Towards a Unified Consideration of Memory and Imagination as Cognitive Constructs

Benjamin Midler
Department of Psychology, Stanford University, Building 420, Stanford, CA 94305, USA

While memory and imagination exist as rhetorically distinct cognitive phenomena, ongoing bodies of research increasingly converge around a unified architecture that provides a common basis for memory and imagination. Specifically, recent research leveraging photon microscopy and optogenetics have demonstrated the sufficiency of neural circuits—or ensembles—for the construction of perceptual states. When merged with the robust bodies of work investigating the engram circuit as the brain’s mnemonic store, and cortico-temporal networks as the mechanism for reactivating and reinstating stored representations, the result is a unified conceptual understanding of how neural imagery and simulations are constructed. This constitutes an updating and integration of previous theories concerning the indexing, recall, and reinstatement of stored representations, and their remixing and recombination as imagined simulations. The present goal is to review the respective literature and proffer the base theoretical framework upon which cognitive imagery is built.

Abbreviations:  PFC – Prefrontal Cortex; MTL – Medial Temporal Lobe; LTP – Long Term Potentiation; LTD – Long Term Depression; DMN – Default Mode Network; fMRI – functional Magnetic Resonance Imaging; SCT – Scene Construction Theory; FM – Functional Module

Keywords: Memory; Imagination; Engram Circuit; Optogenetics; Hippocampus; Prefrontal Cortex

Introduction

Rhetorically, imagination and memory exist as distinct entities. At times, they are even at odds with one another, as memory is the recollection of what has passed while imagination is the conception of what might be, or even of something that will never happen. Despite such differences, an expanding understanding of the neurobiological basis for memory and imagination suggests the two are far more mechanistically similar than what might be expected (see Mullally and Maguire, 2013 for review).

Recent advances in functional imaging, photon microscopy, and optogenetics have yielded unprecedented opportunities to understand how the brain stores and retrieves memories, and how imagination may be recombinations of such stored representations.

These empirical advances build upon the theoretical foundation of connectionism—the hypothesis that cognitive phenomena can be modeled and studied using artificial neural networks (Rumelhart and McClelland, 1986; McClelland et al., 1995)—and, when combined with the developing narrative of the neural ensemble being the core unit of computation (Yuste, 2015), a new understanding emerges: a unification of memory and imagination as cognitive constructs.

This theory draws on robust bodies of literature to establish the subjective and empirical similarity between memory and imagination, and then proffers an explanation for how the engram circuits believed to compose memory (see Josselyn and Tonegawa, 2020 for review) can likewise be recombined to compose imagination.
The conclusion of this work is therefore predicated by the fact that different forms of cognitive imagery are merely permutations of the same neural architecture. Therefore, continuing to treat various cognitions as more distinct than related, while well-rooted in empirical literature and neurobiological tradition, perhaps fails to capture the whole picture. In the analogous example of spatial navigation and memory, while robust bodies of work characterized them as distinct functions (O’Keefe and Dostrovsky, 1971; O’Keefe and Nadel, 1978; Morris et al., 1982; Cohen and Eichenbaum, 1991), emphasizing their extensive neurobiological similarity yielded an improved conceptual understanding (Eichenbaum and Cohen, 2014). Therefore, while treating the rhetorically dissimilar memory and imagination as mechanistic siblings rather than distant cousins may seem radical, the present goal is to demonstrate that doing so is quite a small step to take, and that it may likewise yield conceptual improvements. As research demonstrating the common neurobiological basis of memory and imagination has already been produced (see Klein et al., 2002; Barron et al., 2013 for examples), the need to develop a theoretical framework to explain and contextualize this observation is acute. Furthermore, while theoretical propositions that novel simulations are derived from existing repertoires of knowledge already exist, for instance those of Stein (1953), Koestler (1964, p. 751), and Ingvar (1985), these theories are constrained by their lack of access to the last half century of technological and scientific advancement. As such, they are ill equipped to explain and reconcile recent research demonstrating intriguing mechanistic similarities between memory and imagination. The present theory seeks to do just that.

This hypothetical exploration begins with a review of the literature establishing the known architecture of memory—the engram circuit—along with relevant characteristics and reinstatement. It will then explore the neuroanatomical evidence that provides a basis for the structural linkage between memory and imagination before concluding with a theoretical—but empirically and evolutionarily sound—basis for imagination constituting a redeployment of memory circuits. The proposed theory is that there is a single, unified mechanism for the neural assembly of remembered and imagined imagery, and that this mechanism is heavily implicated in perception.

**Structure of Memory**

**Foundations**

The contemporary basis for memory has its roots in Hebb (1949), as his seminal work established the principle of synaptic plasticity. Plasticity refers to the dynamic ways in which neural nodes can adapt their connections in response to activity. In a simple, two-node system, such changes are not capable of significant representational nuance. When combined across a sprawling network such as a brain, however, the result is an organ capable of representing a number of states orders of magnitude greater than its 86 billion neurons (Azevedo et al., 2009).

