

A Preliminary Theory of the Interactions Between Prefrontal Cortex and Hippocampus that Contribute to Planning and Prospective Memory

Jonathan D. Cohen
*Carnegie Mellon University
University of Pittsburgh*

Randall C. O'Reilly
Carnegie Mellon University

This chapter addresses the neurobiological mechanisms that may underlie prospective memory. In our work, we have exploited the use of computational modeling techniques to help identify the role that specific brain structures play in cognition. In particular, we have used such techniques to characterize the function of prefrontal cortex and hippocampus in terms of specific processing mechanisms. This work suggests that an important function of prefrontal cortex is the representation and maintenance of *contextual information*—information that must be held in mind in such a form that it can be used to mediate an appropriate behavioral response. At the same time, our work supports the idea that an important function of the hippocampus is to rapidly establish novel associations, that can also be used to guide behavior. In our view, prospective memory reflects the interaction between these two systems, allowing established sequences of behavior to be associated with new conditions—in effect, providing a mechanism for planning.

Consider the following situation: You open the refrigerator one morning and discover that you are out of orange juice. You make a mental note to stop by the grocery store on the way home from work, and then “put it out of mind.” Five o'clock rolls around, and as you are leaving your office, you think to yourself: “Right, the grocery store,” and stop by to pick up the orange juice on your way home. The remarkable thing is that during the day you may not have thought about the errand at all. Somehow, you magically remember it, and manage to perform it at the appropriate time. We hypothesize that this reflects an interaction between the prefrontal cortex and hippocampal systems in the

following way. When you first recognize that you need to go to the grocery store, you decide to do so within a particular context—after work. We assume this process involves the activation of representations of the action plan (going to the grocery store) as well as the context in which it is to take place (e.g., the end of the work day). We assume, furthermore, that prefrontal cortex is responsible for eliciting these representations, and that the hippocampus is responsible for encoding an association between them. Once this is done, the activity of the representations fades. As the end of the day comes, the environment elicits representations—darkness outside, clock striking 5:00, and so on—that are closely associated with “end of the work day.” This activates the action plan by way of the hippocampal association that was established that morning. The action plan is then maintained by prefrontal cortex, and the behavior is executed.

This example illustrates how we believe that interactions between prefrontal cortex and hippocampus can support both the scheduling and elicitation of plans of action. Although planning is probably the most common reflection of the operation of the prefrontal cortex–hippocampal system in daily life, we believe this system is centrally involved in a number of standard laboratory tasks, which can be used to test hypotheses concerning this system's function. We will return to these in the general discussion.

In the sections that follow, we review both empirical and computational modeling work that bears on the function of prefrontal cortex and hippocampus. This work provides a basis for the assumptions, both implicit and explicit, present in the example. We begin by reviewing work that suggests prefrontal cortex plays an important role in representing and maintaining context representations. We then turn to the hippocampus, and review research indicating that it is responsible for encoding novel associations, and for using these to activate corresponding representations within association cortex. We conclude by considering how these two systems can be integrated to provide a mechanistic account of prospective memory, discuss how this account can shed light on traditional distinctions between controlled and automatic processes, and describe predictions our account makes for performance in a set of classical learning tasks.

FUNCTIONS OF PREFRONTAL CORTEX AND HIPPOCAMPUS

Functions of Prefrontal Cortex

Prefrontal cortex is the area of the human brain most significantly expanded relative to other animals. There is general consensus that prefrontal cortex is centrally involved in higher cognitive activities such as planning, problem solv-

ing and language. However, despite this consensus, there is little discussion of the specific information-processing mechanisms subserved by prefrontal cortex that contribute to these cognitive activities. Theorists have attributed many general functions to the frontal lobes, such as attention (Ferrier, 1886), abstract intelligence (Hitzig, 1874), and synthesis of percepts (Bianchi, 1922). Attention seems to have recently focused on two information-processing functions—inhibition and working memory (Fuster, 1989; Goldman-Rakic, 1987; Mishkin, 1964; Petrides & Milner, 1982). We briefly review the literature concerning these two functions of prefrontal cortex, and then describe recent work we have conducted using computational models of prefrontal cortex function, suggesting that these two functions reflect the operation of a single underlying information-processing mechanism that is responsible for representing and maintaining context information.

Inhibition. Traditionally, lesions of prefrontal cortex were associated with a clinical syndrome of behavioral disinhibition, in which subjects exhibit impulsive, socially inappropriate behavior (Hecaen & Albert, 1978; Stuss & Benson, 1984). Abnormalities of frontal cortex (mostly orbital) were implicated in clinical disorders involving forms of behavioral or cognitive disinhibition, such as attention deficit disorder with hyperactivity (Gorenstein, Mammato, & Sandy, 1989) and obsessive-compulsive disorder (Swedo et al., 1989). These phenomena were often cited as evidence that prefrontal cortex plays an important role in mediating socially appropriate behavior, by inhibiting compelling but inappropriate behaviors. Neuropsychological and neurodevelopmental data support this view: Patients with damage to prefrontal cortex exhibit perseverative behavior (i.e., a tendency toward prepotent but inappropriate responses) in tests such as the Wisconsin Card Sort Task (WCST; Grant & Berg, 1948) and Stroop Color-Word test (Stroop, 1935). Schizophrenics, who are believed to have deficits of prefrontal cortex, also exhibit prepotent response tendencies in a variety of tasks (see Cohen & Servan-Schreiber, 1992 for a review). Furthermore, Diamond and her colleagues (Diamond, 1990a; Diamond & Doar, 1989; Diamond & Goldman-Rakic, 1989) showed that human and monkey infants fail to inhibit prepotent response tendencies in developmental tasks such as the A-not-B. Diamond cited extensive data from lesion studies in adult monkeys and from developmental studies in human and monkey infants that use a variety of behavioral tasks (including object retrieval, visual paired comparisons, delayed response, and the A-not-B task). Results from these and many previous studies suggest that prefrontal cortex is directly involved in maintaining representations (e.g., the location of a hidden object) that are required to inhibit reflexive, dominant, or habitually reinforced behaviors (e.g., return to the most recently rewarded location) in order to attain a goal.

