

A Computational Architecture for Machine Consciousness and Artificial Superintelligence: Updating Working Memory Iteratively

Jared Edward Reser Ph.D., M.A., M.A.

Program Peace LLC

+ 818-425-2340

jared@jaredreser.com

Abstract

This theoretical article examines how to construct human-like working memory and thought processes within a computer. There should be two working memory stores, one analogous to sustained firing in association cortex, and one analogous to synaptic potentiation in the cerebral cortex. These stores must be constantly updated with new representations that arise from either environmental stimulation or internal processing. They should be updated continuously, and in an iterative fashion, meaning that, in the next state, some items in the set of coactive items should always be retained. Thus, the set of concepts coactive in working memory will evolve gradually and incrementally over time. This makes each state is a revised iteration of the preceding state and causes successive states to overlap and blend with respect to the set of representations they contain. As new representations are added and old ones are subtracted, some remain active for several seconds over the course of these changes. This persistent activity, similar to that used in artificial recurrent neural networks, is used to spread activation energy throughout the global workspace to search for the next associative update. The result is a chain of associatively linked intermediate states that are capable of advancing toward a solution or goal. Iterative updating is conceptualized here as an information processing strategy, a computational and neurophysiological determinant of the stream of thought, and an algorithm for designing and programming artificial intelligence.

Keywords

artificial intelligence, consciousness, focus of attention, information processing, long short-term memory, neural assembly, recurrent neural network, short-term memory, systems neuroscience, working memory

Introduction and Literature Review

0. Working Memory Relies on the Persistence of Information

The animal brain evolved through natural selection to produce adaptive behaviors in response to environmental contingencies. To accomplish this, the brain registers the current state of the environment by representing different co-occurring features with preexisting memory traces. Most animals will register their immediate surroundings, including moving objects and active sounds. By coactivating the brain's representations for these, a search of its associative network is conducted for information that pertains to the current circumstances. However, the sensory inputs from the present instant are not the only representations capable of being coactivated.

Animals are capable of retaining information about the recent past and using it to inform their current behavior. This persistent information often comes in the form of sensory inputs that are no longer present but that retain some contextual relevance because they have a bearing on what can be expected. Information persistence permits the coactivation of the brain's representations for environmental features occurring at different points in time that would never occur simultaneously in the environment. This, in turn, permits the search of associative memory to use related representations (long-term dependencies) from various moments as search parameters. In some animals, this capacity to carry information through time so that it interacts with new information in the near future amounts to a continuous, internal production pipeline known as working memory. The present work will consider how the contents of this pipeline evolve over time and how this evolution structures the process of thought. This will be done to inform the design of a working memory architecture for artificial intelligence.

Current artificial neural networks such as recurrent, long short-term memory, and transformer models use a form of persistent activity and iterative updating to create a form of working memory. This article will explain why these methods are limited and will not lead to artificial general intelligence. It also introduces a number of new concepts, terms (Table 3), and illustrations (Figures 1-22) to explain what a computer needs to make the state space transitions necessary to achieve general intellectual faculties.

1. Iterative Workflow

Contemporary models of working memory generally do not explicitly address gradual change in information persistence. In many discussions, updating of the information held in working memory is considered to be complete rather than partial, meaning that after being updated, the contents from the previous state are entirely replaced. In other discussions, information can be updated without complete replacement, such as when working memory holds three words and then accommodates a fourth word in addition to the first three (e.g., Pina et. al., 2018, Niklaus et al., 2019, Miller et. al., 2018, Manohar et. al., 2019).

In contrast, the present account explores the hypothesis that partial updating occurs continuously. As some representations are added, some are subtracted, and others from the previous state remain, due to persistent neural activity. This cascading persistence allows successive states to share a large proportion of their content in common, creating complex causal relationships between them (Reser, 2011, 2012). This perspective may be useful because it illuminates how the gradually transforming collection of representations in working memory allows iterative

progress as updated states elaborate intelligently on the states that came before them (Reser, 2013, 2016).

A familiar example of the concept of iteration is “iterative design.” It is a method of developing commercial products through a cyclic process of prototyping, testing, analyzing, and refining. With this method, designs are assessed through user feedback and improved in an incremental fashion. Think of the installment histories of a popular product such as a cell phone, operating system, or car. The newest version of the product contains novel features but preserves many aspects of the previous version and even of versions before that. The workflow of human thought is interpreted here in a similar way. As mental representations in working memory are updated, the frame of reference is gradually replaced, and a thought about one scenario incrementally transitions into a thought about a related scenario. The result is a series of intermediate states capable of exploring a problem space and progressing toward a solution state. This article will explore how this general process may make contributions to reasoning, mental modeling, executive processes, and consciousness.

This abstract, high-level model also offers a broad neurophysiological explanation for the iterative transitions that take place during updating. That is, the iterative updating of the set of representations in working memory is made possible by the iterative updating of the set of persistently active neurons in the cerebral cortex. The firing neurons that underlie the representations in working memory spread their combined excitatory and inhibitory effects to other cells throughout the cortex. Thus, the coactivation of all the contents of working memory amounts to an associative search of long-term memory for applicable information (e.g., predictions, probabilities, and motor instructions). The nonactive (baseline) cells that receive the most spreading activation become active and comprise the representation(s) that will update working memory. Similarly, the representations that continue to receive activation energy are maintained in working memory while those that receive reduced energy are subtracted from it. Performing search using a modified version of the previous search, and doing so repeatedly, amounts to a compounded form of search that ultimately enables the compounding of predictions and inferences.

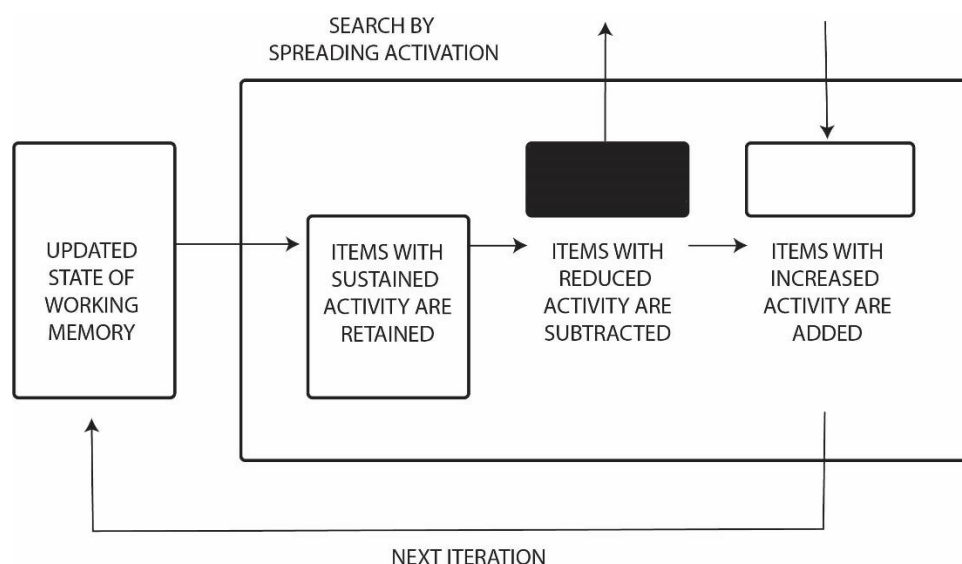


Fig. 1. Flowchart of Iterative Updating

In an iterative process, a set of components is modified repetitively to generate a series of updated states. Each state is an iteration as well as the starting point for the next iteration. One way to accomplish iterative modification is to alter a given state by retaining pertinent elements and then subtracting and adding others. In the brain, the content to be added and subtracted may be determined by spreading activation.

The newly activated representations are added to the representations that have remained in working memory from the previous state, and this updated set is used to conduct the next search. This cycle is then repeated in a loop to produce the thinking process. Thus, there is a direct structural correspondence between the turnover of persistent neural activity, the gradual updating of working memory, and the continuity of the stream of thought. Many of the major features of thought derived from introspection (Hamilton, 1860; Weger et al., 2018) are addressed by this hypothetical explanation such as how mental context is conserved from one thought to the next, how one thought is associated with the next, and how it logically (or probabilistically) implies the next.

This opinion article focuses on ongoing, internally generated activity within working memory and the emergent iterative pattern of information flow. This pattern, introduced in Figure 1, is elaborated on methodically through a series of 20 figures that attempt to illustrate the “shape” of the thought process. Topics considered include the neural basis of items in working memory, variation in the rate of updating, the involvement of multiple working memory stores, and how iterative updating can be implemented within neural network models to enhance the performance of artificial intelligence (AI). This work builds on these issues, while assimilating current theoretical approaches and remaining consistent with prevailing knowledge. Sections 2 through 5 review pertinent literature that forms the foundation of the iterative updating model. Sections 6 through 17 develop said model.

2. Interactions Between Sensory Memory, Working Memory, and Long-term Memory

Working memory has been defined as the components of the mind that temporarily hold a limited amount of information in a heightened state of availability for use in ongoing information processing (Cowan, 2016). It involves holding on to ephemeral sensory and semantic information (e.g., objects, shapes, colors, locations, movement patterns, symbols, rules, concepts, numbers, and words) until they are needed to execute an action or decision. It is one of multiple phases of memory and has been variously referred to as immediate memory and primary memory. It was conceptualized by William James (1842-1910) as the “trailing edge of the conscious present” and a major determinant of which portions of new information will be perceived and which of those will be analyzed (James, 1890). Working memory is thought to facilitate various operations, such as planning, language comprehension, reasoning, decision making, and problem solving (Baddeley, 2012).

The working memory store is constantly updated with new items, which then fade over the course of seconds or minutes (some more quickly than others). Updating allocates processing resources to important information coming from the senses (e.g., novelties, needs, and threats), or from internal states (e.g., intentions, plans, or schemas). Most mental functions require the active maintenance of multiple items at once, along with methodical updating of these items (Baddeley, 2012). Active updating is necessary because the importance of individual items changes as processing demands change (Myers et al., 2017).

Research on working memory has traditionally relied on behavioral investigations (such as memory tasks) to study interactions and dissociations between memory systems. Experimental studies on the topic are concerned with capacity limits, rehearsal, interference, suppression of irrelevant information, removal of active information, and other regular phenomena. Theorists have tried to capture these regularities using abstract models.

From the late 1950s to the 1960s, memory researchers (e.g., Atkinson and Shiffrin, 1968; Broadbent, 1958) developed models that conceptualized memory as being comprised of three interacting systems: (1) a sensory store that briefly holds and preprocesses sensory inputs, (2) an active short-term system capable of attending to this information over a time frame of seconds, and (3) a passive long-term system capable of maintaining information indefinitely (Fig. 2). Current models (including the present model) have retained many aspects of these early models of working memory.

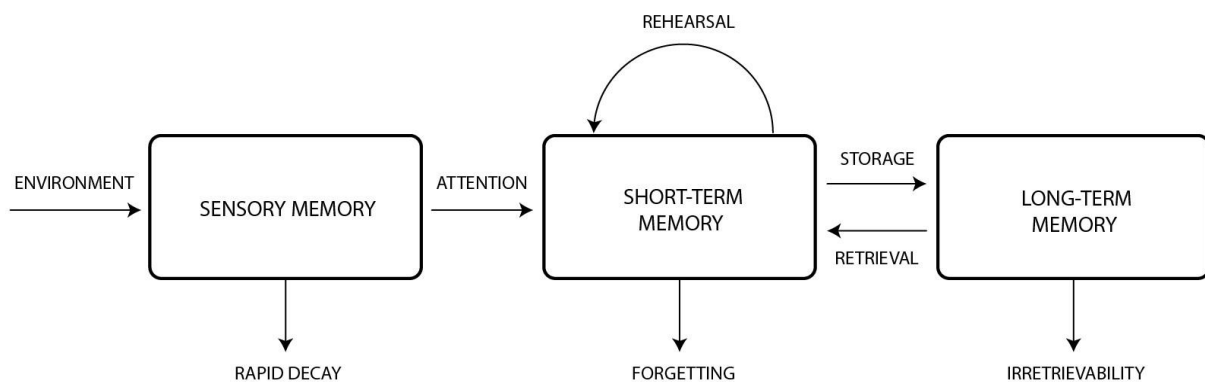


Fig. 2. Atkinson and Shiffrin's (1968) Multi-Store Model

This model depicts environmental stimuli received by the senses and held in sensory memory. If attended to, this stimulus information will enter short-term memory (i.e., working memory). If not rehearsed, it will be forgotten; if rehearsed, it will remain in short-term memory; and if sufficiently elaborated upon, it will be stored in long-term memory, from which it can be retrieved later.

The multi-store model has been expanded upon in several key ways. Studies performed by Alan Baddeley and Graham Hitch (1974, 1986) using dual-task interference experiments indicated that the capacity limitations for visual and verbal working memory are independent, leading the authors to categorize these two modalities as separable stores. This distinction led to the authors' influential multicomponent model, which divided working memory into two domain-specific stores: the visuospatial sketchpad and the phonological buffer (Fig. 3). These stores work in concert to construct, sustain, and modify mental imagery.

Baddeley and Hitch also envisioned a dedicated supervisory subsystem, which they named the "central executive," that selected items for activity, shuttled information from one store to another, and made other processing decisions. Because researchers have not yet explicitly determined how the central executive, visuospatial sketchpad, and phonological buffer cooperate, they remain areas of active research and theory.

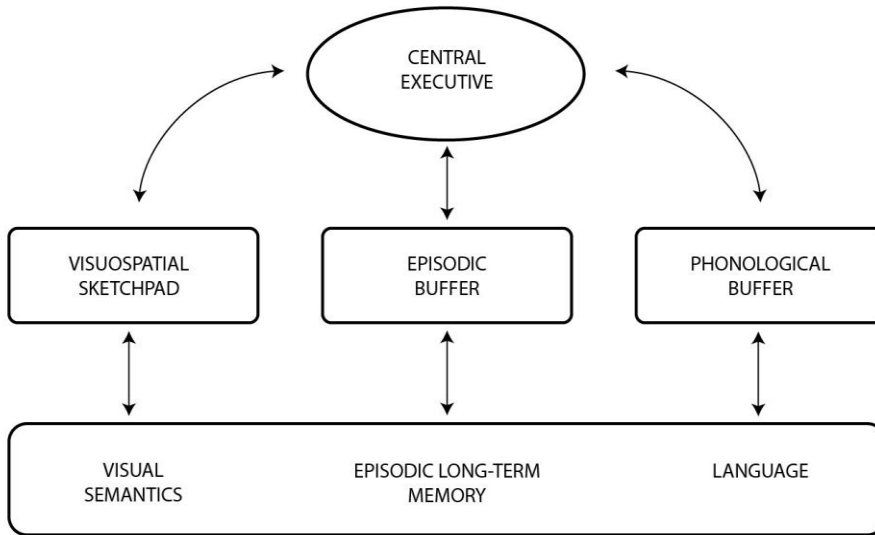


Fig. 3. Baddeley and Hitch's (1974) Multicomponent Model

In this model, the short-term store from Atkinson and Shiffrin's model (1968) is split into four interacting components that together constitute working memory: the visuospatial sketchpad, the phonological buffer, the central executive, and the episodic buffer, which was added later (Baddeley, 2000). These components interact with long-term memory, represented by the bottom rectangle.

More recently, Bernard Baars introduced the global workspace model, which combines the multi-store model with the multicomponent model (Baars, 2007). This framework, adapted in Figure 4, integrates other cognitive constructs such as attention, consciousness, and planning. It also draws further subdivisions within long-term memory. In Baars' model, active contents in working memory are broadcast throughout the brain, resulting in the stimulation of unconscious long-term memories. These long-term memories then compete to enter the global workspace. This type of organization is known as a "blackboard" architecture and can be traced back to Newel and Simon (1961). Many present-day computer science, neural, and psychological models assume a fleeting but centralized working memory capacity that acts as a common workspace where long-term memories become coactive and are exposed to one another (e.g., Dehaene, 2020; Ryan et al., 2019; Glushchenko et al., 2018).

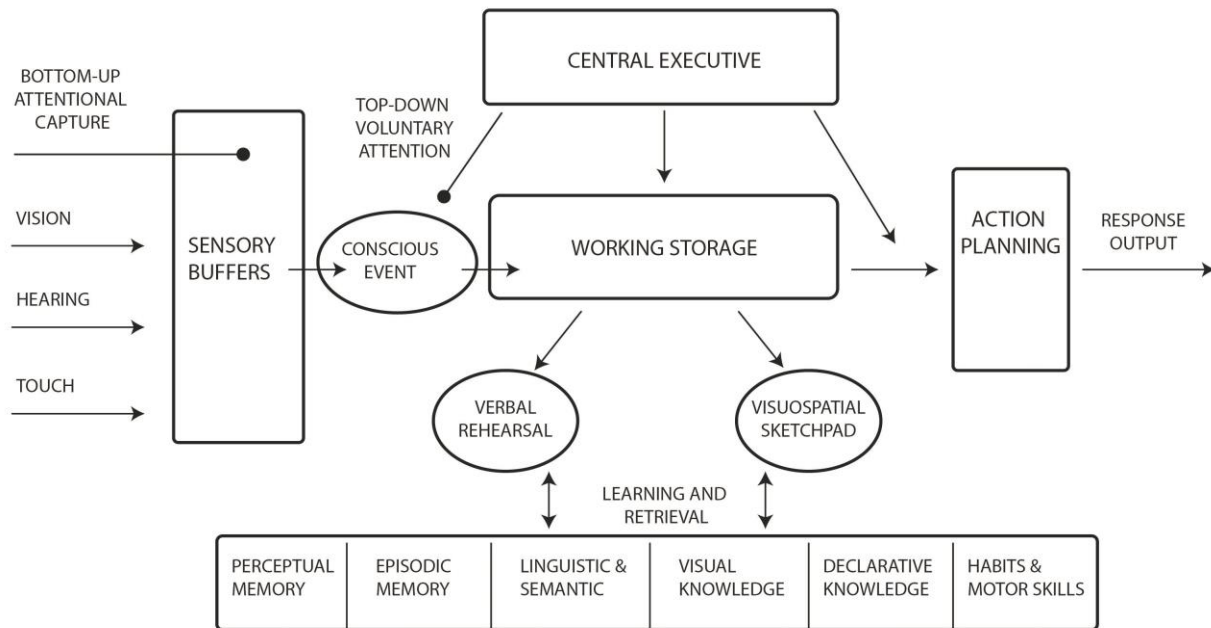


Fig. 4. Baar's Global Workspace Model (2007)

This model incorporates the multi-store model with the multicomponent model. Working memory activates long-term memories, knowledge, and skills, which are shown in the box at the bottom. Spontaneous (bottom-up) attention and voluntary (top-down) attention are symbolized as vectors. The relationship between voluntary attention and conscious events is also depicted.

Early models attempting to explain how long-term memory is transferred into working memory were influenced by computer science. They envisioned long-term memories being copied and transferred from long-term storage to a separate processing substrate (i.e., from the hard drive to random-access memory (RAM) to the central processing unit (CPU)). In a departure from this conception, several theorists (e.g., Cowan, 1984; Norman, 1968; Treisman, 1964) envisioned that information is encoded into working memory when existing units of long-term memory are activated and attended to without being copied or transported. Today, this is commonly referred to as activated long-term memory.

Imaging studies support this view and provide evidence that units of long-term memory reside in the same locations involved in processing this information during non-working memory scenarios (D'Esposito & Postle, 2015). These findings suggest that information is not transferred between dedicated registers, caches, or buffers but activated right where it is (Chein & Fiez 2010; Moscovitch et al., 2007). Thus, although neurons are stationary, as long as they remain active, they continue to broadcast their encoded information to the neurons they project to.

3. The Focus of Attention Is Embedded within the Short-term Memory Store

Nelson Cowan's embedded processes model (1988) reconciles the major features of the multi-store and multicomponent models with the concept of activated long-term memory. In Cowan's model, the short-term memory store is comprised of units of long-term memory that are activated above baseline levels, such as memories that have been primed. This activation can last from

seconds to minutes. Thus, the short-term store of working memory is simply an active subset of the long-term store it is “embedded” within (Cowan, 1999).

The other key component of Cowan’s model is the focus of attention (FoA). The FoA holds consciously attended units of information and is embedded within the short-term store (Fig. 5). Units in the FoA comprise an even more active subset of the short-term store. Their elevated activity lasts from milliseconds to several seconds. Cowan and others consider the short-term store and the FoA together as constituting working memory (Cowan, 2005). In Cowan’s model, these stores interact with a store for sensory memory and a central executive.

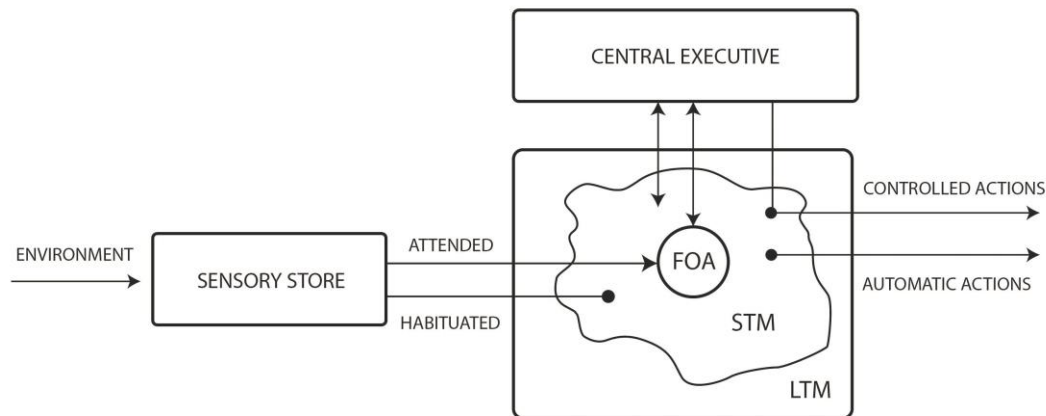


Fig. 5. Cowan’s (1988) Embedded Processes Model

According to this model, short-term storage is an activated subset of long-term storage, and the FoA is an attended subset of short-term storage.

