From cognitive to neural models of working memory

Mark D’Esposito*

Helen Wills Neuroscience Institute and Department of Psychology, University of California, 132 Barker Hall, Berkeley, CA 94720-3190, USA

Working memory refers to the temporary retention of information that was just experienced or just retrieved from long-term memory but no longer exists in the external environment. These internal representations are short-lived, but can be stored for longer periods of time through active maintenance or rehearsal strategies, and can be subjected to various operations that manipulate the information in such a way that makes it useful for goal-directed behaviour. Empirical studies of working memory using neuroscientific techniques, such as neuronal recordings in monkeys or functional neuroimaging in humans, have advanced our knowledge of the underlying neural mechanisms of working memory. This rich dataset can be reconciled with behavioural findings derived from investigating the cognitive mechanisms underlying working memory. In this paper, I review the progress that has been made towards this effort by illustrating how investigations of the neural mechanisms underlying working memory can be influenced by cognitive models and, in turn, how cognitive models can be shaped and modified by neuroscientific data. One conclusion that arises from this research is that working memory can be viewed as neither a unitary nor a dedicated system. A network of brain regions, including the prefrontal cortex (PFC), is critical for the active maintenance of internal representations that are necessary for goal-directed behaviour. Thus, working memory is not localized to a single brain region but probably is an emergent property of the functional interactions between the PFC and the rest of the brain.

Keywords: working memory; short-term memory; executive function; cognitive control; prefrontal cortex; functional magnetic resonance imaging

1. INTRODUCTION

What is the neural basis of working memory? To answer this question, one must begin with a proper definition of the term ‘working memory’. To me, working memory refers to the temporary retention of information that was just experienced but no longer exists in the external environment, or was just retrieved from long-term memory. These internal representations are short-lived, but can be stored for longer periods of time through active maintenance or rehearsal strategies, and can be subjected to various operations that manipulate the information in ways that make it useful for goal-directed behaviour. Thus, working memory is critically important in cognition and seems necessary for many cognitive abilities, such as reasoning, language comprehension, planning and spatial processing. However, it is clear from the numerous empirical studies of working memory that are published each year that it is an evolving construct, one that has come a long way over the past 30 years, since Baddeley & Hitch (1974) introduced their highly influential cognitive model of working memory.

Within a cognitive framework, Baddeley conceptualized working memory as a cognitive system comprising multiple components that support executive control (see also Burgess et al. 2007; Robbins 2007; Stuss & Alexander 2007) as well as active maintenance of temporarily maintained information (Baddeley 1986). Thus, a ‘central executive system’ was proposed as a system that could actively regulate the distribution of limited attentional resources and coordinate information within limited capacity verbal and spatial memory storage buffers. The central executive system, based on the analogous supervisory attentional system introduced by Norman & Shallice (1986), was proposed to take control over cognitive processing when novel tasks are engaged and/or when existing behavioural routines have to be overridden. Empirical studies of working memory using neuroscientific techniques, such as neuronal recordings in monkeys (e.g. Funahashi et al. 1989) and functional neuroimaging in humans (e.g. Curtis et al. 2004), have provided a rich dataset that can be reconciled with behavioural findings derived by testing cognitive models of working memory such as the one proposed by Baddeley. In this paper, I review the progress that has been made towards this effort by illustrating how investigations of the neural mechanisms underlying working memory can be influenced by cognitive models and, in turn, how cognitive models can be shaped and modified by neuroscientific data.

2. TRADITIONAL COGNITIVE MODELS OF WORKING MEMORY

A critical component of Baddeley’s working memory model is the existence of verbal and spatial storage buffers. The cognitive concept of a buffer translated into neural terms would propose that temporary
retain temporary retention of task-relevant information requires transfer of that information to a part of the brain that is dedicated to the storage of information. Presumably, such buffers are analogous to a computer’s RAM, which serves as a cache for information transferred from the hard drive that is processed by a CPU. Consistent with this interpretation of a working memory ‘buffer’, many descriptions of cognitive models of working memory refer to the information being ‘in’ or ‘out’ of working memory. For example, in a recent review of working memory, Repovs & Baddeley (2006) state that ‘the function of the articulatory rehearsal process is to retrieve and re-articulate the contents held in this phonological store and in this way to refresh the memory trace. Further, while speech input enters the phonological store automatically, information from other modalities enters the phonological store only through recoding into phonological form, a process performed by articulatory rehearsal’. Later, the authors refer to ‘focal shifts of attention to memorized locations that provide a rehearsal-like function of maintaining information active in spatial working memory’. Thus, one question that neuroscientific data can address regarding how the brain implements working memory processes is whether such buffers or storage sites exist in distinct parts of the brain to support the active maintenance of task-relevant information.

