

Neurocognitive Architecture of Working Memory

Johan Eriksson,^{1,2,*} Edward K. Vogel,³ Anders Lansner,^{4,5} Fredrik Bergström,^{1,2} and Lars Nyberg^{1,2,6}

¹Department of Integrative Medical Biology, Umeå University, 901 87 Umeå, Sweden

²Umeå Center for Function Brain Imaging (UFBI), Umeå University, 901 87 Umeå, Sweden

³Department of Psychology, Institute for Mind and Biology, University of Chicago, Chicago, IL 60637, USA

⁴Department of Computational Biology, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden

⁵Department of Numerical Analysis and Computer Science, Stockholm University, 106 91 Stockholm, Sweden

⁶Department of Radiation Sciences, Umeå University, 901 87 Umeå, Sweden

*Correspondence: johan.eriksson@umu.se

<http://dx.doi.org/10.1016/j.neuron.2015.09.020>

A crucial role for working memory in temporary information processing and guidance of complex behavior has been recognized for many decades. There is emerging consensus that working-memory maintenance results from the interactions among long-term memory representations and basic processes, including attention, that are instantiated as reentrant loops between frontal and posterior cortical areas, as well as sub-cortical structures. The nature of such interactions can account for capacity limitations, lifespan changes, and restricted transfer after working-memory training. Recent data and models indicate that working memory may also be based on synaptic plasticity and that working memory can operate on non-consciously perceived information.

Introduction

Working memory maintains information in an easily accessible state over brief periods of time (several seconds to minutes). This feature is required for future goal-directed behavior and allows us to act beyond the confines of the here and now. As such, working memory is taxed by numerous laboratory and everyday cognitive challenges. The research literature on working memory is enormous, and in this Perspective we will not provide a comprehensive review. Rather, we aim to present a condensed summary of key facts and features to illustrate the “neurocognitive architecture” of working memory (Box 1; for related accounts, see D’Esposito and Postle, 2015; Fuster, 2009; Jonides et al., 2008).

Although there is no complete consensus on its definition, a basic feature in most conceptualizations of working memory is short-term maintenance of information in the absence of sensory input. Here, too, this definition is at the heart of our treatment of working memory. Information maintenance is considered to be the result of an interaction between basic building blocks of working memory (Figure 1A), notably a selective attention process (Figure 1B) that operates on perceptual information and related long-term memory (LTM) representations. Thus, here and elsewhere, attention is understood to be a cornerstone of working-memory processes (e.g., Baddeley and Hitch, 1974; Cowan, 1995; see Kastner and Ungerleider, 2000; Petersen and Posner, 2012). Figure 1B exemplifies maintenance of object information. First, in orange, the encoding of information into working memory is the result of interactions among selective attention processes and perceptual object representations that trigger related LTM object representations. Working-memory representations are vulnerable to distraction and interference. Therefore, when the perceptual input no longer is present, sustained attention along with a rehearsal process is crucial for maintaining the information in working memory (red contours in Figure 1B). If all the information to be maintained can “fit” within

the focus of attention, an active maintenance process fulfills maintenance through reverberating signals between regions that provide attentional/“top-down” signals (e.g., fronto-parietal areas; see Kastner and Ungerleider, 2000) and regions specifically related to the current content of working memory (i.e., perceptual and LTM representations). If there is more to maintain than fits within the focus of attention, an additional rehearsal process needs to complement the active maintenance process. Finally, at the retrieval phase (white contours), as in a delayed-match-to-sample task, selective attention and pattern completion processes become engaged to match the perceptual information provided at the retrieval stage with information maintained in working memory. Figure 1C exemplifies the situation in which “manipulation operations” are performed on the information currently maintained in working memory. This could be mental arithmetics, e.g., as in the computation-span working-memory task, or it could involve updating of the current content of working memory (e.g., O’Reilly, 2006). The concept of working memory includes the prospective use of information, which has been promoted as a major motivation for using the term “working,” rather than “short-term,” memory (e.g., Fuster, 2009). Purposeful use of the temporarily maintained information depends on the objective (goal) and structure of the task, as well as the context in which the task is performed. Together, these aspects provide the scaffold on which working memory proceeds. Accordingly, task set, prospective planning, and other cognitive control operations are integral parts of working-memory processing (purple field in Figure 1).

According to this “component processes” view of working memory, no processes (and correspondingly no brain structures) are unique or specific to working memory. Rather, working memory is the result of various combinations of processes that in other constellations can be functionally described in other terms than working memory (Figure 1D; cf. Cowan, 2001; D’Esposito and Postle, 2015; Fuster, 2009; Jonides et al., 2008). It should

Box 1. Current Status of the Field

- Working memory results from the interaction between several component processes, including attention, prospective, and perceptual and long-term memory representations.
- Many brain regions interact during working memory and include “executive” regions in the PFC, parietal cortex, and basal ganglia, as well as regions specialized for processing the particular representations to be maintained, such as the fusiform face area for maintaining face information.
- Persistent neural activity in various brain regions accompanies working memory and is functionally necessary for maintenance and integration of information in working memory.
- Working-memory capacity is limited and may only hold a small amount of information (absolute limits remain controversial); capacity can be increased through “chunking” bits of information into more complex units.
- Working-memory functioning changes across the lifespan with an inverted U-shaped trajectory and can be modified by training.
- Working memory may involve short-term plasticity but does not seem to require structural alterations, such as new protein synthesis, as it works by recruiting already existing synapses and ion channels (“activated LTM”).

be emphasized that working memory, as conceptualized here, is a particular state of a representation (temporarily enhanced accessibility), regardless of the kind of representation. That is, working memory can basically involve any kind of representation (verbal, visual, auditory, spatial, etc.), including various procedures or temporally ordered action sequences (e.g., when following a recipe), and by extension engage many different parts of the brain, where these representations are stored. Also, often the information to be encoded into working memory does not exactly match stored representations (e.g., novel configurations of familiar objects in tests of spatial working memory or unfamiliar faces). Therefore, although stored information in LTM can support working-memory maintenance, many tasks will require encoding and maintenance of novel information and, in some cases, even information that has no clear mapping to stored information (Olsson and Poom, 2005). In the latter case, working-memory capacity will be lower and may more or less entirely rely on perceptual representations. Whether or not the encoding of such information and other information into working memory is likely to also foster new LTM representations, and by inference lead to synaptic resculpting, will be further discussed below.

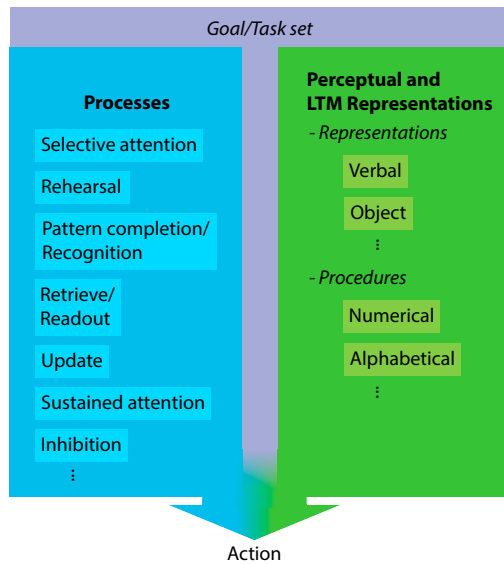
Behavioral Properties of Working Memory **Capacity Limitations**

A fundamental property of working memory is that it is highly limited in how much information can be held active simultaneously (Baddeley, 2003; Cowan, 2001; Luck and Vogel, 1997). Most estimates of the average capacity among healthy

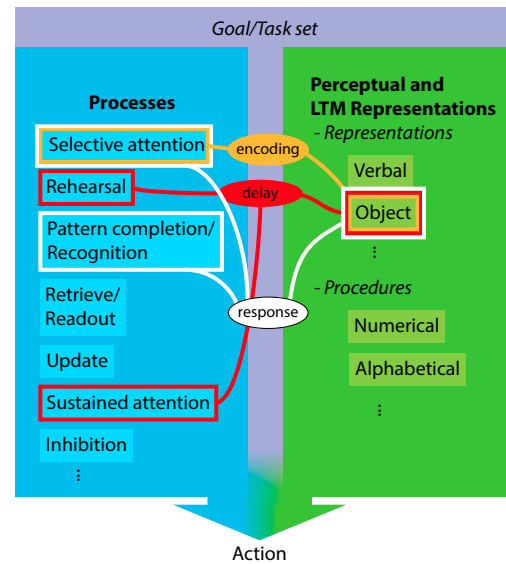
young adults suggest that working memory has a capacity limit of approximately 3 or 4 simple items (Luck and Vogel, 1997). This limitation highlights a sharp contrast between working memory and LTM, which is thought to have a nearly boundless capacity for storing new information from the environment. While there is broad agreement that working memory can maintain only a small amount of information simultaneously, two factors make a simple statement of a maximum limit very challenging. First, the amount of information that can be held depends strongly on whether the items can be grouped into meaningful units, or “chunks.” That is, by clustering information together one can exploit preexisting information about concepts already stored in long-term memory, which allows more efficient storage in working memory, presumably by reducing the number of active elements that must be maintained in working memory. Such chunking can be observed in many domains, from the clustering of letter strings to form acronyms of familiar concepts (Miller, 1956) to the exploitation of visual statistical regularities to form grouped arrays of objects (Brady et al., 2009). Second, objects with high levels of complexity may require additional resources to adequately resolve their details. Thus, working-memory performance may be reduced for such complex items due to insufficient precision at encoding. Indeed, there is evidence for variability in encoding precision even between objects presented within a single array (van den Berg et al., 2012; Fournie et al., 2012). When considering these factors, it becomes apparent that the functional limits to working-memory performance can vary substantially depending upon the nature of the processing demands imposed by the specific working-memory task. The opportunity to utilize either LTM or grouping tends to increase performance, while the requirement to report fine details of complex objects tends to decrease performance.