The performance deficits observed for infants and frontally lesioned mon-

keys on delay tasks are similar to those observed for adult frontal lobe patients on the WCST. In this task, subjects are presented with a series of cards containing figures that vary in shape, color, and number. They are asked to sort the cards into piles according to a rule that the experimenter has in mind (e.g., separate the cards by color). However, subjects are not told the rule for sorting; rather, they are given feedback for each card as to whether or not they sorted it properly. Normal subjects discover the rule quickly. Once they demonstrate they know it (i.e., by correctly sorting a certain number of cards in a row) the experimenter switches the rule, and the subject is required to discover the new rule. Patients with damage to the frontal lobes do poorly on this task (e.g., Milner, 1963; Nelson, 1976; Robinson, Heaton, Lehman, & Stilson, 1980). Although they are able to discover the first rule without much difficulty, they are unable to switch to a new one: They continue to sort according to the old rule. As in delay tasks, this is a failure to overcome a response pattern that was correct on previous trials. Furthermore, there are additional indications from these tasks that a specific failure to control behavior is involved in these tasks, as distinct from a disturbance in declarative, or short-term memory. In both the WCST and in delayed response tasks, some subjects show perseveratory behavior despite indications that they remember the relevant prior information. Thus, subjects in the WCST sometimes comment that they know their perseveratory response is incorrect even as it is carried out (Goldberg, Weinberger, Berman, Pliskin, & Podd, 1987). In the A-not-B task, some subjects look at the cued (new) location, while reaching for the old (incorrect) one (Diamond & Goldman-Rakic, 1989). These kinds of observations support a dissociation between declarative or short-term memory and the kinds of representations stored in prefrontal cortex that are needed to actually control the response. We return to this point in discussing the specific processing mechanisms subserved by prefrontal cortex.

In both the WCST and A-not-B task, subjects with poor prefrontal function are not impaired in their ability to learn the basic elements of the task. Rather, they are impaired in their ability to use an internal representation of context to override the effects of prior experience in the task. This characterization of frontal lobe function fits well with clinical descriptions of the *disinhibition syndrome* that often accompanies frontal lobe pathology (e.g., Stuss & Benson, 1984). It is also consistent with difficulties observed for frontal lobe patients in performing the Stroop task (Perret, 1974) and similar tasks in clinical use (e.g., the "go-no-go" paradigm) that require the subject to use task instructions to inhibit a dominant response tendency.

Working Memory. Studies of nonhuman primates provide strong evidence that certain areas of prefrontal cortex are centrally involved in *working memory*. Working memory is typically defined as the maintenance of active representations of task-relevant information needed to mediate an appropri-

ate response. Specific regions of prefrontal cortex have been implicated in memory for spatial and, more recently, object information. For example, in neurophysiological studies, Fuster (1980, 1985a, 1985b), Goldman-Rakic (1987) and others (e.g., Barone & Joseph, 1989) observed cells in prefrontal cortex that are specific to a particular stimulus and response, and that remain active during a delay between these. They argued that neural patterns of activity are maintained in prefrontal cortex that encode the temporary information needed to guide a response. Furthermore, structural and neuropharmacological lesions along the superior margin of the principal sulcus in the monkey produce spatially specific deficits in memory-guided, but not sensory-guided, performance. That is, deficits occur only when the animal must remember a target location over a delay prior to responding, but not when the location is marked by a continuous visual cue during the delay. Other regions of prefrontal cortex support memory for object-related information, such as the color or shape of a stimulus (Wilson et al., 1993). These authors and others (e.g., Damasio, 1979; Mishkin & Pribram, 1955; Passingham, 1985; Rosenkilde, 1979; Rosvold, Szwarcbart, Mirsky & Mishkin, 1961; Stuss & Benson, 1986) suggested that prefrontal cortex is needed to perform tasks involving delayed responses to stimuli.

Neuroimaging data from normal human subjects also provide support for the role of frontal cortex in working memory. A number of PET studies have shown activation of prefrontal cortex during a variety of tasks that appear to rely on working memory. Thus, studies of digit-span (Becker, Mintun, Diehl, DeKosky, & Dobkin, 1993; Grasby et al., 1993) and spatial stimuli (Jonides et al., 1993), of the ability to maintain information about the temporal order of sequential stimuli (Petrides et al., 1993), and of face recognition (Haxby et al., 1993) all demonstrate activation of prefrontal cortex—typically localized to Brodmann's area 46, which is one of the areas associated with working memory in nonhuman primates (e.g., Goldman-Rakic, 1987). Functional magnetic resonance imaging studies provide similar results, for both spatial (Blamire, McCarthy, Bloch, Rothman, & Shulman, 1993) and nonspatial (Cohen et al., 1994) stimuli.

Computer Simulation Models of Prefrontal Cortex Function: Processing of Context. We used computational models to address the function of prefrontal cortex. Based on this work, we proposed that inhibition and working memory reflect the operation of a single underlying processing mechanism that represents and maintains context information (Cohen & Servan-Schreiber, 1992). By context information, we mean information that must be held in mind in such a form that it can be used to mediate an appropriate behavioral response. By this definition, context information is relevant to but does not necessarily form part of the content of the actual response. This distinguishes context information from the kind of information traditionally thought to be

stored in short-term memory (i.e., recently presented information, the specific identity of which must soon be retrieved). The role of context information is to enhance and/or maintain information relevant to task performance in the pathways responsible for processing that information. Another way of stating this, that we find useful further on, is that context representations serve to bias processing in favor of information relevant to the task at hand.

