

A Biologically-Based Computational Model of Working Memory

Randall C. O'Reilly[†], Todd S. Braver[‡], & Jonathan D. Cohen[‡]

[†]Department of Psychology
University of Colorado at Boulder
Campus Box 345
Boulder, CO 80309-0345

[‡]Department of Psychology and Center for the Neural Basis of Cognition
Carnegie Mellon University
Baker Hall
Pittsburgh, PA 15213
oreilly@psych.colorado.edu, tb2j@crab.psy.cmu.edu, jdcohen@andrew.cmu.edu

December 22, 1997

Draft Chapter for Miyake, A. & Shah, P. (Eds) *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. New York: Cambridge University Press.

Five Central Features of the Model

We define working memory as controlled processing involving active maintenance and/or rapid learning, where controlled processing is an emergent property of the dynamic interactions of multiple brain systems, but the prefrontal cortex (PFC) and hippocampus (HCMP) are especially influential due to their specialized processing abilities and their privileged locations within the processing hierarchy (both the PFC and HCMP are well connected with a wide range of brain areas, allowing them to influence behavior at a global level). The specific features of our model include:

1. A PFC specialized for active maintenance of internal contextual information that is dynamically updated and self-regulated, allowing it to *bias* (control) ongoing processing according to maintained information (e.g., goals, instructions, partial products, etc).
2. A HCMP specialized for rapid learning of arbitrary information, which can be recalled in the service of controlled processing, while the posterior perceptual and motor cortex (PMC) exhibits slow, long-term learning that can efficiently represent accumulated knowledge and skills.
3. Control that emerges from interacting systems (PFC, HCMP and PMC).
4. Dimensions that define continua of specialization in different brain systems: e.g., robust active maintenance, fast vs slow learning.
5. Integration of biological and computational principles.

Introduction

Working memory is an intuitively appealing theoretical construct — perhaps deceptively so. It has proven difficult for the field to converge on a fully satisfying, mechanistically explicit account of what exactly working memory is and how it fits into a larger model of cognition (hence the motivation for this volume). Existing theoretical models of working memory can be traced to ideas based on a traditional computer-like mental architecture, where processing is centralized and long-term memory is essentially passive. In this context, it makes sense to have RAM or cache-like working memory buffers dedicated to temporarily storing items that are needed during processing by the *central executive* (Baddeley, 1986). Alternative processing architectures have been proposed, both within the computational and psychological literatures (e.g., Anderson, 1983; Newell, 1990), in which working memory is defined functionally — as the activated component of long term memory representations — rather than structurally as a dedicated component of the system. However, these typically include a structural distinction between processing and memory. None of these architectures seems to correspond closely to the architecture of the brain, in which processing and memory functions are typically distributed within and performed by the same neural substrate (Rumelhart & McClelland, 1986).

We believe that considering how working memory function might be implemented in the brain provides a unique perspective that is informative with regard to both the psychological and biological mechanisms involved. This is what we attempt to do in this chapter, by providing a biologically-based computational model of working memory. Our goal is not only to provide an account that is neurobiologically plausible, but also one that is mechanistically explicit, and that can be implemented in computer simulations of specific cognitive tasks. We share this goal with others in this volume who have also committed their theories to mechanistically explicit models, at both the symbolic (Lovett, Reder, & Lebiere, this volume; Young & Lewis, this volume; Kieras, Meyer, Mueller, & Seymour, this volume) and neural (Schneider, this volume) levels.

It is possible to identify a core set of information processing requirements for many working memory tasks: 1) Task instructions and/or stimuli must be encoded in such a form that they can either be actively maintained over time, and/or learned rapidly and stored offline for subsequent recall; 2) The active maintenance must be both dynamic and robust, so that information can be selectively maintained, flexibly updated, protected from interference, and held for arbitrary (although relatively short) durations; 3) The maintained information must be able to rapidly and continually influence (bias) subsequent processing or action selection. 4) The rapid learning must avoid the problem of interference in order to keep even relatively similar types of information distinct. In addition to these specifications for an *active memory* system and a *rapid learning* system, we think that the working memory construct is generally associated with tasks that require *controlled processing*, which governs the updating and maintenance of active memory and the storage and retrieval of rapidly learned information in a strategic or task-relevant manner. This is consistent with the original association of working memory with central-executive like functions. Taken together, these functional aspects of working memory provide a basic set of constraints for our biologically based model.

Our approach involves two interrelated threads. The first is a focus on the functional dimensions along which different brain systems appear to have specialized, and the processing tradeoffs that result as a consequence of these specializations. The second is a set of computational models in which we have implemented these functional specializations as explicit mechanisms. Through simulations, we have endeavored to show how the interactions of these specialized brain systems can account for specific patterns of behavioral performance on a wide range of cognitive tasks. We have postulated that prefrontal cortex (PFC), hippocampus (HCMP) and posterior and motor cortex (PMC) represent three extremes of specialization along different functional dimensions important for working memory: sensory and motor processing based on inference and generalization (PMC); dynamic and robust active memory (PFC); and rapid learning of arbitrary informa-

tion (HCMP). Since each of these specializations involve tradeoffs, it is only through interactions between these systems that the brain can fulfill the information processing requirements of working memory tasks.

As an example of how these components work together, consider a simple real-world task that involves contributions from these different brain systems. Imagine you are looking for some information (the name of a college friend's child) contained in an email message you received a year ago and have stored in one of your many message folders. You can remember several things about that email, like who sent it (a good friend who knows the college friend), what else was happening at around that time (you had just returned from a conference in Paris), but you don't remember the subject line or where you filed it. This information about the email is retrieved from the HCMP system, which was able to bind together the individual features of the memory and store it as a unique event or episode. Once recalled, these features must be used to guide the process of searching through the folders and email messages. We think that this happens by maintaining representations of these features in an active state in the PFC, which is able to keep them active for the duration of the search, and protect them from being dislodged from active memory by all the other information you read. Meanwhile the basic abilities of reading information and issuing appropriate commands within the email system are subserved by well-learned representations within the PMC, guided by representations helped active in PFC.

Once initiated, the search requires the updating of items in active memory (college friend's name, good friend's name, Paris conference) and its interaction with information encountered in the search. For example, when you list all of the folders, you select a small subset as most probable. This requires an interaction between the items in active memory (PFC), long-term knowledge about the meanings of the folders (PMC), and specific information about what was filed into them (HCMP). The result is the activation of a new set of items in active memory, containing the names of the new set of folders to search. You may first decide to look in a folder that will contain an email telling you exactly when your conference was, which will help narrow the search. As you do this, you may keep that date in active memory, and not maintain the conference information. Thus, the items in active memory are updated (activated and deactivated) as needed by the task at hand. Finally, you iterate through the folders and email messages, matching their date and sender information with those maintained in active memory, until the correct email is found.

All of this happened as a result of strategic, controlled processing, involving the activation and updating of goals (the overall search) and sub-goals (e.g., finding the specific date). The maintenance and updating of goals, like that of the other active memory items, is dependent on specialized mechanisms in the PFC system. Thus, the PFC system plays a dominant role in both active memory and controlled processing, which are two central components of the working memory construct. However, other systems can play equally central roles. For example, if you were interrupted in your search by a phone call, then you might not retain all the pertinent information in active memory ("Now, where was I?"). The HCMP system can fill in the missing information by frequently (and rapidly) storing snapshots of the currently active representations across much of the cortex, which can then be recalled after an interruption in order to pick up where you left off. Thus, working memory functionality can be accomplished by multiple brain systems, though the specialized active memory system of the PFC remains a central one.

We have studied a simple working memory task based on the continuous performance test (CPT), which involves searching for target letters in a continuous stream of stimuli (typically letters). For example, in the AX version of the CPT (AX-CPT), the target is the letter 'X', but only if it immediately follows the letter 'A'. Thus, the 'X' alone is inherently ambiguous, in that it is a non-target if preceded by anything other than an 'A'. Like the email search task, this requires the dynamic updating and maintenance of active memory representations (e.g., the current stimulus must be maintained in order to perform correctly on the subsequent one), which makes this a working memory task. Active maintenance is even more important for a more demanding version of this task called the *N-back*, in which any letter can be a target if it is identical to the one occurring *N* trials previously (where *N* is pre-specified and is typically 1, 2, or 3). Thus, more items

need to be maintained simultaneously, and across intervening stimuli. The N-back also requires updating the working memory representations after each trial, in order to keep track of the order of the last N letters.

There are several other relevant demands of this task. For example, upon receiving the task instructions, subjects must rapidly learn the otherwise arbitrary association between the letter 'A' and 'X'. We assume that this is carried out by the HCMP. Of course, subjects must also be able to encode each stimulus, and execute the appropriate response, which we assume is carried out by PMC. Together, the rapid association of the cue with the correct response to the target (HCMP), the active maintenance of information provided by the specific cue presented in each trial (PFC), and the use of that information to guide the response (PMC), constitute a simple form of working memory function. In Figure 1, we present a computational model of performance in this task, that illustrates our theory regarding the functional roles of PFC, PMC and HCMP. We have implemented components of this model, and demonstrated that it can account for detailed aspects of normal behavior in the AX-CPT, as well as that of patients with schizophrenia who are thought to suffer from PFC dysfunction (Braver, Cohen, & Servan-Schreiber, 1995; Cohen, Braver, & O'Reilly, 1996; Braver, Cohen, & McClelland, 1997a).

In the model, the PMC layer of the network performs stimulus identification, and response generation. Thus, in panel **a**), when the 'A' stimulus is presented, an unequivocal non-target response (here mapped onto the right hand, but counterbalanced in empirical studies) is generated. However, the PFC is also activated by this 'A' stimulus, since it serves as a cue for a possible subsequent target 'X' stimulus. During the delay period shown in panel **b**), the PFC maintains its representation in an active state. This PFC representation encodes the information that the prior stimulus was a cue, and thus that if an 'X' comes next, a target (left hand) response should be made. When the 'X' stimulus is then presented (panel **c**), the PFC-maintained active memory biases processing in the PMC in favor of the interpretation of the 'X' as a target, leading to a target (left hand) response. We think that the HCMP would also play an important role in performing this task, especially in early trials, by virtue of its ability to rapidly learn associations between the appropriate stimuli (e.g., 'A' in the PMC and 'left-to-X' in the PFC) based on instructions, and provide a link between these until direct cortical connections have been strengthened. However, we have not yet implemented this important component of working memory in this model.

