"Working memory functions of the prefrontal cortex"

Bradley R. Postle

Departments of Psychology and Psychiatry

University of Wisconsin–Madison

Madison, WI  USA

postle@wisc.edu
Abstract

The prefrontal cortex (PFC) plays an important role in many behaviors, including in situations in which actions must be guided by information that is not currently accessible in the environment. Although the construct of “working memory” is often invoked in association with the PFC, imprecise or erroneous specification of which computations relate to which aspect of anatomy or physiology has been the basis of many erroneous ideas about of the functional organization of the PFC. Indeed, the manner in which working memory has been related to the PFC over the past 75 years offers several cautionary tales about the difficulty of relating brain function to behavior. This proposition is supported by consideration of data from lesions and physiological measurements from human and nonhuman primates.
1.0. Introduction

Science progresses through the articulation of models, using these models to generate predictions, and testing predictions with experiments. Most often the outcomes of experiments, whether confirming or disconfirming a prediction, are used to refine the theoretical framework within which the experiments were carried out. Periodically, however, the model itself can be superceded by a different model that provides a better account of the phenomenon under study. One example comes from physics, in which the Newtonian model that held sway during the 19th and early 20th centuries has given way to a theory of general relativity. It is noteworthy in this example that, even though physicists no longer use the Newtonian framework to guide their thinking about the physical universe, Newton’s laws still capture the naïve intuition of most nonphysicists about “how the world works”. If one can overlook concerns about grandiosity on the part of this author, a useful analogy might be drawn to the study of working memory: Although scientific thinking has been dominated by a model of working memory as a multicomponent cognitive system, one that may correspond nicely with intuitions about “how the mind works”, it is being superceded by a new framework, one that understands working memory as an “emergent property” (Postle, 2006) arising from the attentional selection of information that is relevant for the current behavioral context (e.g., Anderson, 1983, Cowan, 1988, Cowan, 1995, Sreenivasan et al., 2014, Desrochers et al., 2015, Lara and Wallis, 2015). From this perspective, the label “working memory” applies to a category of behaviors, and to the tasks that are used to measure performance on these behaviors, but not to a unitary cognitive system whose engagement can be inferred from the “first-order” inspection of levels of activity in one or more regions in the brain.

The prefrontal cortex (PFC) plays an important role in many behaviors, including in situations in which actions must be guided by information that is not currently accessible in the environment. Although the construct of “working memory” is often invoked in association with
the PFC, imprecise or erroneous specification of which computations relate to which aspect of anatomy or physiology has also been the basis of many erroneous ideas about of the functional organization of the PFC. Indeed, the manner in which working memory has been related to the PFC over the past 75 years offers several cautionary tales about the difficulty of relating brain function to behavior. This chapter will be organized in three sections. The first will address the construct of working memory, and how one’s conceptualization of the architecture of high-level cognition can constrain how one goes about studying the brain. The second will address the phenomenon of sustained activity, and how a priori notions of “what a memory signal must look like” can lead to flawed inference about brain-behavior relations. Finally, the third will review some experiments that provide a framework within which we might make further progress in studying the working memory functions of the PFC.