Even when grouping cues and high-precision demands are minimized, there is still debate about the nature of the maximum limit. Specifically, is it best characterized as a maximal upper limit on the number of discrete representations that can be maintained? Or is it better described as a finite pool of resources for representations that can be flexibly allocated to any arbitrary number of items? There is extensive evidence across many tasks and memoranda that individuals can remember only 3–4 simple items with near perfect accuracy, with steep drop-offs in performance for arrays that exceed this number (Figure 2A). However, flexible resource models can mimic such limits by positing that all (or most) items from a display are represented in working memory, but that with greater numbers of items the precision/resolution of each representation dwindles (Bays and Husain, 2008; Wilken and Ma, 2004). Thus, errors for arrays that have more than 3 or 4 items may be due to imprecise representations of each of the items rather than being due to items simply not being stored in working memory. Although both models make highly similar predictions regarding task performance across varying working-memory loads (e.g., that mnemonic precision declines with number of items), they differ in one key component: the role of guessing. Discrete models propose that if a subject is tested on an item from an array that is not held in one of the 3–4 slots, he or she will guess its identity. By contrast, continuous models propose that subjects never truly guess because all items from the display are assumed to be represented in working

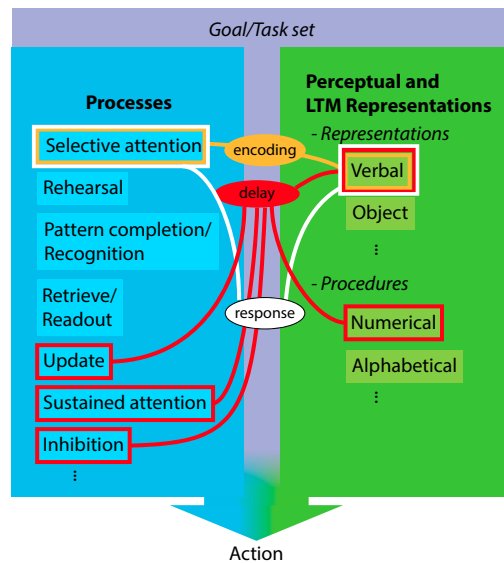
A Building blocks for working memory



B Maintenance, e.g., a simple DMS task



C Manipulation, e.g., mental arithmetics



D Schematic mapping of processes/representations to brain networks during working-memory maintenance

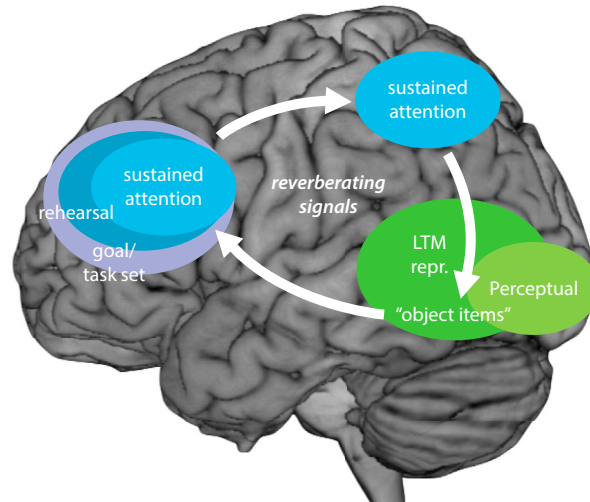


Figure 1. The Component Processes View of Working Memory

(A–D) Schematic overview of how different processing and representational components and associated neuronal networks interact when solving a working-memory task. (A) Suggestions of processes relevant for working memory (blue field) are listed. The list is non-exhaustive and each process may be further analyzed into sub-processes. Working memory is viewed as emerging from the interactions among process components, of which selective attention to perceptual and long-term memory representations (green field; also non-exhaustive list) is central. The purple field connecting processes and representations illustrates the inherent link between goals/task sets and working-memory processing. (B) How various processes and representations interact during a task that requires maintenance of visual information, e.g., a delayed-match-to-sample (DMS) task, and how the involvement of different processes change dynamically throughout task performance are exemplified. (C) Such interactions and dynamics during performance of a manipulation task, e.g., multiplying 42 with 12, are exemplified. Here, procedural long-term memory representations may also support solving the task, e.g., by recollecting procedures for how the multiplication of two-digit numbers can be done efficiently. (D) A schematic mapping of processes during the “delay” phase in (B) to brain regions is shown, demonstrating the distributed nature of processes and representations involved to solve working-memory tasks (see the main text for a more exhaustive account). Across the panels it is highlighted that different processing components and representations come into play depending on task structure (e.g., maintenance or additional manipulation requirements), sensory input (e.g., auditory/verbal or visual/object), and stages of processing (encoding, maintenance, or response). At a higher level of description such mental events can be labeled “working memory.”

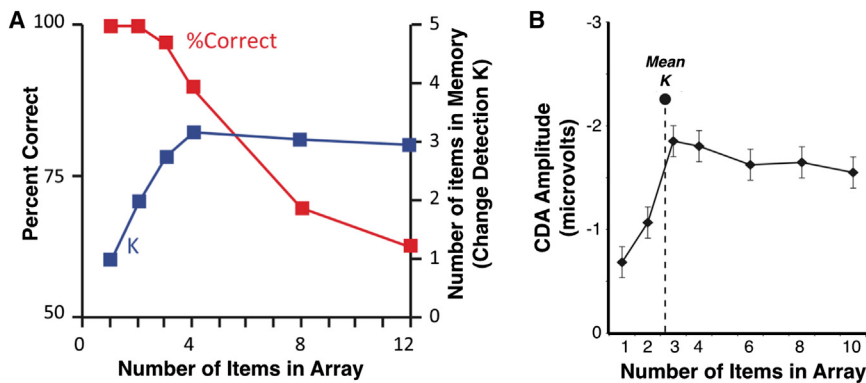


Figure 2. Behavioral and Neural Measures Show Corresponding Working-Memory Capacity Limits

(A) Working-memory performance from remembering arrays of simple objects in a change detection task (Luck and Vogel, 1997). Percent correct (red) and estimated capacity (blue) are shown as a function of number of items in the array.

(B) EEG/CDA amplitude as a function of the number of items to be remembered (Vogel and Machizawa, 2004). The dashed line indicates the average behavioral capacity for the group of participants.

memory. Thus, what appears to be guessing is just the consequence of extremely imprecise information about each object. However, recent work using modern Bayesian approaches has demonstrated that subjects do indeed appear to guess when probed on arrays that exceed typical capacity limits, thus suggesting that even if there is information in the system for items beyond a maximal number limit, subjects do not appear to be able to utilize it (Donkin et al., 2013).

Individual Differences in Working-Memory Capacity

There are substantial differences between individuals in working-memory capacity. These differences are highly stable over time and appear to be a core cognitive trait of the individual (Kane and Engle, 2002). This is because an individual's working-memory capacity is strongly predictive of his or her performance on a wide variety of high-level cognitive measures, such as fluid intelligence, abstract reasoning, mathematics and language abilities, and overall academic performance (Cowan et al., 2005; Daneman and Carpenter, 1980; Unsworth et al., 2014). Extensive work in the past 20 years has indicated that these individual differences are determined primarily by variability in consistently deploying attentional control over what is stored in working memory, rather than the absolute amount of storage space per se (Adam et al., 2015; Kane and Engle, 2002). Low-capacity individuals have more difficulty ignoring distracting information than do high-capacity individuals. This is in part because they are slower at disengaging attention from irrelevant information that captures their attention (Fukuda and Vogel, 2011). Thus, the ability to efficiently deploy attentional control in overloading situations appears to be the common thread that connects working-memory capacity to an individual's ability to perform many complex cognitive tasks.

Brain Areas and Connections

In accordance with the component processing account of working memory outlined in the introduction, brain regions involved in maintaining information in working memory will vary with the type of information to be maintained. Generally, the same brain regions dedicated to sensory processing are believed to store sensory information during delay periods and working-memory task performance (Figure 1D). Accordingly, lesions to the temporal cortex affect visual working memory but leave spatial working memory intact (Owen et al., 1996), whereas parietal-lesion patients show the opposite pattern (Pisella et al., 2004). Similarly,

patients with lesions to regions associated with semantic storage (e.g., the lateral temporal lobes and temporoparietal cortex; Binder et al., 2009) have reduced verbal working-memory performance (Bormann et al., 2015). Material-specific effects are also evident in neuroimaging research, where different stimulus categories, such as faces and houses, activate category-specific regions of the ventral visual cortex during working-memory maintenance (e.g., Ranganath et al., 2004). More recently, research using multivariate pattern analyses has demonstrated that the particular content held in visual working memory can be decoded from activity patterns in visual cortex (see Postle, 2015 for an overview).

Maintaining working-memory item representations in sensory regions makes them potentially vulnerable to task-irrelevant stimulus processing, which may interfere with maintenance of the working-memory content (Miller et al., 1993). The prefrontal cortex (PFC) has been suggested to be critical for resilient information maintenance during working-memory tasks. Single-cell recordings in monkeys (Funahashi et al., 1989; Fuster and Alexander, 1971) and, more recently, human neuroimaging studies (Courtney et al., 1997) have demonstrated sustained neural activity in the PFC during the delay period of working-memory tasks (see below), and studies of PFC lesion patients indicate that an intact lateral PFC is necessary for normal performance in delayed-response tasks, particularly if the task involves distraction (D'Esposito and Postle, 1999) or if task difficulty is increased in other ways. Non-invasive electrical or magnetic stimulation of the PFC has also been shown to affect working-memory performance (Brunoni and Vanderhasselt, 2014; Feredoes et al., 2011).