In the following sections, we first elaborate our theory of working memory in terms of a more comprehensive view of how information is processed within neural systems. While we believe it is important that our theory based on mechanistic models of cognitive function whose behavior can be compared with empirical data, a detailed consideration of individual models or empirical studies is beyond the scope of this chapter. Furthermore, many of the features of our theory have not yet been implemented, and remain a challenge for future work. Thus, our objective in this chapter is to provide a high level overview of our theory, and how it addresses the theoretical questions posed for this volume.

A Biologically-Based Computational Model of Cognition

Our model of working memory is a unified one in that the same underlying computational principles are used throughout. We and others have relied upon these computational principles in previous work to address issues regarding cognitive function and behavioral performance in many other task domains, including ones involving response competition, classical conditioning, and covert spatial attention. Moreover, the functional specializations that we postulate for different brain systems emerge as different parametric variations within this unified framework, giving rise to a continuum of specialization along these dimensions. The basic computational mechanisms are relatively simple, including standard parallel-distributed-processing (PDP) ideas (Rumelhart & McClelland, 1986; McClelland, 1993; Seidenberg, 1993), the most relevant being:

The Brain uses *parallel, distributed processing* involving many relatively simple elements (neurons or

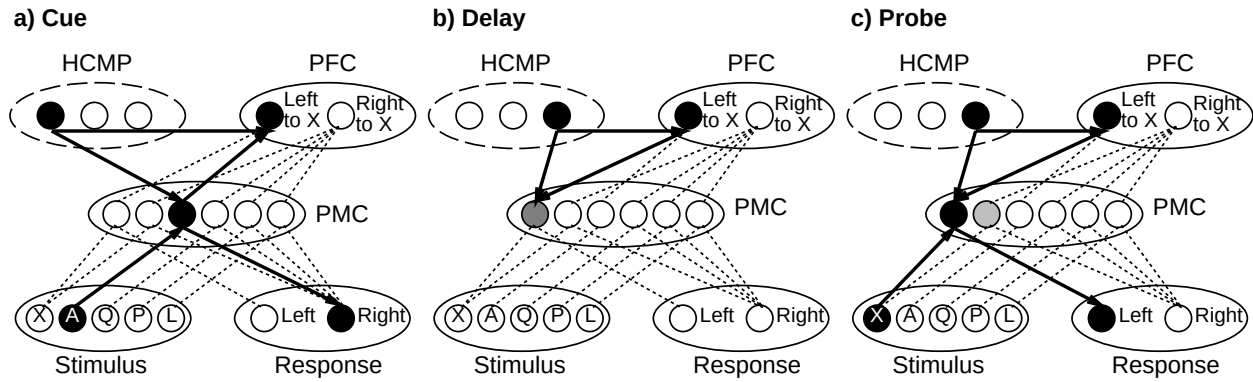


Figure 1: Neural network model of the AX-CPT task, showing roles of PMC (posterior & motor cortex), PFC (pre-frontal cortex) and HCMP (hippocampus and related structures: not actually implemented in our models yet). Activation is shown by black units, and the weights between such units are highlighted to emphasize the flow of information through the network. Lateral inhibition exists within each of the layers. **a)** The cue stimulus ‘A’ is presented, resulting in the activation of a PFC representation for “Respond with left (target) hand if an X comes next.” **b)** During the delay, the PFC representation is actively maintained, providing top-down support for the target interpretation of the ‘X’ stimulus. **c)** When the ‘X’ comes, it results in a target (left) response (whereas a non-target (right) response would have occurred without the top-down PFC activation during the delay period). In early trials, the HCMP provides appropriate activation to various task components, as a result of its ability for rapid learning.

neural assemblies), each of which is capable of performing local processing and memory, and which are grouped into systems.

Systems are composed of groups of related elements, that subserve a similar set of processing functions. Systems may be defined anatomically (by patterns of connectivity) and/or functionally (by specialization of representation or function).

Specialization arises from parametric variations in properties possessed by all elements in the brain (e.g., patterns of connectivity, time constants, regulation by neuromodulatory systems, etc.). Parameter variations occur along continuous dimensions, and thus subsystem specialization can be a graded phenomenon.

Knowledge is encoded in the synaptic connection strengths (*weights*) between neurons, which typically change slowly compared with the time course of processing. This means that neurons have relatively stable (*dedicated*) representations over time.

Cognition results from *activation propagation* through interconnected networks of neurons. Activity is required to directly influence ongoing processing.

Learning occurs by *modifying weights* as a function of activity (which can convey *error* and *reward* feedback information from the environment).

Memory is achieved either by the relatively short-term persistence of activation patterns (*active memory*) or longer-lasting weight modifications (*weight-based memory*).

Representations are *distributed* over many neurons and brain systems, and at many different levels of abstraction and contextualization.

Inhibition between representations exists at all levels, both within and possibly between systems, and increases as a (non-linear) function of the number of active representations. This results in *attention*

System	Function	Internal Relation	External Relation	Act Capacity	Learn Rate
PMC	inference, processing	distributed, overlapping	embedded	many	slow
PFC	maintenance, control	isolated, combinatorial	global	few	slow
HCMP	rapid learning	separated, conjunctive	context sensitive	one	fast

Table 1: Critical parameterizations of the three systems. **Function** specifies the function optimized by this system. **Internal Relation** indicates how representations within each system relate to each other. **External Relation** indicates how representations relate to other systems. **Act Capacity** indicates how many representations can be active at any given time. **Learn Rate** indicates characteristic rate of learning. See text for fuller description.

phenomena, and has important computational benefits by enforcing relatively *sparse* levels of activation.

Recurrence (bidirectional connectivity) exists among the elements within a system and between systems, allowing for *interactive* bottom-up and top-down processing, *constraint satisfaction* settling, and the communication of error signals for learning.

A central feature of this framework, as outlined above, is that different brain systems are specialized for different functions. In order to characterize these specializations (and understand why they may have arisen), we focus on basic tradeoffs that exist within this computational framework (e.g., activity- vs weight-based storage, or rapid learning vs extraction of regularities). These tradeoffs lead to specialization, since a homogeneous system would require compromises to be made, whereas specialized systems working together can provide the benefits of each end of the tradeoff without requiring compromise. This analysis has led to the following set of coincident biological and functional specializations, which are also summarized in Figure 2 and Table 1:

Posterior perceptual and motor cortex (PMC): Optimizes knowledge-dependent *inference* capabilities, which depend on dense interconnectivity, highly distributed representations, and slow *integrative* learning (i.e., integrating over individual learning episodes) in order to obtain good estimates of the important structural/statistical properties of the world, upon which inferences are based (McClelland, McNaughton, & O'Reilly, 1995). Similarity-based overlap among distributed representations is important for enabling generalization from prior experience to new situations. These systems perform sensory/motor and more abstract, multi-modal processing in a hierarchical but highly interconnected fashion, resulting in the ability to perceive and act in the world in accordance with its salient and reliable properties. We take this to be the canonical type of neural computation in the cortex, and view the other systems in reference to it.

Pre-frontal cortex (PFC): Optimizes *active memory* via restricted recurrent excitatory connectivity and an active gating mechanism (discussed below). This results in the ability to both flexibly update internal representations, maintain these over time and in the face of interference and, by propagation of activation from these representations, bias PMC processing in a task-appropriate manner. PFC is specialized because there is a fundamental tradeoff between the ability to actively sustain representations (in the absence of enduring input or the presence of distracting information) and dense interconnectivity underlying distributed (overlapping) representations such as in the PMC (Cohen et al., 1996). As a result, the individual self-maintaining representations in PFC must be relatively *isolated* from each other (as opposed to distributed). They can thus be activated *combinatorially* with less mutual interference or contradiction, allowing for flexible and rapid updating. Because they sit high in the cortical representational hierarchy, they are less embedded and more globally accessible and influential. Because they are actively maintained and strongly influence cognition, only a relatively small number of representations can typically be concurrently active in the PFC at a given time in order for

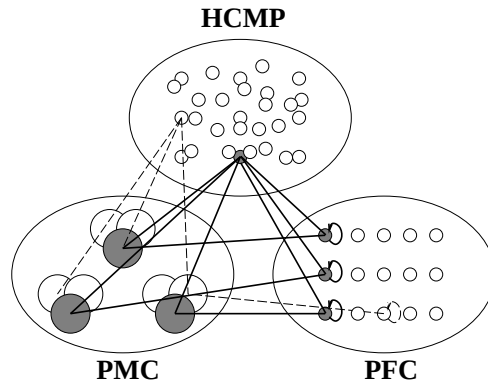


Figure 2: Diagram of key properties of the three principal brain systems. Active representations are shown in grey; highly overlapping circles are distributed representations; non-overlapping are isolated; in between are separated. Weights between active units are shown in solid lines; those of non-active units in dashed. Three active feature values along three separate “dimensions” (e.g., modalities) are represented. **PMC** representations are distributed and embedded in specialized (e.g., modality specific) processing systems. **PFC** representations are isolated from each other, and combinatorial, with separate active units representing each feature value. Unlike other systems, PFC units are capable of robust self-maintenance, as indicated by the recurrent connections. **HCMP** representations are sparse, separated (but still distributed), and conjunctive, so that only a single representation is active at a time, corresponding to the conjunction of all active features.

coherent cognition to result. Thus, inhibitory attentional mechanisms play an important role in PFC, and for understanding the origin of capacity constraints.

Hippocampus and related structures (HCMP): Optimizes rapid learning of arbitrary information in weight-based memories. This permits the binding of elements of a novel association, including representations in PFC and PMC, providing a mechanism for temporary storage of arbitrary current states for later retrieval. There is a tradeoff between such rapid learning of arbitrary information without interfering with prior learning (retroactive interference), and the ability to develop accurate estimates of underlying statistical structure (McClelland et al., 1995). To avoid interference, learning in the HCMP uses *pattern separation* (i.e., individual episodes of learning are separated from each other), as opposed to the integration characteristic of PMC. This separation process requires *sparse, conjunctive* representations, where all the elements contribute interactively (not separably) to specifying a given representation (O'Reilly & McClelland, 1994). This conjunctivity is the opposite of the combinatorial PFC, where the elements contribute separably. Conjunctivity leads to *context specific* and *episodic* memories, which bind together the elements of a context or episode. This also implies that there is a single HCMP representation (consisting of many active neurons) corresponding to an entire pattern of activity in the cortex. Since only one HCMP representation can be active at any time, reactivation is necessary to extract information from multiple such representations.