![Figure 1](image-url)  
*Figure 1* Engram circuits are formed by learning and are re-activated to facilitate memory. (A) Learning is the process of forming engram circuits (multi-colored networks, circles are neurons) that structurally represent memory episodes and components. When an episode is experienced, sensory-evoked neuronal activity induces synaptic plasticity which re-shapes circuit dynamics to represent the constituent mnemonic components of the episode. Together, those ensemble components constitute the engram circuit. (B) After learning, the synaptic dynamics induced by the learning experience constitute the neurobiological basis of the memory and the engram circuit. (C) When the memory is recalled, activity is routed through the engram circuit with its synaptic dynamics guiding the reassembly of the constituent ensemble components of the memory into the sensory imagery of the episode.

Such states—the patterns of activity favored by a particular set of synaptic connections—take the form of dynamically efficient ‘wells’ known as attractors. First described by Hopfield (1982), these attractors are structures whose existence is implicitly based on the assortment of connections between the
neurons in the system. Hopfield (1982) speculated that these attractors are capable of storing such computational phenomena as memories and mathematical solutions, as activity that enters the system will be guided by the weightings of each connection until it finds an energetically stable end point.

In the brain, while the neural structures that facilitate memory are the subject of ongoing debate (see Langille and Gallistel 2020 for review), a preponderance of the evidence holds that memory is composed of attractors, and, by extension, neuronal associations. These structural and anatomical linkages between neurons constitute the corresponding linkages between the episodic and semantic details that form memories. Learning, as the process of forming such associations, is therefore conceptually tied to memory (Figure 1a). It refers to the mechanisms that guide inputs, such as sensory experiences, to a form implicitly constituted by the network of associations between neurons. When such associations are formed conjunctively across a network of neurons to represent a memory, the result is an engram circuit (Poo et al., 2016; Figure 1b), as the engram is a physical structure, composed of synaptic dynamics, that provides the basis for experiencing a memory. When activity is precisely routed through the circuit—guided by the synaptic dynamics of the circuit—the engram is recalled (Figure 1c).

A necessary consequence of this model is constant neural adaptation, which is facilitated by protein synthesis. When activity passes through a synapse and triggers an action potential in the post-synaptic neuron, an intra-cellular cascade is initiated that up-regulates the transport of AMPA receptors to the synaptic membrane (Fonseca et al., 2006). Greater density of AMPA receptors, which are excitatory glutamate receptors, makes the neuron more sensitive to pre-synaptic action potentials as there are now more synaptic targets for a given quantal of transmitter release. Consequently, a more pronounced post-synaptic depolarization is precipitated by the same pre-synaptic transmitter release. This upward tuning of the synaptic weight is known as long term potentiation (LTP) (Bliss and Collingridge, 1993) whereas the inverse process—a downward tuning—is long term depression (LTD) (Massey and Bashir, 2007). Therefore, the subjective qualities of memories are embodied by the structural composition of their attractors and are thus malleable and subject to perturbation.

**Experimental Evidence of Plasticity and its Necessity for Learning and Memory**

Proof of this associative model of memory and the plastic basis for engram circuits extends far beyond the theoretical. For instance, the notion that learning changes synapses has been confirmed by experimental analogues for learning, such as optogenetics (Whitlock et al., 2006). Nabavi et al. (2014) used variable frequencies of optically controlled cellular activation to both construct and decompose memories through the plastic principles of LTP and LTD. They demonstrated that memories represented by an engram circuit formed by LTP can subsequently be reversed by experimentally induced LTD. These results were confirmed by Kim and Cho (2017) and Hayashi-Takagi et al. (2015) who showed that a learned behavior is weakened when synaptic associations are reversed selectively at those engaged in the initial learning—even down to the level of specific dendritic spines. Furthermore, LTD has proven capable of removing dendritic spines (Sanders et al., 2012), which results in the degradation of the engram circuit as such structural changes destroy the synapses that compose the circuit. This is in addition to LTD triggering the removal of AMPA receptors from dendritic spines (Ma et al., 2016), which may be described as a downward tuning of the synaptic weight rather than wholesale removal. That said, such tuning remains capable of impeding circuit activity enough to disrupt expression of the memory or behavior in question (Abdou et al., 2018). Such network and neuronal specificity strengthens the notion that memories are stored in precisely defined circuits characterized by a particular assortment of cells and synaptic weights. The alteration of these dynamics—either experimentally or through naturally occurring processes—results in a degradation of engram fidelity and accuracy of the memory’s subjective qualities.

In conjunction with this evidence that memories are implicitly constituted in plastic engram circuits, further experimentation has disentangled the roles of synaptic adaptation and activity in forming and recalling engrams. For
instance, work by Ryan and colleagues (2015) combined neuronal stimulation with anisomycin, which suppresses synaptic changes, to demonstrate that activation of the engram circuit alone is sufficient to evoke a memory, and that synaptic plasticity, while necessary for the formation of the memory, is not involved in the recall of that memory. Complementary work with optogenetics demonstrated that memories can be evoked by stimulating their respective engram circuits. Specifically, when the hippocampal and retrosplenial neurons recruited during a fearful experience are selectively stimulated, behavioral evidence of recalling that fearful memory, such as freezing, is produced (Liu et al., 2012; Cowansage et al., 2014; Lacagnina et al., 2019). Likewise, when activity in these cells is suppressed, behavior associated with fear memory recall is less pronounced (Denny et al., 2014; Tanaka et al., 2014), demonstrating that, while structural connectivity defines the engram and its associated memory, activity is necessary for recalling the memory.