Another cognitive model of working memory, put forth by Cowan (1988, 1999), proposes that the ‘contents of working memory’ are not maintained within dedicated storage buffers, but rather are simply the subset of information that is within the focus of attention at a given time. He describes an embedded-processes model where working memory comes from hierarchically arranged faculties comprising long-term memory, the subset of working long-term memory that is currently activated and the subset of activated memory that is the focus of attention. These ideas are similar to that put forth by Anderson (1983) who referred to working memory as those representations currently at a high level of activation. Thus, task-relevant representations are not in working memory, but they do have levels of activation that can be higher or lower. After use, for example, representations may be temporarily more active or ‘primed’. In this formulation, working memory does not have a size, or maximum number of items, as a structural feature. Instead, performance on working memory tasks is determined by the level of activation of relevant representations, and the discriminability of activation levels between relevant and irrelevant representations (Kimberg et al. 1997).

Again, in neural terms, Cowan or Anderson’s cognitive model of working memory would predict that information that is represented throughout the brain is not transferred to an independent buffer or storage site, but rather that temporary retention of task-relevant information is mediated by the activation of the neural structures that represent the information being maintained or stored (for a further discussion of these and related ideas see Ruchkin et al. (2003) and the commentary that followed). In other words, the temporary retention of a face, for example, would require activation of cortical areas that are involved in the perceptual processing of faces.

It is not possible in this paper to consider all cognitive models of working memory (for an excellent starting point, the reader is referred to Miyake & Shah 1999). However, based on the models put forth by Baddeley and Cowan as prototypes, one can begin to consider how the cognitive mechanisms proposed by such models are implemented in the brain. Thus, with these cognitive models in mind, I review evidence that begins to provide some insight into the neural mechanisms regarding how relevant information is temporarily stored in the service of goal-directed behaviour.

From a neuroscience perspective, it is counter-intuitive that all temporarily stored information during goal-directed behaviour requires specialized dedicated buffers. Clearly, there could not be a sufficient number of independent buffers to accommodate the infinite types of information that need to be actively maintained to accommodate all potential or intended actions. In a system with only two buffers, such as verbal and visuospatial, how would the retention of odours or tactile sensations, which cannot always be recoded into verbal or visuospatial representations, be accomplished? More recently, an additional episodic buffer has been proposed to be a store capable of multidimensional coding that allows the binding of information to create an integrated episode (Baddeley 2000). However, even with the addition of this buffer, Baddeley’s working memory model cannot accommodate storage of all possible types of information processed by the human brain (it is important to note, however, that this was not probably the original intent of this model). Alternatively, in Cowan’s proposal, which does not rely on the concept of specialized dedicated storage buffers, active maintenance or storage of task-relevant representations could be implemented with a neural system where memory storage occurs in the very same brain circuitry that supports the perceptual representation of information. Such a neural system presumably would be more flexible and efficient than one that transfers information back and forth between dedicated storage buffers. Can studies of brain function, either in animal or man, test these competing hypotheses?

