CHAPTER 12

Working Memory
An Evolving Concept

DEREK EVAN NEE AND MARK D’ESPOSITO

INTRODUCTION

Traditionally, working memory is conceptualized as a system involved in the maintenance and manipulation of information in the absence of external stimulation. Intuitively, working memory might be considered a formalized system for describing thought (Jonides, 1995). When you consider what is “on your mind”—from mental images to inner speech—you are considering your working memory. In this conceptualization, working memory is transient—lasting for however long the thought remains useful, typically on the order of seconds.

Functionally, working memory is a mental work space that is critical for higher-level cognitive abilities. Several tasks have been devised to measure the capacity of working memory or the amount of information one can hold in mind (Luck & Vogel, 1997; Turner & Engle, 1989). Individual differences in performance on these tasks predict reasoning, planning, problem-solving, and a variety of other cognitive abilities (Carpenter, Just, & Shell, 1990; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Fukuda, Vogel, Mayr, & Awh, 2010; Just & Carpenter, 1992), demonstrating the importance of having a large mental work space. As a result of its broad impact, a clear understanding of working memory is essential for understanding cognition.

Because of its central importance, working memory has been a topic of intense interest in psychology and neuroscience. Foundational research in working memory was performed by Baddeley and colleagues, and to this day Baddeley’s multicomponent model remains one of the most influential models of working memory (Baddeley, 1986, 2012; Baddeley & Hitch, 1974). Although a full treatment of the model and the data that support it is outside the scope of this chapter, certain key assumptions of the model lay the groundwork for the issues to come. First, the model assumes that working memory is distinct from long-term memory. That is, the processes that underlie the temporary maintenance of information for working memory can be distinguished from the processes that underlie the more permanent retention of information associated with long-term memory. Second, the model assumes distinct representational bases for different kinds of information. For example, materials that can be verbalized (e.g., words, letters) are represented in a qualitatively distinct way from materials that can be visualized (e.g., faces, scenes). Finally, the model assumes that processes that select, modulate, and manipulate retained information are
fundamentally dissociable from those that retain the information. As we describe, these assumptions have been profoundly influential to the neuroscientific study of working memory.

Before delving into the neural investigation of working memory, it is important to first define the scope of what working memory encompasses. Using Baddeley’s definition, working memory includes not only the retention of information but also the processing of retained information to support ongoing cognition. However, in much of Baddeley’s own work and the psychological and neuroscientific work that followed, retention and processing aspects of working memory have been studied in isolation from one another. This ignores a potential important interaction in that the nature of retention may change depending on the intended processing (e.g., Lee, Kravitz, & Baker, 2013). As such, a strict definition of working memory would require that the information is retained not simply for its own right (e.g., not for a simple match or reproduction) but also for the purpose of some other cognitive activity (e.g., retaining numbers for the purpose of mental arithmetic). The former might instead be referred to as short-term memory and may very well engage distinct mechanisms compared to the latter. Nevertheless, because the lion’s share of neuroscientific research has examined short-term memory assuming that it has implications for working memory, we will consider all forms of short-term retention. Toward the end of the chapter, we revisit this distinction and whether it may have implications for apparent discrepancies in the literature.

BACKGROUND ISSUES

The “Delay Cell”

What is the neurobiological underpinning of working memory? One might begin to address this question by eliminating what might not be the neural instantiation of working memory. For example, suppose we accept Baddeley’s assumption that working memory is distinct from long-term memory. Then it could be the case that (1) distinct neural structures are associated with working memory and long-term memory, (2) distinct neural mechanisms support working memory and long-term memory, or (3) both. Let’s begin with the first possibility. The structure most implicated in long-term memory is the medial temporal lobe because of the profound amnesia observed in patients with damage to this brain region (Cave & Squire, 1992; Scoville & Milner, 1957). Notably, working memory performance is widely reported as spared in these patients, suggesting that different neural structures mediate working memory function (but see Ranganath & Blumenfeld, 2005). Next, consider the second possibility. Although the neural mechanisms associated with long-term memory are widely assumed to involve lasting changes at the synapse (e.g., long-term potentiation), the transient nature of working memory would suggest a mechanism that does not necessarily lead to lasting structural changes. A natural hypothesis that follows would be that working memory is associated with changes in neuronal firing (e.g., Hebb, 1949). Taken together, a strong candidate for a neurobiological mechanism of working memory would be neuronal firing in brain regions other than the medial temporal lobe that is sustained while information is retained for working memory but fades when the information is discarded.

Early clues for where to find such neuronal firing came from lesion studies in monkeys. Jacobsen (1935) observed that following bilateral lesions to the prefrontal cortex (PFC), monkeys were markedly impaired in retaining a location in memory during a brief delay, yet they were spared in the ability to perform visual object discrimination
and the ability to learn new tasks involving visual object discrimination. Studies that followed confirmed a critical role for the PFC in short-term retention with marked behavioral impairments resulting from PFC lesions (see Curtis & D’Esposito, 2004, for a review; Goldman & Rosvold, 1970; Mishkin, 1957; Mishkin, Vest, Waxler, & Rosvold, 1969). Following these lesion studies, it was found that neurons in the PFC showed sustained elevated firing during a delay interval between the presentation of an item and a test of that item (Fuster & Alexander, 1971; Kubota & Niki, 1971). The duration of sustained elevated firing of these cells varied with manipulations in the length of the delay interval leading to the labeling of such cells as delay cells. Given their temporal properties and localization within the PFC, these delay cells appeared to be a strong candidate for the neurobiological locus of representation for working memory.