According to our hypothesis, memory and inhibition reflect the operation of the context processing mechanism under different task conditions. When strong competing response tendencies must be overcome for appropriate behavior, then context mechanisms play an inhibitory role by selectively enhancing processing in the task-relevant pathway relative to processing in the competing pathway. When there is a delay between information relevant to a response and the execution of that response, then context mechanisms play a memory role, by actively maintaining that information, and supporting it against the cumulative effects of noise. In both cases, the context mechanism is still performing the same basic function: supporting representations necessary to perform the task, and that may be stored elsewhere in cortex, against sources of interference. In the case of inhibition tasks, the interference is from an explicit, competing process; in memory tasks, interference comes from the degrading effects of noise.

We used computer simulation models to demonstrate how a single mechanism for representing context can perform both of these functions (Cohen & Servan-Schreiber, 1992). These models were developed within the parallel distributed processing (PDP) framework (Rumelhart & McClelland, 1986), in which information is represented as patterns of activation over simple processing units, and information processing takes place through the spread of activation between modules made up of such units. We used these models to simulate quantitative aspects of human performance in a variety of attention and language tasks that involve overcoming interference to respond appropriately. In each model, prefrontal cortex is represented by a module responsible for representing and maintaining context information, supporting the processing of task-relevant information over sources of interference.

For example, one model simulates human subjects' ability to interpret the meaning of ambiguous words (such as *pen*) given the context of the sentence in which they occur (e.g., "In order to keep chickens you need a pen"). This model (Fig. 14.1) contains modules for representing the individual words of the sentence (the *input* module), the meaning that the subject might report for each (the *output* module), and the topic, or overall meaning of the sentence (the *discourse* module). Words in the sentence are presented to the model one by one. Bottom-up connections in the model allow each word to contribute to the activation of an appropriate discourse representation. Top-down connections allow this representation to influence the interpretation of subsequent words. Thus, when words such as *chicken* or *farmer* are presented,

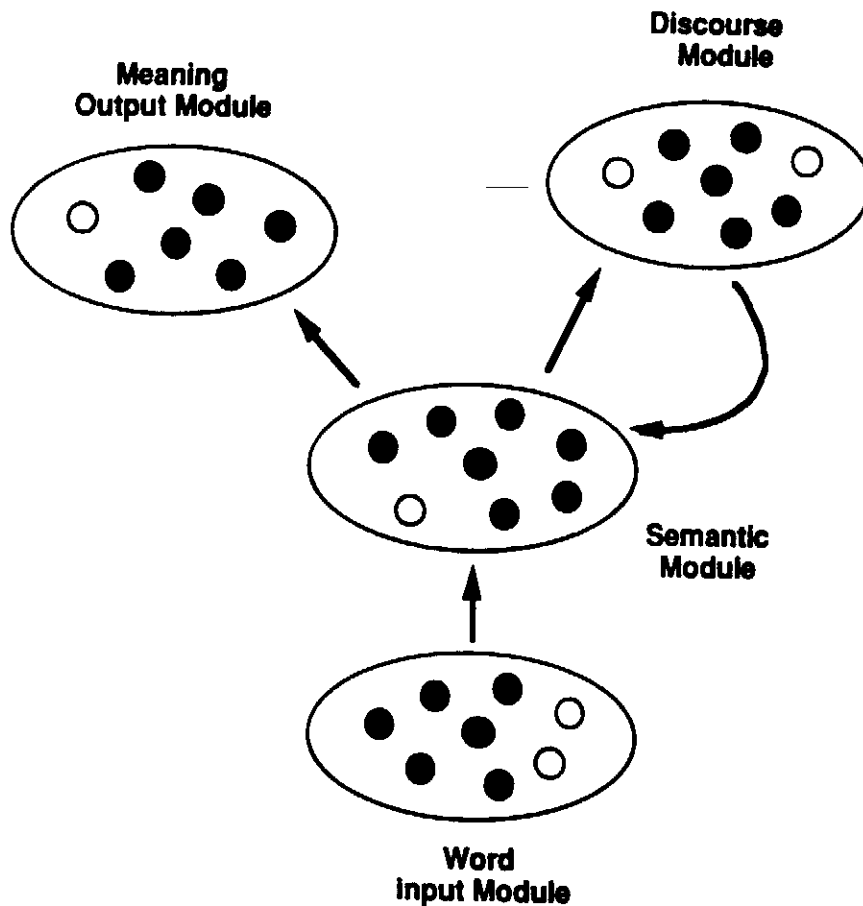


FIG. 14.1. Simulation model of a lexical disambiguation task. Patterns of activation over the units in the input module are assumed to represent the current sensory stimulus (e.g., the orthographic code for a written word), whereas the output module is assumed to represent the information necessary to generate an overt response (e.g., the phonological code needed to pronounce the meaning of the word). Note that the connections between the semantic and discourse modules are bidirectional, allowing stimuli to help establish a discourse representation, acting as context that governs the processing of subsequent stimuli.

they activate a discourse representation related to farming, which then helps to select the weaker, but contextually appropriate meaning of an ambiguous word like *pen* (i.e., fenced enclosure), rather than its more frequent meaning (writing implement). The discourse module serves to represent and maintain context information, and so functions as the context processing mechanism. In models of delay tasks—in which the correct response to a target stimulus depends upon prior information—the module identified with prefrontal cortex is responsible for maintaining this information against the cumulative effects of noise. By introducing a specific lesion to this module, we were able to quantitatively account for the performance in these tasks of patients with schizo-

phrenia, who are believed to have disturbances of frontal lobe function (Cohen & Servan-Schreiber, 1992). Most recently, we used the models to predict specific behavioral deficits that should arise in patients with differing degrees of disturbance to prefrontal cortex. These predictions were corroborated by a comparative study of first episode and multiepisode schizophrenics (Servan-Schreiber, Cohen, & Steingard, in submission).