During perception, task relevant features from the sensory store are used to update the FoA. When attention shifts to new information, these items then pass into the short-term store (Nyberg & Eriksson, 2016). However, information that has been demoted from the FoA to the short-term store can still influence automatic actions and be readily reactivated into the FoA (Manohar et al., 2019). If not reactivated, this information returns to inert long-term memory (through the processes of decay, inhibition, interference, or contamination) (Cowan, 2009). Some items that enter working memory are demoted almost immediately, whereas others remain active for sustained periods (Cowan, 2011). This feature, along with features of the other models discussed thus far, forms critical assumptions subsumed by the present model.

The first of five quotes in this article from the prescient William James complements Cowan’s conception of an FoA store, which interacts with and is embedded inside a short-term store:

My present field of consciousness is a centre surrounded by a fringe that shades insensibly into a subconscious more.... The centre works in one way while the margins work in another, and presently overpower the centre and are central themselves. What we conceptually identify ourselves with and say we are thinking of at any time is the centre; but our full self is the whole field, with all those indefinitely radiating subconscious possibilities of increase that we can only feel without conceiving, and can hardly begin to analyze. (James, 1909, p. 288)

4. Sustained Firing Maintains Information in the Focus of Attention

The neurophysiological basis of the persistent activity responsible for working memory is an area of active research. Single-cell recordings of neurons in primates evince that a form of information retention occurs via a cellular phenomenon known as sustained firing. Glutamatergic pyramidal neurons in the prefrontal cortex (PFC), parietal cortex, and other association cortices are specialized for sustained activity, allowing the cells to generate action potentials at elevated rates for several seconds at a time (Funahashi, 2007; Fuster, 2015). Sustained firing is thought to maintain the signal of information that the neuron encodes. A neuron in the PFC with a background firing rate of 10 Hz (typical for cortical cells) might increase its firing rate to 20 Hz when utilizing sustained firing to temporarily retain mnemonic information.

One of the earliest of these studies provides an illustrative example. In 1973, Joaquin Fuster recorded sustained electrical activity of PFC neurons in monkeys performing a delayed matching task. In the task, a macaque monkey watches the experimenter place food under one of two identical cups. A shutter is then lowered for a variable delay period, so that the cups are not visible. After the delay, the shutter is raised, and the monkey is given one attempt to collect the food. Through training, the animal learns to choose the correct cup on the first attempt. Completing the task requires the animal to hold the location of the food in working memory during the delay period. Presumably, the monkey must sustain either a retrospective sensory representation of the food's location, or a prospective representation of the motor plan needed to retrieve it.

Using implanted electrodes, Fuster was able to record from neurons in the PFC that fired throughout the delay period. He found that the sustained firing subsided once the monkey responded, suggesting that the observed neuronal activity represented the food's location while the cup was out of sight. This landmark study revealed the brain's mechanism for keeping important representations active without external input. It also suggested the presence of a dynamically updated pool of coactive neurons underlying thought and behavior.

Subsequent research has found that the duration of sustained firing predicts whether items will be remembered and that when this delay-period activity is weak, the likelihood of forgetting is greater (Funahashi et al., 1993). Moreover, lesioning of the prefrontal and association cortices (which contain neurons with the greatest capacity for sustained firing) significantly impairs performance in these types of tasks. Consistent with this animal work, functional magnetic resonance imaging (fMRI) studies in humans show that activity in prefrontal and association areas persists during the delay period of similar working memory tasks. In fact, the magnitude of this activation positively correlates with the number of items subjects are instructed to hold in memory (Rypma et al., 2002).

Patricia Goldman-Rakic (1987, 1990, 1995) was the first to suggest that the phenomenon of sustained firing in the PFC is responsible for the retention interval exhibited by working memory. Further work by Fuster (2009), Goldman-Rakic (1995), and others has shown that neuronal microcircuits within the PFC maintain information in working memory via recurrent, excitatory glutamatergic networks of pyramidal cells (Baddeley & Hitch, 1994; Miller & Cohen, 2001). Many researchers now believe that sustained firing plays a critical role in the maintenance of working memory. The evidence backing this assumption is provided by studies reporting positive correlations between sustained firing and working memory performance. For example,

both human and animal subjects can retain information in mind as long as sustained firing persists (Rypma et al., 2002). This has been found using extracellular, electroencephalographic, and hemodynamic approaches (D’Esposito & Postle, 2015).

Sustained firing in the PFC and parietal cortex is now assumed to underlie the capacity to internally maintain and update the contents of the FoA (Braver & Cohen, 2000; Sarter, Givens, & Bruno, 2001). As a result, working memory, executive processing, and cognitive control are now widely thought to rely on the maintenance of activity in multimodal association areas that correspond to goal-relevant features and patterns (Baddeley, 2007; Fuster, 2002a; Moscovich, 1992; Postle, 2007). Sustained rates of action potentials allow responses throughout the brain to be modulated by prior history over multiple timescales, from milliseconds to tens of seconds.

5. Synaptic Potentiation Maintains Information in the Short-term Store

fMRI studies have suggested that the information represented by sustained firing corresponds only to the FoA, not the short-term store as a whole (Lewis-Peacock et al., 2012). This is because neuronal activity corresponding to items that have exited the FoA quickly drops to baseline firing rates. Nevertheless, information about the items may be rapidly and reliably recalled after a brief delay. It is thought that the passive retention of information in the short-term store but outside the FoA may be mediated by a different “activity-silent” neural mechanism, such as changes in synaptic potentiation (short-term synaptic plasticity) (LaRocque et al., 2014; Rose, 2016). The evidence supporting this is strong (Silvanto, 2017; Nairne, 2002). For example, synaptic strength can be temporarily modified by transient increases in the concentration of presynaptic calcium ions or by GluR1-dependent short-term potentiation (Silvanto, 2017). The information potentiated by these changes in synaptic weighting can be converted back into active neural firing if the memory is reactivated by a contextual retrieval cue (Nairne, 2002).

Thus, the maintenance of information in working memory is achieved by at least two neural phenomena operating in parallel that correspond to distinct states of prioritization: sustained firing, which maintains information in the FoA, and synaptic potentiation, which maintains information in the short-term store. Both mechanisms contribute to the initialization of long-term potentiation, including RNA synthesis, protein synthesis, and morphological synaptic changes that underlie the formation and consolidation of new long-term memories (Debanne, 2019).

Sustained firing and synaptic potentiation create a tiered pipeline of continued activity that provides the conceptual basis used by this article to view the previous four models from the perspective of iteration.

	Sensory Memory	Focus of Attention	Short-term Store	Long-term Memory	References
Description	Early sensory processing	Attended content	Temporary store of content	Inactive but retrievable content	Baddeley et al., 2018
Brain Correlate	Firing neuron	Sustained firing	Synaptic potentiation	Long-term potentiation	Christophel et al., 2017
Capacity	Numerous features	1 to 9 items	Very large	Extremely large	Shipstead et al., 2015
Duration	Visual: .25 s. Auditory: 2.5 s.	Up to a minute	Minutes	Days to lifetime	Brydges et al., 2018
Origin	Pre-attentive	Attentive	Recent attention	Memorization	Cowan, 2017




Location	Sensory cortex	Association cortex	Cerebral cortex	Cerebral cortex	Eriksson et al., 2015
Maintenance	Not possible	Continued attention	Rehearsal	Repetition, mnemonics	Chia et al., 2018
Departure	Decay, highly volatile	Replacement	Forgetting	Irretrievability	Constantinidis et al., 2018
Depiction in Figures: 8-16, 18, 20, 21	N/A				

Table 1. General Characteristics of Four Forms of Memory

This table summarizes some of the major comparisons between four different forms of memory. The details addressed in this table are not definitive and are active areas of debate and research.

Working Memory Is Updated Iteratively: A Proposed Model

6. Persistent Activity Causes Successive States to Overlap

The study of sustained firing has shown that the neocortex contains many neurons in persistent coactivity at any instant in time (Goldman-Rakic, 1995). Yet, these coactive neurons could not have all started firing at the same time, nor could they all stop firing at the same time. Because sustained activity has been shown to occur for different durations in different neurons, even in those that were temporarily coactive (Fuster, 2002b), the spans of activity of neurons exhibiting sustained firing must be staggered and must only partially overlap with one another rather than completely coinciding (Fuster, 2008). In other words, the set of neurons demonstrating sustained firing in the neocortex must update incrementally rather than all at once (Reser, 2016).

Short-term synaptic potentiation (or any other neurophysiological mechanism) responsible for the maintenance of non-FoA working memory can be expected to exhibit the same iterative properties as the sustained firing responsible for the FoA. This expectation derives from the fact that the sum total of synaptically potentiated neurons is equivalent to a pool that is constantly being added to as new neurons are potentiated and subtracted from as other neurons lose their time-limited potentiation. Thus, both the short-term store and the FoA demonstrate iterative functionality, albeit on different time scales.

The incremental updating expected at the neurophysiological level may be isomorphic with and provide a substrate for the incremental updating experienced on the psychological level. For example, a given line of thought does not change all at once but rather makes additive transitions that are grounded by content that remains unchanged. The subset of neurons that continue to exhibit persistent activity over the course of these incremental changes should be expected to embody the persisting subject of mental analysis. Stated differently, neurons with the longest-lasting activation likely correspond to the underlying topic of thought that remains as other contextual features come and go. Sustained firing and synaptic potentiation allow pertinent information to be carried over from one state to subsequent ones. This creates coherence and continuity between distinct epochs (Reser, 2016), as depicted in Figure 6. The present article contends that without the continuity made possible by iteration, thought as we know it could not arise.

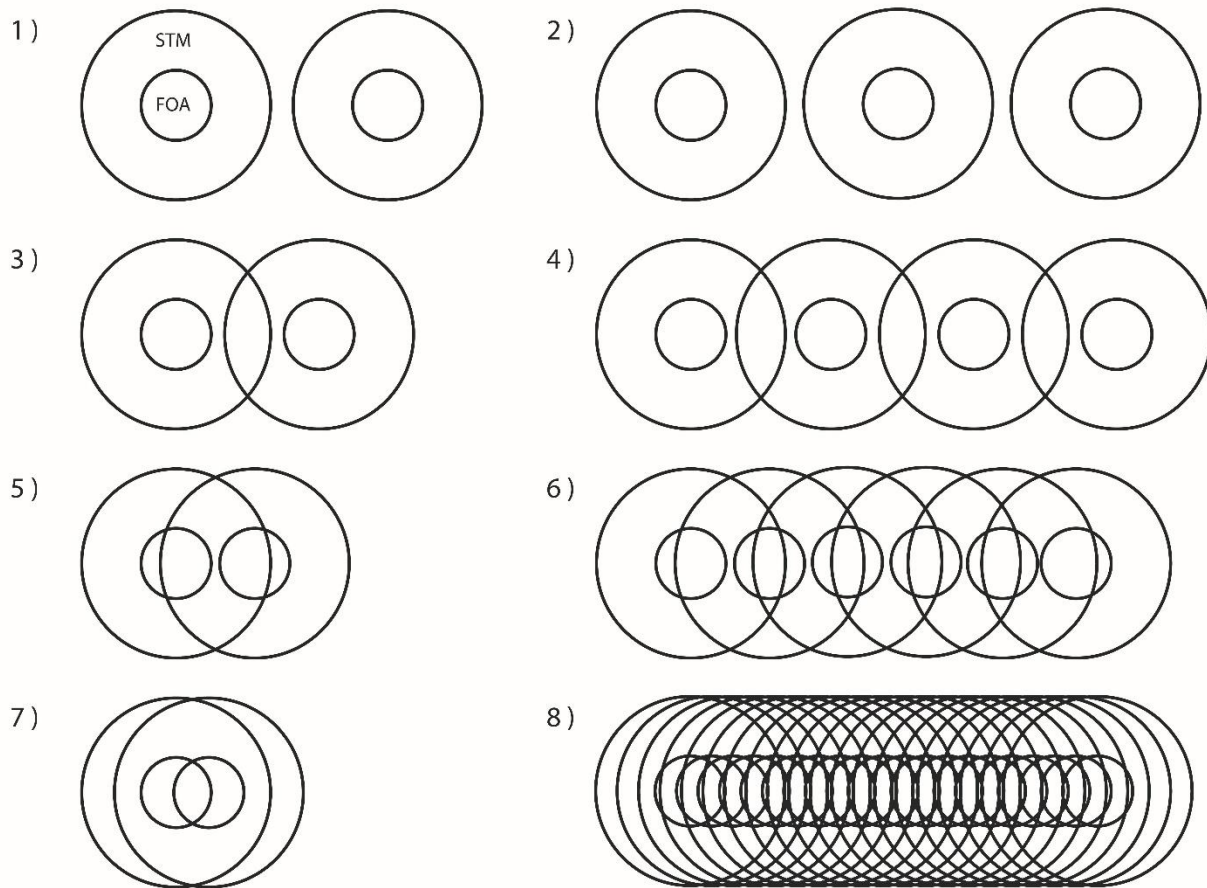


Fig. 6. Venn Diagrams of Information Shared Between Successive States of Working Memory

These Venn diagrams depict informational overlap between successive states of working memory. The horizontal axis represents time. The small circles represent information within the FoA, and the large circles represent information within the short-term store. Diagrams 1 and 2 show no Venn overlap between states from different periods; 3 and 4 show overlap in the short-term store only; 5 and 6 show the short-term store of one state overlapping with the FoA of the neighboring state; and 7 and 8 show the FoA of separate states overlapping, suggesting attentive continuity. It may be plausible that Diagrams 1 and 2 roughly depict sampling of cortical activity hours apart, 3 and 4 depict sampling several minutes apart, 5 and 6 depict sampling every minute, and 7 and 8 depict sampling every second.

Diagram 1 of Figure 6 depicts two states of working memory whose contents do not overlap. We can assume that these states are from separate thoughts. Diagram 7 depicts two states whose contents overlap significantly. It is intended to represent a fractional transition in the thought process, such as two points in a line of reasoning.

The overlapping informational content of the small circles shown in Diagram 7 indicates that the two states share neurons in common that exhibit sustained firing. The overlap of the large circles represents the sharing of potentiated synapses. Thus, the diagrams shown in Figure 6 depict updating as continuous change in active neurons and synapses. However, as the rest of this article will explore, partial change to the FoA may be more realistically depicted as iterative updates in discrete cognitive items.

7. Iteration Causes Consecutive States of Working Memory to Be Interrelated

If updates to the working memory store involve partial rather than complete replacement, then these dynamics indicate an ongoing pattern of recursion and iteration. Recursion is the reapplication of a rule, definition, or procedure to successive results. A recursive function references itself. Self-referential routines are common in mathematics and computer science. Recursion's sister algorithm, Iteration, involves the application of a computational procedure to its own output. It is common in mathematics and computer science and often involves the reapplication of a rule, definition or procedure to successive results. The terms "iteration" and "recursion" uniquely capture different aspects of the present model, and both are used here depending on context.

The principle of recursion and iteration as it pertains to the present model is illustrated in Figure 7. At time 1 (t_1) in Figure 7, neuron "a" has stopped firing. Neurons b, c, d, and e exhibit sustained coactivity. By time 2 (t_2), neuron "b" has stopped firing, while c, d, and e continue to fire and "f" begins to fire. The figure depicts iteration because the set of coactive neurons at time 2 (c, d, e, and f) includes a subset (c, d, and e) of the coactive neurons at time 1. In computer programming, the goal of iteration is to obtain successively closer approximations to the solution of a problem. In later sections, this article will argue that the brain utilizes iteration for the same purpose.

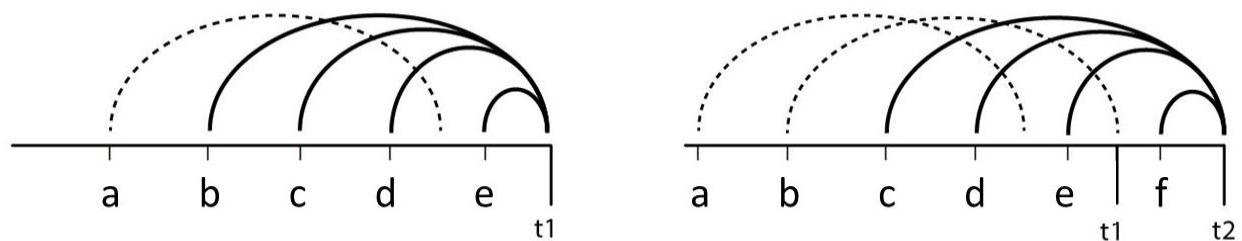


Fig. 7. Hypothetical Depiction of Iteration in Neurons Exhibiting Sustained Firing

Each arc, designated by a lowercase letter, represents the time span during which a neuron exhibiting sustained firing remains active. The x-axis represents time. Dashed arcs represent neurons that have stopped firing, whereas full arcs denote neurons that are still active.

Given that at any point in time we can expect there to be thousands of neurons engaged in sustained firing, we should expect the type of iterative pattern seen in Figure 7 to be ubiquitous. Furthermore, if examined on the order of hundreds of milliseconds, we should expect activity in the brain to be densely iterative. Iteration occurring within the FoA causes consecutive brain states to be interrelated and autocorrelated as a function of the delay between them. Because a delineable subset of the active cells that characterize one brain state remain active in the next, each state is recursively nested within the one that precedes it. This allows the brain to record and keep track of its interactions with the environment (stateful), so that each interaction does not have to be handled based only on the information available at present (stateless).

Consider a hypothetical situation in which the order of neuronal activation portrayed in Figure 7 is applied to the chronological series of events from Joaquin Fuster's 1973 experiment (detailed in Section 6). Let's assume that the activation of neuron "a" corresponds to a monkey seeing the experimenter raise the left cup, neuron "b" corresponds to the monkey watching the

experimenter place food under the left cup, neuron “c” corresponds to the lowering shutter obscuring the monkey’s view of the cups, neuron “d” corresponds to the monkey witnessing a distracting stimulus, neuron “e” corresponds to the shutter raising, and neuron “f” corresponds to the monkey’s awareness that it is time to select one of the two cups. At time 2 of Figure 7, both neurons “a” and “b” have stopped firing, suggesting that the monkey was distracted enough to forget which cup holds the morsel of food (unless the location was encoded into and can be retrieved from the short-term store).

Aside from the FoA and short-term store, information found in other forms of temporary storage may also work iteratively. For example, neural binding may represent a third tier of temporary information storage, which might be embedded in the FoA and take place at even shorter intervals. Neural binding involves synchronized oscillations of network activity that form and dissipate on the order of milliseconds (Opitz, 2010). It is thought that this synchronization integrates different forms of information into cohesive conscious experiences (Pina et al., 2018). However, binding will be left out of the present discussion as it is not yet clear whether state changes in binding are complete or partial. Consequently, this article’s focus on iterative updating may not apply to it. Similarly, dynamic coding, closed recurrent loops, persistent attractor networks, and reentrant neural oscillations may or may not involve iterative updating. While Oberauer’s one-item attentional store (2002) has found considerable empirical support (Niklaus et al., 2019), it could not work iteratively because (as just one item) it is either completely updated or not updated at all.

Term	Definition
Iteration	Repetition of a computational procedure applied to the product of a previous state, used to obtain approximations that are successively closer to the solution of a problem.
Working Memory	A mechanism dedicated to maintaining selected representations available for use in further cognitive processing.
Working Memory Updating	Changes in the contents of working memory occurring as processing proceeds through time.
Iterative Updating	A shift in the contents of working memory that occurs during updating as some representations are added, others are removed, and others are repeated.

Table 2. Definition of Key Terms

It is argued here that iterative updating should be considered inherent in any brain with neurons exhibiting persistent activity and that it is utilized by animals as a fundamental means of information processing. In particular, working memory may harness iteration in a way that allows potentially related representations to accumulate and coactivate despite delays between their initial appearances. This ensures that relevant processing products are temporarily sustained in working memory until a full suite of contextually related items are compiled, so they can be used in aggregate to inform behavior.

8. Iterative Updating of Items Creates Narrative Continuity

Figure 8 depicts an FoA store that holds four psychological items (or chunks) at a time. In this example, one discrete item is updated at each point in time. Thus, it could be conceptualized as a “sliding store.” The depiction of the FoA store as limited to four items is derived from an extensive literature review by Cowan (2001, 2005), which demonstrates that adults are generally

able to recall four items (plus or minus one) in situations when they cannot carry out chunking, rehearsal, or other memory strategies to aid them. This capacity of four items generally holds true, regardless of whether the items are numbers, words in a list, or visual objects in an array. The figures could alternatively feature seven items, rather than four, after less restrictively controlled research by George Miller (1956). While discussing the capacity of the FoA, Cowan remarked,

“When people must recall items from a category in long-term memory, such as states of the United States, they do so in spurts of about three items on average. It is as if a bucket of short-term memory is filled from the well of long-term memory and must be emptied before it is refilled.” (2009, p. 327)

Yet, perhaps this bucket does not need to be emptied to be refilled. While naming states and repeating numbers may not rely on iterative updating, rational thought may.