Thus, previous research has established that the PFC is causally involved in normal working-memory functioning. Moreover, the PFC likely contributes to working memory in several ways, and sub-divisions of PFC have been suggested. However, there is yet no consensus on the details of the functional organization of the PFC. Several meta-analyses have demonstrated regional specificity, in that the left, particularly ventral, PFC is more involved in verbal working-memory tasks, whereas the right, particularly dorsal, PFC is more involved in spatial working-memory tasks (Nee et al., 2013; Owen et al., 2005; Wager and Smith, 2003). Early animal models suggested a dorsal/ventral dissociation based on spatial versus object working memory (Levy and Goldman-Rakic, 2000), but differences in task difficulty between

material types may have influenced results. Correspondingly, tasks involving updating and ordering of working-memory content (i.e., manipulation tasks) are more likely to involve dorsolateral PFC compared with maintenance tasks (Wager and Smith, 2003). A recent meta-analysis found little support for regional specificity within the PFC for different types of executive demands, such as distractor resistance, intrusion resistance, updating, and shifting (Nee et al., 2013).

Although the organizational principle of the PFC remains unclear, it may be more based on content-oriented associations than on process types (Fuster, 2009; Miller and Cohen, 2001). However, the “content” of PFC representations seems to be more complex/abstract compared with representations in posterior brain regions. Specifically, recent research has demonstrated that frontal cortex neurons have “high dimensionality” (Rigotti et al., 2013), in that they code for combinations of item-related and task-related information (see below). This is consistent with a view of the PFC as primarily associated with cognitive control functions related to “goals and means to achieve them” (Miller and Cohen, 2001, p. 167). Such prospective codes have been suggested to underlie the organization of lateral PFC, in that the abstraction level of goals and task rules are suggested to peak in rostral PFC and decrease, i.e., become more and more specific to certain situational contexts, toward posterior parts (Badre and D’Esposito, 2009; Fuster, 2009; Koechlin and Summerfield, 2007).

Together with the PFC, parietal cortex is also strongly involved in working-memory functioning. Superior parietal cortex has been associated with executive aspects of working memory (Collette et al., 2005; Koenigs et al., 2009) and is thought to implement selective attentional control (Figure 1D; see Awh et al., 2006). Interestingly, parietal cortex activity correlates with working-memory capacity, such that activity increases with an increasing number of items to remember until the limit of 3–4 items is reached, at which time activity levels out (Figure 2B; Vogel and Machizawa, 2004). As with the PFC, spatial working-memory tasks commonly activate parietal cortex bilaterally, with some lateralization toward the right hemisphere (Nee et al., 2013; Owen et al., 2005). Such lateralization is also consistent with patient studies, in which lesions to the right parietal cortex impair spatial working memory, whereas left-sided lesions do not (Koenigs et al., 2009). In contrast, verbal working-memory performance is greatly affected in left-lesion patients (Shallice and Warrington, 1970; Vallar and Baddeley, 1984). However, such lesion symptoms are likely due to damage of ventral rather than dorsal parietal cortex, which is also part of the language network (Binder et al., 2009). Thus, together with superior temporal and ventral prefrontal regions, left inferior parietal cortex is critical for verbal working memory (Buchsbaum and D’Esposito, 2008). Although usually not highlighted as a key region of the working-memory network, the cerebellum is commonly activated in working-memory tasks (Nee et al., 2013). The cerebellum has been suggested to support verbal rehearsal, but may contribute to working memory more generally (Stoodley and Schmahmann, 2009).

Basal ganglia, and more specifically striatal, involvement in working-memory tasks is a relatively common finding in neuroimaging research (Wager and Smith, 2003), and basal ganglia are

key structures in computational models (O’Reilly, 2006). Striatal involvement is also supported by human lesion studies (Voytek and Knight, 2010) and findings of reduced working-memory performance related to fMRI BOLD signal and dopamine changes in the striatum in Parkinson’s disease patients (Ekman et al., 2012). According to computational models, the striatum acts as a gating mechanism for representations in the PFC, by controlling when PFC representations should be maintained versus updated (O’Reilly, 2006). This idea is supported by working-memory training studies (see below) and counterintuitive improvements in distractor resistance for Parkinson’s disease patients off dopamine medication; reduced striatal dopamine levels impairs the updating regulation by the striatum, which makes PFC representations more rigid and thereby more distractor resistant (Cools et al., 2010). Also, McNab and Klingberg (2008) showed that PFC and basal-ganglia activity was positively correlated with working-memory capacity and that parietal load effects, here associated with unnecessary storage of distracting information, was negatively correlated with activity in basal ganglia (McNab and Klingberg, 2008). As noted above, low-capacity individuals have more difficulty ignoring distracting information than high-capacity individuals, and the gate-keeping function of fronto-striatal regions and associated dopamine mechanisms is thus one likely source of the capacity limits of working memory (Cools and D’Esposito, 2011; see also the sub-section *Lifespan Changes*).

A key finding that early on motivated the distinction between short-term and long-term memory was that patients with medial temporal lobe (MTL) resection had heavily impaired long-term memory but apparently normal short-term memory and the MTL may therefore not be considered important for working memory. However, neuroimaging research has demonstrated MTL activity during working-memory tasks (e.g., Axmacher et al., 2007), and it has more recently been suggested that the MTL may be needed for working-memory tasks that require binding and relational processing. For example, Olson et al. (2006) demonstrated that patients with bilateral MTL lesions performed normally on tasks that required maintenance of only objects or only locations throughout an 8-s delay, but were impaired when trying to maintain object-location conjunctions. However, a recent MTL patient study did not find working-memory impairment even though the task required binding/relational processing (Allen et al., 2014). An alternative interpretation of the MTL—working memory findings is that the involvement of the MTL system in working-memory tasks depends on load, such that the MTL is required if task load exceeds working-memory capacity (Jeneson and Squire, 2012). Such involvement further highlights the dynamic relation between working and long-term memory.

Consistent with the view of working memory as emerging from the dynamic interaction of a large number of brain regions, connectivity analyses of neuroimaging data show that, during a delayed-response task, activity in sensory regions is correlated with activity in PFC, parietal cortex, striatum, and also the MTL (Gazzaley et al., 2004). Furthermore, the integrity of white matter pathways connecting the PFC, parietal cortex, and temporal cortex correlates with working-memory performance (Charlton et al., 2010). Taken together, previous research demonstrates

that working memory is the result of the interaction among several brain regions (Figure 1D); the specific regions involved depend on a number of factors, including the type of material to be remembered, the task (e.g., simple maintenance or additional manipulation requirements), and also which stage of the dynamic interplay of processing components that is considered (e.g., during encoding, the delay period, or the response phase).

Neuronal Codes and Properties of Neurophysiological Responses

As a consequence of working memory involving multiple interacting processes, a particular item maintained in working memory will be coded in a highly distributed manner (Fuster and Bressler, 2012). For example, a visual working memory item may consist of particular visual feature representations in early (Harrison and Tong, 2009) and/or late (Ranganath et al., 2004) parts of the ventral visual pathway, combined with relevant spatial representations in frontal and parietal cortex (Nee et al., 2013), as well as representations of the behavioral significance of the item in frontal cortex (Rigotti et al., 2013). Moreover, given that working-memory representations are assumed to rely on existing perceptual and LTM representations, the general principle of distributed information storage (Fuster and Bressler, 2012) is inherited by working-memory representations (Emrich et al., 2013). The integration of such distributed components of information is believed to rely on long- and short-range recurrent connections among brain regions that support oscillatory signals during working-memory tasks (e.g., Liebe et al., 2012; see below). The necessity of such reverberant activity was demonstrated by Fuster et al. (1985) by cooling either the PFC or the inferotemporal cortex in awake macaques during a delayed match-to-sample color discrimination task. Cooling of either region changed neural activity in the other region, with concomitant drops in behavioral performance. Consistent with this view, Siegel et al. (2009) have observed that information about the identity of memoranda during the delay period can be decoded in the 40-Hz oscillations of the local field potential in the PFC of monkeys. This work also suggested a “phase-coding” scheme for holding multiple items in working memory simultaneously, in which each item in memory was held in distinct phases of the gamma oscillation (see the sub-section *Neural Network Models of Working Memory*).

Sustained delay activity has also been observed in the human using visual event-related potentials (ERP). The contralateral delay activity (CDA) is a retinotopically organized slow wave that is a negative-going voltage over the hemisphere that is contralateral to the positions of the items that are being remembered on a trial. CDA amplitude is highly sensitive to the number of items that are being remembered, but reaches an asymptote at typical memory capacity limits and is highly predictive of individual differences in working-memory capacity (Figure 2B; Vogel and Machizawa, 2004). The CDA has been observed in a wide variety of task settings that are presumed to necessitate active working-memory representations, such as visual search, perceptual monitoring, mental rotation, and attentional tracking (Drew et al., 2011; Emrich et al., 2009; Prime and Jolicoeur, 2010; Tsubomi et al., 2013).

Analogous sustained delay activity phenomena during working-memory tasks can also be observed in EEG/MEG oscillations in other frequencies, including the alpha (8–12 Hz) and theta (4–7 Hz) frequency bands (Roux and Uhlhaas, 2014). Alpha power changes during the retention interval are dependent upon the current working-memory load (Jensen et al., 2002). Indeed, alpha power has long been associated with active cognitive processing in various attention and working-memory tasks, and contemporary views of alpha suggest that it reflects the allocation of spatial attention to the memoranda, as well as the suppression of distracting information (Klimesch, 2012). Moreover, van Dijk et al. (2010) have recently proposed that sustained activity, such as the CDA, could be created from dynamic modulations of the alpha frequency (8–12 Hz) that are asymmetric in nature. This means that the amplitude is modulated more in the peaks than in the troughs of the alpha cycle. Consequently, such modulations would survive the trial averaging procedures used in ERPs and be revealed as a slow wave. However, while these proposals are compelling, to date there is no direct evidence indicating that these two neural metrics are isomorphic or if they reflect distinct measures of working-memory activity.