While there are undoubtedly many other important specialized brain systems, we think that these three provide central, and critical contributions to working memory function. However, brainstem neuromodulatory systems, such as dopamine and norepinephrine, play an important supporting role in our theory, as a result of their capacity to modulate cortical processing according to reward, punishment, and affective states. In particular, as we discuss further below, we have hypothesized that dopamine activity plays a critical role in working memory function, by regulating active maintenance in PFC (Cohen & Servan-Schreiber, 1992; Cohen et al., 1996).

It should also be emphasized that the above are relatively broad characterizations of large brain systems, which (especially in the case of the neocortical systems) may have subsystems with different levels of conformance to these generalizations. Further, there may be other important differences between these systems that are not reflected in our account. Nevertheless, these generalizations are consistent with a large corpus of empirical data and ideas from other theorists (e.g., Fuster, 1989; Shallice, 1982; Goldman-Rakic, 1987; Squire, 1992). Finally, we note that there are still important portions of this account that have not yet been implemented in computational models, and the sufficiency of these ideas to perform complex cognitive tasks, especially those involving extended sequential behavior, remains untested as of now. Nevertheless, encouraging progress has been made implementing and testing models some of the more basic functions we have described, such as the active maintenance function of PFC and the binding function of hippocampus (see Cohen et al., 1996; McClelland et al., 1995; Cohen & O'Reilly, 1996 for reviews).

In what follows, we will elaborate the ways in which these brain systems interact to produce controlled processing and working memory, and make more clear their relationship to other constructs such as consciousness and active memory. We will then focus on a set of important issues surrounding the operation of the PFC active memory system, followed by an application of these ideas to understanding some standard working memory tasks. This then provides a sufficient set of principles to address the theoretical questions posed in this volume.

Controlled Processing and Brain System Interactions

We consider controlled processing to be an important aspect of our theory of working memory. This has classically been described in contrast with *automatic processing* (Shiffrin & Schneider, 1977; Posner & Snyder, 1975), and has been thought to involve a limited capacity attentional system. However, more recent theories have suggested that a continuum may exist between controlled and automatic processing, (Kahneman & Treisman, 1984; Cohen, Dunbar, & McClelland, 1990), and we concur with this view. Thus, working memory also varies along this same continuum. In particular, we have conceptualized controlled processing as the ability to flexibly adapt behavior to the demands of particular tasks, favoring the processing of task-relevant information over other sources of competing information, and mediating task-relevant behavior over habitual or otherwise prepotent responses. In our models, this is operationalized as the use and updating of actively maintained representations in PFC to bias subsequent processing and action selection within PMC in a task-appropriate manner. For example, in the AX-CPT model described above, the context representation actively maintained in PFC is able to exert control over processing by biasing the response made to an ambiguous probe stimulus.

While it is tempting to equate controlled processing with theoretical constructs such as a central executive (Gathercole, 1994; Shiffrin & Schneider, 1977), there are critical differences in the assumptions and character of these mechanisms that have important consequences for our model of working memory. Perhaps the most important difference between our notion of controlled processing and theories that posit a central executive is that we view controlled processing as emerging from the interactions of several brain systems, rather than the operation of a single, unitary CPU-like construct. We believe that our interactive, decentralized view is more consistent with the graded aspect of controlled processing, as well as the character of neural architectures. However, aspects of our theory are compatible with other models. For example, Shallice (1982) has proposed a theory of frontal function, and the operation of a central executive, in terms of a supervisory attentional system (SAS). He describes this using a production system architecture, in which the SAS is responsible for maintaining goal states in working memory, in order to coordinate the firing of productions involved in complex behaviors. This is similar to the role of goal stacks and working memory in ACT (Anderson, 1983; Lovett et al., this volume). Similarly, our theory of working memory and controlled processing depends critically on actively maintained representations (in PFC). This central role for active maintenance in achieving controlled processing contrasts with a view where active maintenance

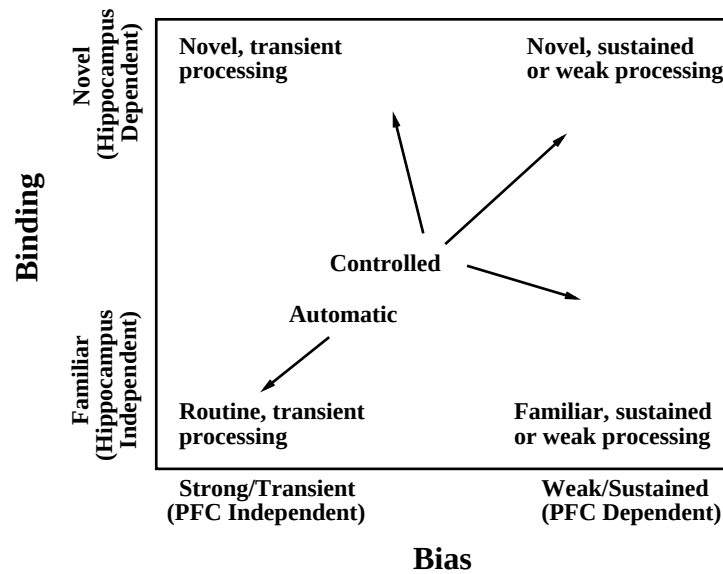


Figure 3: Ways in which the HCMP and PFC contribute to the automatic vs controlled-processing distinction (after Cohen & O'Reilly, 1996). **Bias** is provided by the PFC, and can be used to perform sustained processing, can facilitate the processing of weakly-learned (i.e., relatively infrequent) tasks, and can serve to coordinate processing across different systems. **Binding** is provided by the HCMP, and can be used to rapidly learn and store the information necessary to perform novel tasks or processing. Controlled processing can involve either or both of these contributions, while automatic processing can be performed independent of them.

and executive control are strictly segregated (Baddeley & Logie, this volume).

We consider controlled processing to arise from the interplay between PFC *biasing* and HCMP *binding* (Cohen & O'Reilly, 1996). Figure 3 illustrates the central ideas of this account, which is based on the functional characterizations of the PFC and HCMP as described above. According to this view, the degree to which controlled processing is engaged by a task is determined by the extent to which either or both of the following conditions exist:

- Sustained, weakly-learned (i.e., relatively infrequent), or coordinated processing is required.
- Novel information must be rapidly stored and accessed.

Since the PFC can bias processing in the rest of the system, sustained activity of representations in PFC can produce a focus of activity among representations in PMC needed to perform a given task. This can be used to support representations in PMC over temporally extended periods (e.g., in delayed response tasks), and/or weakly-learned representations that might otherwise be dominated by stronger ones (e.g., in the Stroop task, where highly practiced word-reading dominates relatively infrequent color naming — Cohen et al., 1990). This function of PFC corresponds closely to Engle, Kane, and Tuholski's (this volume) notion of controlled attention and to Cowan's (this volume) notion of focus of attention. In contrast, the HCMP contributes the ability to learn new information rapidly and without interference, binding together task-relevant information (e.g., task instructions, particular combinations of stimuli, intermediate states of problem solutions, etc) in such a form that it be retrieved at appropriate junctures during task performance. This may be relevant for Ericsson and Delaney's (this volume) notion of long term working memory, Young and Lewis' (this volume) production learning mechanism, and Moscovitch and Winocur's (1992) notion of "working-with-memory". We propose that the combination of these two functions (PFC biasing and HCMP binding) can account for the distinction between controlled and automatic processing. On this account,

automatic processing is what occurs via activation propagation through intrinsic PMC connectivity, while controlled processing reflects the additional constraints on the flow of activity brought to bear by the PFC and/or HCMP.

Activation Propagation and Multiple Constraint Satisfaction

While some aspects of behavior can be understood in terms of relatively local processes within the brain, we assume that, under most circumstances, behavior is determined by a rich and dynamic set of interactions involving the widespread propagation of activation to multiple, distributed brain systems. While the detailed outcome of such processing in a particular case may be difficult, if not ultimately impossible, to describe, its general character can be understood in terms of *multiple constraint satisfaction*: the activation state that results from this propagation of information over weighted neuronal connections is likely to be one that satisfies various constraints, including those imposed by three critical components: 1) external stimuli; 2) sustained activity in PFC; and 3) recalled information from the HCMP. Thus, representations in PFC and HCMP act as “control signals,” insofar as these influence the flow of activity and thereby shape the constraint satisfaction process that is taking place in the rest of the brain. Furthermore, their activation states are themselves influenced by similar constraint satisfaction mechanisms based on activations from the PMC (though a presumed gating mechanism in the PFC can make it more or less susceptible to this “bottom-up” influence — see discussion below). All of these constraints are mediated by the synaptic connections between neurons, which are adapted through experience in such a way as to result in better activation states in similar situations in the future. Thus, much of the real work being done in our model (and our avoidance of a homunculus or otherwise unspecified central executive mechanisms when we discuss controlled processing), lies in these activation dynamics and their tuning as a function of experience. Computational models are essential in demonstrating the efficacy of these mechanisms, which may otherwise appear to have mysterious properties.

Accessibility and Consciousness

In our model, one of the dimensions along which brain systems differ is in the extent to which their representations are globally accessible to a wide range of other brain systems, as opposed to embedded within more specific processing systems and less globally accessible. We view this difference as arising principally from a system’s relative position within an overall hierarchy of abstractness of representations. This hierarchy is defined by how far removed a system is from direct sensory input or motor output. Systems supporting high level, more abstract representations are more centrally located with respect to the overall network connectivity, resulting in greater accessibility. Accordingly, because both the PFC and HCMP are at the top of the hierarchy (Squire, Shimamura, & Amaral, 1989; Fuster, 1989) they are more influential and accessible than subsystems within PMC. Like the other dimensions along which these systems are specialized, we view this as a graded continuum, and not as an all-or-nothing distinction. Furthermore, we assume that the PMC has rich “lateral” connectivity between subsystems at the same general level of abstraction (at least beyond the first few levels of sensory or motor processing). Nevertheless, the PFC and HCMP assume a position of greater accessibility, and therefore greater influence, relative to other systems.