3. NEURAL MODELS OF WORKING MEMORY

For over 30 years, the results of experiments in behaving monkeys using recordings from single neurons within the lateral prefrontal cortex (PFC, figure 1a) have consistently found persistent, sustained levels of neuronal firing during the retention interval in tasks that require a monkey to retain information over a brief period of time (e.g. Fuster & Alexander 1971; Kubota & Niki 1971; Funahashi et al. 1989). This sustained activity is thought to provide a bridge between the stimulus cue (e.g. the location of a flash of light) and its contingent response (e.g. a later delayed saccade to the remembered location). These results have been supported by the functional neuroimaging studies in humans, and there is now a critical mass of studies that find lateral PFC activity in humans during delay tasks (for review, see Curtis & D’Esposito 2003).
Curtis observed in the previously mentioned fMRI study task-relevant information. For example, we also region that is active during the temporary retention of memory have also found that the PFC is not the only other brain areas and human fMRI studies of working memories of anticipated action and preparatory set) codes. (e.g. past sensory events) or prospective (e.g. representations of anticipated action and preparatory set) codes. During the performance of this task, participants in the study were biased towards or against the use of a prospective motor code. In one condition (match trials), the participants were able to plan a saccade to the target as soon as the cue appeared and then they could simply postpone the initiation of the saccade until after the delay. During these trials, delay-period activity should reflect this strategy, i.e. the maintenance of a prospective motor code. In a comparison condition (non-match trials), a saccade was made after the retention interval to an unpredictable location that did not match the location of the sample. The participants still had to remember the location of the sample so that they could discern between the matching and non-matching targets. Since a saccade was never made to the sample location and the non-matching location was unpredictable, we expected that, during these trials, the participants were able to plan a saccade to the target location and the non-matching location was unpredictable. We found impaired performance on delay tasks in monkeys with selective lesions of the lateral PFC (Bauer & Fuster 1976; Funahashi et al. 1993).

However, monkey physiology studies recording from other brain areas and human fMRI studies of working memory have also found that the PFC is not the only region that is active during the temporary retention of task-relevant information. For example, we also observed in the previously mentioned fMRI study (Curtis et al. 2004) that different brain regions were involved during the performance of the oculomotor delayed response task. Specifically, different brain regions were active depending on whether the task required the temporary maintenance of retrospective (e.g. past sensory events) or prospective (e.g. representations of anticipated action and preparatory set) codes. During the performance of this task, participants in the

Figure 1. Neural activity in the monkey and the human lateral PFC during the retention interval of a spatial oculomotor delayed response (ODR) task. (a) Macaque: average of single-unit recordings from 46 neurons with delay-period activity from the monkey lateral PFC (brain area (BA) area 46; adapted from Funahashi et al. 1989). C, cue; D, delay; R, response. (b) Human: significant delay-period activity (left) and average (±s.e.) fMRI signal (right) from right lateral PFC (BA area 46; circled) in a human performing an ODR task (unpublished data from my laboratory). The grey bar represents the length of the delay interval. Note that how in both cases the level of PFC activity persists throughout the delay, seconds after the stimulus cue has disappeared.

For example, in a functional magnetic resonance imaging (fMRI) study using an oculomotor delay task identical to that used in monkey studies, we observed not only the frontal cortex activity during the retention interval (figure 1b), but also the magnitude of the activity correlated positively with the accuracy of the memory-guided saccade that followed later. This relationship suggests that the fidelity of the actively maintained location is reflected in the delay-period activity (Curtis et al. 2004). Thus, the existence of persistent neural activity during blank memory intervals of delay tasks is a powerful empirical finding, which lends strong support for the hypothesis that such activity represents a neural mechanism for the active maintenance or storage of task-relevant representations. The necessity of the PFC for the active maintenance of task-relevant representations has been demonstrated by studies that have found impaired performance on delay tasks in monkeys with selective lesions of the lateral PFC (Bauer & Fuster 1976; Funahashi et al. 1993).

Figure 2. Statistical parametric t-maps contrasting oculomotor delayed matching-to-sample versus non-matching-to-sample delay period-specific activity (Curtis et al. 2004). Activity during the (a) early and (b) late delay periods is shown. Warm colours depict regions with greater delay-period activity on matching than non-matching trials. Cool colours depict regions with greater delay-period activity on non-matching than matching trials. BA, brain area; FEF, frontal eye fields; SEF, supplementary eye fields; MFG, middle frontal gyrus; pIFS, posterior inferior frontal sulcus; iPCS, inferior precentral sulcus; IPS, intraparietal sulcus.