2.0. "Working memory” as a cognitive construct

2.1. Working memory as RAM

The idea that the cognitive system requires a working memory derives directly from the metaphor of the brain as a computing machine, with working memory carrying out the function of maintaining multiple action plans in a rapidly accessible state (Miller et al., 1960), as does random access memory (RAM) in many von Neumann computing architectures. Pribram and colleagues (1964), in building on this idea, were the first associate working memory with the PFC. They found that poor performance by frontally lesioned monkeys, on a variety of tasks imposing a delay between cue and response, were better explained as impaired control of behavior, rather than forgetting, per se. For example, one task required first searching through a set of “junk” objects to learn which covered a reward, then returning to the rewarded object until a criterion-level of five consecutive correct choices was achieved, at which time a different object in the set would be selected by the experimenter (baiting of food wells on all trials was concealed from the animal,
thereby adding the "working memory" element). On the first set of trials, the frontal animals made more errors, a pattern that could have been due either to trial-to-trial forgetting of which object had been rewarded, or by an inability to shift from a "explore" strategy to an "exploit" strategy. Once they achieved criterion, however, this ambiguity was resolved, because the frontal animals then also perseverated on the "exploit" strategy longer than did temporal lobe-lesioned and control animals. That is, their impairment wasn’t in the ability to retain a small amount of trial-specific information over a short period of time, but, rather, in the ability to use the discrepancy between the previous trial's stimulus-reward contingency vs. that of the present trial to change behavioral strategy (to change "set", in the parlance of mid-century neuropsychology). In their discussion of this and several other experimental findings, Pribram and colleagues (1964) drew from contemporaneous computer models of problem solving to propose that, rather than reflecting "memory trace formation and decay", the deficits resulting from frontal-lobe damage may have reflected a "mechanism of temporary, flexible stimulus compounding" (p. 51)1, a hypothesized process that is reminiscent of contemporary ideas of establishing trial-unique "bindings" between stimulus features and behavioral repertoires (Oberauer, 2013). Thus, in this first instance in the literature of an association between the construct of working memory and the PFC, the emphasis was on "working with memory" rather than on the storage, per se, of the remembered information

In the analogy to RAM, another factor is relevant, which is that RAM is not inherently time dependent. When, for example, while composing this chapter, I leave my word processing application to open a Web browser and access the precise wording of that quote from Pribram, the

1 The clarity and prescience with which Pribram and colleagues (1964) relate this line of reasoning and, more generally, with which they advocate an approach of “simulation … with the use of computers” is remarkable. Although the edited volume in which their chapter appeared is no longer in print, at the time of this writing, a digitized copy was downloadable from http://www.karlpribram.com/wp-content/uploads/pdf/D-049.pdf.
manuscript file running in the word-processing application will remain immediately accessible whether I return to it as soon as I access the quote, or, instead, if I set the computer down, make breakfast, walk to the beach, and then return to my computer several hours later to resume this work. In biological systems, it’s also the case that trial-specific memories need not be temporally constrained. If a rat explores 3 arms of an 8-arm radial maze, and is then returned to its home cage and only returned to the maze several hours later, it can “pick up where it left off”, knowing which 5 arms remain baited. And this “working memory” of which 3 arms had been visited will be of no use, of course, once the remaining 5 arms are visited and the experimenter rebait all 8 arms of the maze (Olton et al., 1979). Similarly, a memory for where in the lot she parked the car on Monday is of little use to the office worker leaving at the end of the day Tuesday, assuming that Tuesday morning’s choice of parking space was not influenced by previous choices. Furthermore, successful performance on such tasks is known to depend on the hippocampus, not on the PFC (as reviewed by Becker and Morris, 1999).

These examples highlight several important points. The first is that there is not a principled computational reason for working memory to be time-delimited. (Whether there may be biological factors, relating to, for example, decay or interference, will be taken up further along in this chapter.) A second is that there is no a priori reason why working memory functions need to be carried out by a specialized system that is distinct from other categories of cognition – in the “Honig-Olton” scheme (Becker and Morris, 1999), for example, working memory and reference memory can both depend on medial temporal-lobe neural systems. A third, as exemplified by the example from Pribram et al. (1964), is that many computationally distinct operations must be carried out in order to successfully execute even the simplest working memory task, and the retention of stimulus information is only one of these.
2.2 The Multiple Component Model, and its relation to the dorsolateral PFC