Several different cognitive processes likely span the delay between a cue and a test on these tasks. For example, there is the anticipation of the test, vigilance, and so on. How can one be certain that delay cells correspond to the representations maintained for working memory? To provide such evidence, Funahashi, Bruce, and Goldman-Rakic (1989) trained monkeys to remember spatial locations across short delays and examined whether different delay cells maintained information about different spatial locations. Such stimulus specificity would provide strong evidence for a link between cellular activity and working memory. These authors found precisely such tuning in the PFC. Individual cells demonstrated preferences for specific locations in space showing sustained elevated firing for their preferred stimulus. Moreover, a neuron’s firing rate diminished as a function of the distance of a memoranda from the neuron’s preferred location, leading to a tuning curve relating firing to maintained stimuli (Figure 12.1). Furthermore, small PFC lesions produced “mnemonic scotomas,” impairing memory just for the parts of space that were lesioned, providing evidence that PFC neurons play a causal role in stimulus-specific retention (Funahashi, Bruce, & Goldman-Rakic, 1993; but see Tsujimoto & Postle, 2012). Collectively, these data support the idea that sustained neuronal firing in the PFC mediates retention for working memory.

The PFC contains cells sensitive to a variety of properties including visual stimuli, motor responses, reward, and so on (Goldman-Rakic, Funahashi, & Bruce, 1990). What differentiates delay cells from these other cells? The defining feature of delay cell activity is its persistence, which contrasts with the more punctate activity profile of stimulus- and response-driven cells. A fair amount of computational and empirical research has explored the mechanisms underlying persistent elevated firing (see Arnsten, Wang, & Paspalas, 2012, for an in-depth review). Computational work has demonstrated that stable persistent firing requires a delicate balance of excitation and inhibition, whereby excitation operates over a slower timescale than inhibition (Compte, Brunel, Goldman-Rakic, & Wang, 2000; Lim & Goldman, 2013; Wang, 1999). If this balance is not met, activity will either decay to baseline or grow to a maximum that does not resemble the in vivo behavior of delay cells. These computational dynamics implicate a critical role of NMDA receptors whose temporal properties match the appropriately slow excitatory timescales necessary to produce persistent firing (Wang, 2001). Supporting these computational observations, empirical work has demonstrated that blocking NMDA receptors abolishes delay cell persistent firing in the PFC while also greatly diminishing stimulus selectivity (Wang et al., 2013). Intriguingly, disrupted excitatory-inhibitory balance due to deficient NMDA receptor function provides a
cellular model of impaired working memory in patients with schizophrenia (Murray, Anticevic, et al., 2014). Together, these results indicate that precise NMDA-mediated molecular dynamics in the PFC underlie retention for working memory.

In addition to intrinsic cellular mechanisms, recurrent inputs from non-PFC areas are likely to be partly responsible for persistent elevated firing in the PFC (Major & Tank, 2004). The PFC has diverse connections, receiving input from numerous cortical and subcortical areas (Barbas, 1988; Miller & Cohen, 2001; Petrides & Pandya, 1999, 2002). Since the initial observations of persistent, elevated firing in the PFC, there have been abundant demonstrations of persistent, elevated firing in regions other than the PFC including temporal (Fuster & Jervey, 1982; Miyashita & Chang, 1988); parietal...
Background Issues 401

(Chafee & Goldman-Rakic, 1998; Mazzoni, Bracewell, Barash, & Andersen, 1996; Pesaran, Pezaris, Sahani, Mitra, & Andersen, 2002); and subcortical areas (Fuster & Alexander, 1971; Watanabe & Niki, 1985). Hence, understanding the persistent activity that mediates working memory representation requires an approach that examines multiple areas of the brain at once, revealing the networks that support working memory.