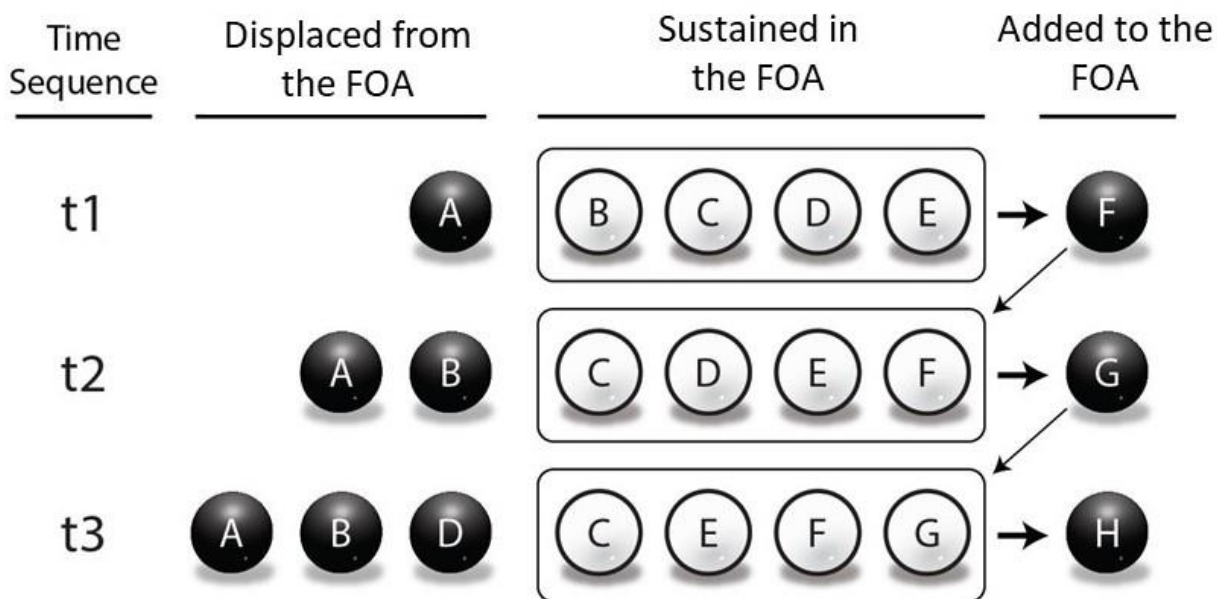


Fig. 8. Abstract Schematic of Iterative Updating in the FoA

As with the following figures in this article, Figure 8 is an emblematic abstraction that uses a state-space model in discrete time. White spheres indicate active items while black spheres indicate inactive ones. At time 1, item A has just been deactivated, while B, C, D, and E are coactive. This echoes the pattern of activity shown in Figure 7 except that the uppercase letters here represent items whereas the lowercase letters in Figure 7 represent neurons. While coactive, these items (B, C, D, and E) spread their activation energy, which results in the convergence of activity onto a new item, F. At time 2, B has been deactivated; C, D, and E remain active; and F has become active. At time 3, D exits the FoA before C, reflecting that the order of entry does not determine the order of exit.

At time 2, three of the items from time 1 (C, D, and E) remain active and are combined with the output from time 1 (F). This new set of items is then used to search for the next update (as illustrated in Figure 1). At time 3, item D has exited the FoA before C (out of alphabetical order). This indicates that items that have been in the FoA for the longest time are not necessarily the first to exit. Items B, C, D, and E are active at time 1, while C, E, F, and G are active at time 3. Thus, items C and E demonstrate reiteration because they exhibit uninterrupted activity from time 1 through time 3. The longer these items are coactive, the more likely they are to become

associated and possibly “chunk” or merge into a single item, “CE.” While C and E remain active, the neural circuits underlying them can be expected to impose sustained, top-down information processing biases on the targets they project to throughout the thalamocortical hierarchy. Items sustained enduringly in this way should be expected to influence the overarching theme of ongoing thought.

Imagine that item B represents your psychological concept of brownies, C represents your friend Cameron, D represents shopping, and E represents a grocery store. With these representations active in your FoA, you may form a mental image of your friend Cameron shopping for brownies at a grocery store. This scenario may cause you to remember Cameron’s preference for drinking milk when he eats brownies. Thus, your next thought may be about your friend shopping in the same store for milk. Some of the contextual factors (the place, person, and activity) remained the same even though another (the object being shopped for) changed. This kind of narrative about the same place and person could take several seconds and many rounds of iteration to play out. This example illustrates how iteration enables continuity by allowing context to shift incrementally, which may be a central hallmark of the thought process. Iteration may make similar contributions to attention, awareness, and subjective experience.

In Reser’s (2016) incremental change model, the subset of neurons demonstrating sustained firing over a series of states (represented by C, D, and E in Figure 8) was said to exhibit “state-spanning coactivity” (SSC). Over time, the set of coactive neurons shifts, creating “incremental change in state-spanning coactivity” (icSSC). According to that model, the content of working memory is effectively in SSC, and as it progresses over time, the content exhibits icSSC. The iterative process of icSSC may apply, not only to working memory, but other constructs such as attention, awareness, thought, and subjective experience.

9. The Rate of Attentional Updating Varies with Demand

Although Figure 8 depicts working memory updating one unit at a time, this likely varies according to processing demands. For instance, when an individual pursues a new train of thought, initiates a different task, or is exposed to a novel or unexpected stimulus, his or her attention shifts entirely from its previous focus. When this happens, the content of the FoA can change completely. In this scenario, attentional resources are reallocated to the new context, and rather than a graduated transition, an abrupt transition occurs without iteration.

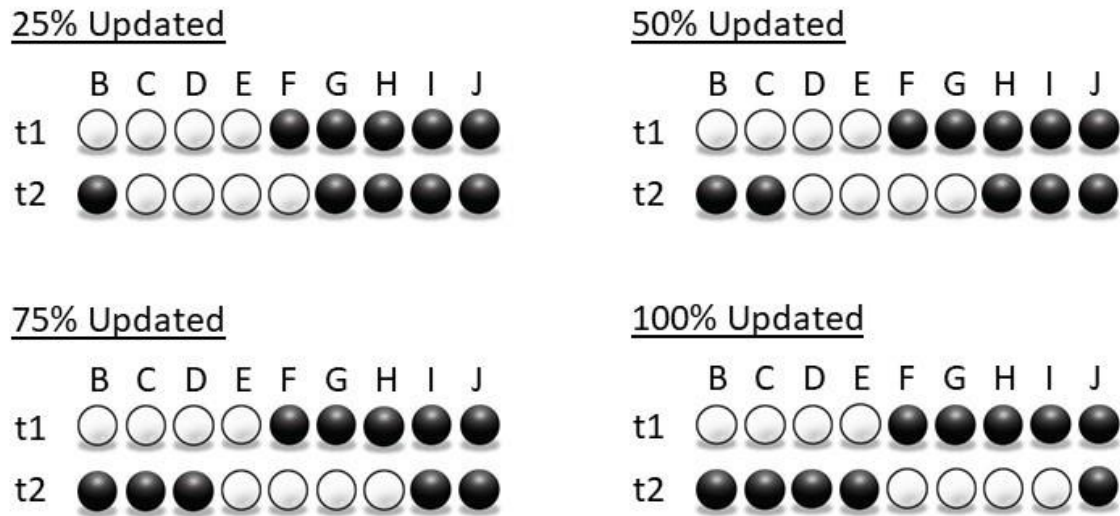


Fig. 9. Four Possible State Transitions in the FoA

In the first scenario, there are four active items at time 1, which are marked as white spheres. At time 2, one of these four items has been replaced, so that one white sphere (B) becomes black (inactive) and a different black sphere (F) becomes white (active). Thus, 25% ($1 \div 4$) of the items have been updated between time 1 and time 2 without any change in the total number of active items. Most other figures in this article feature this 25% updating. However, in a store with four items, updating can occur in three other ways. The other transitions in this figure depict 50%, 75%, and 100% updating.

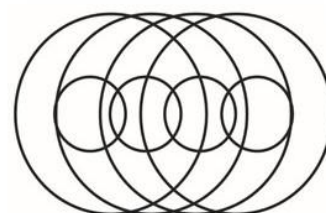
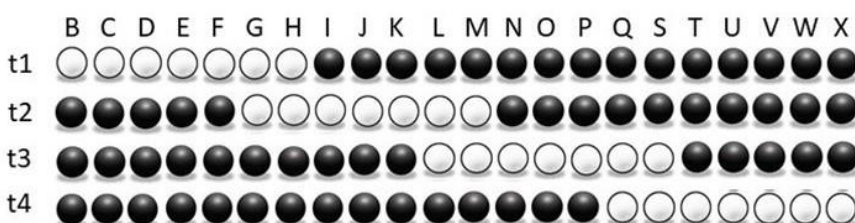
Note that abrupt, noniterative updating is not possible in the short-term store. This is due to the slower nature of turnover in synaptic potentiation. Because the number of active representations is much higher and they subside much more slowly (minutes) than in the FoA (seconds), the short-term store will continue to exhibit substantial iterative overlap, even during complete shifts in focal attention. Thus, the rate of updating from one period to the next is expected to remain relatively stable in the short-term store, whereas the rate of updating in the FoA is expected to fluctuate markedly under different processing requirements.

We should expect the average percentage of updating within the FoA per unit time to be lower in animals with larger, more complex brains. During mammalian evolution, association cortices were greatly enlarged relative to the sensory cortices (Striedter, 2005). This development increased the number of neurons capable of sustained firing, as well as their maximum duration (Sousa et al., 2017), despite increased metabolic costs (Mongillo et al., 2008). For primates, and humans in particular, the presence of highly developed association areas likely leads to (1) more and longer sustained activity, (2) extended coactivity of items, (3) a lower percentage of updating per second, and (4) a corresponding higher degree of continuity between iterations.

In animals, a lower percentage of iterative updating might be correlated with greater working memory capacity as well as higher fluid and general intelligence. This can be conceptualized as a longer working memory half-life. The concept of a half-life could be used to quantify the persistence of information in both the FoA and the short-term store, where generally the shorter the half-life of activity in working memory, the shorter the span of attention. For instance, the

half-life for the diagrams in Figure 9 exhibiting 25% updating is two time intervals, whereas the half-life for 50% updating is only one time interval. Figure 10 addresses the rate of decay using an FoA capacity of seven items. The first diagram illustrates how neural activity in small-brained animals largely models the present and adjusts this model with bits from the recent past. The second diagram illustrates how neural activity in large-brained mammals models the recent past and adjusts it with bits from the present.

71% Updating Over Four Time Periods



29% Updating Over Four Time Periods

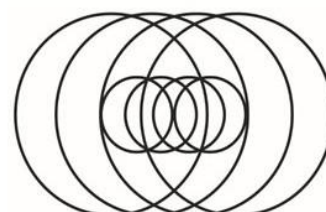
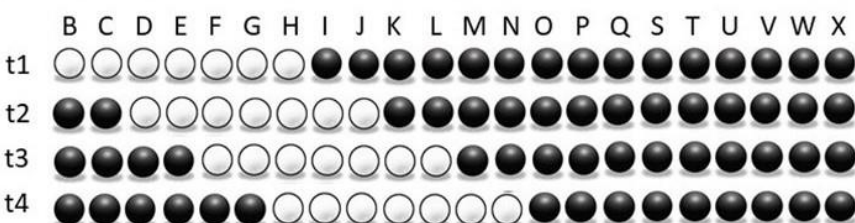


Fig. 10. Two Rates of Updating Carried Out Over Five Time Periods

In the first scenario, 71% ($5 \div 7$) updating is carried out over four different time periods. In the second scenario, 29% ($2 \div 7$) updating is carried out. This comparison delineates the difference between unfocused, minimally overlapping thought (loose iterative coupling) and highly focused, closely overlapping thought (tight iterative coupling). To better illustrate this point, the capacity of the FoA is depicted here as seven items after Miller (1956). The Venn diagrams to the right of each diagram illustrate the percentage of iterative updating in the FoA using the style of Figure 6.

The top diagram in Figure 10 covers a wider breadth of information and proceeds at a faster rate but may be associated with an attention deficit, distractibility, and superficial associations. The bottom diagram is probably more conducive to concentrated attention, effortful/elaborative processing, and structured systematization of knowledge. This is because the search for the next state will be informed by a larger number of conserved parameters. Contrarily, in Diagram 1, more than half of the initial parameters are excluded after only one time interval because they could not be maintained and thus the next search performed loses precision and specificity. For example, it should be more difficult to solve a mathematical word problem in one's head using the updating strategy depicted in Diagram 1 relative to that in Diagram 2 because too many of the problem's crucial elements would be forgotten and thus would not be available to contribute spreading activity in the search for a solution.

These two diagrams may represent the distinction not only between information processing in “lower” and “higher” animals but also between implicit and explicit processing in a single animal. Diagram 1 may be illustrative of implicit or system one processing (i.e., Kahneman's

“thinking fast” (2011)) and its impulsive, heuristic, intuitive approach. Diagram 2 may illustrate explicit or system two processing (i.e., “thinking slow”) in which a problem is encountered that requires multiple processing steps, recruitment of executive attention, the prefrontal cortex, and the prolonged maintenance of intermediate results. Figure 10 is also meant to suggest that implicit and explicit processing exist on a continuum and that implicit processing may transition into explicit when demand, novelty, surprise, curiosity, anticipated reward, or error feedback engage dopaminergic centers and increase the duration of sustained firing in the neurons that represent prioritized contextual variables.

As Figure 11 illustrates, it may be the case that the rate of iterative updating decreases during a thought but then increases during the transition between thoughts. The first diagram in Figure 11 features a larger number (4 vs. 2) of individual instances of continuity (i.e., discrete thoughts). The transitions between thoughts could be conceptualized as intermittent noniterative updating in the FoA. As a cognitive strategy, the processing found in the second diagram is probably more conducive to staying on topic, comprehending complicated scenarios, and solving complex problems.

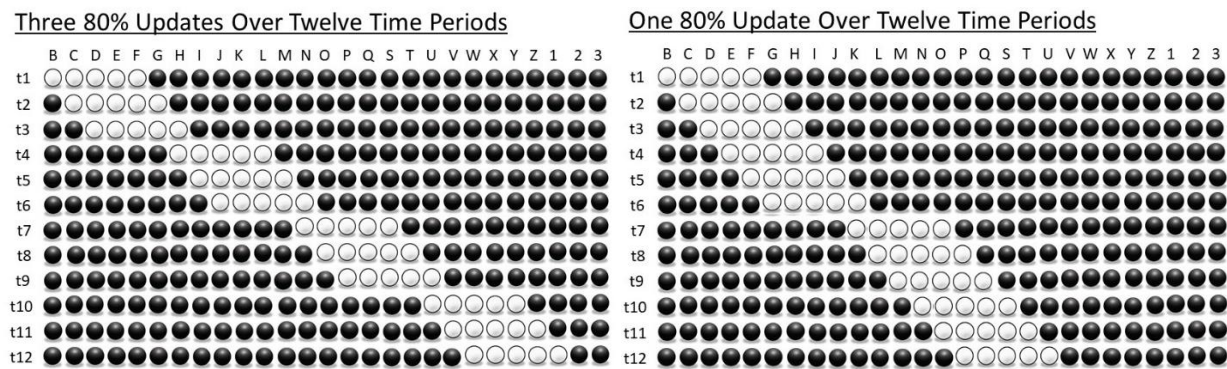


Fig. 11. Intermittent Noniterative Updating Marks a Boundary between Thoughts

In both diagrams, most of the updating is occurring at a rate of 20% (1/5). In the first diagram, there are three intermittent updates of 80% (4÷5). In the second, there is only one intermittent update of 80%. This comparison delineates the difference between four brief thoughts occurring in quick succession and two more prolonged thoughts. The first strategy would result in small islands of associative connections among coactive items. The second strategy would result in longer sequences of iterated associations and the consequent less fragmented learning.

10. Iterative Updating Gives Rise to Mental Continuity

Continuity is defined as the uninterrupted and consistent operation of something over a period of time. According to this model, continuity of thought involves a process in which a set of mental representations demonstrates gradual drift across a series of processing states (Reser, 2016). Continuous, partial updating makes each mental state a reframed version of the last. This reframing process results in an updated group of conditions, modulating rather than replacing the conceptual blend created by the previous set of coactive items. The manner in which iteration permits relevant information from the past to conjoin and assimilate with relevant information from the present may provide the connective tissue for the continuous nature of reflective thought and phenomenal consciousness.

A few analogies may help clarify the nature of iterative continuity. When it demonstrates continuity, we should expect the attentional “spotlight” to move by degrees (e.g., the panning of a video camera) rather than abruptly (e.g., the saccade of an eye). The components within the spotlight vary smoothly. It is like the carousel function used in computer graphical interfaces where a collection of visible objects is updated as individual elements of the collection rotate into and out of view. This is similar to the morphing technique used in computer animation where an image is transformed fluidly into another by maintaining certain features but changing others in small gradual steps. Corresponding points on the before and after images are usually anchored and then incrementally transfigured from one to the other in a process called “crossfading.” It is also like the changes taking place within the set of interlocking teeth of two gears. As the gears turn and a new tooth is added to this set, a different tooth is subtracted, yet other teeth will remain interdigitated. In literary terms, the subset of items that remain interdigitated constitutes the “through line,” connecting theme, or invisible thread that binds elements of a mental experience together. Mental continuity is an evolutionary process, and like natural selection involves non-random retention and elimination of candidate structures leading to incremental modifications to the population.

There are some published articles that utilize iteration in describing various psychological phenomena (e.g., Shastri et al., 1999; Howard & Kahana, 2002; Hummel & Holyoak 2003; Botvinick & Plaut, 2006; Kounatidou et al., 2018). However, these models are not applied to modeling continuity in brain activity, working memory, thought, or consciousness. Although modern research on the present topic appears to be scarce (Reser, 2016), the present model was not the first to consider the role of iteration in the thought process. In fact, William James addressed the continuous nature of consciousness in a number of his writings. In a lecture from 1909 entitled “The Continuity of Experience,” James spoke about the “units of our immediately felt life,” describing how these units blend together to form a continuous sheet of experience:

It is like the log carried first by William and Henry, then by William, Henry, and John, then by Henry and John, then by John and Peter, and so on. All real units of experience overlap. Let a row of equidistant dots on a sheet of paper symbolize the concepts by which we intellectualize the world. Let a ruler long enough to cover at least three dots stand for our sensible experience. Then the conceived changes of the sensible experience can be symbolized by sliding the ruler along the line of dots. One concept after another will apply to it, one after another drop away, but it will always cover at least two of them, and no dots less than three will ever adequately cover it. (James, 1909, p. 287)

The above quote evinces that James had conceived an iterative model of consciousness over a hundred years ago. Moreover, his minimum of three “dots” coincides with Cowan’s four (plus or minus one) items of working memory. The next section adds detail to the present account of the neural basis of the items in working memory and describes how active neurons search long-term memory for the next update.

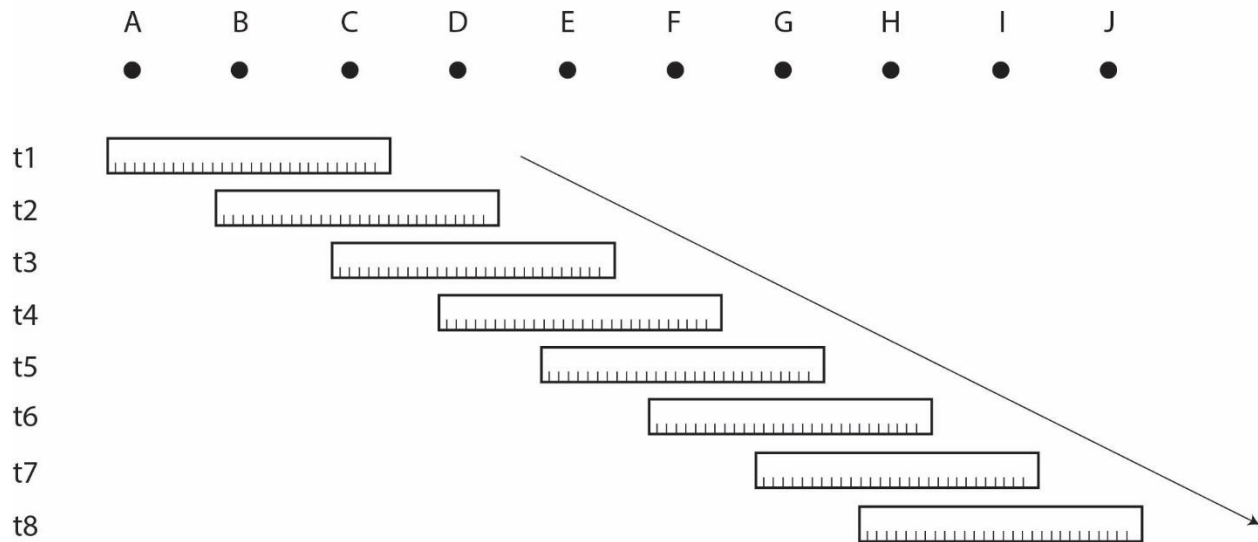


Fig. 11.5. A Representation of William James' Sliding Ruler

This figure is meant to convey William James's ruler analogy for the overlapping units of conscious experience. The ruler encompasses a set of dots. As the ruler slides down a line of equidistant dots the set it encompasses is updated iteratively.