In summary, neurophysiological and neuroimaging data suggest that prefrontal cortex plays an important role in both behavioral inhibition and working memory. Using computer simulation models, we showed that a single mechanism responsible for the representation and maintenance of context information can subserve both inhibition and memory in cognitive tasks. That is, rather than reflecting the operation of two separate mechanisms, memory and inhibition reflect the influence of two factors that characterize cognitive tasks—the relative dominance of task-relevant processes compared to competing ones, and the temporal extent over which these processes operate. The context processing mechanism is sensitive to both of these factors. Its primary function is to protect task relevant processes against the effects of interference, which can come either from competing, dominant processes, or from the effects of noise that accumulate over delays.

We suggest that prefrontal cortex may house such a context processing mechanism. This could account for the involvement of frontal cortex in goal-oriented activity; that is, the planning and sequencing of complex actions (e.g., Bianchi, 1922; Duncan, 1986; Luria, 1969; Shallice, 1982). The construction and maintenance of internal representations of context are an important component of these faculties. The actions associated with a particular goal may, in other situations, be relatively infrequent or weak behaviors. Such actions require the maintenance of an internal representation of the goal, or of goal-related knowledge to favor their execution (i.e., to suppress competing, possibly more compelling behaviors, or to sustain them over temporal delays). Thus, context representations maintained by prefrontal cortex can function as a plan that helps govern the execution of goal-directed behaviors. We think of plans as constituting just such sets of context representations, and wherever we use the term *plan* this is how we mean it—as a set of context representations maintained in prefrontal cortex that support the execution of actions directed toward a goal. We return to this point in a later section, when we discuss the interactions between prefrontal cortex and hippocampus that we believe underlie prospective memory.

In the section that follows, we discuss the information-processing functions of the hippocampus. Before we do so, it will be helpful to say one more thing about the use of the term context. We argue that the representation of context is an important, if not the central function of prefrontal cortex. However, those familiar with the animal learning literature on hippocampal function will know that hippocampus is thought to be necessary for context-sensitive learning.

This use of the term is slightly different than the way we use it. In this case, context refers to features of the environment, that can become associated with an identified stimulus and influence responses to it. As we use the term, with regard to prefrontal cortex function, it refers to internal representations that can influence processing. Of course, these are not unrelated concepts: In both cases, one source of information (whether external stimuli or internal representations) is influencing the processing of another. The relationship between these, and the involvement of hippocampus and prefrontal cortex in processing each will be taken up further on. For now, in order to be clear about which type of context we mean, we refer to external elements of context as the environment, whereas we continue to refer to internal context representations simply as context.

Complementary Memory Systems in Hippocampus and Neocortex

Evidence indicates that the hippocampus plays a unique role in the formation and retrieval of memories (see Squire, 1992 for a review). A theoretical perspective on the hippocampus proposed in McClelland, McNaughton, and O'Reilly (in press; also see McClelland, McNaughton, O'Reilly, & Nadel, 1992) provides a framework for understanding this unique role, especially in terms of the relationship between the hippocampus and the neocortex. This approach follows Marr (1971) and many others in proposing that the hippocampus is part of a dual memory system consisting of cortical and hippocampal components.

In brief, we propose that the cortex is responsible for developing a stable, efficient, and general framework for representing and encoding information, whereas the hippocampus is responsible for rapidly storing the contents of specific episodes or events, thought to correspond to particular states of the cortical system. The critical distinctions between these forms of learning are the temporal duration underlying the formation of the representations, and the relationship to other representations in the system. In the temporal dimension, the hippocampus must form and rapidly store its representations (in order to capture and bind together temporally coincident events in one memory), whereas the cortex must adapt slowly in order to capture the relevant general structure common to different samples of the environment.

How do representations relate to other memories or representations in the system? We assume that hippocampal representations must be kept distinct, so that specific events with many similar features (e.g., where you parked your car today as opposed to yesterday or the day before, etc.) can be recalled without undue interference from similar memories. For the cortex to exploit the shared structure present in ensembles of events and experiences it must assign similar internal representations to similar events, and to do so it must make the representations overlap. To summarize the differences between

these two systems, we label the hippocampal system as performing *rapid, arbitrary learning*, where the term arbitrary denotes that the representation does not capture structural relationships between the components of the representation. Conversely, the neocortical system performs *long-term, integrative learning*.

Experimental data in support of the idea that the hippocampus is responsible for a particular kind of memory is vast and goes back to the seminal work of Scoville & Milner (1957) who described the severe anterograde amnesia produced by bilateral removal of the hippocampus in patient HM. Squire (1992) presented a review of the literature, and argued for understanding the difference between the hippocampus and neocortex in terms of declarative versus nondeclarative memory. Although we argue that this terminology does not capture precisely the distinctions described above, there is a great deal of overlap between them. In particular, Squire (1992) described the role of the hippocampus as one of rapidly binding together representations in the cortex. This view is consistent with the data showing an impairment in amnesic subjects for both *paired-associate* and *episodic memory tasks*, among others (see Squire, 1992). Paired-associate learning requires the subject to rapidly form an association between two words that are otherwise unrelated, so as to be able to recall one of the words when given the other as a cue. Episodic memory (Tulving, 1983) refers to the ability to remember particular episodes or events of daily life that due to their transient nature, must be learned rapidly and require the binding together of the individual features of the event (e.g., visual appearance of the place, representations of what was happening at the time, etc.).