Networks of Working Memory

Based on a combination of cell recordings, anatomical projections, and causal manipulations (e.g., lesion, cooling), Goldman-Rakic (1987) proposed a network model of working memory. In this framework, distinct frontal-posterior networks mediate spatial and object working memory, respectively. This dichotomy has its origins in the well-characterized “what-where” bifurcation observed in temporal or parietal areas subserving visual processing (Ungerleider & Mishkin, 1982). In posterior cortices, dorsal areas process location-based inputs and ventral areas process identity-based inputs. Accordingly, dorsal areas of the PFC coordinate with the parietal cortex to support spatial working memory and ventral areas of the PFC coordinate with the temporal cortex to support object working memory (Goldman-Rakic, 1987; Levy & Goldman-Rakic, 2000). Human neuroimaging research has been broadly consistent with these findings. For example, Courtney, Petit, Maisog, Ungerleider, and Haxby (1998) examined delay-period fMRI signal while human participants sustained memory representations of either faces or spatial locations. These authors found sustained elevated signals in the ventrolateral PFC (area 45) during the maintenance of faces and sustained elevated signals in the caudal superior frontal sulcus (area 8) and posterior parietal cortex during the maintenance of locations. This dorsal-ventral distinction in the PFC has been replicated numerous times in direct comparisons between spatial and object working memory (Courtney, Ungerleider, Keil, & Haxby, 1996; Mohr, Goebel, & Linden, 2006; Munk et al., 2002; Sala & Courtney, 2007; Sala, Rama, & Courtney, 2003), as well as in meta-analyses of neuroimaging data (Nee et al., 2013; Owen, McMillan, Laird, & Bullmore, 2005; Rottschy et al., 2012). These distinctions are reinforced by causal data obtained through transcranial magnetic stimulation of either dorsal or ventral PFC (Mottaghy, Gangitano, Sparing, Krause, & Pascual-Leone, 2002). Dorsal-ventral distinctions between spatial and object working memory have also been observed in posterior cortices (Courtney et al., 1996; Postle & D’Esposito, 1999; Postle, Stern, Rosen, & Corkin, 2000; Sala et al., 2003; Smith et al., 1995) and in meta-analyses (Rottschy et al., 2012; Wager & Smith, 2003). Collectively, these data indicate commonalities between the monkey and human working memory systems and validate the efficacy of human neuroimaging to the study of the neural mechanisms of working memory.

Although monkeys and humans appear to have similar dorsal-ventral networks governing the maintenance of spatial-object working memory, humans have additional capacities not shared with the monkey. To keep information active in mind, humans often rely on verbal codes, covertly rehearsing relevant information (e.g., holding in mind a phone number before dialing). Because of its ubiquitous use in human cognition, the verbal working memory system has garnered intense attention in the psychological literature (Baddeley, 2012). Commensurately, human neuroimaging has repeatedly localized neural correlates of verbal working memory to the left ventrolateral PFC (areas 44 and 45, i.e., Broca’s area), and the left peri-sylvian
cortices (Awh et al., 1996; Chein & Fiez, 2001; Cohen et al., 1997; Fiez et al., 1996; Paulesu, Frith, & Frackowiak, 1993; Postle, Berger, & D’Esposito, 1999; Smith, Jonides, & Koepppe, 1996). Furthermore, damage to these areas causes impairments in verbal working memory (Koenigs et al., 2011). These results dovetail nicely with the visuo-spatial working memory systems previously described. In each case, working memory of a given form engages a particular region of the PFC along with a particular posterior cortical area (Figure 12.2).

What are the respective roles of frontal and posterior cortices in working memory? Taking verbal working memory as an example, the frontal area—Broca’s area—is involved in sequencing phonemes for the purpose of articulation (Flinker et al., 2015). Concurrently, the posterior area—peri-sylvian cortex—is involved in phonological representation (Paulesu et al., 1993), binding acoustic representations of speech with articulatory counterparts (Buchsbaum & D’Esposito, 2008). Put more abstractly, the posterior area maintains representations while the frontal area keeps those representations active via rehearsal or refreshing. Similar coordination between frontal and posterior areas are thought to govern spatial and object working memory as well as other working memory systems (Courtney, 2004; Curtis & D’Esposito, 2003; Jonides, Lacey, & Nee, 2005; Pasternak & Greenlee, 2005; Postle, 2006). Using these ideas, posterior systems involved in the representation of different types of information for the purposes of perception and long-term memory are

**Figure 12.2** Networks involved in the maintenance of working memory representations. Dissociable networks are engaged for the maintenance of different stimulus domains (top—spatial, middle—verbal, bottom—object). In each case, a frontal area (SFS—superior frontal sulcus; Broca’s—Broca’s area; VLPFC—ventrolateral PFC) is involved in maintenance processes including rehearsal or refreshing of posterior representations (PPC—posterior parietal cortex; peri-sylvian—peri-sylvian cortex; IT—inferior temporal cortex). Color version of this figure is available at http://onlinelibrary.wiley.com/book/10.1002/9781119170174.
also those involved in the maintenance of that information across working memory delays. Evidence for these claims comes from studies that examined posterior areas that preferentially process different object types. For example, an area in the fusiform gyrus shows heightened activity during the perception of faces relative to other objects (i.e., fusiform face area [FFA]; Kanwisher, McDermott, & Chun, 1997). It has been observed that the FFA demonstrates persistent activation during delay periods (Druzgal & D’Esposito, 2001; Lepsien & Nobre, 2007; Ranganath, DeGutis, & D’Esposito, 2004; Sala et al., 2003). Furthermore, delay-period FFA activation is positively modulated by the number of faces held for object working memory, suggesting that its activity is a linear function of working memory load (Druzgal & D’Esposito, 2001, 2003). Moreover, working memory activation in the FFA can occur independently of the presentation of faces. When participants learn to associate non-face cues with face stimuli for a delayed paired-association decision, presentation of the non-face cue elicits delay-period activation in the FFA (Ranganath, Cohen, Dam, & D’Esposito, 2004). Finally, these observations are not exclusive to the FFA. For example, areas of temporal cortex that represent scenes show scene-specific working memory effects (Lepsien & Nobre, 2007; Ranganath, DeGutis, et al., 2004). Together, these data indicate that the same areas of the brain that process categorical object information in perception and long-term memory retain that information for working memory (see also Polyn, Natu, Cohen, & Norman, 2005; Squire & Wixted, 2011).