M. D'Esposito 763
domain of spatial information, separable neural mechanisms are engaged for the active maintenance of ‘motor’ plans versus ‘spatial’ codes. Moreover, given that our task only required the oculomotor system, it is probable that distinct neural circuitry will be recruited when the motor act involves other modalities, such as speech or limb output (e.g. Hickok et al. 2003). Thus, this is the first piece of evidence presented in this review that the concept of specialized buffers (for, say, verbal versus spatial information) may not map adequately onto neural architecture. Rather, the findings appear more consistent with a system in which active maintenance involves the recruitment of the same circuitry that represents the information itself, with different circuits for different types of spatial information (e.g. visual versus oculomotor).

Similar findings exist when the ‘visual’ component of working memory is investigated with neuroscientific methods. For example, in another fMRI study (Ranganath et al. 2004), we asked the participants to learn a series of faces, houses and face–house associations and they were scanned while performing a delayed match-to-sample (DMS) and delayed paired-associate (DPA) task with these stimuli. Results showed that delay-period activity within category-selective inferior temporal subregions reflected the type of information that was being actively maintained—the fusiform gyrus showed enhanced activity when participants maintained previously shown faces on DMS trials, and when subjects recalled faces in response to a house cue on DPA trials. Likewise, the parahippocampal gyrus showed enhanced activity when participants maintained previously shown houses on DMS trials and when they recalled houses in response to a face cue on DPA trials (figure 3). These fMRI findings are consistent with several monkey neurophysiological studies which have also shown that temporal lobe neurons exhibit persistent stimulus-selective activity in tasks requiring the active maintenance of visual object information across short delays (Miyashita & Chang 1988; Miller et al. 1993; Nakamura & Kubota 1995). Again, like spatial and motor codes, active maintenance of visual stimuli is mediated by the activation of cortical regions that also support processing of that information, perceptual in this case.

4. IS PHONOLOGICAL WORKING MEMORY SPECIAL?
Neuroscientific studies of verbal working memory, which has been most extensively studied by behavioural methods (Vallar & Shallice 1990), provide a similar...
view regarding the neural mechanisms underlying working memory. Consistently, performance on tasks that tap the ‘phonological loop’, as conceptualized by Baddeley, engage a set of brain regions that are thought to be involved in phonological processing. For example, using functional neuroimaging techniques during verbal working memory tasks, the left inferior parietal lobe, posterior inferior frontal gyrus (Broca’s area), premotor cortex and the cerebellum are typically activated (e.g. Paulesu et al. 1993; Awh et al. 1996).

However, is this network of brain regions also responsible for the active maintenance of non-phonological language representations (e.g. lexical-semantic)? For visual word recognition, a functionally specialized processing stream is thought to exist within inferior temporal cortex, representing visual words at increasingly higher levels of abstraction along a posterior-to-anterior axis (Cohen & Dehaene 2004). Intracranial electrophysiological recordings (Nobre et al. 1994), for example, show that posterior inferior temporal cortex differentiates letter strings from non-linguistic complex visual objects. Brain activity in more anterior inferior temporal cortical regions, in contrast, distinguishes words from non-words and is affected by the semantic context of words, indicating that anterior inferior temporal cortex holds more elaborate linguistic representations (see also Marslen-Wilson & Tyler 2007; Patterson 2007). To demonstrate that there is distinct neural circuitry supporting the active maintenance of non-phonological language representations, we explored the role of language regions within the left inferotemporal cortex (ITC) that are involved in visual word recognition and word-related semantics. Using fMRI, we first localized a visual ‘word form’ area within inferior temporal cortex area and then demonstrated that this area was involved in the active maintenance of visually presented words during a delay task (Fiebach et al. 2006). Specifically, we found that this area was recruited more for the active maintenance of words than pseudowords (i.e. orthographically legal and pronunciable non-words). Maintenance of pseudowords should not elicit strong sustained activation in such brain regions, as no stored representations pre-exist for these items. These results suggest that verbal working memory may be conceptualized as involving sustained activation of all relevant pre-existing cortical language (phonological, lexical or semantic) representations.

If working memory maintenance processes reflect the prolonged activation of the same brain regions that support online processing, evidence for cortical activity in the absence of stimuli should be evident not only in association cortex (as discussed thus far) but also within primary cortical regions. This indeed is the case as such effects have been observed in primary olfactory (Zelano et al. 2005), visual (Klein et al. 2000; Silver et al. 2006) and auditory cortex (Calvert et al. 1997; Kraemer et al. 2005). Thus, the neuroscientific data presented in this paper are consistent with most or all neural populations being able to retain information that can be accessed and kept active over several seconds, via persistent neural activity in the service of goal-directed behaviour.