Another component of memory relevant to the present discussion is functional overlap. The engram model of memory holds that the responsiveness of each cell is shaped by the cells and processes that feed into it or are upstream of the cell in question. This explains why cells that share a high degree of anatomical similarity can come to represent such different phenomena as a bar of light in the occipital lobe (Hubel and Wiesel, 1959), a tone in the temporal lobe (Recanzone et al., 2000), or the memory of either stimulus. Therefore, when it comes to a certain concept or episodic component stored in memory, to the degree that component is shared between memories, the respective engram circuits likewise share the cells that together constitute that component (Langille and Gallistel, 2020). Thus, while the engram circuit is that which corresponds to a particular concept or representation, multiple engram circuits may share particular engram cells—a fact that may underpin semantic or conceptual similarities between multiple stored representations (Figure 2).

The unified theory of cognitive construction rests on two pillars: that engram circuits are the anatomical basis for representational cognitions, such as remembered scenes, and that such representational circuits can be decomposed into their respective cellular and ensemble components and shared between engrams to assemble new representations in imagination. For instance, the network representing a specific concept will be recruited into engram circuits involving that concept irrespective of context and other extraneous factors. As long as a particular memory involves that conception, its representational engram circuit will be recruited (Josselyn and Tonegawa, 2020). This explains why, after recalling a memory with a specific element, it is easier to jump to another memory that shares the element than it would be from a neutral starting point, as the activation of the shared element’s engram circuit provides a basis for pattern-completing the attractor circuit for the new memory, manifesting as semantic priming (Mckoon and Ratcliff, 1979). This piecemeal view of memory recall, which might be compared to constructing a model out of building blocks, explains why describing memory and imagination as wholly separate cognitive phenomena fails to capture the brain’s connectionist nuance, and why a unified understanding of cognitive construction is more
apt. While memory is the re-assembly of the same model, imagination is the construction of a novel model, but it is built from building blocks borrowed from multiple different sets stored throughout memory.

Reinstatement and Imagery

Reinstatement as the Basis for Memory Retrieval

Thus far, a basis for the formation and storage of memories has been presented. The last component of memory to be reviewed is its retrieval. Cortical reinstatement has emerged as the leading theory of how the subjective experience of a memory can be drawn from its engram circuit and experienced as imagery. While there has recently been a swell of rhetorical and experimental support for this theory, (see Danker and Anderson, 2010 for review) it is by no means a new one. James (1890) first proposed that to remember is to reinstate the sensory and motor components of the original event to their respective brain components and has more recently been found consistent with both the encoding specificity (Tulving and Thompson, 1973) and transfer appropriate processing (Morris et al., 1977) theories of memory retrieval. As both theories hold that mechanistic details of memory encoding—such as context and cognitive process respectively—play an important role in recall, the intersection of these theories and James’ is that the neural route a memory takes into the brain is influenced by endogenous factors such as cognitive strategy and exogenous factors such as environment, which, in turn, influences the route a memory takes to be recapitulated in sensory cortex.

In computational work, retroactivation and connectionism support cortical reinstatement (Rumelhart and McClelland, 1986; Damasio 1989; McClelland et al., 1995) as they stipulate neural representations are recorded into memory in the same circuits used to process perception. This understanding has become a core theoretical pillar of models of cortical-hippocampal interaction (Alvarez and Squire, 1994; McClelland et al., 1995; Squire and Alvarez, 1995; Rolls, 2000; Shastri, 2002; Moscovitch et al., 2005), and is consistent with the experimental understanding of cortical reinstatement.

The bulk of this evidence converges on sensory and prefrontal cortices during reinstatement (Nyberg et al., 2000; Wheeler et al., 2000; Vaidya et al., 2002; Wheeler and Buckner, 2003; Gottfried et al., 2004; Wheeler and Buckner, 2004; Wheeler et al., 2005; Tayler et al., 2013; Trouche et al., 2013; Xie et al., 2014; DeNardo et al., 2019), demonstrating that, during the reinstatement of a modality or task specific memory, activity is localized to the brain components implicated in that modality or task (Bosch et al., 2014). Meanwhile, in sub-cortical brain regions, neuronal tagging has revealed significant functional overlap between experience and recall using a diverse range of memory tasks—such as multi-modality fear conditioning and novel object exploration—in the dorsal hippocampus (Tayler et al., 2013; Ramirez et al., 2013; Tanaka et al., 2014; Zelikowsky et al., 2014; Denny et al., 2014; Nakazawa et al., 2016; Lacagnina et al., 2019) and amygdala (Reijmers et al., 2007; Tayler et al., 2013; Trouche et al., 2013; Nonaka et al., 2014; Zelikowsky et al., 2014; Kitamura et al., 2017; Figure 3).

Figure 3 An overlapping neural basis for perception and memory. Both neuropsychological and neurobiological research has demonstrated that the neurons activated (red circles) by the memory of a sensory experience are a subset of those implicated in the original experience. Optogenetics experiments have demonstrated the sufficiency and necessity of these memory engrams for the experiencing of the memory. This line of research constitutes a link between neural ensembles and the storage and recall of memory.