The idea that information for working memory is represented in posterior areas is a marked departure from initial hypotheses that the PFC is responsible for representation (e.g., Goldman-Rakic, 1995). However, human neuroimaging has made it abundantly clear that the PFC is active across a wide variety of cognitive demands, many of which have no or minimal requirement to sustain representations across a delay (Duncan & Owen, 2000; Niendam et al., 2012). As a result, the idea that the PFC serves specifically as a store for working memory representations appears to be somewhat untenable. A more general hypothesis is that the PFC is involved in the top-down selection or biasing of posterior representations (Miller & D’Esposito, 2005; Miller & Cohen, 2001; Sreenivasan, Curtis, & D’Esposito, 2014). This account synergizes well with the idea that PFC areas are involved in the rehearsal or refreshing of posterior representations in order to keep them active for working memory. These maintenance processes can be achieved through repeated selection of representations by way of top-down control. Similar top-down selection processes are presumed to subserve the selection of representations in perception through attention (Desimone & Duncan, 1995), suggesting common selection mechanisms underlying attention and working memory (Gazzaley & Nobre, 2012). Hence, working memory involves the selection of the same representations involved in perception. One salient difference between attention and working memory then is that in the case of the former, the representations persist in the environment, and in the latter, the representations exist solely in the mind.

**RECENT AND EMERGING TRENDS**

**Representational Locus of Working Memory**

The reviewed fMRI literature supports the idea that posterior cortices involved in perception are also involved in working memory maintenance through sustained elevated activity. Although compelling, the presence
of elevated category-specific fMRI signals in the posterior areas related to working memory provides substantially less detailed information than the stimulus-specific tuning of PFC cells documented in the monkey (e.g., Funahashi et al., 1989). That is, activity that demonstrates the representation of categorical information does not necessarily indicate that a given instance of that category is retained for working memory. In particular, elevated activity in category-selective areas can be observed by the mere expectation of perceiving an object of a given category (Esterman & Yantis, 2010; Puri, Wojciulik, & Ranganath, 2009). Presumably, such activity reflects a general modulation (e.g., gain) of category-selective cells without favoring any one instance of the category over another. Given that typical working memory paradigms use a category-specific probe to query retained information, the anticipation of a probe of a given category type may be responsible for the observed elevated activity in category-selective areas (Lepsien, Thornton, & Nobre, 2011). As a result, delay period category selectivity is insufficient to provide evidence for the representational basis of working memory. Activity that can distinguish particular instances of a category would provide more convincing evidence.

More recent data-analytic methods have enabled better examination of stimulus-specific tuning of working memory representations. A particularly elegant approach is the use of forward models built using hypothetical tuning functions (Brouwer & Heeger, 2009, 2011). The idea behind this approach is that the observed signal (e.g., in a voxel for fMRI) is composed of a mixture of neurons each tuned to particular stimulus values, showing maximal responses to their preferred stimulus and smoothly decreasing responses to stimuli increasingly distal from the preferred stimulus (Figure 12.3). Assuming this, signals can be modeled as a weighted sum of tuning functions centered at different stimulus values, with the weights providing estimates of the relative amounts of differently tuned neurons contributing to a given signal. These weights are then estimated based on an observed signal in a subset of the data and then those weight estimates are used to reconstruct representational values on a test set of data. If accurate, the model should faithfully reproduce the hypothetical representation held for working memory. These modeling techniques are often coupled with production tasks wherein participants must recall the originally presented stimulus through a method of adjustment (Wilken & Ma, 2004), such as tilting the orientation of a test grating to match the study grating. As long as the probe is randomized (e.g., a random orientation angle), this procedure ensures a retrospective code for what had originally been presented rather than an anticipatory code about a pending motor response (e.g., Curtis, Rao, & D’Esposito, 2004) or probe stimulus (e.g., Lepsien, et al., 2011).

