, eds. (Elsevier), pp. 525–553.

94. Rugg MD, and Curran T (2007). Event-related potentials and recognition memory. Trends Cogn Sci 11, 251–257. [PubMed: 17481940]

95. Ben-Zvi S, Soroker N, and Levy DA (2015). Parietal lesion effects on cued recall following pair associate learning. Neuropsychologia 73, 176–194. [PubMed: 25998492]

96. Hutchinson JB, Uncapher MR, Weiner KS, Bressler DW, Silver MA, Preston AR, and Wagner AD (2014). Functional heterogeneity in posterior parietal cortex across attention and episodic memory retrieval. Cereb Cortex 24, 49–66. [PubMed: 23019246]

97. Gonzalez A, Hutchinson JB, Uncapher MR, Chen J, LaRocque KF, Foster BL, Rangarajan V, Parvizi J, and Wagner AD (2015). Electrocorticography reveals the temporal dynamics of posterior parietal cortical activity during recognition memory decisions. Proc Natl Acad Sci U S A 112, 11066–11071. [PubMed: 26283375]

98. Rutishauser U, Aflalo T, Rosario ER, Pouratian N, and Andersen RA (2018). Single-Neuron Representation of Memory Strength and Recognition Confidence in Left Human Posterior Parietal Cortex. Neuron 97, 209–220 e203. [PubMed: 29249283]

99. Parvizi J, and Wagner AD (2018). Memory, Numbers, and Action Decision in Human Posterior Parietal Cortex. Neuron 97, 7–10. [PubMed: 29301107]

100. Aflalo T, Kellis S, Klaes C, Lee B, Shi Y, Pejsa K, Shanfield K, Hayes-Jackson S, Aisen M, Heck C, et al. (2015). Neurophysiology. Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. Science 348, 906–910. [PubMed: 25999506]

101. Foster BL, Kaveh A, Dastjerdi M, Miller KJ, and Parvizi J (2013). Human retrosplenial cortex displays transient theta phase locking with medial temporal cortex prior to activation during autobiographical memory retrieval. J Neurosci 33, 10439–10446. [PubMed: 23785155]

102. Pouget A, Drugowitsch J, and Kepecs A (2016). Confidence and certainty: distinct probabilistic quantities for different goals. Nat Neurosci 19, 366–374. [PubMed: 26906503]

103. Yeung N, and Summerfield C (2012). Metacognition in human decision-making: confidence and error monitoring. Philos Trans R Soc Lond B Biol Sci 367, 1310–1321. [PubMed: 22492749]

104. Metcalfe J (2008). Evolution of Metacognition In Handbook of Metamemory and Memory, Dunlovsky J and Bjork R, eds. (New York: Psychology Press), pp. 29–46.

105. Vickers D (1979). Decision processes in visual perception, (New York; London: Academic Press).

106. Bogacz R, Brown E, Moehlis J, Holmes P, and Cohen JD (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychol Rev 113, 700–765. [PubMed: 17014301]

107. Navajas J, Bahrami B, and Latham PE (2016). Post-decisional accounts of biases in confidence. Curr. Opin. Behav. Sci 11, 55–60.

108. Hansen N, and Manahan-Vaughan D (2014). Dopamine D1/D5 receptors mediate informational saliency that promotes persistent hippocampal long-term plasticity. Cereb Cortex 24, 845–858. [PubMed: 23183712]

109. Fried I, Rutishauser U, Cerf M, and Kreiman G (2014). Single Neuron Studies of the Human Brain: Probing Cognition, (Boston: MIT Press).

110. Fried I, Wilson CL, Maidment NT, Engel J, Behnke E, Fields TA, MacDonald KA, Morrow JW, and Ackerson L (1999). Cerebral microdialysis combined with single-neuron and electroencephalographic recording in neurosurgical patients - Technical note. Journal of Neurosurgery 91, 697–705. [PubMed: 10507396]

111. Minxha J, Mamelak AN, and Rutishauser U (2018). Surgical and Electrophysiological Techniques for Single-Neuron Recordings in Human Epilepsy Patients In Extracellular Recording Approaches. (Springer), pp. 267–293.