Implications of the Proposed Model

11. Iterative Updating Provides Structure to Associative Search

Donald Hebb (1949) first posited the idea that a group of cells firing simultaneously could represent a memory fragment in the mind for as long as the neurons remained in an active state. He called these groups of coactive cells "assemblies." Today, many neuroscientists describe cortical architecture as essentially a network of hierarchically organized pattern-recognizing assemblies of this kind (Gurney, 2009; Meyer & Damasio, 2009; Johnson-Laird, 1998; von der Malsburg, 1999). To recognize a complex entity, the network uses hierarchical pattern completion to locate and activate the group of assemblies that best represents the statistical function of the entity's constituent parts (Hawkins, 2004; Kurzweil, 2012).

On this groundwork and that of the foregoing sections, the present model proposes that the engram for an item of working memory consists of a large set (ensemble) of cell assemblies located in multimodal association areas (where cells are capable of encoding complex conjunctive patterns). This ensemble of cells is not a stable, immutable symbol but a fuzzy set that varies every time the concept it encodes is activated (Reser, 2016). Thus, an individual item in the FoA would correspond to an ensemble, a distinct subset of the total set of assemblies active in that instant. The assemblies constituting this ensemble would be strongly interconnected, have strong interactions between them, and have the tendency to be added to (or subtracted from) working memory as a discrete group (Fig. 12).

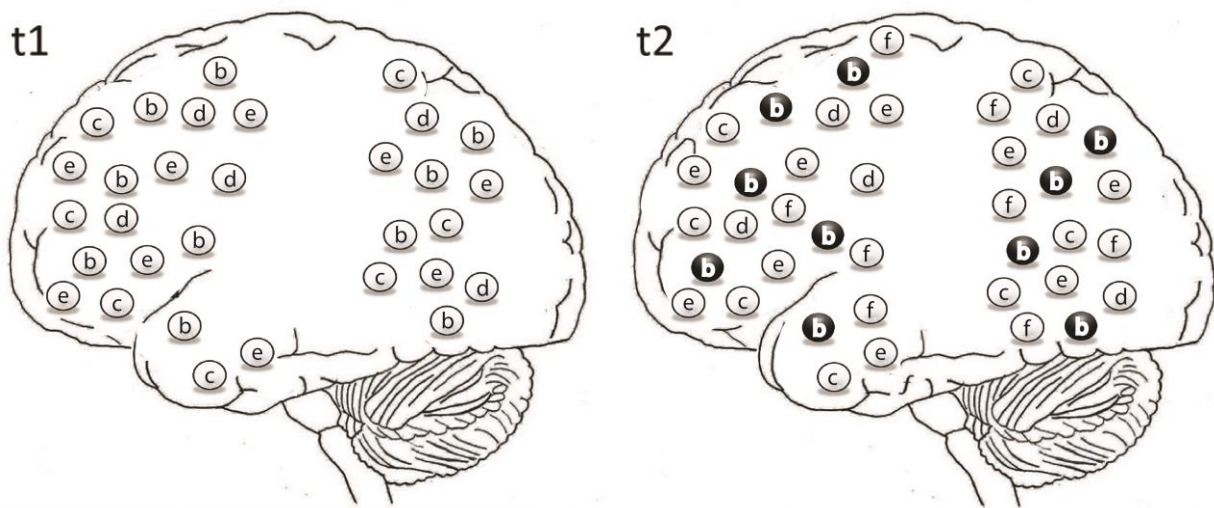


Fig. 12. Two Successive Instances of Coactive Assemblies in the FoA

The engrams for items B, C, D, E, and F are each composed of many assemblies of neurons active in association areas, represented by lowercase letters b, c, d, e, and f, respectively. At time 1, assemblies b, c, d, and e are active. At time 2, the assemblies for b have deactivated and those for f have become active. Thus, time 2 is an iterated update of time 1.

The primate neocortex can hold a number of contextually related items coactive for several seconds at a time. This model proposes that these items are used to perform a global search function by spreading the combined electrochemical activation energy of their neural assemblies throughout the thalamocortical network. This activation energy converges on and activates inactive items in long-term memory that are most highly associated with the current state of activity. This is similar to the case where being exposed to the words “course,” “current,” “wet,” and “bank” might result in the involuntary search for, and activation of, the brain’s representation for the word “river.” This model views each instantaneous state of active items in working memory as both a solution to the previous state’s search and a set of parameters for the next search.

This description of search is compatible with spreading activation theory. According to that theory, the capacity for search in associative networks is derived from activation energy (in the form of action potentials) produced by active neural assemblies (Anderson, 1983). Some of this energy is excitatory, and some is inhibitory. Activation energy from active assemblies spreads in parallel to inactive assemblies that are structurally connected to (i.e., associated with) the active ones due to a history of Hebbian plasticity (Collins & Loftus, 1975).

This activation energy propagates among assemblies through axons and dendrites and follows the weighted links of synapses. Ultimately, multiple alternative pathways, originating from assemblies representing distinct items held in working memory, converge on several of the same inactive items in long-term memory. The number of items that have been converged on may be exceptionally large; however, not all of these can enter the FoA. The item(s) receiving the most excitatory energy is activated, becoming an iterative update to the FoA.

Studies of semantic priming show that either conscious or subliminal exposure to a brief stimulus can temporarily increase the implicit availability of many associated concepts within long-term memory (Bargh & Chartrand, 2000). For instance, in a lexical decision task, merely priming the word “water” will speed up the recognition of various related words such as “fluid,” “splash,” “liquid,” and “drink” (Schvaneveldt & Meyer, 1973). The common interpretation of these findings is that the activation of the engram for “water” unconsciously spreads to the engrams for many semantically related words. This activation is rapid, automatic, and irrepressible, and is theorized to be due to extensive spreading activation in associative networks of conceptual nodes (Reisberg, 2010).

It may be reasonable to assume that updating working memory with a new item has a similar priming effect on spreading activation. This new item, added to the residual items, acts as an additional semantic retrieval cue, uniquely altering the field of potentiated items. The item(s) that receive the most activation energy (i.e., the most strongly associated or connected) will be activated, enter the FoA, and partially update working memory. By assuming that updates to working memory are selected by its current contents, one can explain why new associations are marked by high contextual relevance and specificity. Combining this assessment with the claims made earlier regarding iteration results in a system suited for producing a parade of complementary impressions, views, notions, and ideas.

12. Multiassociative Search Spreads the Combined Activation Energy of Multiple Items

“Associationism” is a longstanding philosophical position advocating the idea that mental states determine their successor states by psychological associations between their contents. According to associationism, the sequence of ideas a person produces is largely a matter of the preexisting links between stored memories. One idea is believed to succeed another if it is associated with it by some shared principle (Shanks, 2010). William James believed that one thought can induce another by way of a logical, correlative connection (1890). The face validity of associationism stems from the commonplace notion that one thought “suggests” the next.

In his discussion of “the succession of memories,” Plato suggests three principles of association: similarity (resemblance), contiguity (in time and place), and contrast (difference). Numerous other principles capable of linking mental states were added to this list by the nineteenth century, including simultaneity, affinity, reinstatement of the remainder, cause and effect, reason and consequence, means and end, and premise and conclusion (Hamilton, 1860). When any of these forms of association occur, they may simply involve an iterative update, selected by spreading activation, to join a global workspace of persistent items.

The associationism school of thought primarily focused on a single logical associative relationship between one thought and another. This may provide only a limited explanation. The model presented here can be read as a version of associationism that escapes this limitation through the assumption that all the neurons currently involved in working memory search cooperatively and probabilistically for the succeeding association. This cooperative search strategy may occur regardless of when the neurons started firing and irrespective of the item to which they belong. Thus, contiguous states are not only interrelated but are also interdependent.

Reser (2016) proposed that the selection of new items to be added to working memory might derive from this pooling of assembly activity in the cortical workspace. This autonomous

process, which is termed “multiassociative search” here, operates as follows: as the activation energy from assemblies representing the items currently in working memory spreads, (1) items that continue to receive sufficient activation energy remain active, (2) items that receive sufficiently reduced activation energy (or are inhibited) lose activity, and (3) inactive items that receive sufficient activation energy become active (Fig. 13). The item(s) receiving sufficient activation energy (through spatial and temporal summation) from both the present constellation of coactive assemblies (FoA) and potentiated synapses (STM) may be recalled autoassociatively (i.e., an active subset of the item’s assemblies is sufficient to activate the rest of the item). This nonlinear, stochastic process should be taken to be responsible not only for finding and activating the next item(s) but also for determining the percentage of items updated in the FoA (Figs. 9 and 10).

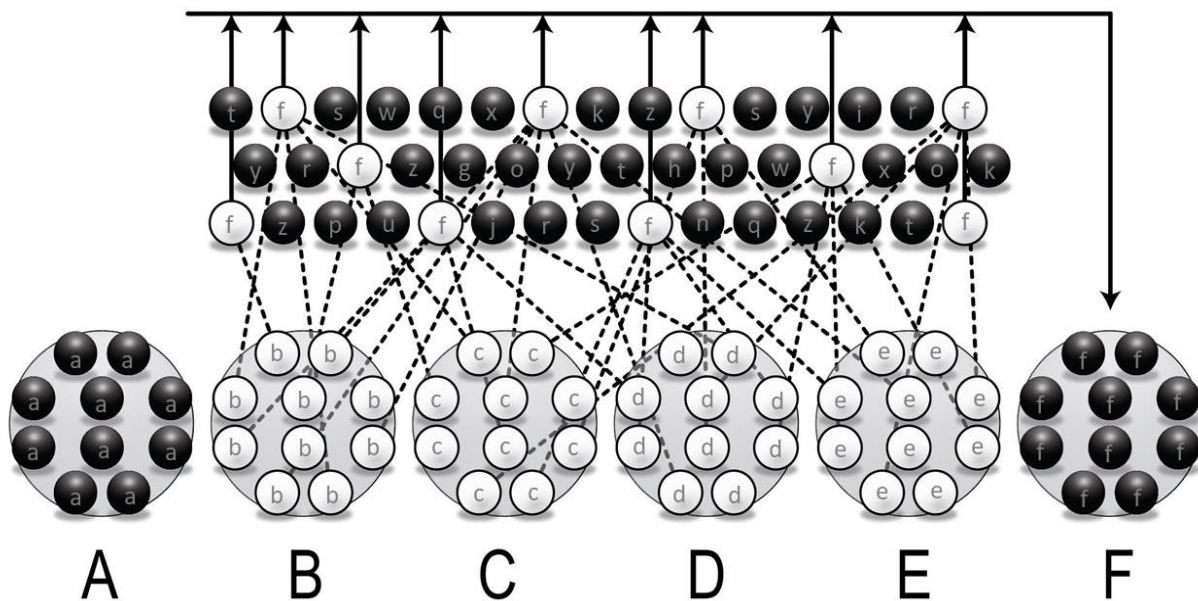


Fig. 13. A Schematic for Multiassociative Search

Spreading activity from each of the assemblies (lowercase letters) of the four items (uppercase letters) in the FoA (B, C, D, and E) propagates throughout the cortex (represented by the field of assemblies above the items). This activates new assemblies that will constitute the newest item (F), which will be added to the FoA in the next state. The assemblies that constitute items B, C, D, and E are each individually associated with a very large number of potential items, but as a unique group, they are most closely associated with item F.

This nonlinear, stochastic, search-and-replace procedure is executed unceasingly during waking consciousness. It takes sets of neural assemblies that have never been coactive before and uses their collective spreading activation to select the most applicable iterative update. At every moment, the set of assemblies in coactivity is unprecedented; however, at times, the set of items in coactivity is not. When the set of coactive items has been coactive at some point in the past, the spreading activity either converges on the same item that was selected the last time (recall) or it may converge on an altogether different item (inference). Regardless of which way this occurs, the process transforms the latent information inherent in the original set into new, manifest information by forcing it to interact with inert long-term memory. Each set of coactive items and

the links between them can potentially be recorded to memory. Thus, multiassociative searching gives rise to multiassociative learning.

It is important to point out that these concepts explain how long-term semantic memory might be updated. New memories don't replace old ones; rather, they retune the connectional strengths between groups of items. For instance, in Figure 13, the associative relationship between F and B is strengthened, but mostly only so in the presence of some or all of the items C, D, and E. As items demonstrate coactivity within working memory, we should expect their assemblies to exhibit a Hebbian propensity to wire together, allowing groups of items to form the kinds of statistical codependencies that would support such learning. Reoccurring examples of coactivity would lead to the formation of heavily encoded associations (Asok et al., 2019), which would persist as iterable procedural and semantic knowledge.

Undoubtedly, many canonical information processing algorithms not mentioned here (see, e.g., Miller et al., 2018; Sreenivasan, 2019) also contribute to this search and play causal roles in this process. However, it may be parsimonious to assume that the subsymbolic components of the symbolic items of working memory work synergistically and in parallel to search for the updates to working memory in this way. In other words, the production sequence of thought is not determined by semantic dependencies between symbols (e.g., rules, utilities, predicates, conditionals, functions, etc.) as in other cognitive architectures (e.g., ACT-R, Soar, Sigma, etc.). Instead, it is determined by syntactic dependencies among subsymbols. These dependencies may reconcile with declarative, symbolic knowledge at the psychological level. Nonetheless, they operate unconsciously below it. In other words, the outcomes of these “blind” statistical searches only appear rational because they are based on a history of structured learning from orderly environmental patterns.

Note that, in the present model, the assemblies constituting items currently in the FoA are not the only contributors to the selection of the next item(s). Rather, all firing neurons that participate in the spreading of activation in the cortical workspace contribute definitions to this global search. Potentiated neurons in the short-term store—as well as active neurons in sensory and motor cortex (semantic), hippocampus (episodic), basal ganglia (procedural), and other cortically connected subcortical brain regions—all contribute to the multiassociative search. Figure 14 depicts this situation, in which a working memory store characterized by iterative updating selects its updates using spreading activation generated by several different neuroanatomical systems.

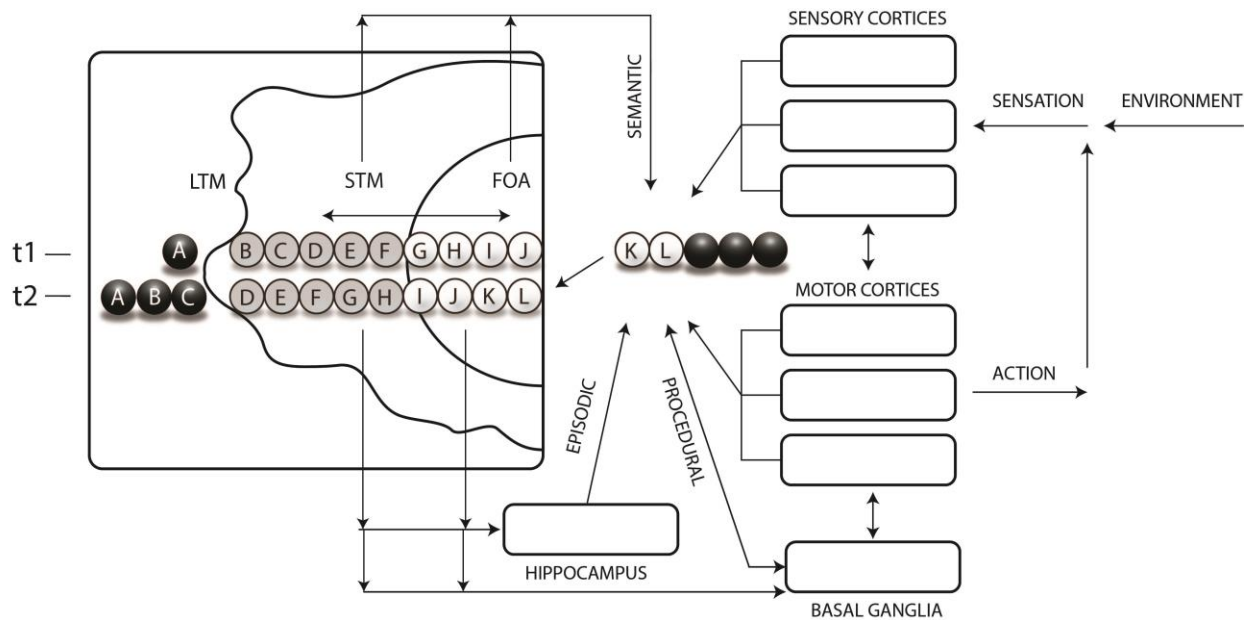


Fig. 14. A Single Cycle of the Iterative Updating Procedure

The FoA, the short-term store, as well as active neurons in the hippocampus, basal ganglia, sensory and motor cortices all contribute to the spreading activation that will select the next item(s) to be added to working memory. At time 1, two (K and L) of a potential five items are converged on, and these update the FoA in time 2.

This concept of multiassociative search demonstrates how prior probability encoded in the network by experience could be used to derive conditional (if-then) rules that remain sensitive to new network weights. The next section will discuss how iterative updating and multiassociative search may work together to formulate not only associations but also predictions.

13. States Updated by the Products of Search Are Predictions

The meaning of an event is determined by the events that came before it and by those that will come after it. Claude Shannon, the founder of information theory, knew this and was interested in predicting events based on their context. He introduced a hypothetical situation in which a person is tasked with guessing a randomly selected letter from a book (Shannon, 1951). Because there is no contextual information available, any response would be highly uninformed and made by chance. But if this person is given the letter that comes before the unknown letter, a more informed guess can be made. The more previous letters are known, the better the guess (Stone, 2015). For instance, if you knew that the sequence of letters that precede an unknown letter was “t,” “h,” “i,” and “n,” then you would know that there is a high probability that the letter you were trying to guess could be either “k” or “g.”

As with letters in a word or words in a sentence, events occurring along a timeline in a natural environment are not independent or equiprobable. Rather, there are correlations and conditional dependencies between successive events. Knowledge of conditional dependencies allows us to predict what other people are going to do next and to finish their sentences for them. The combination of iterative updating and multiassociative search enables working memory to capture and record long-term dependencies. This in turn permits working memory to treat events as causally related variables that can be used to predict future events. By capturing the statistical

structure of a sequence of recent events (including rewards and punishments), working memory provides animals with a way to form an autoregressive interpretation of an unfolding scenario, forming associative expectations about it and responses to it.

The interaction between iterative updating and multiassociative search may form the basis for prediction in the brain. Consider the case in which four environmental stimuli present themselves in quick succession. This could involve a sequence of events involved in finding food. If each stimulus is attended to and persistently activated, then the items representing these stimuli will have the chance to come in the FoA. Their coactivity may cause them to exhibit activity-dependent plasticity even though they never actually occurred simultaneously in the environment. If this sequence of stimuli is repeated frequently (as would be expected if there were conditional dependencies between them found in nature), then they will come to be strongly associated. The next time the first three stimuli appear, their very activity may be sufficient to search for and recruit the item representing the fourth stimulus from long-term memory. This may happen even before the actual appearance of the stimulus the item represents. Consequently, the activation of this fourth item would be a prediction. This may be true whether working memory is modeling external stimuli, internal representations, or some combination of the two. Therefore, internally generated, self-directed thought can be conceptualized as an iterative procession of concatenated, associative predictions, each based on the prediction before it.

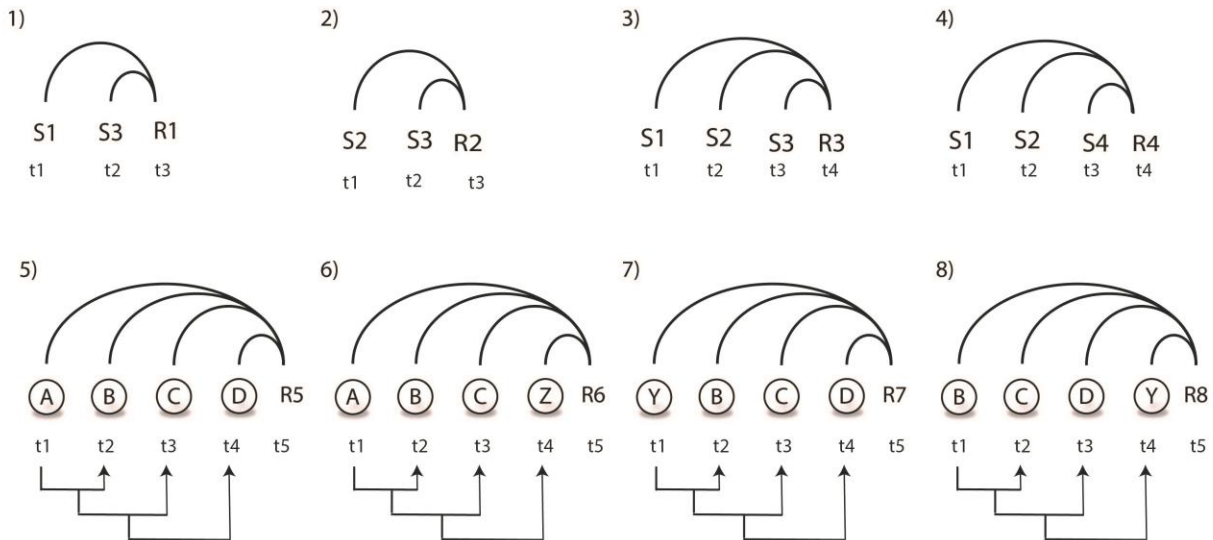


Fig. 14.5. Conditional Dependencies Between Consecutive Events

Each arc represents the span of time since an event occurred. S represents stimuli, R represents responses, and other capital letters represent items. To provide an illustrative example, the variables named above could correspond to the following events: S1 = friend, S2 = enemy, S3 = approach, S4 = depart, R1 = act friendly, R2 = act aggressive, R3 = wait, R4 = follow, A = foraging alone, B = feel hungry, C = find berries, D = not poisonous, Z = poisonous, Y = friend approaching, R5 = eat, R6 = don't eat, R7 = share berries, R8 = eat berries before friend arrives.