Accessibility has many implications, which relate to issues of conscious awareness, and psychological distinctions like *explicit* vs *implicit* or *declarative* vs *procedural*. We view the contents of conscious experience as reflecting the results of global constraint satisfaction processing throughout the brain, with those systems or representations that are most influential or constraining on this process having greater conscious salience (c.f., Kinsbourne, 1997). In general, this means that highly accessible and influential systems like PFC and HCMP will tend to dominate conscious experience over the more embedded subsystems of the PMC. Consequently, these systems are most clearly associated with notions of explicit or declarative processing, while the PMC and subcortical systems are associated with implicit or procedural processing. We

endorse this distinction, but add the important caveat that PFC and HCMP are participants in an extended interactive system, and that, once again, such distinctions should be considered along a continuum. Thus, our theory is not compatible with strong assumptions about informationally encapsulated modules (e.g., Fodor, 1983; Moscovitch & Winocur, 1992).

Active Memory vs Working Memory

In our model, we use the term active memory to refer to information that is represented as a pattern of activity (neuronal spiking) across a set of units (neural assembly) which persists over some (possibly brief) time interval. We view working memory as relying on active memory, by virtue of the need to rapidly and frequently access stored information over short intervals and use this information to bias processing in an on-going way in other parts of the system. However, the HCMP, because it is capable of rapidly forming novel associations and retrieving these in task-relevant contexts, is also useful for working memory. Conversely, we do not assume that actively maintained representations are invoked exclusively within the context of working memory. Sustained activity can occur and play a role in automatic processing as well. For example, it is not difficult to imagine that relatively automatic tasks such as typing would require persistent active representations, and sustained activity has indeed been observed in areas outside of PFC (Miller & Desimone, 1994). We assume that actively maintained representations participate in working memory function only under conditions of controlled processing — that is, when sustained activity is the result of representations currently being actively maintained in the PFC, or retrieved by the HCMP. This corresponds directly to the distinctions, proposed by Cowan (this volume) and Engle et al. (this volume), between controlled or focused attention vs other sources of activation and attentional effects.

Regulation of Active Memory

It has long been known from electrophysiological recordings in monkeys that PFC neurons remain active over delays between a stimulus and a contingent response (e.g., Fuster & Alexander, 1971). Furthermore, while such sustained activity has been observed in areas outside of PFC, it appears that PFC activity is robust to interference from processing intervening distractor stimuli, while activity within the PMC is not (Miller, Erickson, & Desimone, 1996; Cohen, Perlstein, Braver, Nystrom, Jonides, Smith, & Noll, 1997). Although the precise mechanisms responsible for active maintenance in PFC are not yet known, one likely mechanism is strong recurrent excitation. If groups of PFC neurons are strongly interconnected with each other, then strong mutual excitation will lead to both sustained activity, and some ability to resist interference. This idea has been developed in a number of computational models of PFC function (e.g., Dehaene & Changeux, 1989; Zipser, Kehoe, Littlewort, & Fuster, 1993). However, we believe that this simple model is inadequate to account for both robust active maintenance, and the kind of rapid and flexible updating that is necessary for complex cognitive tasks.

The underlying problem reflects a basic tradeoff — to the extent that units are made impervious to interference (i.e., by making the recurrent excitatory connections stronger), this also prevents them from being updated (i.e., new representations activated and existing ones deactivated). Conversely, weaker excitatory connectivity will make units more sensitive to inputs and capable of rapid updating, but will not enable them to be sustained in the face of interference. To circumvent this tradeoff, we think that the PFC has taken advantage of midbrain neuromodulatory systems, which can provide a *gating* mechanism for controlling maintenance. When the gate is opened, the PFC representations are sensitive to their inputs, and capable of rapid updating. When the gate is closed, the PFC representations are protected from interference. Such a gating mechanism can augment the computational power of recurrent networks (Hochreiter & Schmidhuber, 1997), and we have hypothesized that dopamine (DA) implements this gating function in PFC, based on a substantial amount of biological data (Cohen et al., 1996).

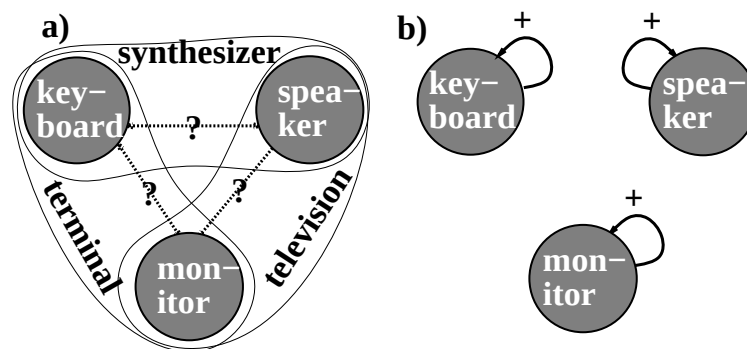


Figure 4: Illustration of difficulties with active maintenance via recurrent excitation with distributed representations. **a)** No value of the excitatory weights will enable an appropriate subset of two features to be maintained, without also activating the third. **b)** If representations are made independent, then maintenance is no problem, but semantic relatedness of the features is lost. One could also just maintain the higher-level items (i.e., synthesizer, terminal and television).

Thus, we propose that the midbrain DA nuclei (the ventral tegmental area, VTA), under control of descending cortical projections, enable the PFC to actively regulate the updating of its representations by controlling the release of DA in a strategic manner. Specifically, we propose that the afferent connections into the PFC from other brain systems are usually relatively weak compared to stronger local excitation, but that DA enhances the strength of these afferents¹ at times when updating is necessary. This would predict that the VTA should exhibit phasic firing at those times when the PFC needs to be updated. Schultz, Apicella, and Ljungberg (1993) have found that indeed, the VTA exhibits transient, stimulus-locked activity in response to stimuli which predicted subsequent meaningful events (e.g., reward or other cues that then predict reward). Further, we argue that this role of DA as a gating mechanism is synergistic with its widely discussed role in reward-based learning (e.g., Montague, Dayan, & Sejnowski, 1996). As will be discussed further in the final discussion section, this learning helps us to avoid the need to postulate a homunculus-like mechanism for controlled processing.

General Nature of Active Memory Representations

Our theory places several important demands on the nature of representations within the PFC, in addition to the rapidly updatable yet robust active maintenance discussed above. In general, we view the PFC's role in controlled processing as imposing a sustained, task relevant, top-down bias on processing in the PMC. Thus, in complex cognitive activities, the PFC should be constantly activating and deactivating representations that can bias a large number of combinations of PMC representations, while sustaining a coherent and focused thread of processing. This means that the PFC needs a vast repertoire of representations that can be activated on demand, and these representations need to be connected with the PMC in appropriate ways. Further, there must be some way of linking together sequences of representations in a coherent way.

Our initial approach towards understanding PFC representations has been dominated by an interesting coincidence between the above functional characterization of the PFC, and a consequence of an active maintenance mechanism based on recurrent excitation. Distributed representations, which are thought to be characteristic of the PMC, are problematic for this kind of active maintenance mechanism, as they rely critically on afferent input in order to select the appropriate subset of distributed components to be activated.

¹In addition, it appears that inhibitory connections are also enhanced, which would provide a means of deactivating existing representations.

In the absence of this afferent selection (e.g., during a delay period), recurrent excitation among the components will spread inappropriately, and result in the loss of the original activity pattern. This is illustrated in Figure 4, where a distributed representation is used to encode three different items, which each share two out of three total features. If these distributions have the strong recurrent excitatory connections necessary for active maintenance, then it will be difficult to keep a unique subset of two features active without also activating the third: activation will spread to the third unit via the connections necessary to maintain it in other circumstances. The alternative (shown in panel **b**) is to use *isolated* representations, which maintain only themselves. However, what is missing from these isolated representations is the rich interconnectivity that encodes knowledge about the relationships between the features, which could be used for performing the knowledge-dependent inference that we think is characteristic of the PMC. We obtain theoretical leverage from this basic tradeoff, which can be avoided by having two specialized systems (PMC and PFC).

The idea that PFC representations are relatively isolated from each other has important functional consequences beyond the active maintenance of information. For example, in order to achieve flexibility and generativity, PFC representations must be useful in novel contexts and combinations. Thus, individual PFC representations should not interfere much with each other so that they can be more easily and meaningfully combined — this is just what one would expect from relatively isolated representations. We think that learning in the PFC is slow and integrative like the rest of the cortex, so that this gradual learning taking place over many years of human development produces a rich and diverse palette of relatively independent PFC representational components, which eventually enable the kind of flexible problem solving skills that are uniquely characteristic of adult human cognition.

This view of PFC representations can be usefully compared with that of human language. In terms of basic representational elements, language contains words, which have a relatively fixed meaning, and can be combined in a huge number of different ways to express different (and sometimes novel) ideas. Words represent things at many different levels of abstraction and concreteness, and complicated or particularly detailed ideas can be expressed by combinations of words. We think that similar properties hold for PFC representations, and indeed that a substantial subset of PFC representations do correspond with word-like concepts. However, we emphasize that word meanings have highly distributed representations across multiple brain systems (Damasio, Grabowski, & Damasio, 1996), and also that the PFC undoubtedly has many non-verbal representations. Nevertheless, it may be that the PFC component of a word's representation approaches most closely the notion of a discrete, symbol-like entity.