Sustained Brain Activity during Working-Memory Maintenance

Since its discovery in the early 1970's, sustained activity during delayed-response tasks in PFC and posterior regions has been suggested as a hallmark for short-term maintenance of information (Fuster and Alexander, 1971; Goldman-Rakic, 1995). Moreover, item-specific information was suggested to be coded through such persistent activity, in that a sub-set of PFC neurons were specifically activated throughout delay periods for particular spatial positions to be remembered (Funahashi et al., 1989). Recent single-neuron recordings in the PFC of monkeys have demonstrated that PFC neurons can have “mixed selectivity” in that they code for both task- and item-related information (Rigotti et al., 2013). Sustained activity has also been recorded in inferotemporal cortex (Fuster and Jervey, 1981; Miller et al., 1993), and in a recent fMRI experiment item-specific working-memory representations were reliably “read out” using multivariate pattern analysis in sensory regions, but not in frontal or parietal regions (see Postle, 2015). Thus, item-specific information may be more readily discernible in posterior regions, but may still be part of the information maintained by PFC activity, albeit in a more complex, higher-dimensional form.

It has been noted that the sustained activity of item-specific cells is often relatively dynamic throughout delay periods, in that they display a range of activity profiles across time (Shafit et al., 2007). Moreover, population coding in PFC neurons have been shown to transition between several representational states during a delayed paired-associates task (Stokes et al., 2013). By using multivariate pattern analysis and retro-cues to indicate which of two items that was going to be task-relevant, Lewis-Peacock et al. (2012) demonstrated that pattern classification performance dropped to chance level for items not in the current focus of attention during a delay, but rose to above-chance level when later cued during the same delay interval (see LaRocque et al., 2013 for similar results using EEG). Together, these findings indicate that stable persistent activity may not be necessary for working memory. Accordingly,

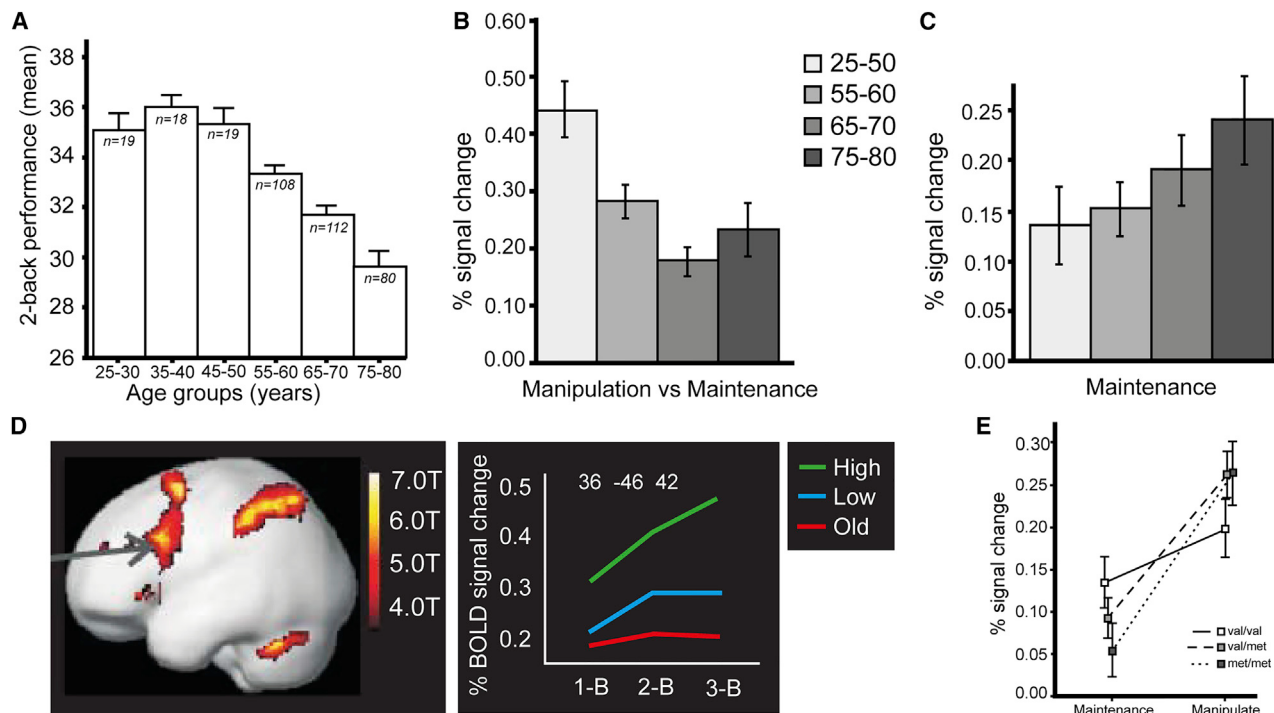


Figure 3. Development and Working Memory

(A) Working-memory performance is stable according to cross-sectional analyses from the 20's to late 50's, after which time a linear decrease is seen. (B) Younger adults show higher upregulation of frontal cortex activity when working-memory demands increase ("manipulation"). (C) By contrast, older adults show elevated prefrontal recruitment during working-memory maintenance. (D) Older adults also tend to show a flatter prefrontal load-dependent response than do younger adults, especially high-performing younger adults. (E) Genetic variation (COMT val/met) modulates the load-dependent prefrontal response in a similar way as aging. (A)–(C) and (E) come from Nyberg et al. (2014), and (D) comes from Nyberg et al. (2009).

alternative coding mechanisms have been suggested (see the sub-section [Neural Network Models of Working Memory](#)).

Plasticity: Lifespan Development and Learning Lifespan Changes

As discussed above in the context of individual differences, working memory is characterized as being a capacity-limited memory system. Several studies suggest that it takes quite some time for individuals to develop full working-memory capacity. Gathercole et al. (2004) examined a large group of children between 4 and 15 years of age, and cross-sectional analyses revealed, across several working-memory measures, an approximately linear increase from 4 to 14 years that leveled off between 14 and 15 years. Similarly, in a longitudinal study of healthy children and adolescents, performance on a visuospatial working-memory task was found to markedly increase from 6 to about 15 years, leveling off thereafter between 15 and 22 (Ullman et al., 2014). Analyses of structural and functional MRI data further revealed that whereas cortical activity in the frontal and parietal lobes were predictive of current working-memory capacity, future capacity could be inferred from structure and activity in basal ganglia and thalamus.

There is relatively little data on adult lifespan changes in working memory, and longitudinal data are particularly scarce. Cross-sectional data from the Swedish *Betula* study (Figure 3A) indicate

stability in working-memory performance between 20 to 50 years and thereafter an apparent linear decline from ages 55–60 to 75–80 years (Nyberg et al., 2014). That study also compared age-related differences on a simpler maintenance task (hold four letters for 3,500 ms) with a more complex working-memory task that required both manipulation and maintenance (mentally go to the next position in the alphabet for two visually presented letters and maintain the result of that operation in working memory). It was found that age differences were more pronounced on the more complex task, and in particular the older (75–80 years) had difficulty with this task. Analyses of fMRI data revealed that older adults showed a weaker increase than younger adults in BOLD signal in DLPFC during manipulation relative to maintenance (Figure 3B). In fact, older adults engaged the DLPFC to a higher degree during the maintenance task (Figure 3C), suggesting a lower frontal cortex efficiency (Nyberg et al., 2014). Correspondingly, some previous studies have suggested that older adults show maximal frontal and parietal BOLD responses at lower levels of working-memory task complexity relative to younger adults (Callicott et al., 1999; Nagel et al., 2009; Nyberg et al., 2009). For example, when performing the *n*-back task with 1-, 2-, and 3-back levels, DLPFC and parietal cortical activity is increased from 1- to 2-back and then levels of or decreases from 2- to 3-back (Figure 3D) for older adults unless they are very high-performing (Nagel et al., 2009).

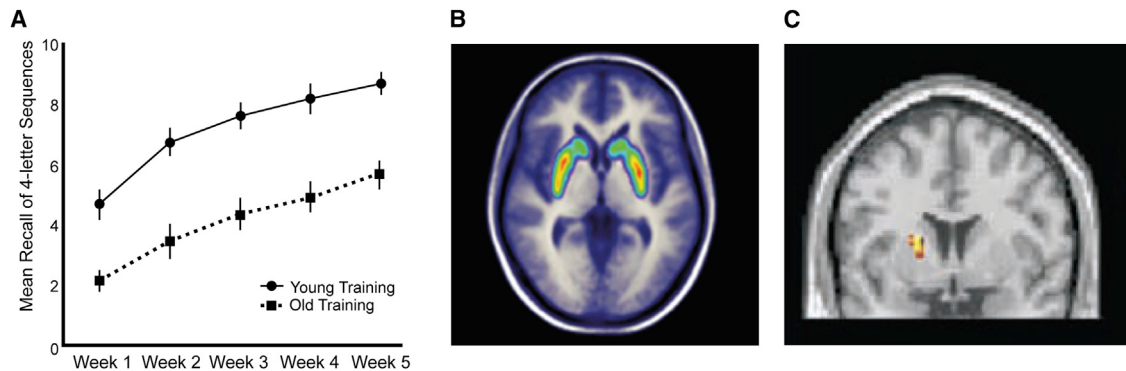


Figure 4. Plasticity and Working Memory

(A) Five weeks of working-memory training strengthened letter-memory performance for both younger and older adults.

(B and C) Such training is associated with dopamine D2 binding potential in the basal ganglia (B) and with striatal activity changes (C). (A) and (C) come from Dahlin et al. (2008b), and (B) is from Bäckman et al. (2011).