In the 1970s and 1980s, Baddeley and colleagues formulated an explicit cognitive model of working memory. It posited a multicomponent architecture whereby the storage function of domain-specific short-term memory buffers was controlled by a domain-general Central Executive (e.g., Baddeley and Hitch, 1974, Baddeley, 1986). Importantly, the Central Executive was construed as a general purpose controller, akin to Norman and Shallice’s (1980) Supervisory Attentional System and, as such, wouldn’t only be engaged by tasks with an overt memory component. (Indeed, consistent with this idea, an early neuroimaging study designed to isolate brain activity attributable to the Central Executive (and identifying it in the PFC) employed a dual-task procedure in which neither of the individual tasks was a memory task (D’Esposito et al., 1995)). Of further importance is that one would also expect an attentional controller to be active even during the simplest tasks that, on the surface, would seem to only require the engagement of a short-term store. This is because, among other things, one can never know when an unexpected change in the environment might render the short-term retention of information, and/or the need to guide behavior with that information, more difficult (e.g., Malmo, 1942, Chao and Knight, 1995, Chao and Knight, 1998, Postle, 2005). From this perspective, the inferential flaw in studies purporting to localize visual working memory storage-related activity to PFC was to assume that sustained, spatially tuned activity in this region corresponded to the operation of the “visuospatial sketchpad” buffer from the multiple component model, instead of to its Central Executive (for more developed argumentation on this point, see D’Esposito and Postle, 2015, Postle, 2015c, Postle, 2015b).
3.0 Sustained Activity in the PFC (and Elsewhere)

The idea that short-term and working memory might depend on sustained, elevated activity dates back at least to Hebb (1949), and is seen in many of Goldman-Rakic's influential writings (Goldman-Rakic, 1987, Goldman-Rakic, 1990). However, the once-popular assumption that such activity in the PFC makes a necessary contribution to the short-term retention, per se, of sensory information is no longer tenable. Empirically, it is well-established that sustained activity in the PFC is neither specific for (e.g., Curtis and Lee, 2010, Riggall and Postle, 2012, Emrich et al., 2013), nor necessary for (e.g., Zaksas and Pasternak, 2006, Lara and Wallis, 2014, Fuster, 2016, Wimmer et al., 2016), the short-term retention of this information (also reviewed in Postle, 2015a). To consider just one type of information in more detail, recent studies have been unsuccessful with multivariate decoding of the direction of motion from the dorsolateral (dl)PFC in humans (Riggall and Postle, 2012, Emrich et al., 2013), successful in the monkey (Mendoza et al., 2014), but, most tellingly, evidence from a lesion study suggests that functions other than sensory storage are supported by the dlPFC: “[B]ecause th[e] deficit [in dlPFC lesioned animals] was independent of stimulus features giving rise to the remembered direction and was most pronounced during rapid shifts of attention, [the] role [of dlPFC] is more likely to be attending and accessing the preserved motion signals rather than their storage (p. 7095)” (Pasternak et al., 2015). With regard to sustained activity, a critical role for the dlPFC may emerge when a task requires the transformation of trial-initiating sensory information into a format that is needed for subsequent guidance of behavior, as well as in the retention of that transformed information (Meyers et al., 2012, Lee et al., 2013, Stokes et al., 2013, Lee and Baker, 2015).

Together with the findings that we have just reviewed, it is important to note that the very relation of sustained activity to working memory is undergoing reconsideration: Just as the
intuition that the behavioral constructs of short-term memory and working memory are inherently time delimited turns out to be flawed (see, e.g., Postle, 2015b, Postle, 2016), so, too, might be the assumed relation between sustained activity working memory. On theoretical grounds, it has been argued that the short-term retention of information might be accomplished via short-term synaptic reorganization (e.g., Mongillo et al., 2008, Barak and Tsodyks, 2014, Stokes, 2015), with elevated activity corresponding, instead, to the focus of attention (Lewis-Peacock et al., 2012, LaRocque et al., 2013, Lewis-Peacock et al., 2015). Empirical evidence that a transient, synaptic weight-based mechanism is the basis for working-memory storage is difficult to assemble, but findings that are consistent with this idea are beginning to emerge (Sugase-Miyamoto et al., 2008, Hayden and Gallant, 2013, Wolfe and Stokes, 2015). Another mechanism for the short-term retention of information that differs from “elevated activity” as it is traditionally construed would be fluctuations in intracellular voltages that can be sustained over tens of seconds (Strowbridge, 2012).