The forward-modeling approach has been used to demonstrate that early visual areas retain a stimulus-specific representation of visual memoranda during short-delay intervals (Ester, Anderson, Serences, & Awh, 2013; Ester, Sprague, & Serences, 2015; Sprague, Ester, & Serences, 2014; Sprague & Serences, 2013). Furthermore, the precision with which stimulus representations can be reconstructed by forward models predict individual differences in behavioral accuracy (Ester et al., 2013), and reconstruction precision and behavioral performance decrease with memory load (Sprague et al., 2014). These data indicate that the reconstructed representations relate directly to overt performance. Hence, these data provide strong evidence that perceptual representations are maintained for working memory in posterior brain areas.
Figure 12.3  Forward modeling approach. Neural responses are hypothesized to show tuning with maximal responses to preferred stimuli and smoothly decreasing responses with increasing distance from the preferred stimulus. Orientation angles are used in this example. Top left: idealized tuning function centered at 0. Top right: a forward encoding model positing eight tuning channels (i.e., eight different populations of neurons) each centered at a different preferred angle. Training data are used to estimate a voxel’s weights across the eight tuning channels. Then for each trial of the test data, the activation pattern across voxels can be combined with the weight estimates to reconstruct the presented stimulus. Color version of this figure is available at http://onlinelibrary.wiley.com/book/10.1002/9781119170174.
As stated, neurophysiological data from monkeys had previously identified posterior loci associated with working memory. However, an important finding was that sustained activity in posterior areas could be disrupted by the presentation of distracting stimuli, and sustained stimulus-selective activity in the PFC persisted (Constantinidis & Steinmetz, 1996; di Pellegrino & Wise, 1993; Miller, Erickson, & Desimone, 1996). Despite the abolishment of sustained posterior activity, behavioral performance can remain well above the chance levels that would be predicted if behavior depended on sustained activity of posterior areas alone (Miller & Desimone, 1994). These data suggested that posterior areas were unsuitable as a locus of representation because their activity could be easily disturbed. By contrast, the robust elevated firing of PFC neurons, even in the presence of distraction, led to the hypothesis that the PFC is a more suitable candidate for representation of working memory than posterior regions. These data are consistent with the temporal rates of decay of activity across the cortical hierarchy. Although decay rates in sensory areas are short, affording the rapid updating of sensory information, decay rates in the PFC are long, enabling the retention and integration of information necessary for working memory (Chaudhuri, Knoblauch, Gariel, Kennedy, & Wang, 2015; Hasson, Chen, & Honey, 2015; Hasson, Yang, Vallines, Heeger, & Rubin, 2008; Honey et al., 2012; Murray, Bernacchia, et al., 2014). Collectively, these data provide an important challenge to the hypothesis that posterior perceptual representations support working memory.

Searching for the Neural Code

Although sustained elevated activity has long been considered the neurobiological underpinning of working memory, it is unclear that such a code is necessary for the representation of information across short timescales. For example, neurons in the early visual cortex show persistent memory-related firing, but this persistent firing is not elevated above pre-stimulus baselines (Super, Spekreijse, & Lamme, 2001). Instead, neuronal firing in early visual areas falls precipitously in the post-stimulus phase, with persistent modulations of the firing seemingly encoding information useful for working memory. Similarly, although representations can be read out from fMRI signals in early visual areas, those areas do not show elevated activity; activation levels during retention are similar to pre-stimulus baselines (Emrich, Riggall, Larocque, & Postle, 2013; Riggall & Postle, 2012). Moreover, it is unclear that sustained elevated activity is sufficient for the representation of information for working memory. Although activations in the PFC often show sustained elevated signals during working memory retention, these signals frequently do not capture stimulus-specific representations (Emrich et al., 2013; Ester et al., 2015; Riggall & Postle, 2012). Even when stimulus specificity is observed in the PFC, it appears to be far less prevalent than in the posterior cortices (Miller et al., 1996; Sprague et al., 2014; Sprague & Serences, 2013; Zaksas & Pasternak, 2006). Finally, retained information can be robustly read out of PFC activity that is not elevated above baseline levels (Stokes et al., 2013). Therefore, although the PFC appears to be better suited for sustained elevated activity than posterior perceptual areas, this sort of activity is neither necessary nor sufficient for sustained representation. Although persistent, elevated signals have garnered the lion’s share of interest in working memory research, there may yet be other neural codes of interest.

Although elevated activity may not be critical for working memory, it remains possible that sustained activity may be important.
That is, even if firing rates and signals are not elevated above baseline, the stable persistence of baseline-level signals may be critical for maintaining a robust representation. However, whether working memory representations are predominantly stable or dynamic remains unclear. Although the wealth of the reviewed literature has looked for and thus reported stable activity, dynamic signals are also prevalent. For example, many information-carrying neurons show ramping activity, falling to baseline levels in the post-stimulus phase of a delay interval and then increasing to high levels of activity just prior to test (Miller et al., 1996; Romo, Brody, Hernandez, & Lemus, 1999). Although there is some indication that ramping activity can reflect preparation for a forthcoming action (Quintana & Fuster, 1992), such activity can be observed even when the forthcoming action is unspecified, suggesting that it can reflect mnemonic, rather than motor, preparation (Romo et al., 1999). At a population level, there have been demonstrations that delay activity is dynamic, corresponding to a consistent trajectory from sample to response (Meyers, Freedman, Kreiman, Miller, & Poggio, 2008; Sreenivasan, Vytalci, & D’Esposito, 2014; Stokes et al., 2013). That is, different patterns of activity correspond to a given representation at distinct points in time (Figure 12.4). Despite these fluctuations, patterns of activity continue to discriminate representations, albeit in distinct ways over time. Finally, recent evidence suggests that working memory may be mediated by brief bursts of activity surrounded by periods of relative silence (Lundqvist et al., 2016). Collectively, these data suggest that activity need not be sustained by individual cells to code information for working memory. Instead, different cells may code for different aspects of memory at distinct time points.