112. Sheth SA, Mian MK, Patel SR, Asaad WF, Williams ZM, Dougherty DD, Bush G, and Eskandar EN (2012). Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. Nature 488, 218–221. [PubMed: 22722841]

113. Bari AA, Mikell CB, Abosch A, Ben-Haim S, Buchanan RJ, Burton AW, Carcieri S, Cosgrove GR, D'Haese PF, Daskalakis ZJ, et al. (2018). Charting the road forward in psychiatric neurosurgery: proceedings of the 2016 American Society for Stereotactic and Functional Neurosurgery workshop on neuromodulation for psychiatric disorders. J Neurol Neurosurg Psychiatry 89, 886–896. [PubMed: 29371415]

114. Fu Z, Wu DJ, Ross I, Chung JM, Mamelak AN, Adolphs R, and Rutishauser U (2019). Single-Neuron Correlates of Error Monitoring and Post-Error Adjustments in Human Medial Frontal Cortex. Neuron 101, 165–177 e165. [PubMed: 30528064]

115. Wittmann BC, Schott BH, Guderian S, Frey JU, Heinze HJ, and Duzel E (2005). Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. Neuron 45, 459–467. [PubMed: 15694331]

116. Ljungberg T, Apicella P, and Schultz W (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. J Neurophysiol 67, 145–163. [PubMed: 1552316]

117. Clausen B, Schachtman TR, Mark LT, Reinholdt M, and Christoffersen GR (2011). Impairments of exploration and memory after systemic or prelimbic D1-receptor antagonism in rats. Behav Brain Res 223, 241–254. [PubMed: 21497169]

118. Rosen ZB, Cheung S, and Siegelbaum SA (2015). Midbrain dopamine neurons bidirectionally regulate CA3-CA1 synaptic drive. Nat Neurosci 18, 1763–1771. [PubMed: 26523642]

119. Medrano P, Nyhus E, Smolen A, Curran T, and Ross RS (2017). Individual differences in EEG correlates of recognition memory due to DAT polymorphisms. Brain and Behavior 7

120. Kishida KT, Saez I, Lohrenz T, Witcher MR, Laxton AW, Tatter SB, White JP, Ellis TL, Phillips PEM, and Montague PR (2016). Subsecond dopamine fluctuations in human striatum encode superposed error signals about actual and counterfactual reward. P Natl Acad Sci USA 113, 200–205.

### Box 1.

## Clinical procedures are a gold mine for cognitive neuroscience at the single-neuron level.

In a limited number of circumstances, the activity of single neurons in awake behaving humans performing cognitive tasks can be recorded [109]. There are three separate clinical scenarios where such work has been performed. First, patients with drug-resistant focal epilepsy undergoing monitoring with depth electrodes [109, 110]. Recordings are performed during the 1–3 week-long stay of such patients in an epilepsy monitoring unit [111] (see Figure I). Second, patients undergoing awake brain surgery for implantation of a deep brain stimulation (DBS) device for treatment of the symptoms of movements disorders [23, 24] or psychiatric indications such as depression or OCD [112, 113]. Third, patients participating in brain machine interface trials with invasive electrodes, such as the Utah array [100].

Each of these approaches provides access to a different set of brain areas with different constraints. The strength of recording in epilepsy patients is the ability to perform experiments in the relative comfort of the hospital room, performing experiments for several days, execution of relatively complex behavioral manipulations, and simultaneous recording of neurons from different brain areas. Limitations include restriction in accessible brain areas due to microwires exiting at the tip, inability to move electrodes, and caveats posed by the underlying seizure disorder. The strength of recording intra-operatively is ability to move the electrode to search for neurons and the ability to record anywhere along the track, which often includes areas not accessible in other settings such as the basal ganglia and striatum. Challenges of this approach include limited experiment time, impaired behavior due to after-effects of anesthesia and inability to perform complex behavioral procedures. The strength of Utah-array recordings is the ability to record large numbers of neurons from a small 4×4 mm patch of cortex for long periods of time during complex behavior. Limitations include inability to access areas not on the cortical surface and inability to move electrodes.