These findings are consistent with the connectionist view of reinstatement, as storing the engram circuit in the brain region responsible for original perception or computation would
logically guide reinstatement activity to that same region—a theory supported by studies showing that disrupting sensory cortex activity similarly impairs sensory perception and sensory memory (Kosslyn et al., 1999; see Harris et al., 2001 for review). Additionally, studies of visual system reinstatement demonstrate that the observed activity produces effects at multiple different mechanistic levels of brain circuits, spanning molecules to the inferotemporal cortex (Sakai and Miyashita, 1991; Rösler et al., 1995; Gratton et al., 1997; Khader et al., 2005; Polyn et al., 2005). This finding is especially significant for cortical reinstatement’s relevance to both connectionism and engram circuits, as such molecular synthesis is the biological basis for synaptic tuning in engram circuits, and analogous to the inter-node weight dynamics that constitute the foundation of connectionist models.

**Imagery as the Subjective Form of Memory and Imagination**

While the recall of an episodic memory can be neurobiologically described as the activation of the relevant engram circuit, that memory manifests subjectively as imagery, or the experience of “seeing with the mind’s eye” (Kosslyn et al., 2001). Generally, imagery refers to the brain’s ability to elicit a sensory experience—or at least something reminiscent of that experience—without the sensory input necessary to produce that experience. Imagery, however, is an internally generated sensory impression, one that, in the context of memory recall, is the subjective manifestation of that recall. Imagery may even suffice in producing a behavioral or physiological response, such as an elevated pulse in response to imagery associated with a frightening episode. Despite its name, imagery is not necessarily confined to visual impressions (Chatterjee and Southwood, 1995; Kosslyn et al., 2001), as it has also been observed in modalities such as audition (Zatorre and Halpern, 1993; Kosslyn et al., 2001).

Imagery, though, is a nonspecific term. It is a subjective cognitive phenomenon and is therefore compositionally agnostic: it can be used to represent both remembered episodes and imagined ones. This is a departure from previous descriptions of imagery, which were confined to imaginative processes (Thomas, 1999). Imagination, though, is not a wholly distinct process. It is a constructive one that facilitates the generation of novel imagery through the remixing of components stored in memory (Agnati et al., 2013). Therefore, while the subjective experience of imagery may be similar between that for memory and imagination, there remains a crucial difference in the origin and composition of the imagery.

What’s more, being capable of constructing novel scenarios, imagination has the ability to both augment a neural experience or task, such as picturing how a piece of furniture will look in another room, and simulate an entirely new one, such as daydreaming. Imagination is therefore a clear example of the phenomena of cognitive construction, as it is quite literally the assembly of a cognitive representation from the individual representational building blocks that manifest as sub-gram—or ensemble—circuits.

**Evidence that Imagination is a Constructive Process**

Insofar as episodic imagery is composed of activity reinstatement in the neural structures implicated in the original experience, imagination can be described as the construction of novel imagery from stored perceptual and conceptual information—a process that can be described as neural tinkering (Agnati et al., 2013).

This fact has proven especially apparent in the simulation function of imagination. Simulation, a theory that encapsulates as diverse concepts as prospection, theory of mind, and self-projection, has been localized to the medial temporal, parietal, and frontal lobes (Buckner and Carrol, 2007). Raichle et al. (2001) assembled these circuits into the default mode network (DMN) and observed an intriguingly elevated metabolic rate at a baseline level of cognitive activity—no sensory input but conscious and without a cognitive task or goal. The DMN is therefore believed to be involved with stimulus independent cognitions such as present and past self-conception, mind-wandering, and theory of mind (Gusnard and Raichle, 2001; Mason et al., 2007; Buckner et al., 2008).
When it comes to the construction of imagined imagery, however, simulation has been found to functionally recruit circuits that, at least in part, overlap with those implicated in experience—reticent of cortical reinstatement and consistent with connectionist philosophies (Jabbi et al., 2008; Rizzolatti and Fabbri-Destro, 2008). Specifically, functional magnetic resonance imaging (fMRI) has been used to establish a high degree of functional overlap in the premotor and parietal cortices and supplementary motor area between performing a finger movement and imagining the same movement (Gerardin, 2000). Addis et al. (2007) also observed an overlap in an extensive network that includes the prefrontal, retrosplenial, temporal, and prefrontal cortices when subjects recalled and imagined detailed episodes.

Building upon these findings, theoretical work has proposed that neural simulation makes use of low-level circuits—such as engrams—for maximum flexibility in sharing neural conceptions between simulations (Gallese, 2008; Hurley, 2008; Deheane, 2009; Anderson, 2010). The convergence of such distributed circuits into a single representation is consistent with the theorized existence of reciprocal connections that span the brain, known as neural re-entry (Edelman and Mountcastle, 1978; Edelman, 1993; Edelman and Gally, 2013), and perhaps is even a manifestation of the recurrent connections and functional reverberations originally described by Rafael Lorente de Nó (1933, 1949).

Empirical and functional experiments have established that overlapping patterns of activity constitute both remembered episodes and novel constructed imagined episodes (Hassabis, Kumaran, and Maguire, 2007; Szpunar et al., 2007; Botzung et al., 2008; Addis et al., 2009; Spreng et al., 2009). While these experiments rely heavily on the use of fMRI, a technology that struggles to resolve circuit dynamics with sufficient resolution, they are consistent with other experiments that leveraged techniques such as photon microscopy and optogenetics. Such research has demonstrated—at a highly specific circuit level—the sufficiency of modality-specific circuits in generating perceptual states (Carrillo-Reid et al., 2019; Figure 4), and of engram activity in recalling a memory (Liu et al., 2012).