In Figure 14, Diagram 1 depicts a situation in which stimulus 1 (S1) is followed by stimulus 3 (S3) and results in the selection of response 1 (R1). This can be contrasted with Diagram 2,

where S3 is preceded by a different stimulus (S2) and a completely different response is selected (R2). The persistent activity of the first stimulus influenced the interpretation of S3, biasing the response accordingly. That is to say, the response to S3 is conditionally dependent on the stimulus that precedes it. Diagrams 1 and 2 have been adapted from a popular model of PFC function (Miller & Cohen, 2001). Diagrams 3 and 4 take this idea further, communicating that when the first two stimuli are the same (S1 and S2) but the subsequent stimulus differs, the responses may also differ. These diagrams underscore the hypothesis that behavior is not merely directed by the differential selection of existing neural pathways underlying stimulus-response pairings (i.e., Miller & Cohen, 2001), but rather by a series of multiassociative searches that utilize sets of stimuli to select the best response at each time step.

Diagrams 5 through 8 communicate that the full complement of items in working memory can be expected to show a pattern similar to that seen with the stimuli in Diagrams 1 through 4: each item affects the interpretation of the items after it and uniquely biases the search for a response to them. Accordingly, the arrows below Diagrams 5 through 8 indicate that, at each time step, the preceding items provide a frame of reference by which subsequent items are interpreted. Note that even though the responses in Diagrams 7 and 8 are reacting to the same four representations (B, C, D, and Y), they react to them differently because the order of items contextualizes the scenario differently. For instance, item C has a different meaning (dependency) when it follows Y versus when it precedes Y. Therefore, it elicits a different response in Diagram 7 relative to Diagram 8.

Consider a situation in which a person is writing with a pencil and the lead breaks. This may cause the long-term memory representations for “writing,” “pencil,” “lead,” and “broken” to become active in the FoA. This combination of coactive items (conditioned from years of writing with a pencil in school) might result in the automatic spreading of activation to the representation for “sharpener.” During another round of updating, the representation for “writing” may exit the FoA and be replaced by the pencil sharpener’s location, such as “desk drawer.” In this way, sets of coactive items can prompt others in advancing sequences capable of producing not only predictions but also adaptive behaviors. Especially when the items that are sustained are task-relevant, this kind of iterative system should be capable of incremental advancement toward a goal.

14. Iterative Updating Allows Progressive Changes to the Contents of Working Memory

In addition to accounting for the serial, cyclic, continuous, narrative, and predictive functions of thought processes, iterative updating may be a fundamental feature of reasoning. This section will provide brief explanations for why this might be the case. According to the present model, iterative updating produces sequences of interdependent states in which each state is capable of representing the current status of a problem-solving procedure and updating it with a prediction. This makes it possible for a starting state to generate a chain of intermediate states that make progress toward a terminal goal state.

When specific items persist in working memory and are reused repeatedly to drive iterative updates, the resulting states will be closely contextually related. When the items that persist are related to the same task or objective, it can be expected that the resulting states will all describe the task in some way. When the associative updates are informed by meaningful causal

dependencies learned from related experiences from the environment, the sequence of resulting states may demonstrate logical progress that approaches a solution.

Iterative updating allows working memory to link a series of rapid, automatic associations so that they can furnish a foundation for each other, resulting in the assembly of complex content. This occurs when a series of linked searches culminates in a higher-order solution that could not otherwise be attained by any single search on its own. A prolonged stretch of tightly recursive searches (where a large proportion of items are retained throughout several states, as in Figure 10, Diagram 2) may be slower and more error-prone but is capable of addressing problems too unfamiliar or complicated to be solved by less iterative, implicit processing.

Generally speaking, short bouts of iteration engage crystallized intelligence and easy-to-reach network states, whereas instances of prolonged iteration access fluid intelligence and highly processed, difficult-to-reach states. Such highly elaborated states are comprised of select subsets of previous states from various points in the recent past. This corresponds to simple thoughts building constructively “on top of” each other to form complex thoughts. William James used the term “compounding” to describe this concept:

...complex mental states are resultants of the self-compounding of simpler ones.... in the absence of souls, selves, or other principles of unity, primordial units of mind-stuff or mind-dust were represented as summing themselves together in successive stages of compounding and re-compounding, and thus engendering our higher and more complex states of mind.” (William James, 1909, p. 185)

The compounding feature of iteration may also enable working memory to implement algorithms for use in reasoning and problem solving. All complex learned behaviors have algorithmic steps that must be executed in a specific sequence to reach completion (Botvinick, 2008). Hunting, foraging, tying shoes, and performing long division each involve following an algorithm. Successive states of working memory could correspond to successive steps in an iterable process.

Iterative updating could be instrumental in implementing learned algorithms because virtually every step of an algorithm relates to the preceding and subsequent steps in some way. A new update could correspond to a behavior or mental operation required in the next step of the sequence of actions that need to be taken. An item that is inhibited or allowed to subside could correspond to an operation that has already been executed or is no longer needed. This update could amount to an action, a memory, a heuristic, or a schema, or provide top-down influence to a perception. Thus, multiassociative search converges on the most appropriate fragment of knowledge at each state of solving a non-routine problem.

Once the associations relevant to an algorithm have been learned and trained, multiassociative search during each state would recruit the items necessary for the next step (Reser, 2016). For instance, performing long division by rote requires many trials, and proficiency may only be reached when the active items in each state have been trained to converge on the items necessary to perform the operation in the next step. For example, after the first digit of the dividend is divided by the divisor, the prevailing state of working memory automatically activates the items necessary to take the whole number result and write it above the dividend. Cognitive algorithms may be constructed in this manner during learning, as trial, error, and repetition link recursive chains of states capable of assembling functional behaviors. At its core, this is a form of

optimization that may use operant conditioning to provide feedback for incremental guess refinement.

Iterative updating could conceivably play a role in the generation of mental models. Mental models are internal representations of external systems and the relationships between their parts (Cheng & Holyoak, 2008). Iteration may afford the incremental modification of a model from its previous state, allowing simple, static models to be elaborated on dynamically. Even dynamic systems could be modeled when their enduring features are held constant by persistent activity and the changing features are updated correspondingly. This enables tweaking of the search parameters of interest to vary the simulation in stages, producing a systematic effort to investigate a structured problem space and possibly generalize from one scenario to another.

Iterative updating may employ this compounding feature during logical or relational reasoning. The item or items that update the FoA create a context to be compared, contrasted, integrated, or otherwise reconciled with the context remaining from the previous state. This may be the same kind of reconciliation that occurs in formal logic. For instance, propositional logic combines simple statements using logical operators (subjects and predicates) and connectives (e.g., and, or, not, if, then, because, etc.) to produce complex rational statements. The operators of such a statement could be instantiated by items (and their subsymbolic assemblies). This group of coactive items could imply a true statement or premise that, when updated in the next state, could invoke another premise or lead to a conclusion. By creating strings of substantiated inferences in this way, multiassociative search could permit the construction of a logical case or argument, form new boundaries and affinities between groups of items, and build expectations about events that have never been encountered.

Solving a complex multistep problem requires the FoA and short-term store to cooperate. For instance, one line of thought developing in the FoA may be temporarily suspended in the short-term store so that the FoA can be made available to solve a related subproblem. The FoA would iterate multiassociatively, progressing toward the solution to this subproblem. When the subsolution is reached, it could then be merged with the pending problem to create a hybrid solution state. This interleaving and eventual merger of states of progress would facilitate the decomposition of a problem that is too computationally taxing to be processed by the FoA alone (Fig. 15).

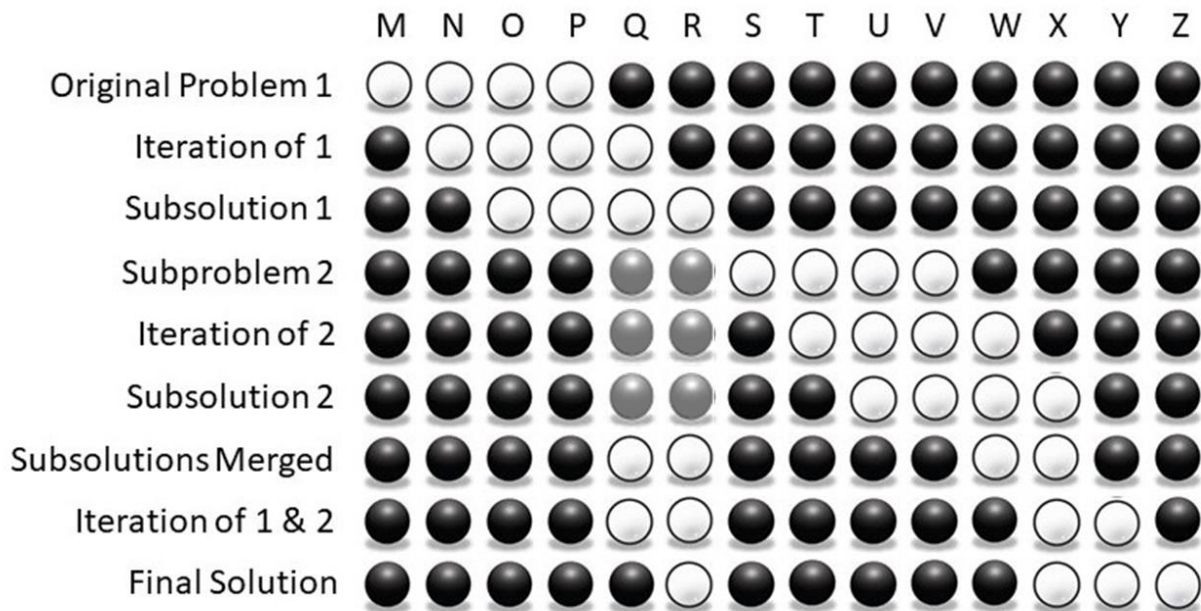


Fig. 15. Merging Subproblems in Working Memory

An original problem is activated (M, N, O, P), and iterative updating is used to reach a subsolution (O, P, Q, R). This subsolution is saved in the short-term store, and a related subproblem (T, U, V, W) is introduced into the FoA. This subproblem iterates until a second subsolution is generated (U, V, W, X). Relevant items from the first subsolution (Q and R) are combined with those from the second subsolution (W and X) and iterated to generate a final solution (R, X, Y, Z).

According to this interpretation, the short-term store holds the animal's present objective, and the FoA is used to query lines of reasoning that interrogate that objective. These lines of reasoning are used to update the objective and bring it closer to resolution using an iterative approach. This allows the animal to keep a present opportunity or threat in mind while it considers possible responses before acting. In effect, previous threads of FoA sequences can be suspended in STM (or LTM) as interim results. These can then be retrieved rapidly if spreading activity reconverges on them. This permits working memory to deviate from its default behavior described thus far and employ a form of backward reference and conditional branching.

Another common pattern found in the updating of working memory may occur when an existing problem-solving process reaches an impasse. The newest addition to working memory is sometimes unhelpful or not task-relevant (e.g., due to prepotent associations formed during a similar but irrelevant task). In this case, it may be inhibited. The same items that recruited it would continue to spread activation energy without being able to reactivate it. Multiple rounds of "iterative inhibition" may be required before an appropriate item can be identified (Fig. 16). This kind of situation might arise as one deliberates over different methods of completing the same task (e.g., "I should fax this letter, no I should email it, no I will text it instead"). Thus, each time a potential coactivate is vetted for exclusion, the search tree is restricted further.

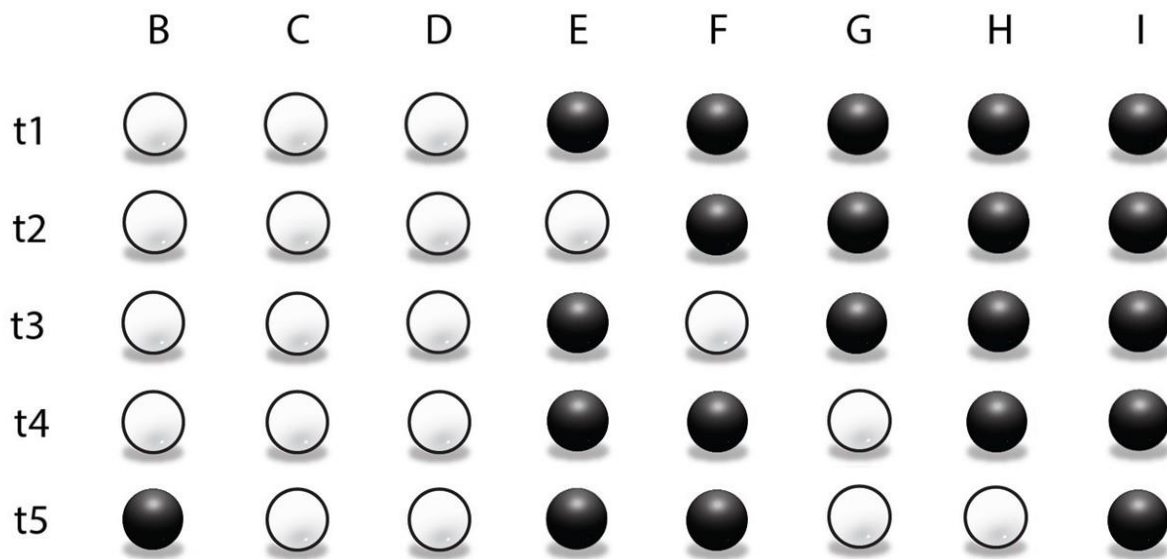


Fig. 16. Iterative Inhibition

An original problem is activated in time 1 (B, C, D), and the spreading activity activates a new item at time 2 (E). Executive processes determine that E is not a suitable behavioral parameter and E is inhibited. With E unavailable, B, C, and D continue to spread activation energy that converges on F (at time 2). The same iterative inhibition occurs with F (at time 4). G is then activated, and iterative updating continues.

This section has considered how iteration of the content of working memory can create progress in information processing. The next section will consider how this form of progress could be used to enhance artificial intelligence.

15. Artificial Intelligence Should Employ Iterative Updating

Many researchers in the field of AI believe brain science will reveal conceptual breakthroughs that will provide essential guidance for the construction of intelligent machines (Haikonen, 2012). Some have suggested that AI may not need to emulate fine-grained molecular or cellular details of the brain to create human-level intellectual function (Bostrom, 2014). Instead, they suggest simulating an abstraction of the neurological mechanisms that produce intelligence (e.g., Hassabis et al., 2017). The present model introduces abstractions that may be useful in the development of intelligent machines. Specifically, the model may help close the “computational explanatory gap,” which is an effort to understand how the parallel, subsymbolic computations involved in low-level neural networks could translate into the serial, symbolic-level algorithms involved in high-level cognition (Reggia et al., 2019). Figures 13 and 14 provide mechanistic accounts of how this could be done. Today, even state-of-the-art AI processing feats are generally only equivalent to a second or less of unconscious human processing (e.g., recognizing objects in a picture) (Goodfellow et al., 2017). To create more generally intelligent AI, these brief processing sessions must be chained together into iterated sequences that more closely resemble thoughts. Iterative updating on its own, however, is not sufficient to elevate computer

information processing to the cognitive domain. Most temporary memory stores, and even random access memory (RAM), are updated iteratively (Comer, 2017). A computer's RAM holds billions of bytes coactive through time, and due to its limited capacity, adds and subtracts from this pool in the same manner as illustrated in Figure 1. However, even though the bytes held in RAM can be considered coactive, they are not "cospreading." Unlike the brain, computers do not use cached information for multiassociative search. There are no modern cognitive architectures or AI systems that do this as described here. Unlike the brain, computers do not make cached information globally accessible for a multiassociative search.

There are advanced AI systems that employ working memory, a global workspace, recursion, and various methods of updating (e.g., Goertzel, 2016). These include cognitive architectures (Gray, 2007), evolutionary computation (Sipper et al., 2018), and soft computing (Konar, 2014) and machine learning. However, such software generally utilizes preprogrammed symbolic rules to transform one state into the next. For this reason, it is usually restricted to formalized, narrow domains of problem-solving (Haikonen, 2003). Artificial neural networks utilize subsymbolic information and do not require preprogrammed rules.

First, we should address the fact that updating a memory store iteratively is quite common in computing and AI. All computers using the Von Neumann architecture routinely update their temporary memory stores (i.e., static RAM, dynamic RAM, virtual memory, etc.). These stores, known as caches, have a strong resemblance to working memory. Cached information includes intermediate results from ongoing processing, as well as data and program instructions from the storage drive. Cache stores have a limited capacity, and because they are constantly tasked with holding new information, they must evict old information. These stores are updated iteratively as the least recently used (LRU) data are replaced (Comer, 2017). However, modern computers do not demonstrate self-directed intelligence, so there must be more to the human thought process than iterative updating. This unexplained factor may be found not in the way cached memory is updated but in the way it is utilized while active.

Digital, rule-based computing systems use cache to speed up the delivery of data to the CPU. However, the next bytes of data processed by the CPU are not determined by the contents of the cache itself. Rather, the instruction sequence is determined by the next line of programmed, executable code. Thus, it is clear that, unlike the brain, computers do not use cache for multiassociative search. The various bytes of data within computer cache memory can certainly be considered coactive, but they are not "cospreading." That is, they do not pool their activation energy to search long-term memory for relevant data as in human working memory.

Artificial neural networks eschew preprogrammed rules. Like the brain, neural networks use parallel, distributed processing to cause connectionist systems of nodes to learn mathematical functions that describe and come to recognize complex patterns. Some neural networks such as recurrent, long short-term memory and transformer networks have nodes capable of persistent activity that permit them to cache previous inputs in a form of working memory (Sherstinsky, 2020). This technology is highly analogous to sustained firing. These networks have been applied successfully in recent years to tasks involving temporal dependencies such as speech recognition, machine translation, and image captioning. However, neural networks are also confined within narrow domains. Each system must be purpose-built and trained for a specific class of unimodal problems (Goodfellow et al., 2017). Most systems are governed by a single

algorithm, and although they can refine that algorithm, they are incapable of developing new algorithms in the manners discussed in previous sections (e.g., Section 16).

To accomplish this, AI working memory should be designed to run iterative updating in lockstep with multiassociative search. If an artificial neural network was engineered to do this in the manner presented in the preceding sections, the resulting system could exhibit some of the qualities and functionality discussed thus far, including association and prediction formation, inference formulation, algorithm implementation, the compounding of intermediate results, progressive modification, and attentive continuity. Because current artificial neural network technology is capable of sustained firing, synaptic potentiation, and spreading activation, everything discussed in this article thus far can potentially be implemented by it.

16. Designing an Artificial Intelligence Capable of Iterative Updating

Implementing iterative updating to a first approximation in an AI system would mean creating a connectionist program with human-like working memory. Spreading activity from information in short-term memory, along with incoming activity from its sensors, would search for related or entailed information from long-term associative memory. This would structure the architecture to be self-organizing and self-sustaining. Consequently, it would not be limited to learning from discrete batches of curated input but could be exposed to continuous data streams from real-world scenarios that unfold through time. Also, the system would not suspend its activity every time it finishes a task. Rather, it would exhibit continuous, endogenous processing. The system's ontological and epistemological development would benefit from embodied, real-time, robotic interactions within physical, social, and intellectual training conditions. During exposure to these conditions, it would engage in unsupervised learning of time-series patterns from unlabeled data on a constant basis.

Iterative updating and multiassociative search may first have to be explicitly programmed into the system using rule-based code until it becomes clear how to design a system in which they will emerge organically as they do in the brain. Hand-coded or not, iterative updating must be defined mathematically and unambiguously to be the basis of computer software.

Iterative/multiassociative search can be expressed as a function (f) that maps input variables (x) of the current state of working memory to an output variable (y) used to update them. Each state of the network would be a search for the update applied to the next state. As a formal algorithm, it could be modeled as a stateless Markov process in discrete time performing a non-deterministic search. As a computable function, it could be instantiated by traditional or neuromorphic machines and executed using brain emulation, hierarchical hidden Markov models, stochastic grammars, probabilistic programming languages, neural networks, or others.

The rest of this section will describe how this system could conceivably be constructed using an artificial neural network architecture. This system could be built using spiking artificial neural networks or fully connected recurrent ones. Either way, network nodes could be used to model the pattern-recognizing assemblies discussed in Section 11. These could be fashioned after cortical columns and perhaps structured in a laminar fashion. Multiple layers of nodes should engage in hierarchical pattern recognition; using nonlinear transformations to build new patterns from more primitive patterns found in nodes lower in the hierarchy (Hawkins, 2004; Kurzweil, 2012). Each level in the hierarchy must build a statistical model of the regularities in the level below it (Eliasmith, 2013).