Examples of forms of learning that are spared in amnesic subjects provide evidence consistent with our characterization of neocortical learning. These include gradually acquired skills that emerge over several sessions of practice, such as the skill of tracing a figure viewed in a mirror (Milner (1966), and reading mirror-reversed print (Cohen & Squire, 1980). At a more cognitive level, amnesic subjects reliably show repetition priming for familiar items such as known words (Graf, Squire, & Mandler, 1984), but not for completely novel items such as in the paired-associate task described (Shimamura & Squire, 1989). Although it was shown that priming can be obtained in hippocampal subjects for novel material of various forms, this is due to the priming of the low-level perceptual representations and not the novel semantic-level information (Schacter, Delaney, & Merikle, 1990). The neocortical representations of familiar things can be facilitated by recent exposure (priming), but the neocortex cannot rapidly form novel representations. Because the visual system is capable of representing novel material without forming new representations, priming is seen at this level but not at a higher level. These results support the characterization of the neocortex as a slow learning system.

Insights as to why the brain might have two complementary learning sys-

tems can be obtained from the successes and failures of neural-network or connectionist models of learning (Rumelhart, Hinton & Williams, 1986). For example, our view of cortical learning and representation is captured by models that develop representations of the overall structure of experience through a gradual learning process, relying on repeated samples of inputs drawn from an environment (White 1989). However, these same networks exhibit *catastrophic interference* (McCloskey & Cohen, 1989) with what is already stored in the system when they are forced to rapidly learn new associations. There are many reasons to think that this result reflects a fundamental trade-off between long-term integrative learning and rapid, arbitrary learning (see McClelland et al., in press).

One of the consequences of a dual memory system of this form is that certain kinds of information (conjunctive information that must be quickly learned) are learned and initially stored in the hippocampus. Consequently, this information must also be retrieved from the hippocampus. Thus, the storage and retrieval properties of the hippocampus will play a large role in tasks that require the rapid storage of conjunctive information. As we discuss in more detail later, prospective memory relies on storing novel conjunctions of various cortical representations, and is therefore partially dependent on the hippocampus. Thus, we consider the properties of storage and recall in the hippocampus.

Pattern Separation and Completion in the Hippocampus: A Trade-off. Pattern separation is the mechanism by which the hippocampus is able to store conjunctive representations rapidly and without undue interference, and pattern completion is the mechanism by which previously stored representations in the hippocampus are recalled. O'Reilly and McClelland (1994) examined these properties using an analytical model that incorporated the broad anatomical and physiological features of the hippocampus. In particular, we focused on the intrinsic trade-off that exists between separation and completion, arising because pattern separation causes similar input patterns to give rise to distinct, separated representations, whereas pattern completion causes similar input patterns to give rise to a common (recalled) representation. We cannot simultaneously perform separation and completion, and so a trade-off between these two processes must be made. In other words, when faced with a stimulus that is similar to, but not identical to some familiar stimulus, a decision must be made. Is the current stimulus a noisy or partial version of the familiar one, in which case completion is the appropriate choice, or is this a genuinely new stimulus, in which case separation is the appropriate choice?

Having a concrete computational framework for examining the performance characteristics of these processes, O'Reilly and McClelland (1994) were able to quantitatively evaluate the ways in which the hippocampus might avoid or minimize the effects of the trade-off. They hypothesized that the dis-

tinctive anatomical and physiological properties of the hippocampus exist so as to avoid (as much as possible) this intrinsic trade-off. The results indicate that indeed the structure of the hippocampus does make sense when viewed in terms of improving the characteristics of this trade-off.

Without going into the anatomical or computational details, we can use this data from O'Reilly and McClelland (1994) as a basis for understanding when patterns of activity over the neocortex will lead to the recall of representations from the hippocampus, and when new hippocampal representations will be stored. In general, the relevant dimension for deciding between storage and recall is the relative similarity of patterns of activity over the region of the brain that provides the primary cortical input into the hippocampus, known as the Entorhinal Cortex (EC). We assume that the EC activity reflects the activity patterns over wide areas of the neocortex. When an input pattern closely resembles a previously stored one, it will trigger the recall of that pattern. When it is sufficiently distinct from other patterns, a new representation will be formed. However, other factors are undoubtedly relevant, including emotional state and arousal level, attention, and task demands.

With the caveat that they form only part of the overall picture, we examine the pattern separation-completion trade-off results from O'Reilly and McClelland. Figure 14.2 shows pattern completion and separation for the best overall compromise between the two. Judging from these results, the hippocampus would be capable of taking a relatively small fragment (around 30% of the original pattern) and completing a substantial portion of the original pattern (around 80%), while being able to keep reasonably separate patterns that overlap to around 60%. These results are just for the feed-forward pathway from the EC to hippocampal area CA3, and it is widely thought that the recurrent collateral pathway within area CA3 would be capable of performing significant pattern completion on top of that performed by the feed-forward pathway (McNaughton & Morris, 1987). This CA3 completion would have the effect of introducing a threshold, above which the input pattern would get drawn into the nearest familiar attractor state. Estimates of the value of this threshold depend on various parameters, most of which are not sufficiently constrained by the known neurophysiology, but it is probably between 50% to 80%.