It is known that when circuits representing perceptual qualities are experimentally activated, they evoke the intended sensory perception (see Carrillo-Reid and Yuste, 2020 for review). Likewise, it is known that imagination and memory occupy overlapping brain regions that correspond to those responsible for the perception in question (Dijkstra et al., 2019). Combining these two established lines of research constitutes support for connectionist theories and underpins the core philosophy of cognitive construction. Circuits, or ensembles, are therefore established as a core component of neural functioning, not only with respect to memory in the form of engram circuits but also with perception and experience. Therefore, two imagery-based aspects of cognition, memory, and imagination, share a common neurological basis. The unified theory of cognitive construction proffers that, for the construction of cognitive representations and imagery, the neural circuit is the common building block used throughout.
Neuroanatomy Links MTL, Memory, and Imagination as Constructive Processes

Neuropsychological Evidence

In the biophysical view of the brain, function and structure are indelibly intertwined. The first evidence that memory and imagination are functionally related, therefore, comes from studies that localize each phenomenon. Studies on human lesion patients suggest a nuanced relationship between damage to anatomical structures in the medial temporal lobe (MTL) and disruption to memory and imagination.

Patient KC, for example, suffered from widespread brain damage—including the MTL—and was observed to struggle imagining future events (Tulving, 1985; Rosenbaum et al., 2005). Patient DB, who likewise displayed extensive episodic memory impairments, also struggled to imagine the future, reporting implausible or nonsensical scenarios when asked about future plans (Klein et al., 2002). These observations are well-replicated (Talland 1965; Korsakoff, 1996; Race et al., 2011), and establish that there is at least a correlational link between MTL-based memory impairments and impeded imagination—a link that spans multiple models of MTL damage (Williams et al., 1996; Driscoll and Sutherland, 2005; D’Argembeau et al., 2008; Schacter et al., 2008; Herold et al., 2013).

However, neither memory nor imagination are binary capabilities, as lesioning has produced nuanced degrees of impairment in both cognitive functions. For instance, Hassabis, Kumaran, Vann, and colleagues (2007) found that scenes imagined by amnesiac patients were less vivid, and that there was a greater degree of spacial fragmentation than for controls. These results were replicated by Mullally, Intraub, et al. (2012) with a separate group of patients. Adding further nuance are studies of amnesic patients with bilateral hippocampus damage who have demonstrated a preserved ability to construct imagined scenes. Patient P01, for instance, who lost 50% of volume of each hippocampus, was nonetheless able to construct imagined scenes (Mullally, Hassabis, et al., 2012). Subsequent fMRI analysis revealed that during imagined scene construction, there was a significant increase in activity in the remaining portion of the right hippocampi. This result suggests that while bilateral hippocampal connectivity is essential for the recall of episodic memory, a smaller portion is sufficient for the construction of basic imagined scenes.

Of further interest is the fact that studies of patients with bilateral hippocampus damage do not necessarily display a general deficit in their capacity for imagination (Hassabis, Kumaran, Vann, et al., 2007). For instance, while patients were able to imagine single concepts or items, they failed to knit multiple imagined concepts into a coherent scene. This inability remained even when the patients were presented with all the scene components and were simply tasked with combining them into a single mental representation (Hassabis, Kumaran, Vann, et al., 2007). This constitutes a particularly interesting caveat in the link between the MTL and imagination. It suggests the hippocampus is involved not as much in producing a representation for each scene component, as this ability was unimpeded by hippocampal lesioning, but in the conjunctive assembly of each component into a single, coherent scene. Considering that this is a necessary function of constructive episodic processes (Bartlett, 1932), it stands to reason that lesioning the hippocampus, thereby disrupting this conjunctive assembly process, similarly impedes memory and imagination.

Proposed Basis for Hippocampus-Guided Constructive Architectures

In addition to sharing the components for storing and recalling imagery, memory and imagination likewise use similar mechanisms to guide the assembly of those components. This is analogized by Tulving (2002), as he described episodic memory as a mental trip into the past, or retrospective self-projection. Likewise, imagination can be construed as prospective, or fictional self-projection. This mental time travel is suspected to heavily depend on the hippocampus (Buckner and Carroll 2007; Suddendorf and Corballis 2007; Szpunar et al., 2007).

A proposed explanation is predicated by the constructive episodic simulation hypothesis.
It holds that episodic memory is the process of constructing simulations from individual components, and that simulating the future is enabled by the flexible extraction of insights and elements from prior experiences (Schacter and Addis, 2007; see Schacter et al., 2012 for review). Therefore, as memory and imagination are similarly constructive processes, the same assembly mechanisms are used for both. Such a view is consistent with current understandings of the hippocampus’ role of associating semantic and episodic elements between experiences, known as the relational theory (Cohen and Eichenbaum, 1993; Konkel and Cohen, 2009), to contextual details about a scene (Ranganath, 2010), or even across temporal episodes (Rashid et al., 2016; Rao-Ruiz et al., 2019; Pignatelli et al., 2019). Relational theory, as the ability to flexibly relate mnemonic details across memory components, therefore refers to a key aspect of memory and imagination being unified cognitive constructs. Namely, that both involve the assembly of coherent mental simulations from memory components.