A Hebbian learning rule would be needed to strengthen the weights between frequently coactive nodes. This would need to work in such a way that groups of highly associated subsymbolic nodes are capable of forming sparse and fuzzy ensembles. These symbolic ensembles would be used to represent invariant, categorical patterns. Such an ensemble would be equivalent to an item and should be made capable of enduring coactivity with other items within a graph-structured global workspace that uses an analog of neural binding (i.e., Klimesch et al., 2010) and synchronized, reentrant oscillations (Edelman, 2004) to integrate (i.e., Tononi, 2004) and unify them into a singular situational representation. This would amount to an emulation of the FoA.

Nodes at the top of the hierarchy would constitute high-order patterns due to having receptive fields composed of various inputs from multiple layers of increasing complexity. These abstract nodes should be capable of persistent activity simulating the sustained firing of pyramidal cortical neurons. This would keep the items these nodes code for active so that they can remain as search parameters and dependency markers, as well as contribute to contextual structuring for extended periods. When the simulated sustained firing abates, the nodes should then simulate synaptic potentiation. This would enable the network to maintain pertinent items in an emulated short-term store, as cached assets. Nodes potentiated in this way would continue to bias the multiassociative workspace until they are either promoted back to the FoA or demoted back to inert long-term memory.

The simulated FoA and short-term memory stores would undergo iterative updating such that the overlap of persistent information is congruent with Figure 6, information replacement is congruent with Figure 14, and information selection is guided by multiassociative search as in Figure 13. Each update would amount to a truth-preserving associative transition in the processing stream underwritten by the structural properties of the network, which are based on past statistical analyses of reliable patterns from the physical world. More than these two levels of working memory storage could be used. They could utilize different methods of item/assembly potentiation and include the following embedded levels of temporary storage: (1) binding, (2) the FoA, (3) the current goal, (4) recent goals, (5) recent situational context, and (6) recently primed concepts.

An implementation of this system would necessitate modular specialization. Each module would correspond to a separate neural network meant to simulate a different specialized cortical or subcortical area of the mammalian brain. These separate networks would interconnect to form a single dynamical system. Coordinating this kind of system to implement the multiassociative algorithm would be a considerable engineering problem. Given that the human brain accomplishes this task, human neuroscience should be used as an archetype. Thus, the system could be constructed biomimetically and arranged according to general neuroanatomical connectivity.

Not only would the nodes of each modular network be organized hierarchically, but the connections between networks would establish an even larger hierarchical structure. This stratified organization, beginning from unimodal networks and progressing up to densely conjunctive multimodal networks, would mirror the gradient seen from sensory cortices to association cortices. Networks higher in the hierarchy would refer to larger space-time regions and multidimensional levels of abstraction. The networks could be designed to emulate specific

human cortical modules if they recreated anatomical connectivity in terms of intrinsic, extrinsic, short-range, and long-range connections, along with the relevant proximities and proportionalities. Multimodal areas that may be pivotal to human thought and, therefore, in need of being reverse engineered in this way, might include the angular gyrus, Wernicke's area, Broca's area, the dorsolateral PFC, the medial PFC, the supplementary motor area, and the frontal pole.

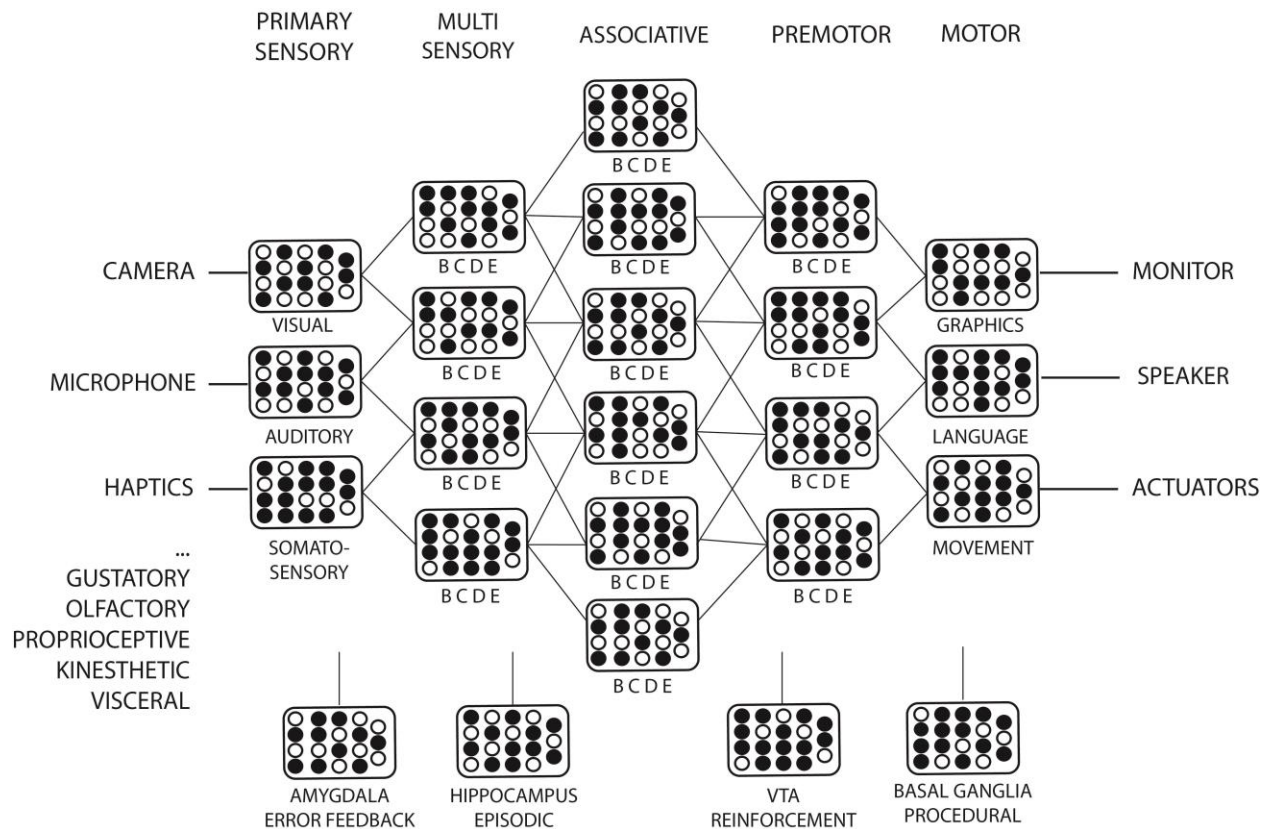


Fig. 17. Artificial Neural Network Implementation of Iterative Updating

Each enclosed set of circular nodes represents a specialized neural network wired to receive a different modality of input. Networks at the bottom (left) of the hierarchy take input of a single modality from the environment. Other networks take input from multiple neural networks below them in the hierarchy. Spreading network activity would oscillate between the top and bottom of the hierarchy while allowing recurrent feedback between and within networks, creating a serial cognitive cycle. This figure features 24 networks, each with 19 nodes. An actual build would necessitate dozens of networks, each with millions of nodes.

Each network, or module, in the system would take inputs from global working memory and use them to create their own unique corresponding set of outputs that have the potential to contribute to the next update. The outputs of primary sensory networks should be constructed as topographic maps (retinotopic for vision, tonotopic for sound, etc.). By creating a series of internally generated maps to match the iterative updating taking place in association areas, these systems could produce sequences of mental images. There are already reliable methods for using

neural networks to generate such “self-organizing” maps (Hameed et al., 2019), and video imagery generation by inverse neural networks is very common today (Byeon et al., 2018). First, let’s look more closely at how this happens in the cerebral cortex.

Neurons in sensory cortex respond to perceptual features from sensory input and fuse them into topographic mappings (Moscovich, Chein, Talmi, & Cohn, 2007). The topographic maps created in this way hold many types of precategorical information (e.g., metric and compositional). In addition to creating topographic mappings from patterns recognized in the external environment, sensory areas are thought to combine top-down inputs from association cortices to generate internally derived scenery (Mellet et al., 1998; Miyashita, 2005). Imaging research supports the idea that imagining something in the “mind’s eye” literally activates existing maps in early perceptual networks (Damasio, 1989; Hasegawa et al., 1998; Ohbayashi et al., 1999).

This kind of top-down control is possible because association areas project densely to several self-contained, topographic mapping areas (Klimesch, Freunberger, & Sauseng, 2010). These include specialized cortical modules for each of the senses (Kaas, 1997). Association areas also interact reciprocally with topologically organized motor areas, including the motor strip, supplementary motor, and premotor areas. This suggests that items in working memory can actively contribute specifications, not only to mental imagery but also to muscular automatisms and behavior in general.

Brain science researchers have long concluded that complex cognitive processes involve reciprocal signaling between early sensory areas and late association areas (Klimesch, Freunberger, & Sauseng, 2010). Information reverberates back and forth between early, bottom-up sensory cortex (where activity is metric, topographic, and transient) and top-down association cortex (where activity is abstract, conceptual, and persistent) (Christophel et al., 2017).

Topographic maps may use low-order perceptual knowledge to depict associative relationships between higher-order items held in persistent activity. This may have happened to you when in Section 10 you imagined your friend shopping at the grocery store. The mental imagery may have injected new content into your train of thought (such as features incidental to the image itself). For example, imagining your friend shopping for milk might have involved the automatic construction of a map depicting a prototypical dairy section by primary and secondary visual areas. This could potentially lead to thoughts about yogurt, cream, or butter. It may be very common for early sensory areas to contribute content to working memory in this way. The output of these modules is determined by a form of multiassociative search carried out by the activity spreading within them from their inputs and also from their own limited capacities for persistent activity.

Based on these assertions, it is reasonable to assume that early sensory cortices produce sequences of mental images that build upon one another, thus contributing to the development of mental models. These sequences of images are symbolically related, specifically because the imagery is controlled by items sustained in association areas that are updated iteratively (Reser, 2011, 2012, 2013). Therefore, as a given set of items is updated, the set of lower-order sensory targets held in synchrony with it is updated correspondingly. In other words, after a mental image is formulated, it is likely to be replaced by another image that uses many of the same working memory items as constraining parameters. This process infuses continuity into both imagery and imagination. Such hierarchical crosstalk, marked by mutual interactions (i.e.,

reciprocal causation), may result in iterative progress, such as when you can see and hear a hypothetical situation play out in your head. This has been termed “progressive imagery modification” (Reser, 2016) and is depicted in Figure 17.

As a given set of items in the FoA is updated, the set of lower-order sensory targets held in synchrony with it would be updated correspondingly (Reser, 2011, 2012, 2013). In other words, after a mental image is formulated, it is likely to be replaced by another image that uses many of the same working memory items as constraining parameters. Consecutive maps formed in this way could infuse video-like continuity into the imagery and could amount to a type of synthetic imagination. This form of hierarchical crosstalk between association and sensory areas, marked by mutual interactions (i.e., reciprocal causation), may result in allowing an AI system to use mental imagery to see, hear, and model a hypothetical situation. This process has been termed “progressive imagery modification” (Reser, 2016) and is depicted in Figure 18.

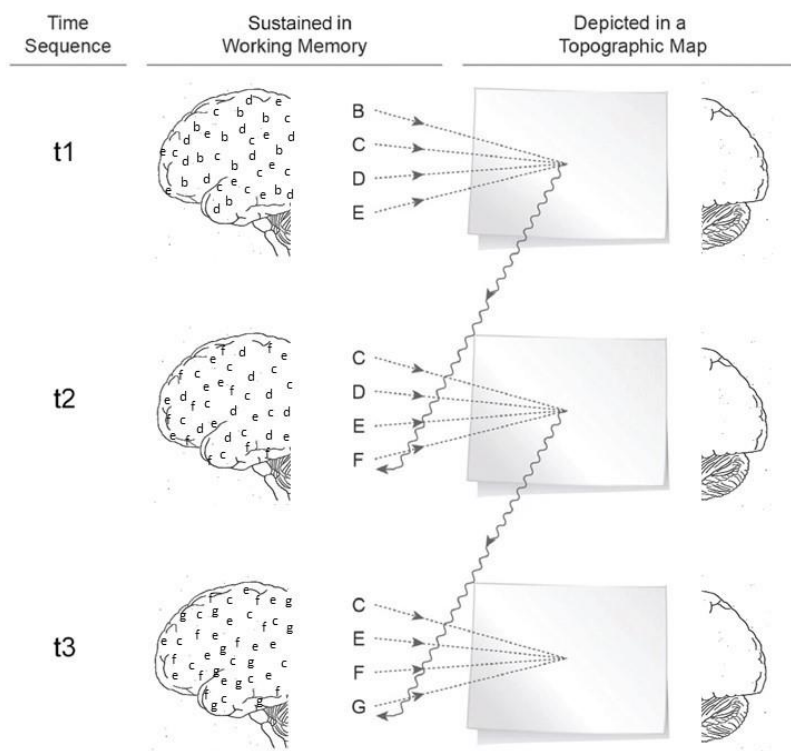


Fig. 18. Progressive Imagery Modification

At time 1, items B, C, D, and E, are active in association networks. The spreading activation from these items provides independent yet interactive top-down bias signals to primary visual networks where a composite topographic map (i.e., sketch) is built based on prior experience with these items. This gestalt sketch will introduce new relevant content to working memory. At time 2, salient features created by the map from time 1 spread activation energy up the hierarchy, converging on the assemblies for item F. Item B becomes inactive while items C, D, E, and F diverge back down the hierarchy toward the primary visual network. Then the process repeats itself. Because this process is informed by prior probability, it is capable of creating a logically connected series of related images.

Brain researchers believe that sensory areas deliver information in the form of fleeting sensory maps, whereas association areas deliver lasting perceptual expectations in the form of templates, and that these interact to construct higher-order cognitive processes (Hawkins, 2004; Carpenter & Grossberg, 2003). It could even enable an AI system to develop the kind of interplay between the central executive and the visuospatial sketchpad/phonological buffer that is characteristic of the Baddeley (2000, 2007) model of working memory (Fig. 3). Further, this process of iterative modification could take place in language areas (where it is involved in the construction of speech), motor areas (where it is involved in action sequencing), and prefrontal areas (where it is involved in planning).

This general design could form the basis of a major security precaution promoting AI safety. To human observers, the representation of knowledge in neural networks is distributed in such a complex manner that it is mostly inscrutable (Castelvecchi, 2016). This lack of transparency heightens fear of superintelligent AI because it would be impossible to tell whether the AI was secretly harboring hostile motives (Bostrom, 2014). However, if the system was inherently obligated to build a composite topographic map of each state of working memory to initiate and inform the next state, then these maps could be displayed on a television monitor for humans to view. A history of all visual and auditory maps could be saved to an external memory drive. This would ensure that all the AI system's mental imagery and inner speech is recorded for later inspection and interpretation. Hostile intentions would not have to be deciphered; they would be plain to see.

16. How to Train an Artificial Intelligence that Employs Iterative Updating

The architecture described in the last section would necessitate a compatible reward function to guide reinforcement learning and credit assignment. Said function should be based on the circuitry of the mammalian dopamine system, including the ventral tegmental area (VTA). In mammals, novel appetitive or aversive events increase dopamine release. In turn, increased concentration of ambient dopamine leads to increases in sustained firing (Seamans & Yang, 2004). It is thought that dopamine neuromodulation acts in this way to allow mammals to prioritize information about unique opportunities and threats (Seamans & Robbins, 2010). Thus, representations of rewarding, punishing, salient, uncertain, or unpredicted events should elicit persistent activity in the AI system's working memory.

An analog of the dopaminergic system's network would be needed to recognize groupings of appetitively stimulating items and prioritize them by sustaining their activity. This would allow pertinent groupings with incentive value to bias processing for extended periods. Mammals are generally driven by predictors of food, sex, and pain. If we want a superintelligent AI that can further human understanding, then we should design its appetitive system to be driven to mine information from literature and databases, then turn that data into knowledge. Thus, this system would need an algorithmic form of curiosity for reading and research.

Functional, programmed instincts (akin to an infant grasping something when its palm is touched or smiling when smiled at) would be built into the direct connections between sensory and motor areas. This innate programming could come in the form of already-trained neural networks that perform useful cognitive tasks (e.g., scene classification, paragraph comprehension, or natural language generation) embedded into the bottom of the hierarchy of this much larger network. These could orient the system toward meaningful tasks, just as reflexes and prepared learning set

animals on the right track during early development. The machine would use operant feedback about its performance on these tasks to bootstrap and then program the rest of the network.

Maturation of the AI's neural network should approximate that observed in the human cortex. It should start by simulating the brain of an infant (Fuster, 2015). Initially, motor output should not be driven by higher-order association areas, but rather by low-order sensory areas. As low-level responses are practiced and refined, and pertinent algorithms are developed through trial-and-error (see Section 14), association networks could be slowly interposed between sensory and motor networks. As in the mammalian brain (Huttenlocher & Dabholkar, 1997), sensory and motor areas should mature (myelinate) early in development, and association areas should mature late. Similarly, the capacity for persistent activity should start low, but increase over developmental time.

Postponing the initialization of sustained firing would allow the formation of low-order associations between causally linked events that typically occur close together in time. This would focus the system on easy-to-predict aspects of its reality (e.g., correlations between occurrences in close temporal proximity). The consequent learning would erect a reliable scaffolding of highly probable associations that could be used to substantiate higher-order, time-delayed associations later in development (Reser, 2016). In other words, the proportion of updating from one state to the next (Fig. 9) would start very high. This would be reversed over weeks to years as an increasing capacity for working memory would be folded into the system.

A working memory store that uses iterative updating would be used to establish associations between clusters of stimuli that appear close together in time from books, articles, lectures, speeches, videos, and experiences. Then, as the length of sustained firing is increased, these temporally proximate contextual representations could be coactivated with other less proximate ones when multiassociative search deems them to be highly probabilistically related (i.e., they share a logical or analogical connection). Thus, two events that were never temporally local in the environment could be selected for joint iterative processing within the FoA (Fig. 15). This kind of reconciliation between separate (previously discrete) threads (subsolutions) of information could build and constantly retune a dynamic knowledge base of interconnected representations. The duration of persistent activity could easily be programmed to outstrip that of humans, allowing the system to capture extremely long-term causal dependencies resulting in the perception of high-order patterns and abstractions that would be imperceptible to humans.

Unlike biological brains, this system would be scalable. There are clear ways to amplify the working memory of such a system beyond the physiological limitations of the human brain. These include: (1) increasing the total number of nodes in LTM, (2) increasing the number of nodes capable of being coactivated in the short-term store, (3) increasing the number of items capable of being coactivated in the FoA, (4) increasing the length of time these can remain active, thereby increasing the half-life and decreasing the rate of updating (Fig. 10), and (5) increasing the number of tightly coupled iterations that can occur before attention is disrupted (Fig. 11).

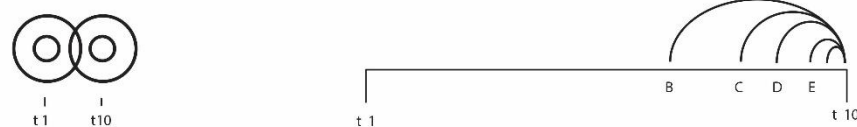
Under conditions of imperfect or incomplete information, the longer the backward memory span and the larger the number of related events that can be used in multiassociative search, the less uncertainty (information entropy) there is about the present state. In information theory, the length beyond which a backward memory span stops providing predictive information is known

as the correlation length (Shannon, 1951; Stone, 2015). The working memory of a species can be seen as having a correlation length beyond which there is little predictive value to be had given its ecological niche. The long correlation length of the human FoA was likely permitted by our cognitively demanding foraging style, selection for social cognition, and the supervised learning, error feedback, and large number of training examples provided by prolonged and intensive maternal investment (Reser, 2006). However, there is no reason to believe that the length or breadth of the human FoA has been optimized for systemizing reality. It was probably constrained by several evolutionary factors that would not apply to computers.

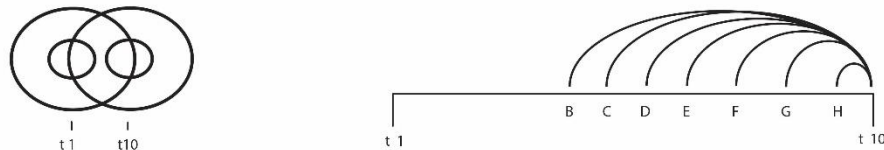
1) Mouse



2) Monkey



3) Human



4) Superintelligent AI

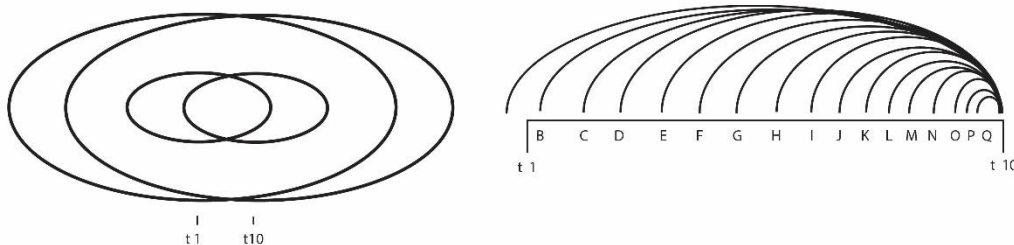


Fig. 19. Venn Diagrams of Working Memory in Different Systems

These diagrams depict informational overlap between states of working memory. The diagrams on the left use the format from Figure 6 while those on the right use the format from Figure 7. Diagram 1 shows zero overlap between working memory at times 1 and 2. This would make it more difficult for system 1, a hypothetical mouse, to make associations between events that are separated by the delay. For example, calling this mouse's name and feeding it 10 seconds later may not condition it to come when called, whereas feeding it one second later might. Training an AI should involve a maturational process where the system begins learning with a very limited working memory span (e.g., Diagram 1) before gradually developing a superhuman capacity for working memory span (Diagram 4) as formative experiences accumulate.