5. NEURAL MECHANISMS OF ACTIVE MAINTENANCE OF TASK-RELEVANT REPRESENTATIONS

The observed persistent neural activity during delay tasks may reflect active rehearsal mechanisms. Active rehearsal is hypothesized to consist of the repetitive selection of relevant representations or recurrent direction of attention to those items. Subvocal articulations probably mediate the rehearsal of verbalizable memoranda (Baddeley 1986) since articulatory suppression (e.g. uttering ‘the...the...the’ during a retention interval), which interferes with rehearsal, degrades memory performance (Murray 1968). In addition, the ventrolateral frontal cortex (i.e. Broca’s area) is often activated in working memory tasks where subvocal rehearsal is the main strategy for maintenance (Bench et al. 1993; Awh et al. 1996). Similar mechanisms may be involved in the active maintenance of visual information such as objects, which may be represented by their visual features (e.g. size, colour, texture, shape) as well as verbal information associated by an individual with a visual stimulus (Posdle et al. 2005). The mechanisms underlying rehearsal of non-verbalizable material like spatial locations have been more difficult to resolve, but are likely to involve related motor and/or attentional processes (Awh et al. 1999; Awh & Jonides 2001). Positional information might be represented in oculomotor coordinates, where the memorized location might be maintained in terms of a saccade vector that acquires the target. Therefore, rehearsal of locations could simply be the reactivations of oculomotor programmes without actually making overt eye movements and can account for consistent activation of the frontal eye fields during spatial working memory tasks (Courtney et al. 1998). Thus, the rehearsal may be one mechanism by which transiently activated representations can be reactivated and refreshed.

Active maintenance (or rehearsal) of task-relevant representations clearly requires interactions between brain regions (Goldman-Rakic 1988; Fuster 1995, 2003). Such interactions could support working memory maintenance processes via synaptic reverberations in recurrent circuits (Durstewitz et al. 2000b; Wang 2001) or synchronous oscillations between neuronal populations (Singer & Gray 1995; Engel et al. 2001). Owing to the limitations in available methodology, only a few studies to date have been able to assess if and how neurons and brain regions interact to facilitate active maintenance processes (Fuster et al. 1985; Chafee & Goldman-Rakic 1998; Tomita et al. 1999; Funahashi & Inoue 2000; Constantinidis et al. 2001; Tallon-Baudry et al. 2001), although this is a critical issue for the future (see also Vuilleumier & Driver 2007). In other words, neither single neuron recordings nor standard univariate analysis of fMRI data, in which each neuron or voxel or brain region is analysed independently of all others, reveal more than the nature of isolated activity within these regions. Thus, only indirect evidence exists to support the assertion that working memory maintenance processes are implemented by the interaction of nodes within a neural network. Human functional neuroimaging, however, is ideally and uniquely suited to explore network interactions, since this method...
The presence of significant delay-period correlations between the FFA and regions of the prefrontal and parietal cortex regions supports models of working memory which suggest that higher-order association cortices interact with posterior sensory regions to facilitate the active maintenance of a sensory percept. The correlation between the FFA and these high-order regions appears to be initially established during the encoding of the cue stimulus, and these correlations are largely sustained during the delay period, despite a dramatic decrease in the level of univariate activity. Similarly, we have also found that language-related visual association areas involved in the maintenance of words (as described earlier), in the absence of visual input, also exhibit increased functional connectivity with the PFC (Fiebach et al. 2006). In summary, all of these empirical findings extend neural mechanisms of maintenance processes from the finding of persistent isolated neural activity within functionally specialized brain regions to encompass the concept of persistent functional connectivity between brain regions.