Notable, across the approaches described above many cortical and subcortical brain areas that are of interest to models of memory can be accessed, including hippocampus, amygdala, DA neurons in the SN, PPC, and the frontal lobe. Indeed, single-neuron data that informs our understanding of human memory has already been obtained from all three clinical scenarios described (as summarized in this review), illustrating the power of utilizing several clinical scenarios to investigate the same scientific question.

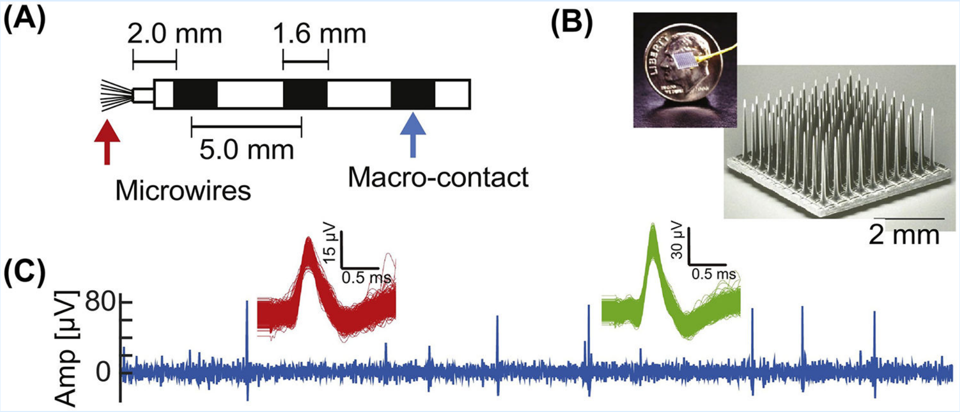

**Figure I (place inside Box 1): Electrodes and recordings of single-neuron recordings in humans.**

(A) Hybrid Depth electrode that is frequently used for recordings in epilepsy patients. Adapted from [114]. (B) Utah array. (C) Example raw recording (0.3–3kHz bandpass filtered) and two isolated clusters on a microwire of the kind shown in (A). Adapted from [111].

## Box 2.

### The role of dopamine in declarative memory.

Does dopamine have a role in the formation of hippocampal-dependent memories? Whilecomparatively little studied in comparison to its role in reward, substantial evidence suggests that it does. First, the extent of DA neuron activity during viewing of novel stimuli correlates with the likelihood that the stimuli will later be remembered. This was originally shown using BOLD-fMRI with ROIs located in areas that contain, among others, DA neurons [115]. But since the BOLD signal is not cell-type selective, it is possible that the changes in BOLD signal observed were not due to changes in DA neuron activity. However, direct recordings from human DA neurons in the SN reveal that DA neurons increase their firing rate in response to novel stimuli and that the degree of this activation is indicative of whether the stimulus will later be remembered or forgotten [24]. Similarly, in macaques, DA neurons are activated by novel stimuli [116]. However, to our knowledge, DA neuron activation and its relationship to behavior has not been tested in a hippocampal-dependent memory tasks in macaques. Second, blocking or otherwise interfering with DA receptors in the hippocampus impairs declarative memory encoding [117, 118] and modulates synaptic plasticity as assessed by LTP/LTD [3]. Thirdly, genetic studies in humans show that polymorphisms in dopamine-related genes explain variance in declarative memory ability and/or results in differential activation measured by BOLD fMRI or EEG [119]. Together, this data suggests that dopamine release is critical for encoding new declarative memories. What remains poorly understood, however, is what activates dopamine neurons in a novelty-dependent manner. How this occurs is what the hippocampus-VTA/SN loop model is concerned with (see main text). A critical experiment that remains to be done is to directly measure levels of dopamine in the human hippocampus at a fast time scale to assess how transient the changes in dopamine are after exposure to novel stimuli and whether the increase in dopamine is selectively target anatomically or to different cell types. Among the potential approaches to pursue this question are fast micro dialysis and cyclic voltammetry, both of which have been utilized in humans for otherpurposes [110, 120].