Recently, whether information for working memory needs to be carried in activity at all has come into question. Another possible coding scheme is offered by synaptic plasticity (Barak & Tsodyks, 2007; Jonides et al., 2008; Mongillo, Barak, & Tsodyks, 2008; Stokes, 2015). By this idea, recent neuronal firing resulting from encoding or early maintenance processes potentiates those synapses involved in the representation of an item. One candidate mechanism is residual presynaptic $\text{Ca}^{2+}$ that can lead to potentiation of firing lasting on the order of 30 seconds to several minutes (Zucker & Regehr, 2002). On an ensemble level, this potentiation sets up an

---

**Figure 12.4** Stable versus dynamic coding. Each colored trace depicts a different representation embodied by the firing rates of three different neurons over time. A. A stable code wherein each trace settles into a distinct area in space. In this case, a similar pattern of activity corresponds to a particular representation at all but the earliest time points. B. A dynamic code wherein each trace corresponds to a unique, but time-varying trajectory over the neuronal firing rate space. Distances between each trace maintain distinctions among representations, but the pattern of activity that identifies a particular representation varies over time. Color version of this figure is available at http://onlinelibrary.wiley.com/book/10.1002/9781119170174.
Collectively, the reviewed data suggest that sustained elevated neuronal firing is not the sole neural code underlying information retention for working memory. Instead, different neural codes appear to be responsible for different aspects of retention. Primary and unimodal association cortices (e.g., visual cortex) tend to show baseline levels of activity during retention intervals that nevertheless retain information about maintained representations. These data suggest that information in these areas is encoded by modulations of low activity levels or sub-threshold potentials (Mendoza-Halliday, Torres, & Martinez-Trujillo, 2014) that consistently distinguish different representations. The low level of activity may help to differentiate memories from percepts, preserving retention while allowing ongoing perception. By contrast, unattended information that remains important for ongoing cognition may be related to synaptic mechanisms rather than neuronal firing. This information remains in a “silent” state, awaiting a retrieval signal to cause reactivation (Mongillo et al., 2008). Residual synaptic potentiation may also provide a source of proactive interference, causing intrusions of no-longer-relevant information (Jonides & Nee, 2006; Nee & Jonides, 2013). However, representations maintained by synaptic potentiation are likely to be susceptible to interference. New incoming stimuli will presumably be coded with greater synaptic potentiation than old, causing retroactive interference to the degree to which similar synapses are involved. Sustained elevated neuronal firing may offer a more robust, distractor-resistant code. Such a code is most strongly associated with the PFC. Though not strictly necessary or sufficient for working memory representation, sustained elevated firing in the PFC may be critical to resist interference (D’Esposito & Postle, 1999; Jonides et al., 2005). Collectively, these data
suggest that different neural codes serve distinct purposes for working memory. This indicates that task demands may ultimately shape the cortical basis and neural code of retention for working memory.

**Top-Down Control, Abstraction, and the Prefrontal Cortex**

Given that the PFC appears to have the appropriate mechanisms to sustain robust, distractor-resistant information, why would posterior areas be involved in retaining working memory representations? One answer may be the nature of the representation to be maintained. Although sustained elevated neuronal firing in the PFC has some stimulus specificity, the specificity is substantially weaker than that of posterior cortices, with one estimate indicating 38% of PFC cells show stimulus specificity compared to 94% in the inferior temporal cortex (Miller et al., 1996; see also Zaksas & Pasternak, 2006). Furthermore, although efforts to reconstruct stimulus-specific memoranda from fMRI signals have primarily focused on early visual areas, memoranda can be reconstructed from signals in parietal and frontal areas (Ester et al., 2015), with evidence of increasingly coarse coding at progressively higher levels of visual hierarchy (Sprague et al., 2014; Sprague & Serences, 2013). This suggests that relative to primary and unimodal association areas, frontal areas contain larger receptive fields favoring more abstract representations. Such properties would make the PFC amenable to retaining abstract information but not as well suited to making fine discriminations. In a direct test of this idea, a study compared the representational locus of categorical versus visual feature information (Lee et al., 2013). Using the same object stimuli, participants were asked to perform a match decision on either the category of a retained object or a fine visual feature of a retained object. Although categorical retention led to a representational locus in the PFC, visual feature retention led to a representational locus in visual association areas. Similarly, recent fMRI studies that have demonstrated information retention in sensory and sensory association areas have used stimuli that require fine visual discriminations (Christophel, Hebart, & Haynes, 2012; Emrich et al., 2013; Ester et al., 2013; Riggall & Postle, 2012). These data suggest that when retained items require fine discrimination, such representations can be distinguished only in primary and unimodal association areas.