As discussed earlier, physiological and behavioural data suggest that each brain region, although forming part of a functional network, may contribute different elements to active maintenance by the nature of the representations that are coded within each region. However, different brain regions within a functional network probably differ only in their degree of participation in a manner that is dependent on the context of the operation being actively performed (Fuster 1995, 2003; McIntosh 2000). This is immediately evident by noting, for example, that the same regions which are involved in temporarily maintaining a representation are often also engaged during the encoding and retrieval of that information (e.g. Funahashi et al. 1989). Further understanding of such neural interactions will require investigating the influence of hypothesis-driven task design manipulations on the functional connectivity between brain regions. For example, we have investigated functional connectivity during a delay task with distracting stimuli aimed at increasing active maintenance demands (Yoon et al. 2006). During the performance of a delayed face recognition task, selective interference was evident behaviourally when face stimuli were presented as distractors during the delay period, relative to a condition in which scene stimuli were presented as distractors. Event-related fMRI data showed that maintenance-related functional connectivity between the lateral PFC and FFAs was perturbed during these face distraction trials. These data provide additional support for the notion that a plausible mechanism for the active maintenance is the coupling of abstracted, higher-order information in the PFC and stimulus-specific sensory information in the visual association cortex through reverberant activity between these areas. Finally, it is important to note that correlational data such as these need to be complemented with techniques to establish the functional necessity of network nodes, through methods such as lesions studies in animals and man (e.g. Fuster et al. 1985; Mottaghy et al. 2002; D’Esposito et al. 2006; see also Vuilleumier & Driver 2007).

6. IS THERE A CENTRAL EXECUTIVE IN THE BRAIN?

Further understanding of the neural mechanisms underlying the active maintenance of task-relevant information may hinge on our ability to resolve the nature of stored representations in addition to the types of operations performed on such representations (Wood & Grafman 2003). ‘Representations’ are
symbolic codes for information activated either transiently or permanently within neuronal networks. ‘Operations’ are processes or computations performed on representations. As we have reviewed thus far, models of working memory (e.g. Fuster 1985; Goldman-Rakic 1987; Petrides 1994; Kieras et al. 1999; D’Esposito & Postle 2000; Miller & Cohen 2001) vary substantially in the relative importance given to representations and operations. Baddeley’s original advance was to move us from the concept of short-term memory that accounted only for the storage of representations, to the concept of working memory as a multi-component system that allows for both storage and processing of temporarily active representations. Likewise, Logie (1995) considered working memory as a ‘mental workspace’ that cannot only hold but is also able to manipulate activated representations. In a recent peer review of an empirical paper we submitted to a journal, the following comment was offered: ‘one concern is conceptual in that the authors describe their tasks as involving working memory when it is basically a letter memory task that primarily emphasizes storage with little or no processing. If working memory is defined as simultaneous storage and processing, then these tasks would probably be considered to assess short-term memory rather than working memory’. Thus, it is probably fair to say that the concept of working memory as a non-unitary system that allows for both storage and processing has gained popular acceptance in our field. However, in my opinion, less progress has been made regarding the neural mechanisms underlying the ‘processing’ component of working memory as compared with the ‘storage’ component (although see Petrides 2005).