Age-related changes in dopaminergic neurotransmission systems have been stressed (Bäckman et al., 2006) and might account for the apparent alterations in working-memory capacity in younger and older age. Indeed, examination of variation in the Val¹⁵⁸Met COMT gene in relation to working-memory manipulation and maintenance revealed a strikingly similar pattern as the age-comparative study (Nyberg et al., 2014). As illustrated in Figure 3E, COMT val carriers, with lower synaptic dopamine levels in prefrontal cortex, showed the strongest DLPFC BOLD response during the easier maintenance condition but the weakest during the more complex manipulation task. By contrast, COMT met carriers showed the strongest upregulation of DLPFC response as a function of task complexity, with an intermediate response for val/met heterozygotes. The similarity in DLPFC activation profiles in aging and as a function of COMT status supports a link between dopamine and DLPFC efficiency (see also Mier et al., 2010), and converges with primate data that have suggested a role of dopamine in working memory (e.g., Wang et al., 2004; for a review see Bäckman et al., 2006). More generally, and consistent with the present process-component framework (cf. Figure 1), a crucial role of fronto-striatal circuits and dopamine might, at least in part, reflect the critical role of attention in working memory (cf. LaHoste et al., 1996; Nieoullon, 2002).

Is It Possible to Strengthen Working Memory by Directed Training?

Traditionally, as noted above, individual differences in working memory have been seen as highly stable over time and considered as a core cognitive trait of an individual. Yet, in recent years, a number of studies have examined training-related changes of working memory, and the findings show that working memory can indeed be improved by certain training programs (Figure 4A) and that such training is associated with cortical and sub-cortical activation changes as measured with fMRI (Dahlin et al., 2009; Klingberg, 2010). Also, positron-emission-tomography-based studies of the dopamine D1 and D2 systems have observed training-related changes in cortical D1 receptors (McNab et al., 2009) and in striatal D2 binding (Figure 4B; Bäckman et al., 2011).

Arguably, the issue that has attracted most interest in the context of training of working memory is the degree to which

such training may transfer to untrained working-memory tasks or even beyond (e.g., to increased long-term memory, intelligence, or reduction of ADHD symptoms). Transfer after working-memory training has been observed in behavioral (Harrison et al., 2013; Klingberg et al., 2002) and brain-imaging (Olesen et al., 2004) studies, and it was reported that demanding working-memory training may even transfer to improve fluid intelligence (Jaeggi et al., 2008). The results of other studies have called into question the possibility of obtaining broad transfer after working-memory training (Redick et al., 2013; Thompson et al., 2013; see Shipstead et al., 2012), and a large-scale study of more than 11,000 participants indicated that this may be the case for cognitive or brain training in general (Owen et al., 2010).

In our training studies we have focused on updating of information in working memory (Dahlin et al., 2008a). Before and after 5 weeks of computerized updating training, fMRI was used to assess training-related changes in brain activity. Three different tasks were scanned; a letter memory updating task (the criterion task) and two transfer tasks: *n*-back working memory and Stroop. All three tasks were expected to engage executive control processes and fronto-parietal circuits, and our fMRI findings supported this prediction. Additionally, the letter-memory task and the *n*-back task were expected to involve updating (Miyake et al., 2000) and engage the striatum (O'Reilly, 2006), whereas the Stroop inhibition task was not. Again, the fMRI results confirmed this prediction. Thus, to the degree that transfer was based on a shared fronto-parietal network, letter memory training should transfer to both *n*-back and Stroop. By contrast, if transfer was more restricted and based on the striatal-updating network, it would be seen for *n*-back only. The findings supported the latter prediction by showing a highly selective behavioral transfer effect to *n*-back along with a training-related modulation of the fMRI signal in the striatum (Figure 4C).

Thus, these and related (Dahlin et al., 2008b) findings support the notion that transfer is highly restrictive and depend on the transfer task taxing the same basic process as is strengthened by the intervention. This notion of process specificity as a basis for transfer is well in agreement with the current processing component framework. Still, it remains an exciting possibility that some forms of working-memory training, possibly by

enhancing attention, can transfer more broadly within working memory (e.g., Brehmer et al., 2012; Harrison et al., 2013), and even to general cognitive domains, such as fluid intelligence (see Au et al., 2015), and to reduce inattention in daily life (Spencer-Smith and Klingberg, 2015).

Models of Working Memory

One of the most influential working-memory models is Baddeley and Hitch (1974)'s multi-store model. A key feature in this model—and one that continues to be a basic feature in all models of working memory (including the current component processes description)—is that working memory emerges from the interaction of several cognitive elements. The multi-store model includes a “central executive” that controls and coordinates information flow in and out of dedicated working memory stores (the phonological loop, the visuospatial sketchpad, and the episodic buffer; see Baddeley, 2012 for a recent overview). Based foremost on neuropsychological dissociations (Scoville and Milner, 1957; Shallice and Warrington, 1970), the working-memory content currently “in” working memory was suggested to be confined to such dedicated buffers and thus separate from LTM, although working memory-LTM interactions are also a core feature of the model. Neurophysiological findings support a separation of executive and storage mechanisms, and of verbal and visuospatial storage (see the sub-section [Brain Areas and Connections](#)), but also show that the regional activity related to working-memory content overlaps with regions involved with perception and LTM of the same content, suggesting a common location for working memory, LTM, and perceptual processes (Lewis-Peacock and Postle, 2008). Furthermore, the MTLs seem to be important for working memory during longer delays, high load, and novel associations, and the PFC is involved in both working memory and LTM, which contradicts the assumption of distinctness between working memory and LTM (Ranganath and Blumenfeld, 2005).

More recent models of working memory are, for the most part, consistent with the multi-store model developed by Baddeley (2012), but they tend to elaborate on particular aspects of working-memory processing and assume that working-memory representations overlap with perceptual and LTM representations (Cowan, 2008; D'Esposito and Postle, 2015; Fuster, 2009; Jonides et al., 2008; McElree, 2006; Oberauer, 2005). Accordingly, such “state-based” models postulate, as we do, that the content of working memory is defined by perceptual and LTM representations being in a particular state of accessibility, largely maintained by persistent neural activity controlled by attentional processes. However, based on empirical findings (see the sub-section [Sustained Brain Activity during Working-Memory Maintenance](#)) and computational simulations (see below), representations outside the current focus of attention can also be considered key working-memory elements.

State-based models differ in terms of how many states of accessibility working-memory content can be in, but generally describe a state of focused attention and one or more states of “activated LTM” (aLTM), that are less accessible than attended representations but more accessible than LTM representations in general (Cowan, 2008; McElree, 2006; Oberauer, 2005). Critically, while attentional processes generally determine working-

memory capacity limits (either by how much information that can be attended simultaneously, or how recently a representation was attended), aLTM representations remain “in” working memory by virtue of the (relatively) increased state of accessibility. This conjecture is supported by findings that if a sub-set of the working memory items is deprioritized by attentional processes, they nonetheless cause an intrusion effect (slower response times) if probed within a few seconds (Oberauer, 2005). Also, if a list of items is held in working memory, the asymptotic accuracy decreases monotonically as a function of the position it had in the list, with the last item having highest accuracy. However, response times are uniform across items in the list, except for the last item, indicating a unique state for the last (attended) item (McElree, 2006). Interestingly, multivariate neuroimaging studies suggest that aLTM might not be active, but rather be latent short-term representations (LaRocque et al., 2013; Lewis-Peacock et al., 2012). Mechanisms other than sustained neural activity may thus contribute to the short-term retention of non-attended working-memory content.

Neural Network Models of Working Memory

Several different computational models at the neural network level have been proposed. Most of the recent models concern working-memory maintenance. This type of model is still far from explaining all behavioral observations, but it has a critical role in connecting the computational level and conceptual theories to neural networks connected by plastic synapses. Early computational models were mainly of a connectionist type, i.e., composed of non-spiking abstract neural units, whereas most later models have been spiking neural network models and have focused on the primary role of the PFC in working memory. The neural mechanisms suggested have been either persistent activity or fast synaptic plasticity, or variants and combinations thereof. The currently prevailing view of the neural mechanisms underlying working memory is based on persistent activity in a recurrent network with fixed connectivity in the PFC.

Working-Memory Models Based on Persistent Activity

It has long been suggested that elevated persistent activity of neurons in the prefrontal cortex is a primary mechanism behind the storage of items in working memory (see above). An early influential computational model of working memory is the non-spiking network developed by O'Reilly et al. (1999). Working-memory maintenance based on persistent activity has also been studied in several spiking attractor memory network models (e.g., Compte et al., 2000). A challenge for the persistent activity working-memory model is how to support multiple-item working memory. The strong lateral inhibition between memory representations causes a winner-take-all dynamics that typically leaves only the winning representation active. Nevertheless, specific network architecture and tuning can achieve a balance that allows for a few simultaneously active items in working memory (Macoveanu et al., 2006). Another challenge is the contrast between the quite unstable intra-trial persistent activity observed experimentally and the very stable persistent activity displayed by such models (Shafi et al., 2007). In fact, this kind of simulated memory is easily erased by a brief dip in the persistent activity, which is inconsistent with empirical findings demonstrating that item representations can be maintained in the absence of sustained activity during the delay (e.g., LaRocque

Box 2. Future Directions

- While there is a consensus that working memory is capacity limited, we have a sparse understanding of why these limits exist in the brain. One popular idea is that the limits arise from difficulty in keeping multiple active representations segregated from one another in neural activity with minimal interference. However, direct evidence for this or other related proposals has not yet been demonstrated.
- The relationship between working memory and conscious experience is not well understood. Is all working-memory content always consciously experienced, and can working memory operate on non-consciously perceived information?
- The relationship is still unclear among more specific processes in working memory (e.g., updating) and more global processes (e.g., those that underlie task sets and intentions). Also, it will be crucial to further consider how motivation influences working memory.
- Can training of working memory lead to broad transfer effects, including affecting long-term memory, fluid intelligence, and academic abilities?
- Although persistent neural activity during working-memory tasks are presumed to result from reentry through recurrent circuits, its precise mechanism is not known and may involve fast plastic changes at the synaptic level.

et al., 2013). Thus, there is reason to look for alternative or complementary mechanisms and explanations for working memory.