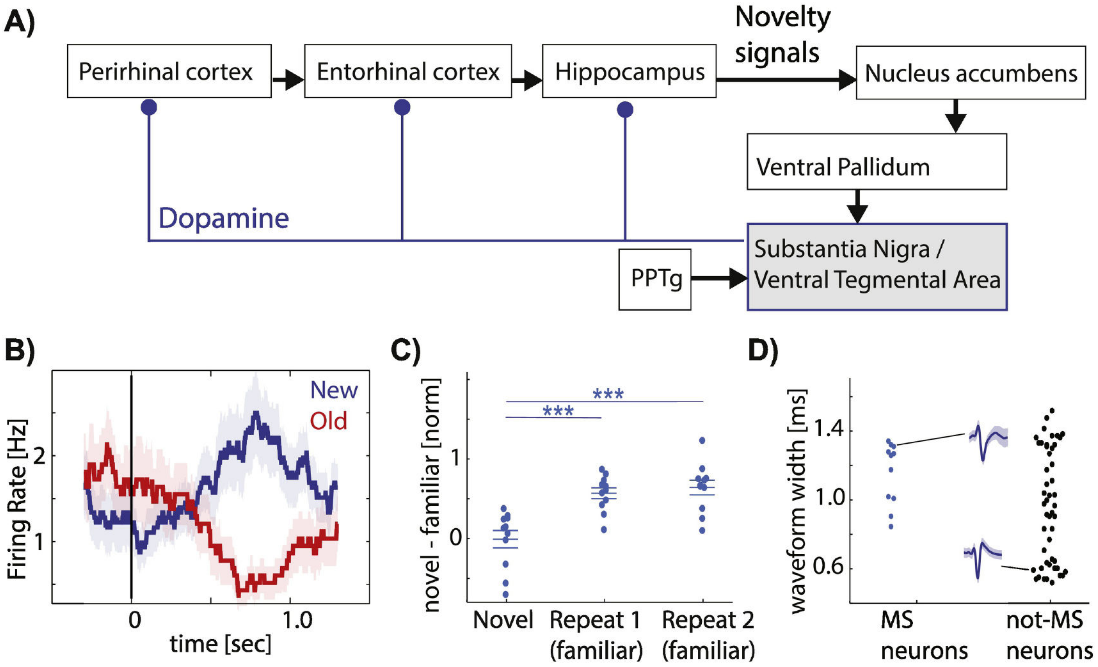

**Figure 1: The Lisman Hippocampus VTA/SN loop model and novelty signaling human DA neurons.**

(A) Schematic of interactions and flow of novelty signals within the hippocampus-VTA/SN loop. Adopted from [2, 3, 108]. (B-D) Novelty signaling human DA neurons. (B) Example neuron that increases its firing when a stimulus is novel (blue) and decreases when the same stimulus is shown again (red). (C) Population summary. Novelty-sensitive DA neurons change their firing rate between the first (left) and second (middle) time the same image is seen in a continuous recognition memory task. Each dot is one neuron. (D) Analysis of extracellular waveforms of neurons recorded in the human SN indicates a population of wide-and narrow waveform neurons, which are putatively dopaminergic and GABAergic, respectively. Note that the novelty-signaling neurons (blue) had wide waveforms. (B-D) adjusted from [24]. Abbreviations: SN – substantia nigra, VTA – ventral tegmental area, PPTg - pedunculopontine tegmental nucleus.