Even when fine visual discriminations are required, the PFC continues to demonstrate persistent, elevated activity (Emrich et al., 2013; Ester et al., 2015; Riggall & Postle, 2012). What purpose does this activity serve? As alluded to previously, one hypothesis is that it provides top-down biasing, which prevents interference. Early research in monkeys revealed that short-term retention is spared following PFC lesions if distractions are minimized by darkening the visual environment (Malmo, 1942). Similar observations have been made based on lesions in human PFC indicating spared retention performance for simple retention but impaired performance during more demanding situations (D’Esposito & Postle, 1999). This work suggests that although the PFC is not essential for the representation of information because information can be retained when distractions are minimized, it is essential to prevent distractions from interfering with stored content. Maintenance of goal-relevant representations during distraction may be accomplished through the representation of abstract information in the PFC. For example, the context of a situation dictates what representations are and are not relevant for working memory. While reading this chapter, you are assembling recent
sentences to form ideas. That information should be distinguished from certain irrelevant thoughts (e.g., what you had for dinner last night) and words that may be present in other windows open on your computer screen. By maintaining a context (e.g., this chapter), you are allowing relevant information to be temporarily maintained while filtering other irrelevant information. The chapter you are reading can be considered an abstract code that indicates what sources of information are permitted to enter sensory and sensory association areas where the details are represented. In this way, abstract PFC signals can direct the contents of working memory (Nee et al., 2013).

There is a wealth of data consistent with the idea that sustained elevated PFC signals reflect abstract information that guides information representation for working memory. Several studies have demonstrated that whereas the contents of working memory can be decoded only from posterior signals, contextual signals dictating what sorts of stimuli are relevant for a particular trial can be decoded from the PFC (Chen et al., 2012; Riggall & Postle, 2012; Sreenivasan, Vytalci, et al., 2014). This contextual information is sustained in the PFC and is likely transmitted to perceptual areas to guide task-relevant processing (Feredoes, Heinen, Weiskopf, Ruff, & Driver, 2011; Siegel, Buschman, & Miller, 2015). Furthermore, disrupting processing in the PFC reduces stimulus-selectivity in posterior cortices (Lee & D’Esposito, 2012; Miller, Vytalci, Fegen, Pradhan, & D’Esposito, 2011) causing an impairment in resisting category-specific distraction (Zanto, Rubens, Thangavel, & Gazzaley, 2011). Notably, these impairments can be state-specific, affecting top-down but not bottom-up processing in posterior areas (Higo, Mars, Boorman, Buch, & Rushworth, 2011). These data indicate that the PFC tunes posterior neurons, providing pattern separation that facilitates top-down selection. Collectively, these data support the idea that abstract contextual signals enable top-down biasing from the PFC that mitigates interference.

Contexts are merely one abstract code that the PFC retains in the service of directing working memory. Contexts can themselves be directed by more abstract representations. For example, consider a multitasking situation wherein you need to switch between reading this chapter and monitoring your e-mail for an important message. Here, a superordinate prioritization signal is required to switch between contexts and admit the relevant sensory information for working memory. It has been suggested that this sort of coordination relies on different levels of PFC representation so that progressively rostral areas of the PFC represent progressively abstract codes for the control of behavior (Badre, 2008; Badre & D’Esposito, 2009; Koechlin, Ody, & Kouneiher, 2003; Koechlin & Summerfield, 2007). Evidence for these ideas comes from functional imaging studies demonstrating that as cognitive control demands require increasingly higher levels of control signals, increasingly rostral areas of the PFC are engaged (Badre & D’Esposito, 2007; Bahlmann, Blumenfeld, & D’Esposito, 2014; Koechlin et al., 2003; Nee & Brown, 2012, 2013; Nee & D’Esposito, 2016). These observations are bolstered by lesion data demonstrating that progressively rostral areas of the PFC are necessary for progressively higher levels of cognitive control (Azuar et al., 2014; Badre, Hoffman, Cooney, & D’Esposito, 2009). These data indicate that working memory and cognition more generally are controlled via abstract signals that are ordered along the rostral-caudal axis of the PFC (see also Nee & D’Esposito, 2016).

Collectively, the picture that emerges from these data is that representations in
Recent and Emerging Trends

Executive memory

Perceptual memory

Conceptual
Plans
Programs
Acts
Phyletic
motor

Conceptual
Semantic
Episodic
Polysensory
Phyletic
sensory

Actions
(behavior, language)

Perceptual representations are maintained for working memory (blue colors). Depending on demands, such memories may correspond closely to the original percept involving primary cortices or more-abstract representations involving associational cortices. These perceptual representations interact with frontal representations that form executive memories (red colors). The interactions between perceptual and executive memory embody cognitive control. Frontal representations are also organized according to abstraction, with concrete representations localized in caudal areas (e.g., motor cortex) and progressively abstract representations localized progressively rostrally. Color version of this figure is available at http://onlinelibrary.wiley.com/book/10.1002/9781119170174.

Source: Reproduced from (Fuster, 2001) with permission.

The brain can be conceptualized as proceeding from concrete to abstract from sensory areas to the rostral-most portions of the PFC (Fuster, 1995, 2001) (Figure 12.5). For working memory, we can consider the representations of retained information along with the representations of abstract information that governs the processing of that information. Information can be retained in sensory cortices for the purposes of fine discrimination. For categorical information, retention will likely proceed further downstream to posterior association areas or even caudal areas of the frontal lobes. The retention of this information is orchestrated by contextual representations in the mid-lateral PFC, which themselves can be informed by the rostral-most portions of the PFC. Hence,
working memory can involve virtually the entire cortex from sensory areas to areas throughout the PFC (D’Esposito, 2007).