Taking this language idea one step further, it may provide useful insights into the kinds of updating and sequential linking that PFC representations need to undergo during processing. For example, language is organized at many different levels of temporal structure, from short phrases, through longer sentences to paragraphs, passages, etc. These levels are mutually constraining, with phrases adding together to build higher-level meaning, and this accumulated meaning biasing the interpretation of lower-level phrases. This same interactive hierarchical structure is present during problem solving and other complex cognitive activities, and is critical for understanding the dynamics of PFC processing. We think that all of these different levels of representation can be active simultaneously, and mutually constraining each other. Further, it is possible that the posterior-anterior dimension of the PFC may be organized roughly according to level of abstraction (and correspondingly, temporal duration). For example, there is evidence that the most anterior area of PFC, the frontal pole, is only activated in more complicated problem solving tasks (Baker, Rogers, Owen, Frith, Dolan, Frackowiak, & Robbins, 1996), and that posterior PFC receives most of the projections from PMC, and then projects to more anterior regions (Barbas & Pandya, 1987, 1989). Finally, this notion of increasingly abstract levels of plan or internal task context is consistent with the progression from posterior to anterior seen within the motor and premotor areas of the frontal cortex (Rizzolatti, Luppino, & Matelli, 1996).

Specialization of Active Memory Representations

An important issue both within the working memory literature and with regard to theories of PFC function is that of specialization along functional and/or representational dimensions. For example, Baddeley (1986) proposed that there are two separate working memory buffers: a *phonological loop* and a *visuospatial sketchpad*, which might itself be subdividable into object and spatial components. It has been proposed that this functional specialization reflects an underlying specialization in the brain systems subserving the different buffer systems (Gathercole, 1994). For example, there may be specific brain systems subserving verbal rehearsal (e.g., Broca's area and/or angular gyrus). There is also a well recognized segregation of processing of object and spatial information into ventral (temporal) and dorsal (parietal) streams within the PMC (Ungerleider & Mishkin, 1982), that may correspond to the two subdivisions of Baddeley's visuospatial sketchpad. At a somewhat more general level, Shah and Miyake (1996) have found evidence consistent with the idea of separable spatial and verbal capacities.

Specialization may also play a role in PFC organization and function. Arguments in the literature have centered around two dimensions along which PFC may be organized: functional and content-based. However, we argue that it is difficult to draw a clean distinction between function and content. Indeed, a basic principle of neural networks is that processing and knowledge (content) are intimately intertwined. For example, the functional distinction between memory (in the dorsolateral PFC) and inhibition (in the orbital areas; Fuster, 1989; Diamond, 1990), could also be explained by a content-based distinction in terms of the representation of affective, appetitive and social information in the orbital areas, which might be more frequently associated with the need for behavioral inhibition. Similarly, the functional dissociation between manipulation (in dorsolateral areas) and maintenance in (ventrolateral areas; Petrides, 1996) may be confounded with the need to represent sequential order information in most tasks that involve manipulation. To further complicate the issue, one type of functional specialization can often give rise to other apparent functional specializations. For example, we have argued that the memory and inhibitory functions ascribed to PFC may both reflect the operation of a single mechanism (i.e., inhibition can result from maintained top-down activation of representations which then inhibit other competing possibilities via lateral inhibition; Cohen & Servan-Schreiber, 1992; Cohen et al., 1996).

Given these problems, we find it more useful to think in terms of the computational motivations described above in order to understand how the PFC is specialized (i.e., in terms of the tradeoff between active maintenance and distributed representations). Thus, we support the idea that the PFC is specialized for the function of active maintenance. Consequently, representations within the PFC must be organized by the content of the representations that are maintained. A number of neurophysiological studies have suggested a content-based organization that reflects an anterior extension of the organization found in the PMC, with dorsal regions representing spatial information (Funahashi, Bruce, & Goldman-Rakic, 1993) and more ventral regions representing object or pattern information (Wilson, Scalaidhe, & Goldman-Rakic, 1993). Other content dimensions have also been suggested, such as sequential order information (Barone & Joseph, 1989), and "dry" cognitive vs affective, appetitive and/or social information (Cohen & Smith, 1997). However, the data does not consistently support any of these ideas. For example, Rao, Rainer, and Miller (1997) have recorded more complex patterns of organization in neurophysiological studies, with significant degrees of overlap and multimodality of representations. Recent findings within the human neuroimaging literature are also confusing, as early reports that indicated distinctions in the areas activated by verbal vs object and verbal vs spatial information (e.g., Smith, Jonides, & Koeppel, 1996) have not been reliably replicated (as reported in a number of recent conference proceedings and in unpublished data from our lab).

In light of this data, we suggest that the PFC may be organized according to more abstract, multimodal, and less intuitive dimensions that have been considered to date (i.e., that do not correspond simply to sensory modalities or dimensions). This seems likely, given the relatively high-level position of the PFC in the processing hierarchy (see discussion above), which would give it highly processed multi-modal inputs.

Further, this type of input may interact with the learning mechanisms and other constraints on the development of representations within PFC. For example, we have shown that task demands and training parameters (i.e., blocked vs interleaved exposure) can play an important role in determining whether a simulated PFC develops uni- or multimodal representations of object and spatial information (Braver & Cohen, 1995).

Example Working Memory Tasks

Earlier, we provided an example of how the mechanisms we have proposed are engaged in a simple working memory task (the AX-CPT). Here, we consider how they may come into play in two tasks that are commonly used to measure working memory capacity, and contrast them with ones that are thought to *not* involve working memory. The verbal working memory span task (Daneman & Carpenter, 1980) involves reading aloud a set of sentences and remembering the final words from each sentence for later recall. Thus, these final words must be maintained in the face of subsequent processing, which makes this task heavily dependent on the robust PFC active maintenance mechanisms. A spatial version of this task (Shah & Miyake, 1996) involves identifying letters presented at different non-upright orientations as being either normal or mirror-reversed, which appears to require some amount of mental rotation to the upright orientation, while remembering the orientations of the letters for later recall. This mental rotation requires the driving of PMC-based visual transformations (learned over extensive experience seeing visual transformations such as rotation, translation, etc.) in a task-relevant manner, presumably via actively maintained PFC top-down biasing. Further, the orientation information must be maintained in the face of subsequent processing of the same kinds of information, which again requires robust PFC active maintenance.

These working memory span tasks have been contrasted with others that are not considered to involve working memory. For example, the verbal working memory span task has been compared with a simple digit span task, which presumably only requires active maintenance, but not controlled processing. Behaviorally, the verbal working memory span task is better correlated with other putative verbal working memory tasks that also involve controlled processing, compared to this simple digit span task (Daneman & Merikle, 1996). The other half of this argument has been made in the case of a spatial equivalent of the digit span task (which involved remembering the orientations of a set of arrows), that was significantly correlated with a simple visual processing task, while the spatial working memory span measure was not (Shah & Miyake, 1996). See Engle et al. (this volume) for a more detailed discussion of this issue and other relevant experimental results.

Another example, involving the use of the HCMP system, is the comprehension of extended written passages. Because of limited capacity in the PFC active memory system, it is likely that some of the representations activated by the comprehension of prior paragraphs are encoded only within the HCMP, and must be recalled as necessary during later processing (e.g., when encountering a reference like, “this would be impossible, given Ms. Smith’s condition,” which refers to previously introduced information that may not have remained active in the PFC). The idea is that this later reference can be used to trigger recall of the previous information from the HCMP, perhaps with the addition of some strategic activation of other relevant information which has persisted in the PFC (e.g., the fact that Ms. Smith lives in Kansas). A successful recall of this information will result in the activation of appropriate representations within the PFC and PMC, which combined with the current text results in comprehension (e.g., Ms. Smith was hit by a tornado, and can’t come into work for an important meeting). In contrast with theories that draw a strong distinction between active memory and HCMP weight-based memories (e.g., Moscovitch & Winocur, 1992), we think that a typical cognitive task analysis may not distinguish between these types of memory in many situations, making the generic working memory label more appropriate for both. Finally, Young and Lewis (this volume) present what appears to be a roughly similar role for rapid learning in their theory of working memory, and Ericsson and Delaney (this volume) describe relatively long-lasting working memory representations which would seem to involve the HCMP (as well as the effects of extensive experience on

underlying cortical representations).

Answers to Theoretical Questions

This section summarizes our answers to a set of eight basic questions about our theory of working memory. The questions are summarized by the section headers, and Table 2 provides a concise summary of our answers to these questions.

Basic Mechanisms and Representations in Working Memory

Active maintenance (for which the PFC is specialized), rapid learning (for which the HCMP is specialized), and controlled processing (biasing and binding based on these) are the basic mechanisms of working memory in our account. Controlled processing emerges from the interactions between all three primary brain systems (PFC, HCMP, PMC), but is most strongly influenced by the PFC and HCMP. For purposes of comparison, we describe our basic mechanisms in terms of standard memory terminology of encoding, maintenance, and retrieval:

Encoding: Due to slow learning, the cortical systems (PFC and PMC) have relatively stable representational capability. Thus, encoding in these systems relies on the selection and activation (via constraint satisfaction processing operating over experience-tuned weights) of those pre-existing representations that are most relevant in a particular context. In the HCMP, encoding involves the rapid binding together of a novel conjunction of the representations active in the rest of the brain. An important influence on this process, and a critical component of controlled processing, is the strategic activation (under the influence of the PFC) of representations that influence HCMP encoding in task-appropriate ways (e.g., activating distinctive features during elaborative encoding in a memory task).

Maintenance: Only the PFC is thought to be capable of sustaining activity over longer delays and in the face of other potentially interfering stimuli or processing. However, under conditions of shorter delays and the absence of interference, PMC can exhibit sustained active memories (Miller et al., 1996). We include in our definition of the PFC the frontal language areas which have been shown to be active in neuroimaging studies involving active memory as discussed above. For example, considerable evidence supports the idea that maintenance in this system is implemented by an *phonological loop* (Baddeley & Logie, this volume; Baddeley, 1986), which may involve more highly specialized mechanisms than those hypothesized to exist in other areas of PFC. We do not think that the HCMP maintains information in an active form, but rather through rapid weight changes made during encoding. These weight-based HCMP memories can persist over much longer intervals than active memories (c.f., Ericsson & Delaney, this volume).

Retrieval: For active memories, retrieval is not an issue, but for HCMP weight-based memories, retrieval typically requires multiple cues to trigger a particular hippocampal memory (due to its conjunctive nature). As with encoding, the strategic activation of such cues constitutes an important part of controlled processing.

As for the nature of the representations in our model of working memory, we have characterized the distinctive properties of representations in each of the three main brain systems (see Table 1 and Figure 2). However, since we do not adhere to a buffer-based or any other distinct substrate view of working memory, this question is difficult to address. Essentially, the space of possible different representations for working memory is as large as the space of all representations in the neocortex and hippocampus, since any such representation could be activated in a controlled manner, thus satisfying our definition of working memory.