To summarize, the O'Reilly and McClelland (1994) results show that the hippocampus could be capable of either recalling a stored pattern with a reasonably small portion of the original pattern as a cue, or separating full input patterns that are similar so that they may be learned rapidly without interference. Although the precise numerical parameters governing these processes are not known, the qualitative relationship between completion and separation can be estimated from results such as those shown in Fig. 14.2. It is important to emphasize the following points. First, whether the system performs recall or storage of new patterns is determined by the properties of the pattern on the EC, and not by some other unspecified control system (i.e., a "homuncu-

Hippocampal Pattern Completion and Separation

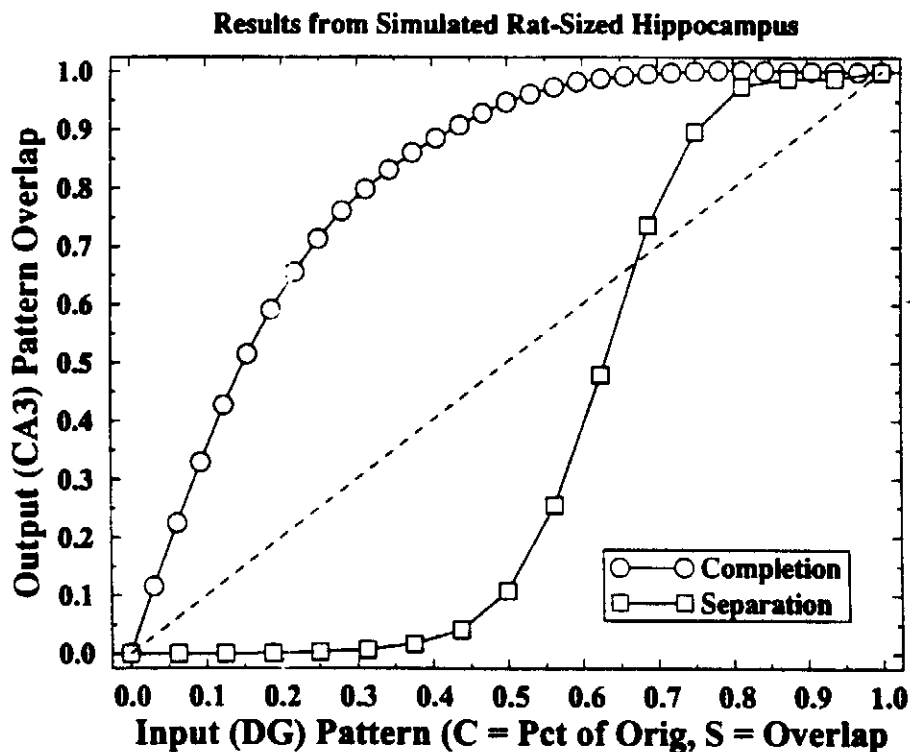


FIG. 14.2. Pattern completion and separation in the hippocampus as simulated in O'Reilly & McClelland (1994). For completion, the X axis represents the percent of the original pattern as a cue for retrieving the stored pattern. For separation, the X axis represents the degree to which a full input pattern overlaps with a stored one. In both cases, the Y axis represents the degree to which the output pattern matches the original stored one. This figure corresponds to the FMSEPO condition with WID learning of .2 from O'Reilly and McClelland (1994).

lus"). Second, separation is driven by parts of the input pattern that do not overlap with stored patterns, to which the system is very sensitive. Thus, it is important that recall cues be partial input patterns, whereas new patterns to be stored should be complete.

Relative Properties of and Interactions Between Prefrontal Cortex and Hippocampus

We previously discussed theories concerning the functioning of prefrontal cortex and hippocampus in detail. In Table 14.1 we compare and contrast their functions along two dimensions that we believe highlight their relative contributions to memory.

Memory Mechanisms: Activation Versus Weight Change. Note first that Table 14.1 distinguishes between two types of memory mechanisms—activa-

TABLE 14.1.
Properties of Neocortex and Hippocampus

<i>Structure</i>	<i>Activation</i>	<i>Weight Changes</i>
Prefrontal Cortex	Sustained	Long-term, integrative
Nonprefrontal Neocortex	Transient	Long-term, integrative
Hippocampus	Transient	Rapid, arbitrary

tion and weight changes. We think of activation as corresponding to the concepts of short-term or working memory, whereas weight changes provide the basis for long-term memory. Thus, activation represents the current state of the system, whereas its connection weights represent its long-term knowledge. In other words, maintenance of activation offers one temporary way to "remember" a state, and weight changes offer another, more enduring way. Note, however, that the latter requires the activation of some cue state in order to retrieve the to-be-remembered one.

From our review of the literature concerning prefrontal cortex, it seems clear that this structure plays an important role in activation-based memory. A growing number of studies suggest that it is required for performance in tasks that rely on the maintenance of task-relevant representations, and thus is characterized by the ability for sustained activation. In contrast, hippocampus and nonprefrontal neocortex seem to be better characterized by transient activity: Lesions of prefrontal cortex impair performance on tasks requiring sustained activation, suggesting that other regions are incapable, on their own, of sustained activity (e.g., Barone & Joseph, 1989; Goldman-Rakic, 1987).

A second distinction that we make in Table 14.1 is between rapid versus long-term, integrative learning. As discussed in the section on the hippocampus, there is strong evidence that this structure is required for the encoding of episodic associations—that is, associations that have been observed only once or a limited number of times. Conversely, there is growing evidence to suggest that, even in the absence of a hippocampus, it is possible to learn new associations. We assume that this occurs through neocortical mechanisms, which require extensive, repeated experience. As discussed, this difference suggests that the hippocampus may be specialized for the rapid encoding of arbitrary new associations, whereas the neocortex is adapted to the encoding and integration of repeatedly observed associations over extended periods of time.

Dimensions of "Top-Down" Control: Bias and Binding. Although our discussions of prefrontal cortex and hippocampus suggest that both of these structures play an important role in the top-down control of behavior, it should also be obvious by now that their contributions are different. In considering how these structures contribute to the control of planful behaviors, such as prospective memory, we think it is helpful to further clarify these differences

by identifying two dimensions of top-down control that reflect the two principles by which prefrontal cortex and hippocampus operate. The first, which we refer to as *bias*, has to do with top-down influences on the execution of and competition between learned, or familiar neocortical processes; the second, which we refer to as *binding*, has to do with mediation of the associations necessary for performing novel tasks.