Programming AI working memory stores to maintain relevant items for longer periods than their biological counterparts could increase their ability to make valid associative connections between causally linked events separated by long delays in time (Fig. 19). In other words, an extended attention span could increase the recognition of long-term dependencies, resulting in the perception of high-order patterns and abstractions that are imperceptible to humans.

Prolonging persistent activity in this way could also allow each search to be more specific and informed. This is because searches would be apprised by a larger number of specifications that stretch further back in time. It would also ensure that the system is less likely to allow crucial intermediate solutions to decay from working memory (i.e., a cache miss) before they are needed to form higher-order, compound inferences. The “thoughts” of such a system would be lengthy, highly focused, and tightly interwoven.

Now may be the time to start building large, state-of-the-art, iterative updating networks and training them as we would train a child with the expectation that aspects of intelligence will emerge. It is hoped that through exposure to experiences with systematic patterns, a system like that described above would construct an associative network capable of producing updates to its states of working memory that build functionally on previous states. This could lead to the capacity to make associations between probabilistically related events (Fig. 15), resulting in the discovery of patterns hidden by separations in space and delays in time. Simulating iterative updating and multiassociative search and enhancing them beyond human capacities could be instrumental in the effort to construct artificial intelligence capable of common sense, insight, creativity, machine consciousness, and superintelligence.

17. Discussion and Conclusions

This article aims to introduce a plausible and internally consistent framework for working memory dynamics. It is intended to inspire more detailed hypotheses that can be tested experimentally. Future work should search for the neural signature of iteration within the brain (see Figs. 6, 7, 8, and 11). As shown in Figure 20, this search for a stable yet gradually updated pattern could utilize simultaneous recordings (electrodes inserted into live cortical tissue) to produce time-series analyses of incremental change in populations of coactive cortical neurons.

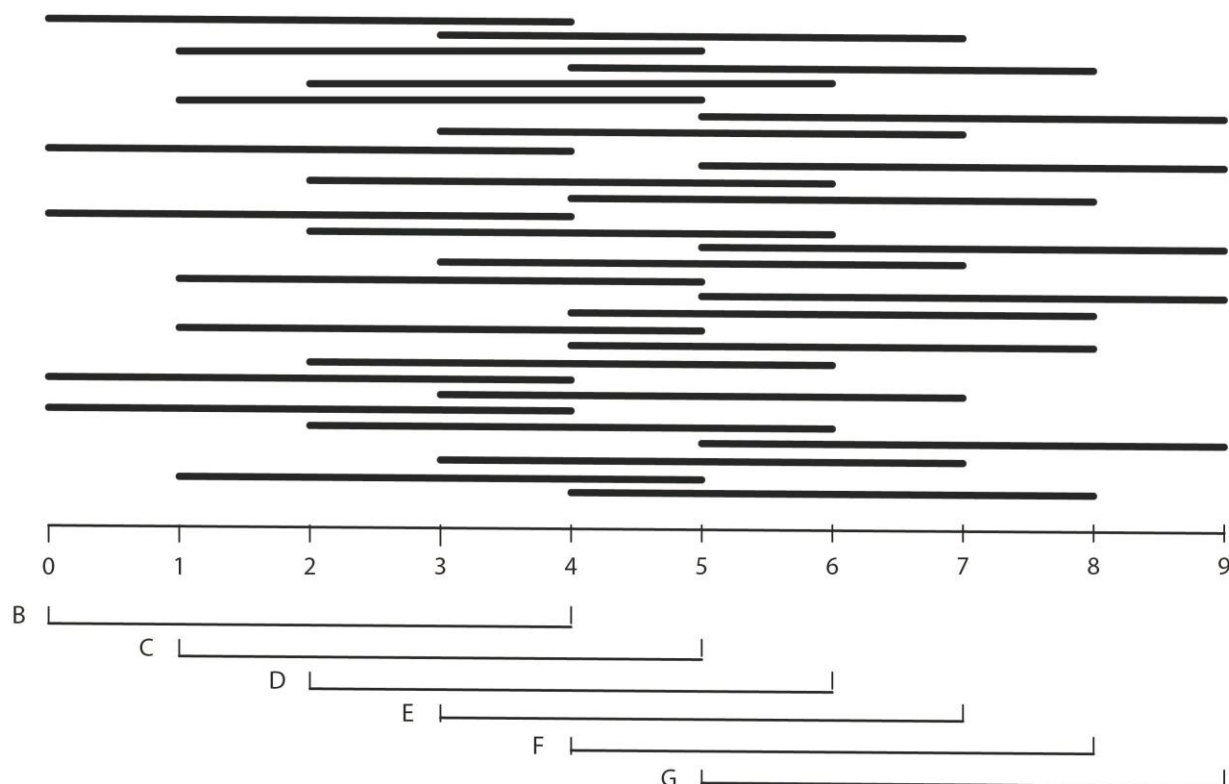


Figure 20. A Hypothetical Example of How Iterative Updating Could Be Found Using Electrodes

Single-cell recording from a large number of cells in association cortex could produce an activity profile exhibiting iterative updating. In this simplified figure, the *x*-axis represents time in seconds while the *y*-axis represents the recorded activity of 30 individual neurons, each of which remains active for four seconds. Five neurons become active each second. Each group of five neurons that begin and end their period of activity at the same time is taken to constitute an individual ensemble, or item, of working memory. Brackets at the bottom of the figure indicate the item to which each group of neurons belongs. This profile coincides precisely with the pattern introduced in Figure 8. Searching new and existing data for this kind of iterative pattern could provide strong support for the present model.

It is not clear whether it is possible to derive conclusive support for the present model using existing neuroimaging technology. Basic fMRI recording reveals the degree to which specialized brain modules exhibit involvement during a specific task but does not reveal the identity of the items or concepts involved. However, advanced recording techniques can demonstrate the onset and duration of brain responses to prepared stimuli, which could result in data like that in Figure 20. It should be possible to use the gain in temporal and spatial resolution to observe how the pattern of working memory activation changes over time. To that end, factorial designs that allow for the measurement of the BOLD signal for each volumetric cell should be able to test for differential activation in response to partial, as opposed to complete, updating of working memory.

Such findings could be derived from neuroimaging experiments in which brain activity is recorded while participants complete a task that requires an algorithmic sequence of steps (e.g., long division). Each step of the task would need to be modeled separately. As the participant

moves from one step to the next, the BOLD activity would be estimated for that particular step. A mixed model of a duration regressor covering the entire span of the problem along with individual regressors for each step would be needed to capture both the sustained attention required to solve the problem and the individual steps needed to progress from one stage to the next. It would be necessary to show that the sequence of mental representations posited as necessary to complete the task has a one-to-one correspondence with the time course of underlying cellular or hemodynamic changes. This may necessitate using multiple methods simultaneously such as fMRI and EEG together or using multivoxel pattern analysis, which has been used to resolve the addition and subtraction of individual cognitive items from working memory (Lewis-Peacock et al., 2012).

First, it would be necessary to show that the activity in association areas underlying working memory contents can be partially rather than completely updated. Next, the goal would be to show that this partial updating happens constantly. Future studies should be able to resolve whether the iterative updating of cortical activity is continuous (at the level of neurons) or incremental, where entire items (and all their comprising neurons) are added or subtracted at once. The line of reasoning suggested by this article predicts that the former may be true of the short-term store (i.e., Figure 6) while the latter may be true of the FoA (i.e., Figure 8).

Modern cognitive neuroscience is limited in its ability to match the components of brain states to the components of mental states. However, matching the iterative updating of ensembles to that of their corresponding items may provide a means to do so. The markers of iterative updating may establish an ordinality and translation strategy to decode the nature of the correspondence between temporary neural traces and their psychological manifestations.

Previous models of working memory have attributed various functions to the central executive (e.g., updating of items, coordination of modules, shifting between tasks, selective attention, gating, the construction of imagery, and others). Because the neural substrate of these functions has never been delineated, the central executive remains a mysterious black box. This article has presented the case that executive functions emerge from collective processing interactions among specialized subsystems guided by iterative updating. If shown to have a tenable neural basis by future research, the concepts introduced in this article (Table 3) may amount to a viable alternative to the notion of the central executive found in other models.

Term	Definition
Iteration	Repetition of a computational procedure applied to the product of a previous state used to obtain approximations that are successively closer to the solution of a problem.
Working Memory	A mechanism dedicated to maintaining selected representations available for use in further cognitive processing.
Working Memory Updating	Changes in the items held in working memory occurring as processing proceeds through time.
Iterative Updating	A shift in the contents of working memory that occurs during updating as some items are added, others are removed, and still others are maintained.
Coactive	A group of items that are active in the same instantaneous state.
Cospreading	A group of coactive items that combine their spreading activation energy to search the same global network (not all coactivity in the brain is cospreading).
Multiassociative Search	A type of search where all of the coactive, cospreading contents converge in parallel on the update for the next state.

State-spanning Coactivity (SSC)	Sustained coactivity exhibited by a set of two or more items that span two or more consecutive brain states.
Incremental Change in State-spanning Coactivity (icSSC)	The process in which a set of items exhibiting SSC undergoes a change in group membership, where some items remain in SSC and others are deactivated and replaced.
Rate of Updating	The proportion of items that are updated as a function of time.
Mental Continuity	The recursive interrelatedness of consecutive mental states made possible by iterative updating.
Iterative Compounding	Search results from a previous state are reused to update the system, incorporating them into the present search, thus prolonging their influence.
Progressive Modification	The logical or algorithmic progress in information processing made possible by iteration, continuity, or compounding.
Iterative Thread	A chain of iteratively updated states that underlies a line of thought.
Merging of Subsolutions	When select contents from two separate instances of iteration or separate lines of thought are coactivated in a new state and used together for multiassociative search.

Table 3. Definition of Terms Used and Introduced in This Article

This article applied iterative updating to the traditional model of working memory items, but it can similarly be applied to a large number of compatible frameworks that model item-like constructs including: ACT-R's "symbols" (Anderson, 1996), adaptive resonance theory's "templates" (Grossberg, 2013), global workspace theory's "processes" (Baars, 2005), the pattern recognition theory of mind's "pattern recognizers" (Kurzweil, 2013), hierarchical temporal memory's "time-based patterns" (Hawkins et al., 2007), SPAUN's "semantic pointers" (Eliasmith, 2013), SOAR's "operators" (Laird, 2012), LIDA's "codelets" (Baars & Franklin, 2007), OpenCog's "atoms" (Goertzel, 2014), Fuster's "cognits" (2005), Hofstadter's "simmballs" (2007), Edelman's "neuronal groups" (2004), Minsky's "agents" (1986), and Damasio's "convergent-divergent zones" (1989). These models, along with many others, provide detailed mechanistic explanations for critical neurocognitive components underspecified by the present model (Table 4).

1. The cerebral cortex is a hierarchy of pattern-recognizing neural assemblies that encode subsymbolic fragments of long-term memory.
2. An item in working memory corresponds to a persistently active group of neural assemblies that have been associated through experience.
3. Items first enter the focus of attention (FoA), which is associated with sustained firing. From there, they move toward the unattended short-term store, which is associated with synaptic potentiation. Lastly, they subside into inert long-term memory.
4. Items remain active in working memory as long as their neural assemblies demonstrate persistent activity. The activity of items is staggered and overlapping. Thus, the set of coactive items changes incrementally.
5. Active items in the FoA and the short-term store serve as search parameters for the next additions to working memory by spreading activation energy throughout the cortex.
6. Newly activated items are added to the set of remaining items from the previous state, completing the previous state's pattern and forming an updated set of search parameters for the next state.
7. This iterative updating process ensures that the next search is not an entirely new search but a modified version (updated iteration) of the previous search.
8. Iterative updating may play a fundamental role in event concatenation, progressive modification, learning and implementation of learned algorithms, mental modeling, inductive inference, rational thought, mental continuity, and consciousness.

Table 4.

Fundamental Features of the Iterative Updating Model

Working memory items in the FoA have been considered to be isomorphic with the contents of consciousness (Baars & Franklin, 2003; Buchsbaum, 2013). This suggests that the contents of conscious thought are held in working memory and operate according to the same (or similar) rules and capacity limitations. In the classic paradigm for working memory testing, subjects can retain approximately four items in mind. However, they are holding much additional declarative content. This is because they are also maintaining the task requirements, active sensory perceptions, and ongoing personal thoughts (which may be limited due to cognitive load). The iterative updating function applies to all this conscious content, not just to the four items described by Cowan (2017) and others. The previous figures in this article have mostly used only four items, but this was done for illustrative purposes only. Figure 21 uses an arbitrarily larger number of items as an alternative. The figure also places these items in an arbitrarily larger number of functionally specialized stores intended to indicate that items may exist along a graded continuum of activation.

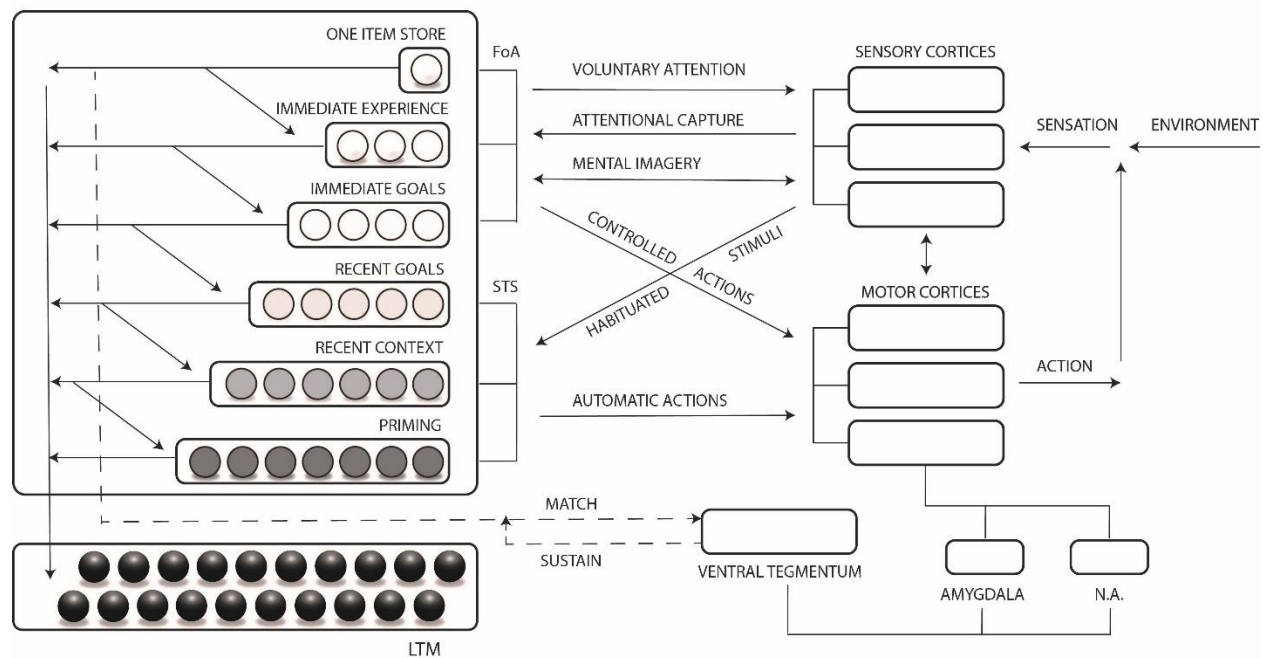


Fig. 21. Imagery and Behavior in the Iterative Updating Model

Iteratively updated items in working memory interact with sensory cortices to progressively construct mental imagery, and they interact with motor cortices to progressively construct behavior. In the next state, the items in working memory will undergo partial replacement. The parameters used in the sensory and motor cortices will reflect this change, making their output an advancement on their previous output, resulting in a capacity to support complex behavior. Related cognitive processes are included as arrows. The ventral tegmentum uses inputs from working memory as well as from the amygdala and nucleus accumbens (N.A.) to determine which patterns of items should be sustained.

“Higher-order” theories of consciousness hold that conscious thought is made possible when a mental state is concerned with a previous mental state. This includes a thought about a perception or a thought about a thought (Rosenthal, 2004). According to this logic, thoughts that are iterations of previous thoughts are backward-referential and could be considered “higher-

order thoughts.” Due to its role in generating a continual production line of higher-order thoughts, iterative updating should be considered a candidate for the neural basis of consciousness. It ensures that the chain or train of thought does not stop and go in discrete steps but is instead driven by the items that endure through time. All items will exit the FoA within seconds, but the common content shared between successive states sustain associative bridges that connect the advancing sequence of thoughts.

In his book *The River of Consciousness* (2017), Oliver Sacks asks, “But how then do our frames, our momentary moments, hold together? How, if there is only transience, do we achieve continuity?” This article has argued that our momentary moments overlap in their set of active representations and that this constant overlap assembles a seamless progression of states in which each state does not die off at once but continues into the next state. After asking the question, Sacks quotes William James. Each thought, in James’s words, is an owner of the thoughts that went before and “dies owned, transmitting whatever it realized as itself to its own later proprietor.” James expounds further on this analogy:

“Consciousness, then, does not appear to itself chopped up in bits. Such words as 'chain' or 'train' do not describe it fitly as it presents itself in the first instance. It is nothing jointed; it flows. A 'river' or a 'stream' are the metaphors by which it is most naturally described. In talking of it hereafter let us call it the stream of thought, of consciousness, or of subjective life. [...] As the brain-changes are continuous, so do all these consciousnesses melt into each other like dissolving views. Properly they are but one protracted consciousness, one unbroken stream.” (James, 1890, p. 239)

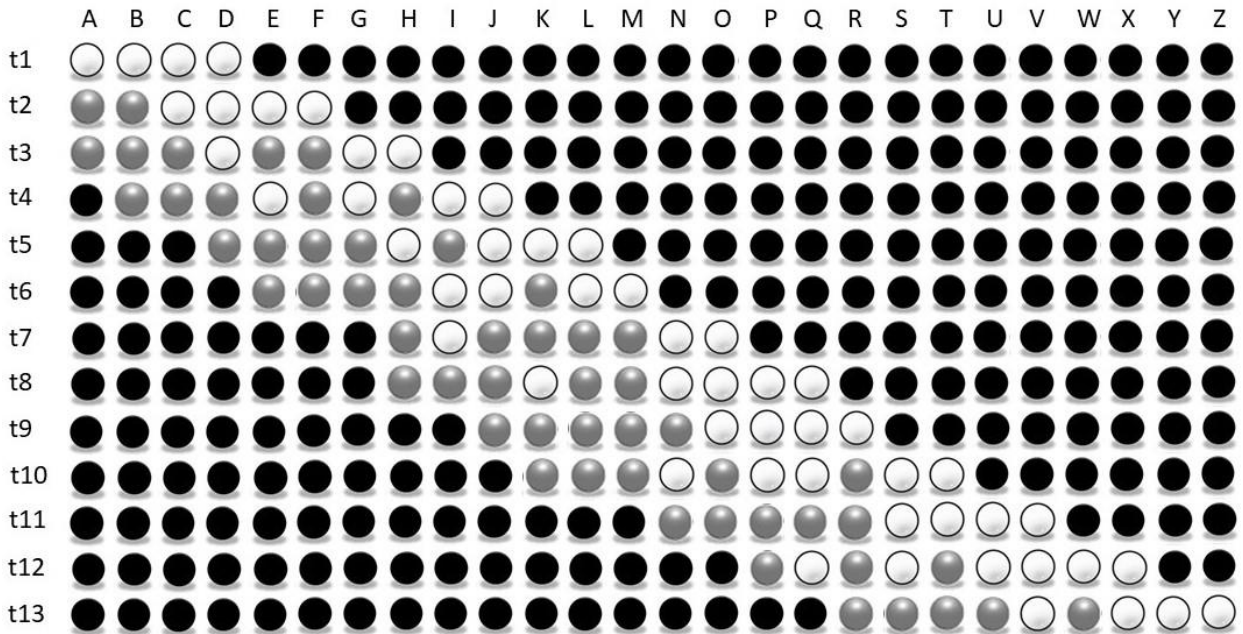


Fig. 22. Schematic Representation of Ongoing Iteration in the FoA and Short-term Memory Store

This graphic expands on previous figures, incorporating a larger number of the present model’s theoretical features. These include the following: (1) the number of items coactive in the FoA (white spheres) at any point in time varies between three and five, (2) the percentage of updating in the FoA varies between 25% and 100%, (3) the order of entry into the FoA does not determine the order of exit, (4) items that exit the FoA briefly enter the short-term store

(gray spheres) before deactivating completely (black spheres), and (5) items that have exited the FoA are capable of reentering the FoA.

The present model bears a resemblance to James's conception of a "stream of consciousness." A stream is a distribution of points that slides through space and time. Figure 22 extends the activity schematized in this article's other figures over 13 points in time. This results in a depiction of brain activity, working memory, and thought that, shifting gradually, appears very much like a stream.