Based on the data we have reviewed thus far, we propose that any population of neurons within primary or unimodal association cortex can exhibit persistent neuronal activity, which serves to actively maintain the representations coded by those neuronal populations. Areas of multimodal cortex, such as PFC and parietal cortex, which are in a position to integrate representations through connectivity to unimodal association cortex, are also critically involved in the active maintenance of task-relevant information (see also Burgess et al. 2007; Stuss & Alexander 2007). Miller & Cohen (2001) have proposed that in addition to the recent sensory information, integrated representations of task contingencies and even abstract rules (e.g. if this object then this later response) are also maintained in the PFC. This is similar to what Fuster (1997) has long emphasized, namely that the PFC is critically responsible for temporal integration and the mediation of events that are separated in time but contingent on one another. In this way, the PFC may exert ‘control’ in that the information it represents can bias posterior unimodal association cortex in order to keep neural representations of behaviourally relevant sensory information activated when they are no longer present in the external environment (Fuster 2000; Ranganath et al. 2004; Miller & D’Esposito 2005; Postle 2005). In a real world example, when a person is looking at a crowd of people, the visual scene presented to the retina may include a myriad of angles, shapes, people and objects. However, if that person is a police officer looking for an armed robber escaping through the crowd, some mechanism of suppressing irrelevant visual information while enhancing task-relevant information is necessary for an efficient and effective search. Thus, neural activity throughout the brain that is generated by input from the outside world may be differentially enhanced or suppressed, presumably from top-down signals emanating from integrative brain regions such as PFC, based on the context of the situation. Thus, in this formulation, the processing component of working memory is that the control of actively maintained representations within primary and unimodal association cortex stems from the representational power of multimodal association cortex, such as the PFC, parietal cortex and/or hippocampus. If the PFC, for example, stores the rules and goals, then the activation of such PFC representations will be necessary when behaviour must be guided by internal states or intentions. As Miller & Cohen (2001) elegantly state, putative top-down signals originating in PFC may permit ‘the active maintenance of patterns of activity that represent goals and the means to achieve them. They provide bias signals throughout much of the rest of the brain, affecting visual processes and other sensory modalities, as well as systems responsible for response execution, memory retrieval, emotional evaluation, etc. The aggregate effect of these bias signals is to guide the flow of neural activity along pathways that establish the proper mappings between inputs, internal states and outputs needed to perform a given task’. Computational models of this type of system have created a PFC module (e.g. O’Reilly et al. 2002) that consists of ‘rule’ units whose activation leads to the production of a response other than the one most strongly associated with a given input. Thus, ‘this module is not responsible for carrying out input-output mappings needed for performance. Rather, this module influences the activity of other units whose responsibility is making the needed mappings’ (e.g. Cohen et al. 1990). Thus, there is no need to propose the existence of a homunculus (e.g. central executive) in the brain that can perform a wide range of cognitive operations which are necessary for the task at hand (for a further discussion of this issue, see Shallice 1988; see also Hazy et al. 2006).

We have used a delay task to directly study the neural mechanisms underlying top-down modulation by investigating the processes involved when participants were required to enhance relevant and suppress irrelevant information (Gazzaley et al. 2005b). During each trial, participants observed sequences of two faces and two natural scenes presented in a randomized order. The tasks differed in the instructions informing the participants how to process the stimuli: (i) remember faces and ignore scenes, (ii) remember scenes and ignore faces, or (iii) passively view faces and scenes without attempting to remember them. In each task, the period in which the cue stimuli were presented was balanced for bottom-up visual information, thus allowing us to probe the influence of goal-directed behaviour on neural activity (top-down modulation). In the two memory tasks, the encoding of the task-relevant stimuli requires selective attention and thus permits the dissociation.
of physiological measures of enhancement and suppression relative to the passive baseline. Also in the memory tasks, after a short-delay period, the participants were tested on their ability to recognize a probe stimulus as being one of the task-relevant cues, yielding a behavioural measure of memory performance. These experiments were performed using both event-related fMRI and electroencephalography (event-related potentials (ERP)) to record correlates of neural activity while the participants performed the task. This allowed us to capitalize on the high spatial resolution of fMRI and the high temporal resolution of ERP.

We investigated activity measures of enhancement and suppression obtained from the visual association cortex of young healthy participants. For fMRI, we used an independent functional localizer to identify both the stimulus-selective face and scene regions in the FFA and the parahippocampal/lingual gyrus, respectively. For ERP, we used a face-selective ERP component, the N170, which is localized to posterior occipital electrodes and is thought to reflect visual association cortex activity with some face specificity (Bentin et al. 1996). Our fMRI and ERP data revealed top-down modulation of both activity magnitude and processing speed that occurred above or below the perceptual baseline depending on task instruction (figure 5). In other words, during the encoding period of the delay task, FFA activity was enhanced, and the N170 occurred earlier, when faces had to be remembered as compared with a condition where they were passively viewed. Likewise, FFA activity was suppressed, and the N170 occurred later, when faces had to be ignored (with scenes now being retained instead across the delay interval) compared with a condition where they were passively viewed.

Thus, there appears to be at least two types of top-down signal, one that serves to enhance task-relevant information and another that serves to suppress task-relevant information. It is well documented that the nervous system uses interleaved inhibitory and excitatory mechanisms throughout the nervous system (e.g. spinal reflexes, cerebellar outputs and basal ganglia movement control networks). Thus, it may not be surprising that enhancement and suppression mechanisms may exist to control cognition (Knight et al. 1999; Shimamura 2000). By generating contrast via both enhancements and suppressions of activity magnitude and processing speed, top-down signals bias the likelihood of successful representation of relevant information in a competitive system.