However, we think that brain systems specialized for language may provide an exceptionally powerful and general-purpose representational system for encoding arbitrary information, and are likely to be used to encode even superficially non-verbal information. Similarly, it may also be that abstract spatial and/or numerical representations are useful for encoding relational and perhaps temporal information.

The Control and Regulation of Working Memory

In our theory, control results from the biasing function of the PFC, and the binding function of the HCMP. These systems, in turn, are regulated by each other, the PMC, and ascending brainstem neuromodulatory systems. Thus, control and regulation are interactive and distributed phenomena, that involve all parts of the system. While these interactions are necessarily complex, it is possible to identify characteristic contributions made by each component of the system. The PFC plays a dominant role in controlled processing by virtue of its characteristic features: its ability to maintain activation over time; the flexible and rapid updating of representations due to their combinatorial and active nature; and its position high in the cortical processing hierarchy. Note that unlike models which separate control (e.g., a central executive) from active storage (e.g., buffers), active maintenance plays a central role in control in our model.

Representations within the PFC are themselves subject to the influence of processing within the PMC and HCMP, by way of specialized control mechanisms that regulate access to the PFC. As described above, we suggest that the midbrain dopamine (DA) system provides an active gating of PFC representations, controlling when they can be updated, and protecting them from interference otherwise. We think that the PFC, together with the PMC and possibly the HCMP, controls the firing of the DA gating signal, through descending projections. Furthermore, we assume that these projections are subject to learning, so that the PFC and PMC can learn how to control the gating signal through experience.

The Unitary vs Non-Unitary Nature of Working Memory

We take the view that working memory is not a unitary construct — instead we suggest that it is the combination of active memory, rapid learning, and emergent controlled processing operating over distributed brain systems. Instead of the moving of information from long-term memory into and out of working memory buffers, we think that information is distributed in a relatively stable configuration throughout the cortex, and that working memory amounts to the controlled activation of these representations. As we noted at the outset, this view shares some similarities with the view of working memory offered by production system accounts (e.g., ACT — Anderson, 1983; Lovett et al., this volume). However, it does not include the structural distinction between declarative and procedural knowledge assumed by such accounts.

This non-unitary view is consistent with findings like those of Shah and Miyake (1996), who found no significant correlation between an individual's verbal and spatial working memory capacities. We would further predict that working memory capacity will vary along a variety of dimensions, depending on the quality of the relevant PMC and PFC representations developed over experience (c.f., Ericsson & Delaney, this volume). However, working memory is also affected by more domain-general, controlled processing mechanisms (such as those supported by brainstem neuromodulatory systems), so that some characteristics of working memory function might exhibit more unitary-like features (c.f., Engle et al., this volume). Thus, the actual performance of a given subject, under a given set of task conditions, will depend on a combination of both domain specific and more general factors.

The Nature of Working Memory Limitations

The presence of capacity limitations seems to be one of the few points about which there is consensus in the working memory literature. Despite this agreement, there is relatively little discussion of *why* such limitations exist. Are these the unfortunate by-product of fundamental limitations in the underlying mechanisms

(e.g., insufficient metabolic resources to sustain additional mental activity), or do they reflect some more interesting computational constraint? We adhere to the latter view. We believe that capacity limitations in working memory reflect a tradeoff between two competing factors: the accessibility, and wide-spread influence of PFC representations — necessary to implement its biasing function as a mechanism of control — and the need to constrain the extent of activation throughout the PMC, to avoid “runaway” activity, and promote focused and coherent processing. We assume that this tradeoff is managed by inhibitory mechanisms, that constrain spreading activation, and prevent the runaway activity that would otherwise result from the positive feedback loops within the cortex. This is particularly important in the PFC, because of its widespread and influential projections to the rest of the brain. We have begun to explore this possibility in explicit computational modeling work (Usher & Cohen, 1997). This account emphasizes the potential benefits of what otherwise might appear to be arbitrary limitations (c.f., Lovett et al., this volume). Note that Young and Lewis (this volume) and Schneider (this volume) present functional motivations similar to our own.

As we stated above, we think there are both domain specific and more general contributions to working memory function. Similarly, there are likely to be both experience dependent and genetically-based contributions. Further, it is likely that there are interactions between these factors. For example, extensive experience will produce a rich and powerful set of domain-specific representations that support the ability to encode more domain-specific information both in active memory and via rapid learning in the HCMP. However, this is unlikely to affect more general factors (e.g., neuromodulatory function, or the overall level of inhibition within the PFC). This is consistent with the general lack of cross-domain transfer from experience-based working memory capacity enhancements as discussed in Ericsson and Delaney (this volume), and with the relatively domain general limitations observed by Engle et al. (this volume).

The role of experience-based learning (which is an important component of our overall model) in enhancing domain specific working memory capacity can be illustrated by considering the following phases of experience:

Novel phase: HCMP is required to store and recall novel task-relevant information, so that capacity is dominated by the constraint of having only one HCMP representation active at a time, with significant controlled processing required to orchestrate the use of this information with ongoing task processing. This is like the first time one tries to drive a car, where complete attention is required, everything happens in slow serial order, and many mistakes are made.

Weak phase: PFC is required to bias the weak PMC representations underlying task performance, so that capacity is dominated by the relatively more constrained PFC. Thus, it is difficult to perform multiple tasks during this phase, or maintain other items in active memory. This is like the period after several times of driving, where one still has to devote full attention to the task (i.e., use PFC to coordinate behavior), but at the basic operations are reasonably familiar and some can be performed in parallel.

Expert phase: Weights within the PMC have been tuned to the point that automatic processing is capable of accomplishing the task. Since the PMC representations are relatively more embedded, they can happily coexist with activity in other areas of PMC, resulting in high capacity. This is the case with expert drivers, who can carry on conversations more effectively than novices while driving. Note that slow improvements within this phase occur with continued practice, resulting in experience-based differences in strength and sophistication of underlying representations, which contribute to individual differences in capacity and performance. This is true in the PFC as well, where fewer active representations need be maintained if a more concise (e.g., “chunked”) representation has been learned over experience.

The Role of Working Memory in Complex Cognitive Activities

Complex cognitive activities involve controlled processing and thus, by our definition, involve working memory. According to our account of the roles of the PFC (biasing) and HCMP (binding), controlled processing occurs under conditions of temporally extended and/or novel tasks, and in cases which require coordinated processing among multiple systems. Typically, complex tasks involve the temporally-extended coordination of multiple steps of processing, often in novel combinations and situations, and the storage of intermediate products of computation, subgoals, etc. Active memory together with the controlled encoding and retrieval of HCMP memories can be used to retain the intermediate results of these processing steps for subsequent use.

We have yet to apply our model to specific complex tasks, as we have yet to produce satisfactory implementations of the entire set of neural systems that would be required. Our overarching goal in developing such models is the ability to account for complex task performance without resorting to a homunculus of one form or another. While many accounts of executive control remain purely verbal and are obviously susceptible to the homunculus problem, even mechanistically explicit accounts of complex task performance in production system architectures (e.g., Young & Lewis, this volume; Lovett et al., this volume) have a hidden homunculus in the form of the researcher who builds in all the appropriate productions in order to enable the system to solve the task. As we discuss in greater detail below, our current modeling efforts are focused on developing learning mechanisms that would give rise to a rich and diverse palette of PFC representations (and corresponding PMC subsystems), which should be capable of performing complex tasks without resorting to a homunculus of any form.

The Relationship of Working Memory to Long-Term Memory and Knowledge

We view working memory as being the active portion of long-term memory, where long-term memory refers to the entire network of knowledge distributed throughout the cortex, HCMP, and other brain systems. As noted earlier, this is similar to production system theories of working memory (such as ACT — Anderson, 1983; Lovett et al., this volume). However, we also specify that the term working memory only applies to those representations that are activated as a result of controlled processing. Thus, it is possible to have active representations that exist outside of working memory (c.f. Cowan, this volume; Engle et al., this volume, for similar views). Because of this intimate relationship between working memory and long-term memory, we expect working memory to be heavily influenced by learning in the long-term memory system (see the discussion in the capacity section above and Ericsson & Delaney, this volume).

We do not think that all components of long-term memory are equally likely to be represented in working memory. As discussed previously, language provides a particularly useful means of encoding arbitrary information, and is thus heavily involved in working memory. In contrast, more embedded, low-level sensory and motor processing is less likely to come under the influence of controlled processing, and is not typically considered to be involved in working memory. Thus, the general level of accessibility associated with a given brain system is correlated with the extent to which it is likely to be involved in working memory. As discussed earlier, this means that more *declarative* or *explicit* long term knowledge is likely to be involved in working memory, whereas *implicit* or *procedural* knowledge is more associated with automatic processing.

The Relationship of Working Memory to Attention and Consciousness

Working memory, attention, and consciousness are clearly related in important ways. We view the underlying constraint that gives rise to attention as resulting from the influence of competition between representations, implemented by inhibitory interneurons throughout the cortex (and possibly also by subcortical mechanisms in the thalamus and basal ganglia). This inhibition provides a mechanism that causes some things to be ignored while others are attended to, and is a critical aspect of attention that is not strictly part

of working memory. However, assuming this constraint, controlled processing plays an important role in determining what is active in a given context (and via competition and inhibition, also what is ignored). Thus, working memory and attention are related in that they are both defined in part by the mechanisms that determine what is activated in a particular context.

Consciousness is also related to both working memory and attention. As stated previously, we view conscious experience as reflecting the outcome of global constraint satisfaction processing, where salience is a function of the degree of influence over this process attributable to a given representation. Thus, systems which are globally accessible like the PFC and HCMP are also highly influential, and thus likely to dominate conscious experience. This means that the controlled processing-based activation (attention) mediated by these systems is most relevant for consciousness, and that the contents of conscious experience are likely to reflect that of working memory as we have defined it (see also previous section and Kinsbourne, 1997).

The Biological Implementation of Working Memory

Our model is based largely on biological data, and its neural implementation has been described both in terms of basic properties such as activation, inhibition, and learning, and in terms of the interactions of the specialized brain systems described above (PFC, HCMP, PMC). By virtue of these biological foundations, there is a wealth of data which is consistent with our model, from anatomy and physiology to neuroimaging and neuropsychological work. We will just review some of the most relevant data here.