Another explanation of hippocampal involvement is scene construction theory (SCT), which proposes that the hippocampus is indelibly involved in atemporal scene construction and facilitates the assembly of such scenes from relevant circuits—or neural building blocks (Maguire and Mullally, 2013). SCT is therefore consistent with the distributed model of memory embodied by connectionism and cortical reinstatement and proffers an explanation for the hippocampal link between memory and imagination construction. The precise mechanism through which the hippocampus achieves this is known as hippocampal indexing theory (Teyler and DiScenna, 1986; Teyler and Rudy, 2007), which proposes that hippocampal neurons do not contain the long-term memory representation themselves, but instead contain traces to the constituent engram circuits that together reconstruct an episodic memory, thereby driving re-activation in cortical neurons corresponding to the neural representation in question (Yassa and Reagh, 2013; Figure 5).

Recent investigations have added nuance to our understanding of hippocampal function, as Kitamura et al. (2017) used a combination of optical and genetic tagging techniques to reveal that hippocampal engram circuits, rather than eroding over time, transition to a state of silence while cortical engrams do the opposite: maturing from silence to an active state. Further experimentation found that inhibiting the connections between hippocampal and cortical engrams blocked this maturation process, indicating that hippocampal engrams play an important role in the maturation of the cortical engram circuits that compose more stable, remote memories (see Josselyn and Tonegawa, 2020 for review). While this finding is an addendum to classic theories of hippocampal indexing, such as multiple trace and trace transformation, the notion that the hippocampus acts as a memory hub with multiple traces connecting disparate episodic and semantic memories throughout the brain remains consistent (Nadel and Moscovitch, 1997; Frankland and Bontempi, 2005; Sekeres et al., 2018). Thus, the hippocampus’ functional role in establishing and reinstating engrams in cortex is affirmed.
The theory of cognitive construction, therefore, not only unifies our empirical understanding of how cognitive imagery is assembled, but also provides a coherent explanation for disparate hippocampus functions; this suggests there is much conceptual benefit to abandoning previous beliefs in a strictly mnemonic hippocampus (Mullally and Maguire, 2013). Instead, hippocampal involvement in memory and imagination can be unified around the notion of cognitive construction.

Prefrontal Guidance of Episodic Imagery and Recombination

While the hippocampus has come to be almost synonymous with memory, carefully designed experimental paradigms have revealed that the prefrontal cortex (PFC) is likewise vital to memory (see Eichenbaum, 2017 for review)—albeit in specific functional contexts. Namely, the PFC has been found to introduce a top-down, goal-directed element to memory recall (Moscovitch, 1992; Dobbins et al., 2002; Preston and Eichenbaum, 2013), meaning that, during memory tasks that involve multiple competing traces (for instance, when new learning overlaps with old), the PFC is instrumental to guiding memory evocation to the trace that is congruent with current goals.

A-B, A-C memory tests present a classic example, as subjects first learn an association between two cues (A and B), and then a new association that overlaps with the original (A and C; Figure 6a). Patients with PFC lesions display significant impairment in learning the new association (A-C), as there is persistent interference from the first (A-B) (Shimamura et al., 1995; Figure 6b). This is consistent with further research indicating that the PFC is instrumental for preventing the intrusion of goal-irrelevant memories into tests of extinction learning in a fearful context (Giustino and Maren, 2015), and selective attention to specific cues (Chudasama et al., 2012). Taken together, this evidence suggests a key role for the PFC in biasing memory recall towards the specific traces that are most relevant for the current task and goal state (Figure 6c). This is consistent with rodent studies showing PFC lesions inhibit the ability to switch between memory strategies for different tasks (Ragozzino et al., 1999; Brown and Bowman, 2002; Ragozzino et al., 2003; Marquis et al., 2007; Rich and Shapiro, 2007; Rich and Shapiro, 2009).

While there is some lesion evidence that the PFC is likewise implicated in imagination (Berryhill et al., 2010; Irish and Piguet, 2013), there is far less direct anatomical support for the PFC being instrumental to imagination than there is for the hippocampus. With that said, given the extensive degree of anatomical (Jay et al., 1989; Jay and Witter, 1991; Dolleman-Van Der Weel and Witter, 1996; Vertes, 2001; Vertes, 2006; Hoover and Vertes, 2007; Vertes et al., 2007; see Eichenbaum, 2017 for review) and functional (Wang et al., 2015; Schmidt et al., 2019; see Preston and Eichenbaum, 2013 for review) connectivity between the PFC and hippocampus—particularly the PFC’s ability to “focus” hippocampal cell activity (Hok et al., 2013)—it is not inconceivable that, for imagination, the PFC exerts a similar goal-directed influence on the hippocampus as it does for memory (Barry et al., 2019). In other words, just as the PFC biases memory recall towards the hippocampal trace most congruent with current goals, it may likewise provide a top-down control.
signal to the hippocampus during the construction of an imagined simulation. Thus, as imagination is the recombination of disparate episodic details into novel, simulated imagery, the PFC potentially provides the goal-directed input to the hippocampus for the reinstatement of specific episodic components into the desired, goal-compatible imagined simulation (Figure 7).