Though it has been proposed that the PFC provides a major source of the types of top-down signals that we have described, this hypothesis largely originates from suggestive findings rather than direct empirical evidence. However, a few studies lend direct causal support to this hypothesis (see also Vuilleumier & Driver 2007). For example, Fuster et al. (1985) investigated the effect of cooling inactivation of specific parts of the PFC upon spiking activity in ITC neurons, during a DMS colour task. During the delay interval in this task—when persistent stimulus-specific activity in ITC neurons is observed—inactivation caused attenuated spiking profiles and a loss of stimulus specificity of ITC neurons. These two alterations of ITC signalling strongly implicate the PFC as a source of top-down signals necessary for maintaining robust sensory representations in the absence of bottom-up sensory activity. Tomita et al. (1999) isolated top-down signals during the retrieval of paired associates in a visual memory task. Spiking activity was recorded from stimulus-specific ITC neurons as cue stimuli were presented to the ipsilateral hemifield. This experiment's unique feature was the ability to separate bottom-up sensory signals from a top-down mnemonic reactivation, using a posterior split-brain procedure that limited hemispheric crosstalk to the anterior corpus callosum connecting each PFC. When a probe stimulus was presented ipsilaterally to the recording site, thus restricting bottom-up visual input to the contralateral hemisphere, stimulus-specific neurons became activated at the recording site approximately 170 ms later. Since these neurons received no bottom-up visual signals of the probe stimulus, with the only route between the two hemispheres being via the PFC, this experiment showed that PFC neurons were sufficient to trigger the reactivation of object-selective representations in ITC regions in a top-down manner. The combined lesion/electrophysiological approach in humans has rarely been implemented (though see Vuilleumier & Driver 2007). However, Chao & Knight (1998) studied patients with lateral PFC lesions during DMS tasks. It was found that when distracting stimuli are presented during the delay period, the amplitude of the recorded ERP from posterior electrodes was markedly increased in patients compared with controls. These results were interpreted to show disinhibition of sensory processing and support a role of the PFC in
suppressing the representation of stimuli that are irrelevant for current behaviour.

Clearly, there are other areas of multimodal cortex such as posterior parietal cortex, and the hippocampus, that can also be the source of top-down signals. For example, the hippocampus has been proposed to be specialized for rapid learning of arbitrary information which can be recalled in the service of controlled processing (O’Reilly et al. 1999). Moreover, input from brainstem neuromodulatory systems probably plays a critical role in modulating goal-directed behaviour (see also Robbins 2007). For example, the dopaminergic system probably plays a critical role in cognitive control processes (for a review, see Cools & Robbins 2004). Specifically, it is proposed that phasic bursts of dopaminergic neurons may be critical for updating currently activated task-relevant representations whereas tonic dopaminergic activity serves to stabilize such representations (e.g. Durstewitz et al. 2000a; Cohen et al. 2002). Empirical studies in animals and man have provided support for a role of dopamine in working memory (e.g. Sawaguchi 2001; Gibbs & D’Esposito 2005).

7. CONCLUSIONS
The overall goal of cognitive neuroscience as a discipline is to determine the biological basis of the mind. As an interdisciplinary discipline that has evolved from both neuroscience and psychology, cognitive neuroscientists consume data derived from each of these disciplines in their attempt to advance cognitive theory as well as determine how the brain implements cognitive function. Advances made in understanding the cognitive and neural basis of working memory provide examples of this synergy. In my opinion, future studies must continue to consider both cognitive and neural data in the way that has been briefly illustrated in this review. Research thus far suggests that working memory can be viewed as neither a unitary nor a dedicated system. A network of brain regions, including the PFC, is critical for the active maintenance of internal representations that are necessary for goal-directed behaviour. Thus, working memory is not localized to a single brain region but probably is an emergent property of the functional interactions between the PFC and the rest of the brain.

I would like to express my sincere appreciation to all of my former and current students and postdocs for their invaluable contributions to the formulation of the ideas presented in this paper.

REFERENCES