FUTURE DIRECTIONS

The neuroscientific study of working memory has evolved from the focus on sustained elevated activity in the PFC to examination of multiple neural codes across the entire brain. This expansion of areas of investigation has opened numerous questions that will require systematic experimentation to address.

Stable Versus Dynamic Coding

We have reviewed classical data demonstrating stable representation of information for working memory via sustained elevated activity as well as recent data revealing dynamic coding of information for working memory. At the present, it is unclear under what circumstances information is coded in one form or another. Stability versus dynamism may be properties of the particular circuits investigated, task parameters, or both. For example, the mechanisms supporting persistent, elevated firing in the reviewed PFC seem ideal for maintaining stable representations over time. However, in many cases, retrospectively encoding what was previously observed may be less optimal than preparing to act conditionally on what was previously observed. Given that the PFC is also involved in such preparatory processes (Braver, Gray, & Burgess, 2007), it seems reasonable that stable, retrospective coding will occur primarily in situations when the forthcoming action or decision closely resembles the memoranda. Delayed saccades to spatial locations are one such example, which have shown stable coding (Funahashi et al., 1989). Conversely, dynamic coding has been observed when decisions are based on transformations of the encoded stimulus (Meyers et al., 2008; Stokes et al., 2013). In posterior areas, there has been evidence of stable (Riggall & Postle, 2012) and dynamic coding (Meyers et al., 2008; Sreenivasan, Vytlacil, et al., 2014) as well. It has been speculated that dynamic coding may prevent the processing of new stimuli from interfering with memory (Meyers et al., 2008). Such coding may not be necessary when only a single item is encoded and retained and tests of memory involve match decisions or reproduction. Alternatively, dynamic coding may be important when multiple items must be maintained, especially following sequential presentation, and when tests involve transformations based on the encoded information.

If these speculations turn out to be correct, then the distinction between stable and dynamic coding of retained information may mirror the distinction between short-term memory and working memory. Using this idea, short-term memory involves the maintenance of a representation as close to the original percept as possible. By contrast, working memory involves a transformation of that representation in the service of some other cognitive goal. If so, it is likely that dynamic coding underlies the bulk of the higher-level cognitive activities for which working memory is deemed essential. Hence, better understanding of the mechanisms governing dynamic coding is of paramount importance.

Synaptic Plasticity and Levels of Analysis

We have suggested that synaptic plasticity may play an important role in the retention of information for working memory, especially when attention is directed elsewhere. At the present, these ideas are largely theoretical based on computational work because it is
challenging to provide direct evidence that synaptic mechanisms are directly tied to behavior. However, as suggested by computational modeling, an undifferentiated retrieval signal should be able to reinstanciate synthetically potentiated traces (Mongillo et al., 2008). This suggests that stimulation techniques directed toward areas mediating synaptic plasticity-based retention could reveal an echo of the otherwise silent memory trace (Stokes, 2015). This would provide important, albeit indirect, support for the theory. New techniques developed to examine synaptic plasticity in vivo hold promise for providing more direct evidence in animal models (Hayashi-Takagi et al., 2015). An important question to answer will be whether the synaptic mechanisms governing short-term retention for working memory are distinct from those that create enduring changes at the synapse responsible for long-term memory. Although there is preliminary evidence that different induction parameters can dissociate short-term plasticity from long-term plasticity (Erickson, Maramara, & Lisman, 2010), it will be important to demonstrate that such mechanisms translate to cognition and behavior. Furthermore, documenting how synaptic weight-based memory traces are changed by interference as compared to active neural traces will provide important knowledge regarding the practical limits of different memory codes.

Finally, precise investigation and modeling of neurotransmitter receptor dynamics responsible for sustained, elevated firing in the PFC have provided deep insight into the mechanisms of such retention. Examination of neuromodulatory influences including dopamine and norepinephrine have granted further insights (Arnsten & Li, 2005; Cools & D’Esposito, 2010). As additional areas of the brain and coding schemes have proven important for working memory, so too grows the need for examination of molecular mechanisms and detailed circuit level modeling. Since the pioneering work of Goldman-Rakic (see Arnsten, 2013, for a review), few researchers have taken on the necessary bridging of multiple levels of analysis needed to understand working memory. The integration of network, neuron, and molecular levels coupled with causal intervention will be necessary to address remaining questions about working memory.

CONCLUSION

We began our review of the neurobiological basis of working memory by attempting to rule out what is not working memory. As we have seen, however, the more that working memory is considered, the more encompassing its study becomes. Depending on the information retained, working memory can engage virtually any cortical area. Information retained for working memory can be coded by the brain in multiple forms from persistent, elevated activity to dynamic trajectories to activity-silent forms presumed to be mediated by synaptic plasticity. Given its encompassing nature, future progress in understanding working memory will require a multifaceted approach, including macro-level examination of the whole-brain and network interactions, targeted examination of local population dynamics, and molecular approaches investigating synaptic potentiation and receptor contributions.

REFERENCES

414 Working Memory


Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension:


Tsujimoto, S., & Postle, B. R. (2012). The prefrontal cortex and oculomotor delayed response:
A reconsideration of the “mnemonic scotoma.” *Journal of Cognitive Neuroscience*, 24(3), 627–635. doi:10.1162/jocn_a_00171