With respect to the involvement of the PFC in working memory tasks, our lab has focused on neuroimaging and schizophrenic patient performance on the AX-CPT task described in the introduction. By making the target A-X sequence very frequent (80%), and the delay between stimuli longer (5 secs), we predicted that schizophrenic patients suffering an impairment of PFC function would make a relatively large number of false alarms to B-X sequences (where 'B' is any non-'A' stimulus) due to a failure of PFC-mediated working memory for the prior stimulus. This was confirmed, with unmedicated schizophrenic patients showing the predicted increase in false alarms, while medicated schizophrenics and control subjects did not (Servan-Schreiber, Cohen, & Steingard, in press). In addition, neuroimaging of healthy subjects performing the AX-CPT showed that PFC increased activity with increases in delay interval (Barch, Braver, Nystrom, Forman, Noll, & Cohen, 1997). Neuroimaging during N-back performance revealed that PFC activity also increases with working memory load (Braver, Cohen, Nystrom, Jonides, Smith, & Noll, 1997b), and is sustained across the entire delay period (Cohen et al., 1997). These data together with other consistent findings from monkey neurophysiology (e.g., Fuster, 1989; Miller et al., 1996), frontally-damaged patients, (e.g., Damasio, 1985) all support the idea that the PFC is critically important for working memory. Also, Engle et al. (this volume) discuss the importance of the PFC in working memory.

With respect to the role of the HCMP in working memory, it has long been known that the HCMP is critical for learning new information (Scoville & Milner, 1957; see Squire, 1992; McClelland et al., 1995, for recent reviews). Recent neuroimaging data suggests that the controlled encoding and retrieval of information in the HCMP depends on interactions between the PFC and the HCMP (e.g., Tulving, Kapur, Craik, Moscovitch, & Houle, 1994). Further, patients with frontal lesions show impaired ability to perform strategic encoding and retrieval on standard memory tests (Gershberg & Shimamura, 1995). All of this is consistent with our view that PFC and HCMP interactions are important for the controlled processing of memory storage and retrieval, which can be used as a non-active form of working memory.

Recent Developments and Current Challenges

Our theory of working memory represents an attempt to understand this construct in terms of a set of biologically-based, computational mechanisms. This has resulted in a novel set of functional principles

- (1) **Basic Mechanisms and Representations in Working Memory:**
Active memory and rapid learning via controlled processing, as implemented by the pre-frontal cortex (PFC), hippocampus and related structures (HCMP), and the posterior perceptual & motor cortex (PMC). Representations, distributed throughout system, are encoded by controlled activation; maintained by robust PFC mechanisms and weight-based HCMP learning; and retrieved in the case of HCMP by controlled activation of cues. Verbal and perhaps spatial and/or numerical representations are especially useful ways of encoding.
- (2) **The Control and Regulation of Working Memory:**
Working memory is not separated from control, since controlled processing and active memory are intimately related. Control is also not centralized, emerging instead from interactions between different brain systems. PFC plays important role due to: robust maintenance capabilities; flexible and rapid updating of representations; position at the top of the cortical processing hierarchy (with HCMP).
- (3) **The Unitary vs Non-Unitary Nature of Working Memory:**
Working memory is not unitary: consisting of active memory, rapid learning and controlled processing, and distributed over several brain systems. Common use of controlled processing mechanisms may contribute a unitary-like component to performance.
- (4) **The Nature of Working Memory Limitations:**
Two mechanisms: inhibition, and interference. PFC has greater inhibition to promote coherent processing, thus lower capacity. Capacity has domain specific and general components (see 3), and corresponding experience and genetic bases. Capacity is highly dependent on amount and type of controlled processing necessary, and efficiency of underlying representations learned over experience.
- (5) **The Role of Working Memory in Complex Cognitive Activities:**
Working memory is critical, as such activities are defined by the involvement of controlled processing, and require active memory/rapid learning to maintain intermediate results. Distributed brain systems are involved as relevant in particular tasks, with more common involvement of PFC and HCMP.
- (6) **The Relationship of Working Memory to Long-Term Memory and Knowledge:**
Working memory is largely just the active portion of long-term memory, which is itself distributed over many brain areas. More globally accessible systems and those that provide particularly useful representations (e.g., language) are more likely to be involved in working memory, leading to a bias towards *declarative* or *explicit* representations instead of *implicit* or *procedural* ones.
- (7) **The Relationship of Working Memory to Attention and Consciousness:**
Working memory is the subset of representations attended to by virtue of controlled processing. Attention also refers to a constraining mechanism (inhibition), and can be influenced by automatic processing. Consciousness reflects the global constraint satisfaction process, which is disproportionately influenced by controlled-processing systems. Thus, the contents of conscious experience are likely to reflect that of working memory.
- (8) **The Biological Implementation of Working Memory:**
Our model is based on the biology, including neural-level properties like activation, inhibition, and learning, and a computational account of specialized brain system function, including the PFC, HCMP, and PMC. A large amount of empirical data from patients, neuroimaging, neurophysiology, and animal studies is consistent with our model.

Table 2: Summary of our answers to the eight designated questions.

that explain many of the same phenomena as traditional working memory constructs, but in a manner that contrasts with existing theoretical ideas in important ways. Our existing computational work has instantiated and validated a number of aspects of our theory, including: the graded nature of controlled processing (Cohen et al., 1990); the ability of PFC representations to bias subsequent processing (Cohen & Servan-Schreiber, 1992); the role of PFC in active maintenance (Braver et al., 1995); and the role of the HCMP in rapid learning (O'Reilly, Norman, & McClelland, 1998; O'Reilly & McClelland, 1994). However, we have not yet implemented a computational model that captures all of our ideas regarding working memory and controlled processing. Moreover, there are a number of important issues raised by our overall model that have not been properly addressed in our prior work, and which form the current focus of our research.

These unresolved issues can be described at two general levels of analysis — one level involves the development of better models of each of the individual brain systems that play a role in our overall model (PMC, PFC, HCMP), and the other level involves characterizing the nature of interactions between these systems. Obviously, the latter effort depends critically on the success of the former, which is where we have been primarily focused. Underlying the entire endeavor are issues of the computational sufficiency necessary to learn and perform temporally extended controlled-processing tasks using neural network models.

Models of the PMC, PFC, and HCMP

Because it represents the canonical form of cortical processing, our model of the PMC lies at the foundation of the other models. We have recently made important advances in characterizing the nature of processing and learning in cortex, and now have a computational framework (called *Leabra*; O'Reilly, 1996b, 1996a) which contains all of the basic mechanisms and properties required by our model. In particular, the *Leabra* framework combines recurrence, inhibition, and integrated error-driven and Hebbian learning mechanisms in a simple, principled, robust, and biologically plausible manner. While these properties have been implemented separately in different models before, *Leabra* integrates them all in a unified, coherent framework. Our model of the HCMP system is relatively well-developed conceptually, and parts of it have been modeled at a very detailed level (O'Reilly & McClelland, 1994). Recently, we have created a complete HCMP model using the *Leabra* framework (O'Reilly et al., 1998), and have modeled the recollective contribution to many of the basic recognition memory phenomena (list length, list strength, etc).

It is the PFC which has received most of our recent theoretical attention, building on previous work that establishes a basic framework for understanding the computational role of PFC in controlled processing. There are two primary threads: 1) The role of a dopamine (DA) mediated active gating mechanism as described previously; and 2) The nature of PFC representations necessary to accomplish controlled processing in complex tasks. We have recently implemented a DA-like gating mechanism in a computational model of PFC, and shown that it can successfully account for all of the phenomena accounted for by our previous models, while making new predictions based on the phasic nature of the DA gate (Braver et al., 1997a). This model is being extended to more complex tasks that will better test the gating mechanism by requiring both rapid updating and sustained maintenance in the face of interference. Our current work on the PFC representations is investigating the tradeoff between distributed representations and active maintenance as a function of different task demands.

Reward-based Learning, Goals, and the PFC

One of the most important unresolved challenges to models of working memory (and cognition more generally) is specifying the mechanistic basis of executive control (controlled processing) in a way that does not resort to a homunculus. While we have generally characterized our view of how controlled processing emerges from constraint satisfaction and the specialized properties of the PFC and HCMP, actually showing that this works in real tasks remains a challenge. We think that the solution to this problem requires a powerful learning mechanism which is capable of developing something like the “productions” that underlie

the performance of complex cognitive tasks (thus avoiding the hidden homunculus of the researcher who builds in the appropriate productions for each task). The following is one set of ideas regarding the nature of this learning mechanism, which emerges from a synthesis of our basic ideas about a DA-based mechanism for active gating, and the nature of the representations in the PFC.

These ideas can be motivated by thinking about the essential difference between human cognition, and that of even our closest primate relatives. It is obvious that language, abstraction, problem solving, and tool use are important behavioral differences. However, we suggest that these may all be facilitated by the ability to internalize, abstract, and chain together representations of reward (and punishment). In short, the real difference between humans and other primates may be that we can establish elaborate systems of internalized reward that motivate us to learn and engage in these more abstract behaviors, whereas other primates, who can learn impressively complex and abstract tasks, must nevertheless be constantly motivated by external forces (e.g., food, juice) to do so. Thus, instead of being a “pure” cognitive system divorced from all emotional or motivational concerns, the PFC may instead be centrally involved in the dirty business of motivation, emotion, pleasure, and pain (Davidson & Sutton, 1995; Bechara, Tranel, Damasio, & Damasio, 1996).

This observation proves tantalizing in the context of our ideas about the role of dopamine (DA) in the PFC. In particular, if the critical specialization of the PFC is that it has taken control over the DA system in order to regulate its own active maintenance function, then it is also in a position to take over and internalize the deployment of DA-mediated reinforcement. It is well known that DA plays a critical role in reinforcement-based learning (Schultz et al., 1993; Montague et al., 1996). If the activation of PFC representations correspond essentially to goals which are maintained in an active and relatively protected state in the absence of DA firing, then the act of satisfying a goal should simultaneously result in reinforcement and gating (i.e., the deactivation of that goal representation and the opportunity to activate a subsequent one). The firing of DA under PFC control would provide both, and the influence of this DA signal on learning should result in more effective elicitation and efficient execution of that goal in the future.