A different possible working-memory mechanism, which also addresses the issue of working-memory capacity limits, is that objects in working memory are represented by synchronized firing at the gamma frequency across populations of cells coding for the various features of the object (color, shape, position, etc.; Jensen and Lisman, 1996). The capacity limit of working memory emerges because there is a finite amount of phase space available within a theta cycle to keep multiple items active and separated. Though some experimental support exists (Siegel et al., 2009), the robustness of this mechanism remains to be demonstrated by means of computational models.

Working-Memory Models Based on Fast Synaptic Plasticity

The possibility that working memory could also rely on fast induced and expressed synaptic plasticity has been proposed several times in the past (e.g., O'Reilly et al., 1999; see Barak and Tsodyks, 2014; D'Esposito and Postle, 2015). The hypothesized mechanisms are closely related to those behind LTM, where Hebbian synaptic plasticity (LTP/LTD) is assumed to play a key role. Working memory requires fast expressing and volatile synaptic changes, whereas LTM traces develop slower and lead to long-term stable connectivity. Memory readout in such a network is manifested as persistent activity, but with a duration limited to a few hundred milliseconds, terminated by cellular adaptation and synaptic depression. Since the memory itself is encoded in the connectivity matrix and not in ongoing activity, it is maintained even after activity terminates and can readily be reactivated, as long as the enhanced synaptic con-

nections remain. More recent work on synaptic working memory includes a computational model with non-spiking units (Sandberg et al., 2003; see also Lansner et al., 2013). The synaptic working-memory hypothesis has further been explored using spiking network models with non-Hebbian synaptic plasticity (Lundqvist et al., 2011; Mongillo et al., 2008). Working-memory networks with fast Hebbian plasticity (Erickson et al., 2010) can readily store novel items and associations, but the non-Hebbian forms are limited to already established LTM representations.

Conclusions

The concept of working memory has been enormously influential in a variety of areas and continues to be treated as a single construct, and indeed in some cases at an even higher level, such as “g” (Davies et al., 2015). While such a level of analysis likely will be fruitful in some cases, here we have argued for a more process-based level of analysis. We argue that the latter approach is more likely to yield a detailed understanding of the neural basis, cellular and systems-level, of working memory. Moreover, a process-based approach will lend itself more readily to attempts at detailed simulations of brain-behavior aspects of working-memory functioning. It may also better serve future studies of the basis of working memory deficits in various populations, including children with learning disorders, psychiatric populations, and various forms of age-related mnemonic deficits, including Parkinson's disease. Needless to say, the processing component framework presented here is far from complete, as is the mapping between component processes and associated brain regions (Box 2). For instance, several of the core processes we outlined in Figure 1 may themselves be described in greater detail, and some key processes are likely missing from the list of processes in Figure 1. Also, it has proven challenging to identify unifying principles for the functional organization of the frontal cortex (e.g., Badre and D'Esposito, 2009; Koechlin and Summerfield, 2007). Still, we argue that future work should continue to refine and revise the process-based approach.

At the systems level, working memory has been linked to most areas of the brain. From a component processing perspective, this is to be expected, because different processes that in various constellations implement working memory will by extension involve many different brain regions. A higher level of anatomic specificity is emerging for specific working-memory processing components. For working-memory maintenance per se, fronto-parietal cortical regions make up a core circuit, but it remains to be examined further how such activity reflects key maintenance processes as sustained attention and rehearsal. Moreover, at least in part, fronto-parietal activity could be related to more task-general processing components, such as maintaining goals and task sets (Fuster, 2013; Miller and Cohen, 2001).

The emerging notion of synaptic working memory puts further emphasis on the intimate relation between working memory and LTM (see Fiebig and Lansner, 2014; Teyler and Rudy, 2007). As noted above, functional imaging studies of working memory have observed enhanced activity in MTL regions, and versions of the synaptic working-memory theory suggest that MTL activity could reflect similar Hebbian neurophysiological processes

on an intermediate timescale between working memory and LTM. Indeed, it remains unresolved why some acts of encoding information into working memory may lead to durable LTM representations and others not. Thus, a question for future research will be to examine what fosters both fast and slow synaptic activity.

The notion of close functional relations between working memory and LTM has implications for work on capacity constraints and individual differences. Working-memory capacity constraints predict learning difficulties also in academic settings, and as noted above, attentional capacity may be a critical factor that bridges working memory and complex LTM tasks (cf. Figure 1). Indeed, individual differences in working-memory capacity are strong predictors of retrieval success in many LTM tasks (Cowan et al., 2005). Thus, one critical link between these two memory systems may be the operation of the same attentional control mechanisms to select task-relevant representations, either from the external world or from the vast landscape of internally stored memories in LTM (Unsworth et al., 2014).

Although it is commonly held that working memory and conscious experience are intimately linked, their relationship remains unclear. Most models posit that only attended working-memory content is consciously experienced (Baddeley, 2003; Cowan, 2001; Fuster, 2009; McElree, 2006; Oberauer, 2005), and working-memory content can by this view transition in and out of conscious experience as a function of attentional deployment. However, it is difficult to measure the conscious experience of items outside the focus of attention without transitioning them into attentional focus. An alternative is to investigate the short-term retention of items never consciously perceived. Until recently, the retention of non-consciously perceived information was assumed so short-lived that the notion of non-conscious working memory was given little consideration. However, new findings suggest that non-conscious retention can last at least 15 s, be distractor resistant, engage the prefrontal cortex, and depend on prospective action (Bergström and Eriksson, 2014; Pan et al., 2014; Soto et al., 2011), which together implicate working-memory operations (Soto and Silvano, 2014). Whether the retention of non-conscious representations is most accurately conceptualized as working memory is still controversial and needs further inquiry. If non-conscious working memory indeed exists, it is critical to understand the underlying neural mechanisms (e.g., persistent activity or synaptic plasticity), how non-conscious and conscious working-memory processes interact, and the range of operations that can be used on non-conscious working-memory content, to accurately integrate the notion of non-conscious working memory with current working-memory models.

Not long since the introduction of the concept of working memory and forward to this day, the “Baddeley model” has had enormous influence. We argue here that continued progress in understanding working memory will in no small part depend on picking apart such over-arching descriptions into more detailed descriptions of the component processes that together, in various constellations, can be labeled working-memory processes. We have highlighted information maintenance as a key aspect of working memory and have suggested how this process can be further analyzed into component processes

(Figure 1). We urge for refinements of this and other processing descriptions that together can be described as the “neurocognitive architecture” of working memory, where no doubt further elaborated models that include biophysical details, exemplified here with theories of synaptic working memory, will play a big part. More generally, while we have here focused rather exclusively on working memory, many features of the current neurocognitive architecture are also applicable to most other higher-order (“executive”) functions with a future dimension (Fuster, 2013; Ingvar, 1985).

ACKNOWLEDGMENTS

This work was supported by the Swedish Science Council (to J.E., L.N., and A.L.), the Torsten and Ragnar Söderberg's Foundation (to L.N.), the European Union Seventh Framework Program (FP7/2007–2013) under grant agreement no. 604102 (Human Brain Project) (to L.N. and A.L.) and by grants from the NIH (R01-MH087214) and Office of Naval Research (N00014-12-1-0972) (to E.K.V.).

REFERENCES

- Adam, K.C., Mance, I., Fukuda, K., and Vogel, E.K. (2015). The contribution of attentional lapses to individual differences in visual working memory capacity. *J. Cogn. Neurosci.* 27, 1601–1616.
- Allen, R.J., Vargha-Khadem, F., and Baddeley, A.D. (2014). Item-location binding in working memory: is it hippocampus-dependent? *Neuropsychologia* 59, 74–84.
- Au, J., Sheehan, E., Tsai, N., Duncan, G.J., Buschkuhl, M., and Jaeggi, S.M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychon. Bull. Rev.* 22, 366–377.
- Awh, E., Vogel, E.K., and Oh, S.H. (2006). Interactions between attention and working memory. *Neuroscience* 139, 201–208.
- Axmacher, N., Mormann, F., Fernández, G., Cohen, M.X., Elger, C.E., and Fell, J. (2007). Sustained neural activity patterns during working memory in the human medial temporal lobe. *J. Neurosci.* 27, 7807–7816.
- Bäckman, L., Nyberg, L., Lindenberger, U., Li, S.-C., and Farde, L. (2006). The correlative triad among aging, dopamine, and cognition: current status and future prospects. *Neurosci. Biobehav. Rev.* 30, 791–807.
- Bäckman, L., Nyberg, L., Soveri, A., Johansson, J., Andersson, M., Dahlin, E., Neely, A.S., Virta, J., Laine, M., and Rinne, J.O. (2011). Effects of working-memory training on striatal dopamine release. *Science* 333, 718.
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839.
- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29.
- Baddeley, A.D., and Hitch, G.J. (1974). Working memory. In *Recent Advances in Learning and Motivation*, G.A. Bower, ed. (Academic Press), pp. 47–89.
- Badre, D., and D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat. Rev. Neurosci.* 10, 659–669.
- Barak, O., and Tsodyks, M. (2014). Working models of working memory. *Curr. Opin. Neurobiol.* 25, 20–24.
- Bays, P.M.P., and Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science* 321, 851–854.
- Bergström, F., and Eriksson, J. (2014). Maintenance of non-consciously presented information engages the prefrontal cortex. *Front. Hum. Neurosci.* 8, 938.
- Binder, J.R., Desai, R.H., Graves, W.W., and Conant, L.L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796.