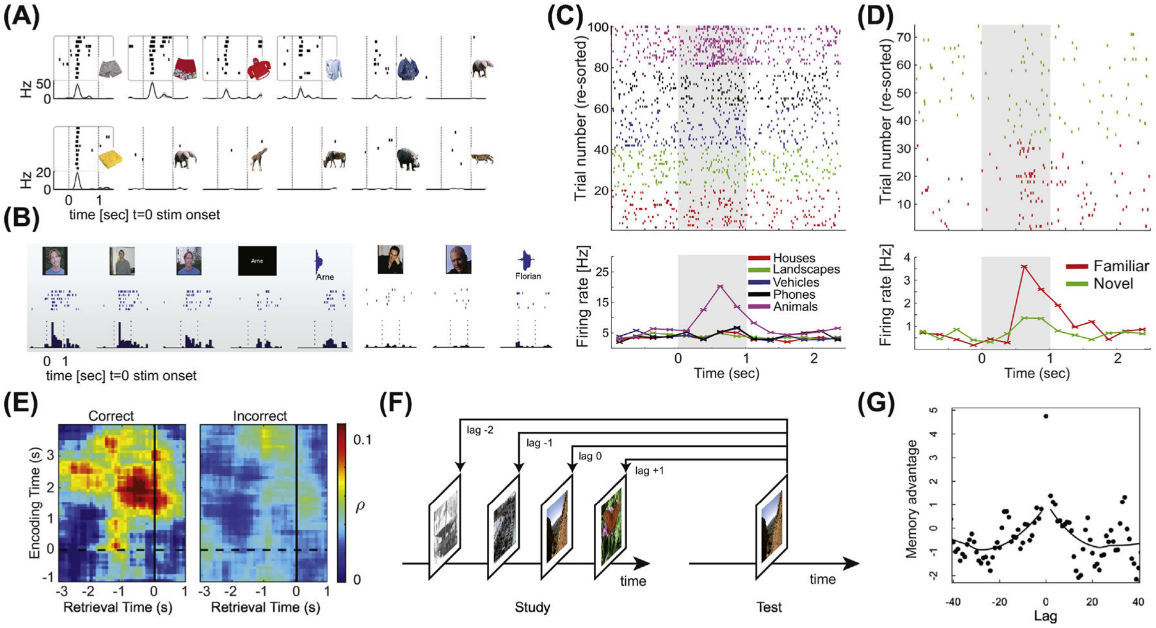

**Figure 2: Sparse and selective coding of declarative memory content and reinstatement by single neurons.**

(A-C) Example of visually selective neurons. (A) Two visually selective neurons, one responding to many different images showing clothes (top, from hippocampus) and one only responding to a single image of a food item (bottom, from amygdala). Adapted from [43]. (B) Example of a highly invariant multimodal concept neuron that responds to images and written and spoken name of an experimenter, but not many other images (only examples are shown). Adapted from [44]. (C) Visually selective category neuron. Trials are ordered by visual category from which the images are chosen (all images shown are different). Stimulus on/offset is shown with dashed lines or a grey box (C). (D) Example memory-selective neuron. Note that the images shown during novel (green) and familiar (red) trials are the same Stimulus on/offset is shown with a grey box. Adapted from [57]. (E) Contextual reinstatement during free recall by middle temporal gyrus neurons. Adapted from [81]. (F-G) Contextual reinstatement by VS neurons during recognition memory. Adapted from [77].