4.0 Working-memory functions of the dlPFC

An expedient rhetorical device for launching a discussion of the working-memory functions of the dlPFC is to consider the idea that a punctate lesion of this region will produce a “mnemonic scotoma”, whereby memory guided saccades to a restricted area of the visual field are impaired, despite the sparing of visually guided saccades into that same region (Funahashi et al., 1993). When an independent group of researchers (Wajima and Sawaguchi) sought to replicate this finding several years later, however, they obtained results that are intriguingly reminiscent of those from Pribram et al. (1964) that were reviewed earlier in this chapter – when testing and scoring procedures were refined, the impairment was revealed to be attributable to factors other than memory. Specifically, the procedure of Funahashi et al. (1993) was to score each trial in which the initial saccade did not land within the cued location as an error (S. Funahashi, personal
communication). Wajima and Sawaguchi (reported in Tsujimoto and Postle, 2012), in contrast, allowed their animals to make multiple saccades on each trial, and rewarded them if they eventually landed in the target location. Although they replicated the earlier finding -- that a disproportionate number of misguided initial saccades were made on trials targeting the critical region of the visual field – they also observed that erroneous initial saccades were almost invariably followed by a second, corrective saccade that acquired to to-be-remembered target location. Furthermore, the erroneous saccades were noted to have often been made to a region of space that had been relevant on the previous trial, either as that trial’s cued location or as the target of that trial’s saccade (or both). Thus, the animal’s errors were better classified as perseverative, or as influenced by proactive interference, than as mnemonic, per se. In a conceptually related finding, Mackey and colleagues (in press) have recently shown that deficits on the oculomotor delayed response task are only seen in human patients when their dIPFC lesions invade the territory of the frontal eye fields. Errors on tests of working memory that result from damage to the dIPFC may be qualitatively similar to those that we know, from decades of behavioral neurology, are characteristic of these patients in situations that make no overt demands on working memory (Tsujimoto and Postle, 2012).

Over the past decade, my group has used (f)MRI-guided repetitive transcranial magnetic stimulation (rTMS) to dissociate mnemonic from nonmnemonic factors in working and short-term memory performance. Several findings are consistent with the assertion that ended the previous paragraph. For example, rTMS of the dIPFC during a delay period does not affect delayed-recognition for locations – nor does delay-period rTMS of the postcentral gyrus – whereas rTMS of the intraparietal sulcus and of the frontal eye fields\(^2\) does affect performance (Hamidi et al.,

\(^2\) Although there are compelling reasons to classify the frontal eye fields as “prefrontal” from the standpoint of the evolution of neural systems (Passingham and Wise, 2012), this author nonetheless finds it obfuscating when physiological studies that are limited to the frontal eye fields are labeled as
A key role for the dLPFC emerges on this task, however, when rTMS is instead delivered concurrent with the onset of the stimulus that initiates the memory-guided response (whether it be recall or recognition, Hamidi et al., 2009). Delay-period rTMS of the dLPFC also does not disrupt the simple short-term retention of verbal information, unless subjects are required to mentally reorder it during the delay period (Feredoes et al., 2006, Postle et al., 2006, Feredoes et al., 2007). When it is applied during the response period, in contrast, rTMS reveals important roles for subregions of the PFC in such functions as controlling the effects of proactive interference (Feredoes et al., 2006), perhaps by adjudicating the influence of various memory signals (e.g., familiarity vs. recollection) on decision processes (Feredoes and Postle, 2010). These findings are consistent with more recent work in the monkey, which also emphasize the role of PFC dynamics in memory-guided decision making and action planning (Wimmer et al., 2016).

5.0. Conclusion

And so, is the PFC important for working memory function? Without a doubt. But a clear understanding of the ways in which PFC circuits do, versus do not, contribute to these behaviors will be important if we are to make meaningful progress in addressing “the riddle of frontal lobe function in man” (Teuber, 1964).
References


Mackey, W., O. Devinsky, W. Doyle, M. Meager and C. E. Curtis (in press). "Human dorsolateral prefrontal cortex is not necessary for spatial working memory." The Journal of Neuroscience.


Strowbridge (2012). "get it. get it get it: get it.