- Bormann, T., Seyboth, M., Umarova, R., and Weiller, C. (2015). "I know your name, but not your number"—patients with verbal short-term memory deficits are impaired in learning sequences of digits. *Neuropsychologia* 72, 80–86.
- Brady, T.F., Konkle, T., and Alvarez, G.A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *J. Exp. Psychol. Gen.* 138, 487–502.
- Brehmer, Y., Westerberg, H., and Bäckman, L. (2012). Working-memory training in younger and older adults: training gains, transfer, and maintenance. *Front. Hum. Neurosci.* 6, 63.
- Brunoni, A.R., and Vanderhasselt, M.-A. (2014). Working memory improvement with non-invasive brain stimulation of the dorsolateral prefrontal cortex: a systematic review and meta-analysis. *Brain Cogn.* 86, 1–9.
- Buchsbaum, B.R., and D'Esposito, M. (2008). The search for the phonological store: from loop to convolution. *J. Cogn. Neurosci.* 20, 762–778.
- Callicott, J.H., Mattay, V.S., Bertolino, A., Finn, K., Coppola, R., Frank, J.A., Goldberg, T.E., and Weinberger, D.R. (1999). Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cereb. Cortex* 9, 20–26.
- Charlton, R.A., Barrick, T.R., Lawes, I.N.C., Markus, H.S., and Morris, R.G. (2010). White matter pathways associated with working memory in normal aging. *Cortex* 46, 474–489.
- Collette, F., Van der Linden, M., Laureys, S., Delfiore, G., Degueldre, C., Luxen, A., and Salmon, E. (2005). Exploring the unity and diversity of the neural substrates of executive functioning. *Hum. Brain Mapp.* 25, 409–423.
- Compte, A., Brunel, N., Goldman-Rakic, P.S., and Wang, X.-J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* 10, 910–923.
- Cools, R., and D'Esposito, M. (2011). Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biol. Psychiatry* 69, e113–e125.
- Cools, R., Miyakawa, A., Sheridan, M., and D'Esposito, M. (2010). Enhanced frontal function in Parkinson's disease. *Brain* 133, 225–233.
- Courtney, S.M., Ungerleider, L.G., Keil, K., and Haxby, J.V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature* 386, 608–611.
- Cowan, N. (1995). *Attention and Memory: An Integrated Framework* (Oxford University Press).
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114, discussion 114–185.
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Prog. Brain Res.* 169, 323–338.
- Cowan, N., Elliott, E.M., Scott Saults, J., Morey, C.C., Mattox, S., Hismjatullina, A., and Conway, A.R.A. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cognit. Psychol.* 51, 42–100.
- D'Esposito, M., and Postle, B.R. (1999). The dependence of span and delayed-response performance on prefrontal cortex. *Neuropsychologia* 37, 1303–1315.
- D'Esposito, M., and Postle, B.R. (2015). The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* 66, 115–142.
- Dahlin, E., Nyberg, L., Bäckman, L., and Neely, A.S. (2008a). Plasticity of executive functioning in young and older adults: immediate training gains, transfer, and long-term maintenance. *Psychol. Aging* 23, 720–730.
- Dahlin, E., Neely, A.S., Larsson, A., Bäckman, L., and Nyberg, L. (2008b). Transfer of learning after updating training mediated by the striatum. *Science* 320, 1510–1512.
- Dahlin, E., Bäckman, L., Neely, A.S., and Nyberg, L. (2009). Training of the executive component of working memory: subcortical areas mediate transfer effects. *Restor. Neurol. Neurosci.* 27, 405–419.
- Daneman, M., and Carpenter, P.A. (1980). Individual differences in working memory and reading. *J. Verb. Learn. Verb. Beh.* 19, 450–466.
- Davies, G., Armstrong, N., Bis, J.C., Bressler, J., Chouraki, V., Giddaluru, S., Hofer, E., Ibrahim-Verbaas, C.A., Kirin, M., Lahti, J., et al.; Generation Scotland (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). *Mol. Psychiatry* 20, 183–192.
- Donkin, C., Nosofsky, R.M., Gold, J., and Shiffrin, R.M. (2013). Discrete-slots models of visual working-memory response times. *Psychol. Rev.* 120, 873–902.
- Drew, T., Horowitz, T.S., Wolfe, J.M., and Vogel, E.K. (2011). Delineating the neural signatures of tracking spatial position and working memory during attentive tracking. *J. Neurosci.* 31, 659–668.
- Ekman, U., Eriksson, J., Forsgren, L., Mo, S.J., Riklund, K., and Nyberg, L. (2012). Functional brain activity and presynaptic dopamine uptake in patients with Parkinson's disease and mild cognitive impairment: a cross-sectional study. *Lancet Neurol.* 11, 679–687.
- Emrich, S.M., Al-Aidroos, N., Pratt, J., and Ferber, S. (2009). Visual search elicits the electrophysiological marker of visual working memory. *PLoS ONE* 4, e8042.
- Emrich, S.M., Riggall, A.C., Larocque, J.J., and Postle, B.R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *J. Neurosci.* 33, 6516–6523.
- Erickson, M.A., Maramba, L.A., and Lisman, J. (2010). A single brief burst induces GluR1-dependent associative short-term potentiation: a potential mechanism for short-term memory. *J. Cogn. Neurosci.* 22, 2530–2540.
- Feredoes, E., Heinen, K., Weiskopf, N., Ruff, C., and Driver, J. (2011). Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proc. Natl. Acad. Sci. USA* 108, 17510–17515.
- Fiebig, F., and Lansner, A. (2014). Memory consolidation from seconds to weeks: a three-stage neural network model with autonomous reinstatement dynamics. *Front. Comput. Neurosci.* 8, 64.
- Fougny, D., Suchow, J.W., and Alvarez, G.A. (2012). Variability in the quality of visual working memory. *Nat. Commun.* 3, 1229.
- Fukuda, K., and Vogel, E.K. (2011). Individual differences in recovery time from attentional capture. *Psychol. Sci.* 22, 361–368.
- Funahashi, S., Bruce, C.J., and Goldman-Rakic, P.S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* 61, 331–349.
- Fuster, J.M. (2009). Cortex and memory: emergence of a new paradigm. *J. Cogn. Neurosci.* 21, 2047–2072.
- Fuster, J.M. (2013). Cognitive functions of the prefrontal cortex. In *Principles of Frontal Lobe Function*, D. Stuss and R. Knight, eds. (Oxford University Press), pp. 11–22.
- Fuster, J.M., and Alexander, G.E. (1971). Neuron activity related to short-term memory. *Science* 173, 652–654.
- Fuster, J.M., and Jervey, J.P. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* 212, 952–955.
- Fuster, J.M., and Bressler, S.L. (2012). Cognit activation: a mechanism enabling temporal integration in working memory. *Trends Cogn. Sci.* 16, 207–218.
- Fuster, J.M., Bauer, R.H., and Jervey, J.P. (1985). Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain Res.* 330, 299–307.
- Gathercole, S.E., Pickering, S.J., Ambridge, B., and Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Dev. Psychol.* 40, 177–190.
- Gazzaley, A., Rissman, J., and D'Esposito, M. (2004). Functional connectivity during working memory maintenance. *Cogn. Affect. Behav. Neurosci.* 4, 580–599.
- Goldman-Rakic, P.S. (1995). Cellular basis of working memory. *Neuron* 14, 477–485.