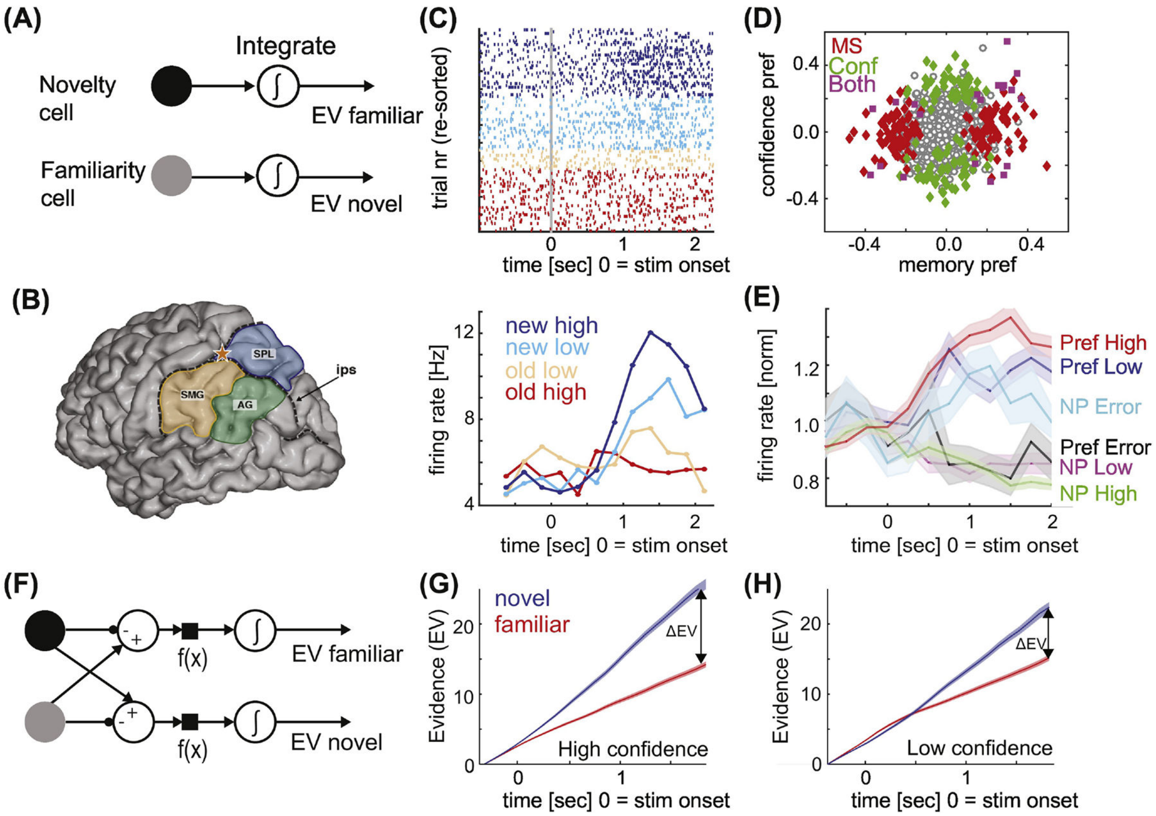

**Figure 3: Mnemonic evidence accumulation model and memory-choice cells.**

(A) Mnemonic evidence accumulation as a race process, with the choice made as the integrator (EV) that first reaches a preset threshold. (B) Anatomy of the PPC. Star marks the recording location for the data shown in (C-E). Adapted from [99]. (C) Example PPC neuron that increases its firing rate for a “new” decision in a graded manner modulated by confidence. Top shows the response of this neuron in individual trials, bottom the average response of this neuron. (D) Population summary of memory-sensitive neurons in the PPC reveal that some neurons signal the confidence of the decision (green), whereas others only signal the new vs. old decision (red). (E) Group PSTH of memory-choice cells in PPC, grouped according to their preferred stimulus (new or old). The preference is defined according to ground truth. Note that during errors, neurons increase their firing rate for their non-preferred stimulus, indicating that they signal choices regardless of whether they are correct or not. Adapted from [98]. (F) The balance of evidence model. Note that this model integrates the difference (new-old or old-new), which is not the same due to rectification. (G-H) Cumulative firing rate of MS neurons, shown separately for high (G) and low (H) confidence trials. The balance of evidence ΔEV scales as a function of confidence. Adapted from [58].