As discussed earlier, we assume that prefrontal cortex is responsible for actively supporting context, or plan representations, and that these serve to support task-relevant processes over extended delays, and to adjudicate the competition between task-relevant and task-irrelevant processes. Thus, the prefrontal cortex comes into play when either a plan is associated with a temporally extended set of processes, or it must "inhibit" competing, more dominant processes. Again, these both reflect the operation of a single mechanism within prefrontal cortex—to actively support task-relevant processes against interference (i.e., from cumulative effects of noise, over time, or from direct competition). Conversely, we assume that tasks requiring only transient activation of a plan, or that involve only dominant sequences of actions, can be performed independently of prefrontal cortex. We refer to this dimension of control as the *bias dimension* to convey the ability of representations within prefrontal cortex to bias processing elsewhere within the system in favor of task-relevant information (see Cohen, Dunbar, & McClelland, 1990, for a detailed discussion and examples of how bias can have a modulatory influence in systems using nonlinear processing units).

Depending upon the circumstances in which a task is performed, it may require a greater or lesser degree of bias; that is, its reliance on prefrontal cortex may vary. Thus, a task relying on a sequence of dominant processes may ordinarily require little bias from prefrontal cortex (other than, perhaps, an initial prefrontal cortex representation to launch it). However, if it is interrupted, it may then rely on prefrontal cortex for completion. The fact that it was interrupted indicates that it has come into competition with another, more dominant process; it now requires top-down input from prefrontal cortex, or bias, to overcome this competition if it is to be completed.

The second dimension of top-down influence has to do with the degree of experience associated with a task. Thus, a task may be novel, it may be highly rehearsed, or somewhere in between. We believe this dimension is closely related to the role of the hippocampus. We assume that the hippocampus is necessary to mediate the performance of novel tasks. It does this by binding together the neocortical representations necessary to perform the task, for which associations have not yet formed. These may or may not include representations within prefrontal cortex, depending upon the task. Thus, a task that involves a simple response to a simple stimulus, even if it is novel, may require hippocampal binding of relatively simple representations, all of which are outside of prefrontal cortex. In contrast, execution of a novel task that involves a

temporally extended sequence of processes, many or all of which are subject to competition from more dominant processes, will require hippocampal binding of representations within prefrontal cortex as well as nonprefrontal neocortex. Of course, with practice, weight changes within neocortex will progressively encode the appropriate associations, so that eventually the hippocampus will no longer be needed to bind together the representations needed to perform the task. We refer to this dimension of top-down control as the *binding dimension*, to convey the role that hippocampus plays in mediating new associations until these are encoded as weight changes within neocortex.

We consider bias and binding to be continuous dimensions along which processes vary by degree. Thus, processes differ by the extent to which they rely on control by prefrontal cortex and hippocampus, and may show different combinations of reliance on each. It is important to distinguish between and consider processes along each of these dimensions. This is particularly clear when considering the relationship between practice on a task and performance. The relationship between practice and binding is straightforward: With more practice, a task becomes less reliant on binding, consequently, increasingly independent of hippocampus. However, practice also affects the degree to which a task relies on bias. As a task becomes more practiced, it also becomes more dominant, and therefore relies less on bias, that is, on prefrontal cortex mediation. Because practice has effects along both the bias and binding dimensions, the patterns of performance that result from practice may be complex. We feel that dissociating these two dimensions can help clarify the effects of practice, particularly with regard to traditional psychological theories of automaticity and control.

Automaticity and Control in Terms of Binding and Bias

Automaticity is usually considered as either a dichotomous variable, or as a unitary dimension that is directly related to practice. The traditional definition of an automatic process is one that does not rely on and is not subject to attentional control, whereas controlled processes rely on the allocation of limited-capacity attention (Posner & Snyder, 1975; Shiffrin & Schneider, 1977). Some authors point out that this distinction should really be considered a continuum, with some processes being more automatic than others (e.g., Kahneman & Treisman, 1984), and with greater practice leading to greater automaticity (e.g., MacLeod & Dunbar, 1988). However, a careful consideration of learning phenomena suggests that there are still reasons to believe in a distinction between novel tasks and practiced ones that is qualitative and different from the variations that are observed among practiced tasks. That is, something different seems to be happening during the earliest stages of practice than during later stages. This is not a new idea. For example, Anderson (1983) distinguished between two procedural learning mechanisms in his ACT* theory—

compilation, which occurs during the initial stages of practice; and strengthening, which continues to occur even for very familiar procedures (and accounts for the power law of practice effects). Similarly, we argued that the mechanisms underlying performance during the earliest stages of practice may be qualitatively different than those involved after moderate amounts of practice (Cohen et al., 1990; Cohen, Servan-Schreiber, & McClelland, 1992). We distinguished between indirect, or mediated processes involved in the performance of novel tasks, and direct processes that underlie performance of more practiced tasks. We took care to point out the differences between this distinction and the more traditional one between controlled and automatic processes. Processes traditionally considered to be controlled include both indirect (i.e., truly novel) as well as weak, but direct processes (e.g., color naming in the Stroop task). We believe that we can now explain the overlap between these two distinctions (controlled vs. automatic, and direct vs. indirect) in terms of the relative contribution of prefrontal cortex and hippocampus; that is, according to the two dimensions introduced earlier—novelty and interference. Thus, the distinction between novelty versus familiarity reflects the influence of the binding dimension, whereas differences in the strength of direct processes reflects the influence of the bias dimension. Processes that are considered to be “controlled” in the traditional sense may be so either because they rely on binding (i.e., hippocampal mediation), or bias (i.e., prefrontal cortex mediation), or both. These relationships are illustrated in Fig. 14.3.