- Harrison, S.A., and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458, 632–635.
- Harrison, T.L., Shipstead, Z., Hicks, K.L., Hambrick, D.Z., Redick, T.S., and Engle, R.W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychol. Sci.* 24, 2409–2419.
- Ingvar, D.H. (1985). “Memory of the future”: an essay on the temporal organization of conscious awareness. *Hum. Neurobiol.* 4, 127–136.
- Jaeggi, S.M., Buschkuhl, M., Jonides, J., and Perrig, W.J. (2008). Improving fluid intelligence with training on working memory. *Proc. Natl. Acad. Sci. USA* 105, 6829–6833.
- Jenkinson, A., and Squire, L.R. (2012). Working memory, long-term memory, and medial temporal lobe function. *Learn. Mem.* 19, 15–25.
- Jensen, O., and Lisman, J.E. (1996). Novel list of 7±2 known items can be reliably stored in an oscillatory short-term memory network: interaction with long-term memory. *Learn. Mem.* 3, 257–263.
- Jensen, O., Gelfand, J., Kounios, J., and Lisman, J.E. (2002). Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. *Cereb. Cortex* 12, 877–882.
- Jonides, J., Lewis, R.L., Nee, D.E., Lustig, C.A., Berman, M.G., and Moore, K.S. (2008). The mind and brain of short-term memory. *Annu. Rev. Psychol.* 59, 193–224.
- Kane, M.J., and Engle, R.W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychon. Bull. Rev.* 9, 637–671.
- Kastner, S., and Ungerleider, L.G. (2000). Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23, 315–341.
- Klimesch, W. (2012). α -band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617.
- Klingberg, T. (2010). Training and plasticity of working memory. *Trends Cogn. Sci.* 14, 317–324.
- Klingberg, T., Forssberg, H., and Westerberg, H. (2002). Training of working memory in children with ADHD. *J. Clin. Exp. Neuropsychol.* 24, 781–791.
- Koechlin, E., and Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci.* 11, 229–235.
- Koenigs, M., Barbey, A.K., Postle, B.R., and Grafman, J. (2009). Superior parietal cortex is critical for the manipulation of information in working memory. *J. Neurosci.* 29, 14980–14986.
- LaHoste, G.J., Swanson, J.M., Wigal, S.B., Glabe, C., Wigal, T., King, N., and Kennedy, J.L. (1996). Dopamine D4 receptor gene polymorphism is associated with attention deficit hyperactivity disorder. *Mol. Psychiatry* 1, 121–124.
- Lansner, A., Marklund, P., Sikström, S., and Nilsson, L.-G. (2013). Reactivation in working memory: an attractor network model of free recall. *PLoS ONE* 8, e73776.
- LaRocque, J.J., Lewis-Peacock, J.A., Drysdale, A.T., Oberauer, K., and Postle, B.R. (2013). Decoding attended information in short-term memory: an EEG study. *J. Cogn. Neurosci.* 25, 127–142.
- Levy, R., and Goldman-Rakic, P.S. (2000). Segregation of working memory functions within the dorsolateral prefrontal cortex. *Exp. Brain Res.* 133, 23–32.
- Lewis-Peacock, J.A., and Postle, B.R. (2008). Temporary activation of long-term memory supports working memory. *J. Neurosci.* 28, 8765–8771.
- Lewis-Peacock, J.A., Drysdale, A.T., Oberauer, K., and Postle, B.R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *J. Cogn. Neurosci.* 24, 61–79.
- Liebe, S., Hoerzer, G.M., Logothetis, N.K., and Rainer, G. (2012). Theta coupling between V4 and prefrontal cortex predicts visual short-term memory performance. *Nat. Neurosci.* 15, 456–462, S1–S2.
- Luck, S.J., and Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281.
- Lundqvist, M., Herman, P., and Lansner, A. (2011). Theta and gamma power increases and alpha/beta power decreases with memory load in an attractor network model. *J. Cogn. Neurosci.* 23, 3008–3020.
- Macoveanu, J., Klingberg, T., and Tegnér, J. (2006). A biophysical model of multiple-item working memory: a computational and neuroimaging study. *Neuroscience* 141, 1611–1618.
- McElree, B. (2006). Accessing recent events. *Psychol. Learn. Motiv.* 46, 155–200.
- McNab, F., and Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nat. Neurosci.* 11, 103–107.
- McNab, F., Varrone, A., Farde, L., Jucaite, A., Bystritsky, P., Forssberg, H., and Klingberg, T. (2009). Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science* 323, 800–802.
- Mier, D., Kirsch, P., and Meyer-Lindenberg, A. (2010). Neural substrates of pleiotropic action of genetic variation in COMT: a meta-analysis. *Mol. Psychiatry* 15, 918–927.
- Miller, G.A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97.
- Miller, E.K., and Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202.
- Miller, E.K., Li, L., and Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J. Neurosci.* 13, 1460–1478.
- Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., and Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100.
- Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science* 319, 1543–1546.
- Nagel, I.E., Preuschhof, C., Li, S.-C., Nyberg, L., Bäckman, L., Lindenberger, U., and Heekeren, H.R. (2009). Performance level modulates adult age differences in brain activation during spatial working memory. *Proc. Natl. Acad. Sci. USA* 106, 22552–22557.
- Nee, D.E., Brown, J.W., Askren, M.K., Berman, M.G., Demiralp, E., Krawitz, A., and Jonides, J. (2013). A meta-analysis of executive components of working memory. *Cereb. Cortex* 23, 264–282.
- Nieoullon, A. (2002). Dopamine and the regulation of cognition and attention. *Prog. Neurobiol.* 67, 53–83.
- Nyberg, L., Dahlin, E., Stigsdotter Neely, A., and Bäckman, L. (2009). Neural correlates of variable working memory load across adult age and skill: dissociative patterns within the fronto-parietal network. *Scand. J. Psychol.* 50, 41–46.
- Nyberg, L., Andersson, M., Kauppi, K., Lundquist, A., Persson, J., Pudas, S., and Nilsson, L.-G. (2014). Age-related and genetic modulation of frontal cortex efficiency. *J. Cogn. Neurosci.* 26, 746–754.
- O’Reilly, R.C. (2006). Biologically based computational models of high-level cognition. *Science* 314, 91–94.
- O’Reilly, R.C., Braver, T.S., and Cohen, J.D. (1999). A biologically based computational model of working memory. In *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, A. Miyake and P. Shah, eds. (Cambridge University Press), pp. 375–411.
- Oberauer, K. (2005). Control of the contents of working memory—a comparison of two paradigms and two age groups. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 714–728.
- Olesen, P.J., Westerberg, H., and Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nat. Neurosci.* 7, 75–79.
- Olson, I.R., Page, K., Moore, K.S., Chatterjee, A., and Verfaellie, M. (2006). Working memory for conjunctions relies on the medial temporal lobe. *J. Neurosci.* 26, 4596–4601.
- Olsson, H., and Poom, L. (2005). Visual memory needs categories. *Proc. Natl. Acad. Sci. USA* 102, 8776–8780.

- Owen, A.M., Morris, R.G., Sahakian, B.J., Polkey, C.E., and Robbins, T.W. (1996). Double dissociations of memory and executive functions in working memory tasks following frontal lobe excisions, temporal lobe excisions or amygdalo-hippocampectomy in man. *Brain* 119, 1597–1615.
- Owen, A.M., McMillan, K.M., Laird, A.R., and Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* 25, 46–59.
- Owen, A.M., Hampshire, A., Grahn, J.A., Stenton, R., Dajani, S., Burns, A.S., Howard, R.J., and Ballard, C.G. (2010). Putting brain training to the test. *Nature* 465, 775–778.
- Pan, Y., Lin, B., Zhao, Y., and Soto, D. (2014). Working memory biasing of visual perception without awareness. *Atten. Percept. Psychophys.* 76, 2051–2062.
- Petersen, S.E., and Posner, M.I. (2012). The attention system of the human brain: 20 years after. *Annu. Rev. Neurosci.* 35, 73–89.
- Pisella, L., Berberovic, N., and Mattingley, J.B. (2004). Impaired working memory for location but not for colour or shape in visual neglect: a comparison of parietal and non-parietal lesions. *Cortex* 40, 379–390.
- Postle, B.R. (2015). The cognitive neuroscience of visual short-term memory. *Curr. Opin. Beh. Sci.* 1, 40–46.
- Prime, D.J., and Jolicoeur, P. (2010). Mental rotation requires visual short-term memory: evidence from human electric cortical activity. *J. Cogn. Neurosci.* 22, 2437–2446.
- Ranganath, C., and Blumenfeld, R.S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends Cogn. Sci.* 9, 374–380.
- Ranganath, C., Cohen, M.X., Dam, C., and D'Esposito, M. (2004). Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *J. Neurosci.* 24, 3917–3925.
- Redick, T.S., Shipstead, Z., Harrison, T.L., Hicks, K.L., Fried, D.E., Hambrick, D.Z., Kane, M.J., and Engle, R.W. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *J. Exp. Psychol. Gen.* 142, 359–379.
- Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–590.
- Roux, F., and Uhlhaas, P.J. (2014). Working memory and neural oscillations: α - γ versus θ - γ codes for distinct WM information? *Trends Cogn. Sci.* 18, 16–25.
- Sandberg, A., Tegnér, J., and Lansner, A. (2003). A working memory model based on fast Hebbian learning. *Network* 14, 789–802.
- Scoville, W.B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* 20, 11–21.
- Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., and Bodner, M. (2007). Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience* 146, 1082–1108.
- Shallice, T., and Warrington, E.K. (1970). Independent functioning of verbal memory stores: a neuropsychological study. *Q. J. Exp. Psychol.* 22, 261–273.
- Shipstead, Z., Redick, T.S., and Engle, R.W. (2012). Is working memory training effective? *Psychol. Bull.* 138, 628–654.
- Siegel, M., Warden, M.R., and Miller, E.K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proc. Natl. Acad. Sci. USA* 106, 21341–21346.
- Soto, D., and Silvano, J. (2014). Reappraising the relationship between working memory and conscious awareness. *Trends Cogn. Sci.* 18, 520–525.
- Soto, D., Mäntylä, T., and Silvano, J. (2011). Working memory without consciousness. *Curr. Biol.* 21, R912–R913.
- Spencer-Smith, M., and Klingberg, T. (2015). Benefits of a working memory training program for inattention in daily life: a systematic review and meta-analysis. *PLoS ONE* 10, e0119522.
- Stokes, M.G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78, 364–375.
- Stoodley, C.J., and Schmahmann, J.D. (2009). Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage* 44, 489–501.
- Taylor, T.J., and Rudy, J.W. (2007). The hippocampal indexing theory and episodic memory: updating the index. *Hippocampus* 17, 1158–1169.
- Thompson, T.W., Waskom, M.L., Garel, K.-L.A., Cardenas-Iniguez, C., Reynolds, G.O., Winter, R., Chang, P., Pollard, K., Lala, N., Alvarez, G.A., and Gabrieli, J.D. (2013). Failure of working memory training to enhance cognition or intelligence. *PLoS ONE* 8, e63614.
- Tsubomi, H., Fukuda, K., Watanabe, K., and Vogel, E.K. (2013). Neural limits to representing objects still within view. *J. Neurosci.* 33, 8257–8263.
- Ullman, H., Almeida, R., and Klingberg, T. (2014). Structural maturation and brain activity predict future working memory capacity during childhood development. *J. Neurosci.* 34, 1592–1598.
- Unsworth, N., Fukuda, K., Awh, E., and Vogel, E.K. (2014). Working memory and fluid intelligence: capacity, attention control, and secondary memory retrieval. *Cognit. Psychol.* 71, 1–26.
- Vallar, G., and Baddeley, A.D. (1984). Fractionation of working memory: neuropsychological evidence for a phonological short-term store. *J. Verb. Learn. Verb. Beh.* 23, 151–161.
- van den Berg, R., Shin, H., Chou, W.-C., George, R., and Ma, W.J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proc. Natl. Acad. Sci. USA* 109, 8780–8785.
- van Dijk, H., van der Werf, J., Mazaheri, A., Medendorp, W.P., and Jensen, O. (2010). Modulations in oscillatory activity with amplitude asymmetry can produce cognitively relevant event-related responses. *Proc. Natl. Acad. Sci. USA* 107, 900–905.
- Vogel, E.K., and Machizawa, M.G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature* 428, 748–751.
- Voytek, B., and Knight, R.T. (2010). Prefrontal cortex and basal ganglia contributions to visual working memory. *Proc. Natl. Acad. Sci. USA* 107, 18167–18172.
- Wager, T.D., and Smith, E.E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* 3, 255–274.
- Wang, M., Vijayaghavan, S., and Goldman-Rakic, P.S. (2004). Selective D2 receptor actions on working memory. *Science* 303, 853–856.
- Wilken, P., and Ma, W.J. (2004). Selective D2 receptor actions on the functional circuitry of working memory. *J. Vis.* 4, 1120–1135.