**Figure 7** PFC provides a top-down signal to assemble disparate neural representations into novel imagined simulations. Just as the PFC guides activity to the goal-congruent trace in memory paradigms, it is likewise able to draw upon multiple disparate memory components (left) to coherently assemble imagined imagery to simulate a novel episode (right). Imagination is therefore constrained by episodic repertoire and the cortico-hippocampal network’s ability to conjunctively draw-upon those disparate representations.

### Imagination as an Extant Redeployment of Memory Substrates

**Theoretical Basis for Imagination Being an Exaptation of Memory**

The precise origin of the engram circuits’, hippocampus’, and PFC’s ability to be used both for memory and imagination remains open for exploration, but a compelling explanation is proffered by Anderson (2007, 2010), who introduced the concept of multiple redeployment to neuroscience. In broad terms, redeployment refers to the acquisition of new capabilities or functions while preserving the original. In this view, therefore, memory is the core function of engram circuits, and their co-opting into imagination constitutes an exaptation.

A survey of neurological literature yields many examples of similarly redeployed structures and functions. Broca’s area, for instance, has long been known to play an important role in numerous cognitive tasks—functions adapted from its constituent motor neurons vital for the act of speech production (Grodzinsky and Santi, 2008; Anderson, 2010). Additionally, mirror neurons, which activate in response to actions performed both by the organism itself and by others, constitute a redeployment of motor neurons as they are tasked with neurologically representing the actions of other organisms while preserving their ability to generate such actions themselves (Rizzolatti and Craighero, 2004; see Case et al., 2015 for review). Recent functional investigations have expanded our understanding of mirror neurons beyond the premotor and inferior parietal cortices (Rizzolatti et al., 2008; Ferrari et al., 2009), and are now suggesting that, as opposed to individual mirror neurons, humans have mirror neuron systems that exist in diverse and widespread brain regions (Gazzola and Keysers, 2008; Iacoboni, 2009). Such findings likewise expand the scope of mirror neurons and their potential to constitute simulator neural circuits, as Rizzolatti and colleagues (2009) have demonstrated the existence of mirror properties in cortical areas responsible for emotion and affect-related behavior. As such, mirror neurons constitute an example of how neural structures can adapt and expand their functional scope to include internally constructed representations in addition to their original functions.

How does this manifest in the redeployment of traditionally mnemonic circuits for imagination? Agnati et al. (2013) propose that there is a substantial degree of computational potential in the dynamic assembly and re-assembly of neural circuits known as functional modules (FMs) in the brain. It is this metanetwork’s potential that enables the extraordinary flexibility in cognitive function and construction observed in memory and imagination (Agnati et al., 2012). Therefore, just as a mirror neuron can be called upon to represent another’s motor action, FMs can be recruited as part of a highly adaptable and diverse network to represent
subjectively diverse imagery. As such, the mosaic of FM recruitment may be a manifestation of multiple selectivity, as each unique combination of modular activation and inhibition precipitates a different imagined simulation. The concept of multiple selectivity, which explains how the non-linearity of neural systems can yield extremely high-dimensional and rich representational nuance, has already been shown to facilitate representational diversity in hippocampal CA1 ensembles (Stefanini et al., 2020), as incorporating multiple selectivity into the decoding of navigational variables yielded higher accuracy than simply focusing on the monotonically selective neurons. Furthermore, the ability of neural systems to creatively recombine FMs into novel simulations may have developed as a result of evolution’s inherent ability to explore functional and mechanistic adaptations of physiology and biological structures, coined evolutionary tinkering by Jacob (1977). It stands to reason that a similar functional adaptation may link the hippocampus’ roles in memory and imagination.

**Functional Experimentation Links Memory and Imagination**

Experimental validation for memory circuit redeployment being the mechanism of imagination can be divided into two distinct tracts. The first centers on the mechanism for recalling and assembling the distributed representations that comprise a memory or imagined simulation, and the second around the engram circuit constituting the store of representational information—a store capable of being recombined and remixed to form novel simulations. For the first, multiple experiments employing a diverse range of paradigms ranging from future thinking (Addis et al., 2007; Hassabis, Kumar, and Maguire, 2007; Szpunar, 2010) to literary creativity (Liu et al., 2015), openness to novel experiences (Beaty, et al., 2017), musical improvisation (Pinho et al., 2015; Chen et al., 2020), and novel idea generation (Benedek et al., 2018) converge on a cortico-temporal network playing an instrumental role in guiding memory and imagination construction. This network, composed predominantly of the medial PFC and MTL, is consistent with the anatomical evidence that has implicated both regions in memory and imagination, and with the notion that the PFC biases episodic recall towards the hippocampal traces congruent with current goal states.

As for the second tract, implicating engram circuits in both memory and imagination presents a technical challenge. While genetic techniques have proven capable of resolving specific memory circuits with single-cell resolution in rodents (Reijmers et al., 2007; Denny et al., 2014; DeNardo et al., 2019), doing the same for imagination presents an operational challenge, as animal models for which there exists the requisite genetic accessibility may lack suitable imaginative capacity (Shevlin, 2021). Barron and colleagues (2013) circumvented this problem by utilizing the well-characterized fMRI phenomena of voxel activity being suppressed following repeated activation, known as neural adaptation, to study the sharing of conceptual representation circuits in humans. By demonstrating that compound concepts induce adaptation in the brain regions responsible for representing the constituent components, they demonstrate that imagined simulations are indeed composed of disparate representations that were formed separately but could be cognitively merged.