Further, we have argued above that the PFC has the capacity for the simultaneous representation of many levels of temporal extent and abstraction, which would be needed to account for the goal structures underlying complex human cognition. Since reward is under the descending control of the PFC itself, the need for external reward is reduced, allowing for the development of elaborate means (intervening goals) to accomplish remote and abstract ends. In contrast, other animals depend to a much greater extent on constant external input to drive the DA reward system, and thus cannot build these elaborate internal goal structures.

There are many different ways in which the internalized control of DA could be implemented in the PFC, but unfortunately little is known about the relevant biological details. Thus, we are using computational models to determine the relative advantages and disadvantages of different implementations. Another important implementational issue has to do with learning on the basis of actively-maintained, isolated representations like those in the PFC, which have a more discrete, binary character and thus do not appear to be amenable to the types of gradient-based learning mechanisms that work so well in the distributed, graded representations characteristic of the PMC. In summary, the specializations of the PFC will likely require specialized learning mechanisms, which are the focus of our current research.

Conclusion

In summary, our overall model of the brain systems underlying working memory, including the PMC, PFC, and HCMP, is still under construction, but we have a broad and compelling blueprint for future exploration. This model provides many examples where computational principles (e.g., basic tradeoffs) are used to understand biological properties, in ways that, while consistent with existing ideas in many cases, can achieve a new level of synthesis and clarity. We hope that this approach will continue to prove useful,

despite the inevitable revision of many of the specific ideas proposed herein.

Acknowledgments

We would like to thank Andrew Conway, Peter Dayan, Yuko Munakata, Ken Norman, Mike Wolfe, and all of the conference participants for useful comments. We were supported by NIH Grant MH47566-06, and R.O. was also supported by a McDonnell Pew Postdoctoral Fellowship at MIT in the Department of Brain and Cognitive Science.

References

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.
- Baker, S. C., Rogers, R. D., Owen, A. M., Frith, C. D., Dolan, R. J., Frackowiak, R. S. J., & Robbins, T. W. (1996). Neural systems engaged by planning: A PET study of the Tower of London task. *Neuropsychologia*, 34, 515.
- Barbas, H., & Pandya, D. N. (1987). Architecture and frontal cortical connections of the premotor cortex (area 6) in the rhesus monkey. *Journal of Comparative Neurology*, 256, 211–228.
- Barbas, H., & Pandya, D. N. (1989). Architecture of intrinsic connections of the prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology*, 286, 353–375.
- Barch, D. M., Braver, T. S., Nystrom, L. E., Forman, S. D., Noll, D. C., & Cohen, J. D. (1997). Dissociating working memory from task difficulty in human prefrontal cortex. *Neuropsychologia*, 35, 1373.
- Barone, P., & Joseph, J. P. (1989). Prefrontal cortex and spatial sequencing in macaque monkey. *Experimental Brain Research*, 78, 447–464.
- Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex*, 6, 215–225.
- Braver, T. S., & Cohen, J. D. (1995). A model of the development of object and spatial working memory representations in prefrontal cortex. *Cognitive Neuroscience Abstracts*, Vol. 2 (p. 95).
- Braver, T. S., Cohen, J. D., & McClelland, J. L. (1997a). An integrated computational model of dopamine function in reinforcement learning and working memory. *Society for Neuroscience Abstracts*, 23 (p. 775). San Diego, CA: Society for Neuroscience.
- Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997b). A parametric study of frontal cortex involvement in human working memory. *NeuroImage*, 5, 49–62.
- Braver, T. S., Cohen, J. D., & Servan-Schreiber, D. (1995). A computational model of prefrontal cortex function. In D. S. Touretzky, G. Tesauro, & T. K. Leen (Eds.), *Advances in neural information processing systems* (pp. 141–148). Cambridge, MA: MIT Press.
- Cohen, J. D., Braver, T. S., & O'Reilly, R. C. (1996). A computational approach to prefrontal cortex, cognitive control, and schizophrenia: Recent developments and current challenges. *Philosophical Transactions of the Royal Society of London Series B (Biological Sciences)*, 351, 1515–1527.
- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing model of the stroop effect. *Psychological Review*, 97(3), 332–361.
- Cohen, J. D., & O'Reilly, R. C. (1996). A preliminary theory of the interactions between prefrontal cortex and hippocampus that contribute to planning and prospective memory. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). Temporal dynamics of brain activity during a working memory task. *Nature*, 386, 604–608.
- Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, 99, 45–77.
- Cohen, J. D., & Smith, E. E. (1997). Response to Owen AM, Tuning in to the temporal dynamics of brain activation using functional magnetic resonance imaging (fMRI). *Trends in Cognitive Sciences*, 1, 124.

- Damasio, A. R. (1985). The frontal lobes. In K. M. Heilman, & E. Valenstein (Eds.), *Clinical neuropsychology* (pp. 339–375). New York: Oxford University Press.
- Damasio, H., Grabowski, T. J., & Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature*, 380, 499.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Daneman, M., & Merikle, P. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3, 422–433.
- Davidson, R. J., & Sutton, S. K. (1995). Affective neuroscience: the emergence of a discipline. *Current Opinion in Neurobiology*, 5, 217–224.
- Dehaene, S., & Changeux, J. P. (1989). A simple model of prefrontal cortex function in delayed-response tasks. *Journal of Cognitive Neuroscience*, 1, 244–261.
- Diamond, A. (1990). The development and neural bases of memory functions as indexed by the a-not-b task: Evidence for dependence on dorsolateral prefrontal cortex. In A. Diamond (Ed.), *The development and neural bases of higher cognitive functions* (pp. 267–317). New York: New York Academy of Science Press.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT/Bradford Press.
- Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1993). Dorsolateral prefrontal lesions and oculomotor delayed-response performance: Evidence for mnemonic 'scotomas'. *Journal of Neuroscience*, 13, 1479–1497.
- Fuster, J. M. (1989). *The prefrontal cortex. anatomy, physiology and neuropsychology of the frontal lobe*. New York: Raven Press.
- Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science*, 173, 652–654.
- Gathercole, S. E. (1994). Neuropsychology and working memory: A review. *Neuropsychology*, 8(4), 494–505.
- Gershberg, F. B., & Shimamura, A. P. (1995). Impaired use of organizational strategies in free recall following frontal lobe damage. *Neuropsychologia*, 33, 1305.
- Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. *Handbook of Physiology — The Nervous System*, 5, 373–417.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short term memory. *Neural Computation*, 9, 1735–1780.
- Kahneman, D., & Treisman, A. (1984). *Changing views of attention and automaticity*. San Diego, CA: Academic Press, Inc.
- Kinsbourne, M. (1997). What qualifies a representation for a role in consciousness? In J. D. Cohen, & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 335–355). Mahway, NJ: Lawrence Erlbaum Associates.
- McClelland, J. L. (1993). The GRAIN model: A framework for modeling the dynamics of information processing. In D. E. Meyer, & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 655–688). Hillsdale, NJ: Lawrence Erlbaum Associates.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419–457.

- Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science*, 263, 520–522.
- Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, 16, 5154.
- Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience*, 16, 1936–1947.
- Moscovitch, M., & Winocur, G. (1992). The neuropsychology of memory and aging. In T. A. Salthouse, & F. I. M. Craik (Eds.), *The handbook of aging and cognition*. Hillsdale, NJ: Erlbaum.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge: Harvard University Press.
- O'Reilly, R. C. (1996a). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, 8(5), 895–938.
- O'Reilly, R. C. (1996b). *The leabra model of neural interactions and learning in the neocortex*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA.
- O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a tradeoff. *Hipocampus*, 4(6), 661–682.
- O'Reilly, R. C., Norman, K. A., & McClelland, J. L. (1998). A hippocampal model of recognition memory. In M. I. Jordan (Ed.), *Advances in neural information processing systems 10*. Cambridge, MA: MIT Press.
- Petrides, M. E. (1996). Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Philosophical Transactions of the Royal Society of London, Series B.*, 351, 1455–1462.
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition* (pp. 55–85). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science*, 276, 821.
- Rizzolatti, G., Luppino, G., & Matelli, M. (1996). The classic supplementary motor area is formed by two independent areas. *Avances in Neurology*, 70, 45–56.
- Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & PDP Research Group (Eds.), *Parallel distributed processing. volume 2: Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.
- Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *The Journal of Neuroscience*, 13, 900–913.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20, 11–21.
- Seidenberg, M. (1993). Connectionist models and cognitive theory. *Psychological Science*, 4(4), 228–235.
- Servan-Schreiber, D., Cohen, J. D., & Steingard, S. (in press). Schizophrenic performance in a variant of the CPT-AX: A test of theoretical predictions concerning the processing of context. *Archives of General Psychiatry*.
- Shah, P., & Miyake, A. (1996). the separability of working memory resources for spatial thinking and language processing: an individual differences approach. *Journal of Experimental Psychology: General*, 125, 4.

- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London*, 298, 199–209.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127–190.
- Smith, E. E., Jonides, J., & Koeppel, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex*, 6, 11–20.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 195–231.
- Squire, L. R., Shimamura, A. P., & Amaral, D. G. (1989). Memory and the hippocampus. In J. H. Byrne, & W. O. Berry (Eds.), *Neural models of plasticity: Experimental and theoretical approaches*. San Diego, CA: Academic Press, Inc.
- Tulving, E., Kapur, S., Craik, F. I. M., Moscovitch, M., & Houle, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: Positron emission tomography findings. *Proceedings of the National Academy of Sciences, USA*, 91, 2016–2020.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *The analysis of visual behavior*. Cambridge, MA: MIT Press.
- Usher, M., & Cohen, J. D. (1997). Interference-based capacity limitations in active memory. *Abstracts of the Psychonomics Society*, Vol. 2 (p. 11).
- Wilson, F. A. W., Scalaidhe, S. P. O., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, 260, 1955–1957.
- Zipser, D., Kehoe, B., Littlewort, G., & Fuster, J. (1993). A spiking network model of short-term active memory. *Journal of Neuroscience*, 13, 3406–3420.