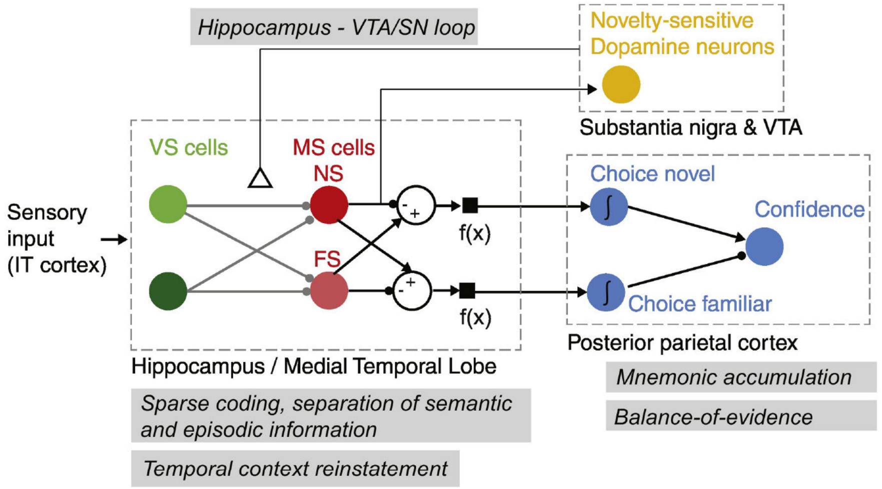

**Figure 4 (Key Figure): Summary of functional cell types and their putative interactions during recognition memory encoding and retrieval.**

Summarized are four cell types (filled colored circles): visually selective (VS) cells, memory selective (MS) cells, novelty-sensitive dopamine neurons, and choice neurons. There are two types of MS cells: Novelty and familiarity selective (NS and FS). Arrows indicate direction of information flow, but do not indicate monosynaptic connections. Anatomical areas are indicated in dashed boxes. The theoretical concepts/models (gray boxes) discussed are: i) the Hippocampus-VTA/SN loop model, which proposes that hippocampal novelty signals excite dopamine neurons in the VTA/SN, which in turn release dopamine in the hippocampus, which leads to long-lasting plasticity. ii) the sparse coding memory model, which suggests that distinct neurons encode semantic and episodic aspects of declarative memories in a sparse but distributed manner. iii) the temporal context model, which suggests that the neural state present at encoding is re-instated at retrieval. iv) mnemonic accumulation, which suggests that neurons exist that integrate memory signals to make choices. v) the balance-of-evidence model, which describes how the difference between two mnemonic integrators can be used to make metacognitive confidence judgments about memories. Jointly, the body of single-neuron experiments discussed provides evidence for key aspects predicted by these models, including novelty cells that exhibit rapid single-trial learning, representations of memory strength that predict subjective confidence judgments, sparse coding of semantic memories, reinstatement of neural state by VS cells during retrieval, and representations of memory-based choices that are putative mnemonic accumulators.

## Table 1.

Models of memory, their predictions and related human intracranial studies.

| Model class | Key predictions | Related studies |
|---|---|---|
| Hippocampus VTA/SN loop model | Dopamine neurons respond to novel stimuli, this novelty response is indicative of memory formation success, and exhibits single-trial learning. | [24, 27] |
| Sparse and separate encoding of semantic and episodic aspects of declarative memories | Highly visually selective cells (response sparsity), separate neurons encoding whether a stimulus has been seen before or not and other episodic aspects, | [35–39, 58] |
| Temporal context model, contextual reinstatement | Jump-back in time during retrieval, slowly drifting neural state, recollection/high confidence recognition predicted by reinstatement | [77, 78, 81] [79, 80] |
| Mnemonic evidence accumulation | Separate accumulators for “new” and “old” decisions. Accumulator output signal is modulated by memory strength. Accumulators predict choices rather than ground truth. | [97, 98] |
| Balance-of-evidence for metacognitive/ confidence judgments | Magnitude of difference between the separate integrators for “new” and “old” choices is proportional to confidence, whereas the sign of the difference is indicative of the choice. Both decisions can be made at the same point of time. | [58] |