References

- Anderson, John R. (1983). "A spreading activation theory of memory." *Journal of Verbal Learning and Verbal Behavior*. 22 (3): 261–295.
- Asok A, Leroy F, Rayman JB, Kandel ER. 2019. Molecular mechanisms of the memory trace. *Trends in Neurosciences*. 42(1): 14-22.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Chapter: Human memory: A proposed system and its control processes. In Spence, K. W., & Spence, J. T. *The psychology of learning and motivation* (Volume 2). New York: Academic Press (pp. 89–195).
- Atkinson, R. C. & Shiffrin, R. M. (1969). Storage and retrieval processes in long-term memory. *Psychological Review*, 76(2), 179-193.
- Averell L & Heathcote A. 2011. The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology*. 55(1): 25-25.
- Baars BJ, Franklin S. 2003. How conscious experience and working memory interact. *Trends Cogn Sci*. 7: 166-172.
- Baars, B. J. A Framework. In: B.J. Baars, N.M. Gage (Eds.), *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*, Academic Press, London, UK (2007), p. 30.
- Baars BJ, Franklin S. 2007. An architectural model of conscious and unconscious brain functions: Global Workspace Theory and IDA. *Neural Networks*. 20(9): 955-961.
- Baddeley, A. (1986). *Working memory*. Oxford, UK: Clarendon Press.
- Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Science*, 4, 417–423.
- Baddeley AD, Hitch GJ & Allen RJ. 2018. From short-term store to multicomponent working memory: The role of the modal model. *Memory and Cognition*. 1-14.
- Baddeley, A. D. (2007). *Working Memory, Thought and Action*. Oxford: Oxford University Press.
- Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology*, 8(4), 485-493.

- Baddeley AD. Hitch GJ. (1974). Working memory. G.A. Bower (Ed.), *Recent Advances in Learning and Motivation*, Vol. 8, Academic Press, New York (1974), pp. 47-89.
- Baddeley AD. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*. 4(11): 417-423.
- Baddeley AD. (2012). Working memory: Theories, models and controversies. *Annual Review of Psychology*. 63: 1-29.
- Bargh, J. A., & Chartrand, T. L. (2000). Reis, H., & Judd, C., ed. *Studying the Mind in the Middle: A Practical Guide to Priming and Automaticity Research*. In *Handbook of Research Methods in Social Psychology*. New York, NY: Cambridge University Press. pp. 1–39.
- Baronett, S. 2008. *Logic*. Upper Saddle River, NJ: Pearson Prentice Hall. pp. 321–325.
- Bostrom N. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press: Oxford UK.
- Botvinick MM. 2008. Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*. 12(5): 201-208.
- Botvinick MM & Plaut DC. 2006. Short-term memory for serial order. A recurrent neural network model.
- Braver, T. S. & Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In Monsell, S. & Driver, J. (Eds.), *Attention and Performance XVIII; Control of cognitive processes* (pp.713-737).
- Broadbent, D (1958). *Perception and Communication*. London: Pergamon Press.
- Brydges C, Gignac GE, Ecker UKH. 2018. Working memory capacity, short-term memory capacity, and the continued influence effect: A latent-variable analysis. *Intelligence*. 69: 177-122.
- Buchsbaum BR. 2013. The role of consciousness in the phonological loop: hidden in plain sight. *Frontiers in Psychology*. 4: 496
- Byeon W., Wang Q., Srivastava R. K., Koumoutsakos P. 2018. ContextVP: Fully Context-Aware Video Prediction. *The European Conference on Computer Vision (ECCV)*, pp. 753-769
- Carpenter, G. A. & Grossberg, S. (2003). Adaptive Resonance Theory, In Michael A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks, Second Edition* (pp. 87-90). Cambridge, MA: MIT Press.
- Castelvecchi D. 2016. Can we open the blackbox of AI: Artificial intelligence is everywhere. But before scientists trust it, they first need to understand how machines learn. *Nature*. (538): 7623.
- Chein, J.M., & Fiez, J.A. (2010). Evaluating models of working memory through the effects of concurrent irrelevant information. *Journal of Experimental Psychology: General*, 139, 117-137.

- Cheng, P.C. and Holyoak, K.J. (2008) Pragmatic reasoning schemas. In Reasoning: studies of human inference and its foundations (Adler, J.E. and Rips, L.J., eds), pp. 827–842, Cambridge University Press
- Chia WJ, Hamid AIA, Abdullah JM. 2018. Working memory from the psychological and neurosciences perspectives: A review. *Frontiers in Psychology*. 27.
- Christophel TB, Klink PC, Spitzer B, Roelfsema PR & Haynes J. 2017. The distributed nature of working memory. *Trends in Cognitive Sciences*. 21(2): 111-124.
- Cohen, G. (2000). Hierarchical models in cognition: Do they have psychological reality? *European Journal of Cognitive Psychology*, 12(1): 1-36.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6) 407-428.
- Comer D. (2017). *Essentials of Computer Architecture*. Chapman and Hall. New York NY.
- Constantinidis C, Funahashi S, Lee Daeyeol, Murray JD, Qi X, Wang M, & Arnsten AFT. 2018. Persistent spiking activity underlies working memory. *Journal of Neuroscience*. 38(32): 7020-7028.
- Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*. 96(2): 341-370
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87-185.
- Cowan, N. (2005). *Working memory capacity*. New York, NY: Psychology Press.
- Cowan N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*. 104 (2): 163-191.
- Cowan, N. 2009. What are the differences between long-term, short-term, and working memory? *Prog. Brain Res*. 169: 323-338.
- Cowan, N. 2011. The focus of attention as observed in visual working memory tasks: making sense of competing claims. *Neuropsychologia*. 49: 1401-1406.
- Cowan, N. 2017. The many face of working memory and short-term storage. *Psychonomic Bulletin & Review*. 24(4): 1158-1170.
- Crick, F. & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, 6(2):119-126.
- D'Esposito M, Postle BR. The cognitive neuroscience of working memory. *Annual Review of Psychology*. 66: 115-142.
- Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems level proposal for the neural substrates of recall and recognition. *Cognition*, 33: 25–62.

- Debanne D, Inglebert Y, Russier M. 2019. Plasticity of intrinsic neuronal excitability. *Current Opinion in Neurobiology*. 54: 73-82.
- Dehaene S. 2020. *How We Learn: Why Brains Learn Better Than Any Machine... for Now*. Penguin Random House. New York NY.
- Ecker, U. K., Oberauer, K., & Lewandowsky, S. (2014). Working memory updating involves item-specific removal. *Journal of Memory and Language*, 74, 1-15.
- Edelman G. *Wider than the sky*. Yale University Press. New York, NY. 2004.
- Eriksson J, Vogel EK, Lansner A, Bergstrom F, Nyberg L. 2015. Neurocognitive architecture of working memory. *Neuron*. 88(1): 33-46.
- Fuji, H., Ito, H., Aihara, K., Ichinose, N., & Tsukada, M. (1998). Dynamical Cell Assembly Hypothesis – Theoretical possibility of spatio-temporal coding in the cortex. *Neural Networks*, 9(8):1303-1350.
- Funahashi S, Bruce CJ, Goldman-Rakic PS. (1993). Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic ‘scotomas’. *Journal of Neuroscience*. 13(4): 1479-1497.
- Funahashi S. 2007. The general-purpose working memory system and functions of the dorsolateral prefrontal cortex. In *The Cognitive Neuroscience of Working Memory*. Eds Naoyuki Osaka, Robert H Logie, and Mark D’Esposito. Oxford, Oxford University Press.
- Fuster, J. M. (2002a). Frontal lobe and cognitive development. *Journal of Neurocytology*, 31(3-5): 373-385.
- Fuster, J. M. (2002b). Physiology of executive functions: the perception-action cycle. In Stuss, D. T., Knight, R. T. (Eds.), *Principles of frontal lobe function*. (pp. 96-108). Oxford: Oxford University Press.
- Fuster JM (1973). "Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory". *Journal of Neurophysiology*. 36 (1): 61–78.
- Fuster, J. M. (2009). Cortex and Memory: Emergence of a new paradigm. *Journal of Cognitive Neuroscience*, 21(11): 2047-2072.
- Fuster J. (2015). *The Prefrontal Cortex: Fifth Edition*. Academic Press, Elsevier: Oxford, UK.
- Glushchenko A. et al. (2018) Unsupervised Language Learning in OpenCog. In: Iklé M., Franz A., Rzepka R., Goertzel B. (eds) *Artificial General Intelligence. AGI 2018. Lecture Notes in Computer Science*, vol 10999. Springer, Cham
- Goertzel B. 2016. *The AGI Revolution*. Humanity Press.
- Goertzel B, Pennachin C & Geisweiller N. 2014. *Engineering General Intelligence. A Path to Advanced AGI via Embodied Learning and Cognitive Synergy*. Atlantis Press.
- Goodfellow I, Bengio Y & Courville A. 2016. *Deep Learning*. MIT Press. Cambridge MA.

- Goldman-Rakic, P. S. (1987). Circuitry of the prefrontal cortex and the regulation of behavior by representational memory. In Mountcastle, V. B., Plum, F., & Geiger, S. R., (Eds.), *Handbook of Neurobiology*, (pp. 373-417). Bethesda: American Physiological Society.
- Goldman-Rakic, P. S. (1990). Cellular and circuit basis of working memory in prefrontal cortex of nonhuman primates. In Uylings, H. B. M., Eden, C. G. V., DeBruin, J. P. C., Corner, M. A., & Feenstra, M. G. P. (Eds), *Progress in brain research*, (vol. 85, pp. 325-336). Elsevier Science Publications.
- Goldman-Rakic PS (1995). "Cellular basis of working memory". *Neuron*. 14 (3): 447–485.
- Goodfellow I, Bengio Y, Courville Aaron. (2017). *Deep Learning*. Cambridge, MA: MIT Press
- Gurney, K. N. (2009). Reverse engineering the vertebrate brain: Methodological principles for a biologically grounded programme of cognitive modeling. *Cognitive Computation*, 1(1) 29-41.
- Gray W.D. 2007. *Integrated Models of Cognitive Systems*. Oxford University Press. NY, NY.
- Haikonen P.O. (2003). *The cognitive approach to conscious machines*. Exeter UK: Imprint Academic.
- Haikonen P.O. (2012). *Consciousness and Robot Sentience*. Hackensack, NJ: World Scientific Publishing.
- Hameed AA, Karlik B, Salman MS, Eleyan G. 2019. Robust adaptive learning approach to self-organizing maps. *Knowledge-Based Systems*. 171(1): 25-36.
- Hamilton W. (1890). (Henry L. Mansel and John Veitch, ed.), 1860 *Lectures on Metaphysics and Logic*, in Two Volumes. Vol. II. *Logic*, Boston: Gould and Lincoln.
- Hawkins, J. 2004. *On Intelligence*. New York, NY. Times Books.
- Hasegawa, I., Fukushima, T., Ihara, T., & Miyashita, Y. (1998). Callosal window between prefrontal cortices: cognitive interaction to retrieve long-term memory. *Science*, 281, 814-818.
- Hassabis D, Kumaran D, Summerfield C & Botvinick M. 2017. Neuroscience-inspired artificial intelligence. *Neuron*. 95: 245-258.
- Hebb, Donald (1949). *The Organization of Behavior*. New York: Wiley.
- Hofstadter D. 2007. *I Am a Strange Loop*. Basic Books. New York NY.
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 269 –299.
- Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological review*, 110(2), 220.

Huttenlocher P.R., Dabholkar AS. 1997. Developmental Anatomy of Prefrontal Cortex. In Development of the Prefrontal Cortex. Edited by Krasnegor NA, Lyon GR, Goldman-Rakic PS. Paul H. Brookes Publishing Co. Baltimore, Maryland.

James W. 1909. A Pluralistic Universe. Hibbert Lectures at Manchester College on the Present Situation in Philosophy. Longmans, Green, and Co. London England.

James, W. (1890). The principles of psychology. New York, NY: Henry Holt.

Johnson-Laird, P. N. (1998). *Computer and the Mind: An Introduction to Cognitive Science*. Harvard University Press.

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Research Bulletin*, 44(2): 107-112.

Kahneman. 2011. Thinking Fast and Slow. Farrar, Straus, and Giroux. New York.

Klimesch, W., Freunberger, R., & Sauseng, P. (2010). Oscillatory mechanisms of process binding in memory. *Neuroscience and Biobehavioral Reviews*, 34(7): 1002-1014.

Konar A. 2014. Artificial Intelligence and Soft Computing: Behavior and Cognitive Modeling of the Human Brain. CRC Press. Boca Raton, Florida.

Kounatidou, P., Richter, M., & Schöner, G.. (2018). A Neural Dynamic Architecture That Autonomously Builds Mental Models. In T. T. Rogers, Rau, M., Zhu, X., & Kalish, C. W. (Eds.), Proceedings of the 40th Annual Conference of the Cognitive Science Society (pp. 643–648).

Kurzweil, R. 2012. How to Create a Mind. New York, NY. Penguin Group.

Laird, John E. (2012). *The Soar Cognitive Architecture*. MIT Press. Cambridge, Massachusetts.

Lansner, A. (2009). Associative memory models: From the cell-assembly theory to biophysically detailed cortex simulations. *Trends in Neurosciences*, 32(3):179-186.

LaRocque JJ, Lewis-Peacock JA, Postle BR. Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in Human Neuroscience*. 8: 1-14.

Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR. 2012. Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience*. 24(1): 61-79.

Manohar SG, Zokaei N, Fallon SJ, Vogels TP, & Husain M. (2019). Neural mechanisms of attending to items in working memory. *Neuroscience and Biobehavioral Reviews*. 101: 1-12.

Mellet, E., Petit, L., Mazoyer, B., Denis, M., & Tzourio, N. (1998). Reopening the mental imagery debate: Lessons from functional anatomy. *Nueroimage*, 8(2):129-139.

Meyer, K., Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neurosciences*, vol. 32, no. 7, 376–382.

- Myers NE, Stokes MG, Nobre AC. 2017. Prioritizing information during working memory: Beyond sustained internal attention. *Trends in Cognitive Sciences*. 21(6): 449-461.
- Miller, GA. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*. 63(2): 81-97.
- Miller, E.K., & Cohen, J.D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Ann Rev Neurosci*, 24:167-202.
- Miller EK, Lundqvist M, Bastos AM. 2018. Working memory 2.0. *Neuron*. 100: 463-475
- Mongillo G, Barak O, Tsodyks M. 2008. Synaptic theory of working memory. *Science*. 319: 1543-1546
- Moscovich, M. (1992). Memory and Working-with-memory: A component process model based on modules and central systems. *Journal of Cognitive Neuroscience*, 4(3):257-267.
- M. Moscovitch, J.M. Chein, D. Talmi, M. Cohn. Learning and memory. B.J. Baars, N.M. Gage (Eds.), *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*, Academic Press, London, UK (2007), p. 234
- Miyashita, Y. (2005). Cognitive memory: cellular and network machineries and their top-down control. *Science*, 306, 435-440.
- Nairne JS. 2002. Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*. 53: 53-81.
- Newell A, Simon HA. Computer simulation of human thinking. *Science*. (134)3495: 2011-2017.
- Niklaus M, Singmann H, Oberauer K. 2019. Two distinct mechanisms of selection in working memory: Additive last-item and retro-cue benefits. *Cognition*. 183: 282-302.
- Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, 75(6), 522-536.
- Nyberg L, Eriksson J. (2016). Working memory: maintenance, updating, and the realization of intentions. *Cold Spring Harbor Perspectives in Biology*. 8(2): a021816.
- Oberauer K (May 2002). "Access to information in working memory: exploring the focus of attention". *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 28 (3): 411–21.
- Opitz B. 2010. Neural binding mechanisms in learning and memory. *Neuroscience and Biobehavioral Reviews*. 34(7): 1036-1046.
- Pina J.E., Bodner M., Ermentrout B. 2018. Oscillations in working memory and neural binding: a mechanism for multiple memories and their interactions. *PLOS Computational Biology*. 14(11): e1006517.

- Postle, B. R. (2007). Activated long-term memory? The bases of representation in working memory. In Osaka, N., Logie, R. H., & D'Esposito, M. (Eds.). *The Cognitive Neuroscience of Working Memory*. Oxford, UK: Oxford University Press.
- Postle B, Ferrarelli F, Hamidi M, Feredoes E, Massimini M, Peterson M et al. (2006). Repetitive transcranial magnetic stimulation dissociates working memory manipulation from retention functions in the prefrontal, but not posterior parietal, cortex. *Journal of Cognitive Neuroscience*. 18, 1712-1722.
- Reggia JA, Katz GE, Davis GP. 2019. Modeling working memory to identify computational correlates of consciousness. *Open Philosophy*. 2: 252-269.
- Reser J. E. (2006). Evolutionary neuropathology & congenital mental retardation: Environmental cues predictive of maternal deprivation influence the fetus to minimize cerebral metabolism in order to express bioenergetic thrift. *Medical Hypotheses*. 67(3): 529-544.
- Reser, J. E. (2011). *What Determines Belief: The Philosophy, Psychology and Neuroscience of Belief Formation and Change*. Saarbrücken, Germany: Verlag Dr. Muller.
- Reser, J. E. (2012). *Assessing the psychological correlates of belief strength: Contributing factors and role in behavior*. (Doctoral Dissertation). Retrieved from University of Southern California. Usctheses-m2627.
- Reser, J. E. *The Neurological Process Responsible for Mental Continuity: Reciprocating Transformations between a Working Memory Updating Function and an Imagery Generation System*. Association for the Scientific Study of Consciousness Conference. San Diego CA, 12-15th July 2013.
- Reser, J.E. 2016. Incremental change in the set of coactive cortical assemblies enables mental continuity. *Physiology and Behavior*. 167 (1): 222-237.
- Reisberg D. 2010. *Cognition: Exploring the Science of the Mind*. WW. Norton & Co. New York NY.
- Rose, N.S. LaFocque JJ, Riggall AC, Gosseries O, Starrett MJ, & Meyering EE. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, 354 (6316): 1136-1139.
- Rosenthal, David M. 2004. "Varieties of Higher-Order Theory," in *Higher-Order Theories of Consciousness*. Editor Rocco Gennaro, 17-44. Amsterdam: John Benjamins.
- Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences*, 26, 709–777.
- Rushworth MF, Nixon PD, Eacott MJ, Passingham RE. (1997). Ventral prefrontal cortex is not essential for working memory. *Journal of Neuroscience*. 17(12): 4829-4838.
- Ryan K, Agrawal P, Franklin S. 2019. The pattern theory of self in artificial general intelligence: A theoretical framework for modeling self in biologically inspired cognitive architectures. *Cognitive Systems Research*. In Press.

- Rypma B, Berger JS, D'Esposito M. (2002). The influence of working-memory demand and subject performance on prefrontal cortical activity. *Journal of Cognitive Science*. 14(5): 721-731.
- Sacks O. 2017. *The River of Consciousness*. Vintage Books. New York, NY.
- Salmon, M. 2012. "Arguments from analogy", *Introduction to Logic and Critical Thinking*, Cengage Learning, pp. 132–142.
- Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain Research Reviews*. 35(2): 146-160.
- Schvaneveldt, R.W.; Meyer, D.E. (1973), "Retrieval and comparison processes in semantic memory", in Kornblum, S., *Attention and performance IV*, New York: Academic Press, pp. 395–409.
- Shanks, D., 2010, "Learning: From Association to Cognition", *Annual Review of Psychology*, 1, 273–301.
- Shannon, C. 1951. Prediction and entropy of printed English. *Bell System Technical Journal*. 30:47-51.
- Shastri, L. (1999). Advances in Shruti—A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony. *Applied Intelligence*, 11(1), 79-108.
- Sherstinsky A. 2020. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*. 404, 132306.
- Shipstead Z, Harrison TL, Engle RW. 2015. Working memory capacity and the scope and control of attention. *Attention, Perception, & Psychophysics*. 77(6): 1863-1880.
- Seamans JK, Robbins TW. 2010. Dopamine Modulation of the Prefrontal Cortex and Cognitive Function. *The Dopamine Receptors*. 373-398.
- Seamans, J.K., & Yang, C.R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in Neurobiology*. 74(1): 1-58.
- Silvanto J. 2017. Working memory maintenance: Sustained firing or synaptic mechanisms? *Trends in Cognitive Sciences*. 21(3): 152-154.
- Sipper M, Fu W, Ahuja K, & Moore JH. 2018. Investigating the Parameter Space of Evolutionary Algorithms, *BioData Mining*, 11:2.
- Sreenivasan KK, D'Esposito M. 2019. The what, where and how of delay activity. *Nature Reviews Neuroscience*. May 13
- Sousa AMM, Meyer KA, Santpere G, Gulden FO, Sestan N. 2017. Evolution of the human nervous system function, structure, and development. *Cell*. 170 (2): 226-247.

- Sperling, G. (1960). The Information Available in Brief Visual Representations. *Psychological Monographs*, 74(1960), 1–29.
- Stone, JV. 2015. Information Theory: A Tutorial Introduction. Sebtel Press.
- Striedter G. 2005. Principles of Brain Evolution. Sinauer Associates, Sunderland M.A.
- Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, 401, 699-703.
- Tononi G. 2010. An information integration theory of consciousness. *BMC Neuroscience*. 5: 42.
- Treisman, A.M. (1964). Selective attention in man. *British Medical Bulletin*, 20, 12-16.
- von der Malsburg, C. (1999). The what and why of binding: The modeler's perspective. *Neuron*, 24, 95-104.
- Weger U, Wagemann J, Meyer A. Introspection in psychology: Its contribution to theory and method in memory research. 2018. *European Psychologist*. 23: 206-216.
- Zanto TP, Rubens MT, Thangavel A, Gazzaley A. 2011. Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience*. 14, 656-661.