While future experimentation will be needed to close the gap between neurobiological tests of engram circuit specificity and behavioral validations of imagination, sufficient evidence already exists to substantiate the theory of a unified mechanism providing for memory and imagination. Specifically, that distributed circuits throughout the brain are tasked with representing stored conceptual information (Josselyn and Tonegawa, 2020)—information retrievable through activation of the corresponding hippocampal trace, guided by the top-down direction of the PFC (Eichenbaum, 2017). When the trace and its associated stored information constitute a consolidated episode, this process is the recollection of a memory. When multiple disparate traces are co-activated, and their respective concepts combined in a novel way, the result is imagination.
Conclusion

The proposed theory of cognitive construction draws upon diverse bodies of experimental and theoretical literature to present a comprehensive case for memory and imagination sharing both the engram circuit as the store of information, and cortico-temporal networks as the mechanism to selectively recall, reinstate, and, in the case of imagination, recombine stored information. While further experimentation is needed to confirm the links between the disparate components of this theory, thus building on existing functional and anatomical studies, each component individually comes with a strong foundation of neurobiological, psychological, and physiological support. As such, the proposed theory, while perhaps unintuitive on account of the rhetorical differences between memory and imagination, does not constitute a large leap from widely accepted neuroscientific tenets.

It is also possible that the proposed theory be construed as an updating of those from Stein (1953), Koestler (1964, p. 751), and Ingvar (1985). They first suggested that, rather than forming in a vacuum, the generation of new ideas are contingent on pre-existing knowledge, and that cortico-temporal networks are responsible for future thinking and simulation. Likewise, creativity was described by Mednick (1962) as being a highly associative process—an understanding that may now be explained in the context of assembling disparate engram circuits into increasingly unconventional combinations. Insofar as there is a neurobiological explanation for differences in creativity, that explanation may well be characterized by a more robust mechanism for forming mosaics of representational circuits.

Much of the present theory’s conceptual strength is derived from the breadth of the research that supports it. Lesion studies, for instance, provide an important piece of the puzzle, demonstrating that prefrontal and MTL regions are similarly implicated in memory and imagination, and even going so far as to suggest that imagination is a highly conjunctive and constructive process. Engram circuit studies, while still a young area of inquiry, have advanced rapidly over the last decade, and proffer a cellular and circuit basis for mnemonic stores—a neurobiological foundation for the conceptual remixing and recombination that has long been theorized as the basis for imagination. As such, there are bountiful routes to test, potentially disprove, and thus hone the present theory of memory and imagination sharing common neurobiological bases. On the behavioral front, experiments linking divergent thinking (Madore, et al., 2015), creativity, and similar cognitive phenomena to episodic repertoires can help strengthen the understanding of memory’s crucial role as the basis for imagined simulations. Likewise, extensive neuroanatomical experimentation can expand on existing lines of research to confirm the role of cortical reinstatement in memory and imagination—demonstrating that functional disruptions in sensory cortices impair corresponding memories and simulations. Lastly, continued biological exploration of engram circuits promise to resolve the question of conceptual recombination at the cellular level. While a preponderance of engram studies have thus far utilized fear conditioning paradigms, future methodological advancements may make a broader range of memory tasks experimentally tractable, facilitating the characterization of engram and cellular sharing between memories and simulations, perhaps manifesting as memory schemas (Tse et al., 2007; McKenzie and Eichenbaum, 2011; Schlichting and Preston, 2015; Schlichting and Preston, 2016; Gilboa and Marlatte, 2017). The present theory, therefore, constitutes not just an incrementally modified framework for understanding the links between memory and imagination, but an eminently disprovable theory that may constitute a starting point for subsequent experiments from myriad disciplines and lines of research.

The present theory proposes that the mechanisms used to store, recall, and reinstate memories are likewise the same mechanisms that provide for imagination. As such, the conjecture is that while memory is the recall of stored information, imagination is the remixing and recombination of mnemonic elements drawn from memory and assembled in goal-congruent simulations. This theory is synthesized from broad bodies of literature spanning theoretical
accounts of hippocampal functions, functional studies of cortico-temporal networks, and biological characterizations of engram circuits. This theory thus provides a potent basis for future exploration of memory, imagination, and related cognitive and computational processes.

Acknowledgements

I’d like to thank Anthony Wagner, Brian Knutson, and Span Lab for providing invaluable critiques, and to Jacqueline Fong for her counsel and support.

Corresponding Author

Benjamin Midler
Stanford University
Department of Psychology, Stanford University, Building 420, Stanford, CA 94305, USA
bmidler@stanford.edu

References


Towards a Unified Consideration of Memory and Imagination as Cognitive Constructs


Rumelhart DE, MacClelland JL (1986) Parallel distributed processing explorations in the
microstructure of cognition. Cambridge, MA: MIT Pr.


Talland GA (1965) Deranged memory: a psychonomic study of the amnesic syndrome. Academic P.


Teyler TJ, Rudy JW (2007) The hippocampal indexing theory and episodic memory:
Updating the index. Hippocampus 17:1158–1169.