The dimensions of bias and binding make a number of predictions regarding performance on prospective memory tasks, specifically with respect to factors that affect task difficulty and types of errors. For example, the extent of delay over which items must be maintained or remembered and the number of novel associations (bindings) required to perform the task will contribute to task difficulty. There are some empirical findings consistent with such predictions (Einstein, Holland, McDaniel, & Gynn, 1992; Einstein & McDaniel, 1990).

Interactions Between Bottom-Up and Top-Down Influences in Neocortex. In the previous section, we considered the two ways in which prefrontal cortex and hippocampus can influence behavior through top-down mediation of nonprefrontal neocortical processes. Before we consider a full account of planful behavior, however, we need to consider how these top-down influences interact with the processing of bottom-up inputs within neocortex. We assume that different inputs to neocortex activate different representations, and that these are in relative competition. However, to the extent that no single representation is significantly more active than competing ones, then multiple representations can coexist. Under such circumstances, top-down influences become important: input to a representation from prefrontal cortex or the hippocampus will not only support that representation, but will allow it to compete more effectively, and suppress other representations. This suggests how

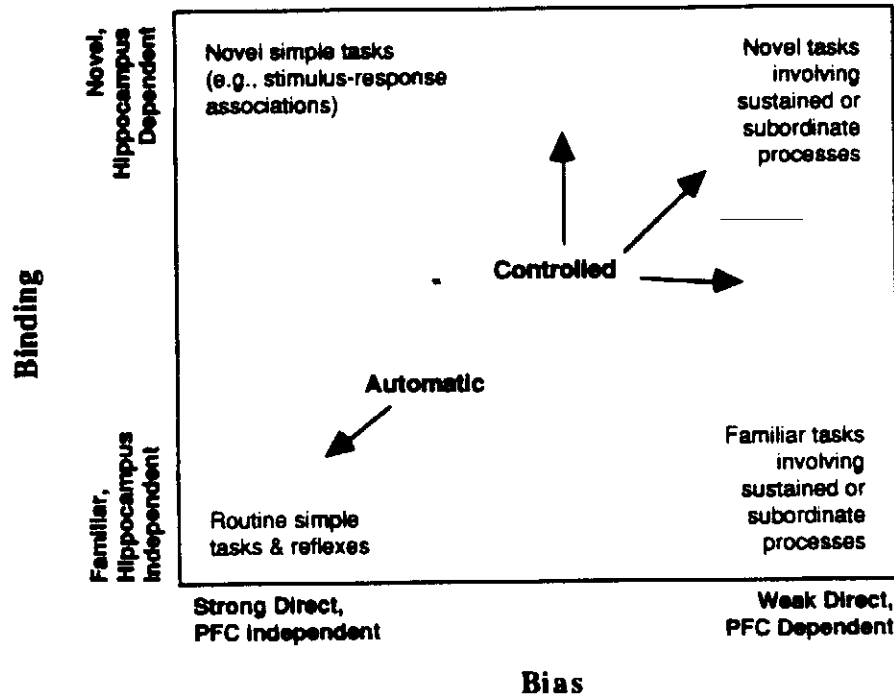


FIG. 14.3. Theorized relationship among the several distinctions that have been used to describe the relationship between practice and underlying processing mechanisms. Automatic processes are assumed to be both familiar and highly practiced, relying neither on hippocampus nor on prefrontal cortex. Processes traditionally considered to be controlled rely either on binding (hippocampus) or bias (prefrontal cortex) or both.

top-down support for a representation can also appear to produce inhibition of competing ones (as with the inhibitory function of prefrontal cortex discussed earlier). Evidence from recent neurophysiological studies (Desimone & Duncan, 1995) supports our view of the interaction between prefrontal cortex and other neocortical areas; that is, that attentional effects occur by the top-down support of task-relevant representations, allowing them to compete more effectively with, and inhibit task-irrelevant ones.

DISCUSSION

In the remainder of this chapter, we consider how these concepts can be used to explain the processes underlying prospective memory, and how these ideas can be applied to performance in more traditional laboratory tasks used to empirically test them.

A Processing Account of Prospective Memory

Recall the example at the beginning of this chapter. You wake up and the environment (time of day, alarm clock going off) as well as internal cues (growl-

ing belly, foggy head) dictate that you get ready to go to work. This occurs via existing cortical associations; that is, you have long-standing experience waking up and going to work, and thus have developed "direct" associations between the environmental conditions and your actions. Although getting ready for work involves a sequence of relatively routine actions, as noted in the discussion of the bias dimension, this will still rely on a plan in prefrontal cortex because it is temporally extended, and may need to be supported in the face of distractions.

In the process of executing the plan, you notice the missing orange juice and the plan is violated. At this point, several things happen. First, the missing orange juice elicits a plan to address this—that is, go to grocery store. However, this plan conflicts with the already active plan of going to work. Attempting to carry both of these out would lead to conflicting sequences of actions. One of the general principles of neural network models is to solve problems (i.e., conflicts) through constraint satisfaction. We assume that this principle applies here. There is an attempt to find a state of activation that accommodates both the current plan (go to work) as well as the plan to get more orange juice. Constraint satisfaction operates by preferentially activating and stabilizing those features that are associated with both plans, or that are associated with one that do not conflict with the other. Through association, the plan to get more orange juice will activate the representation of the trip to the grocery store, sharing some activation with the trips to and from work. At the same time, the current plan to go to work will activate associated representations like "get to work on time." A plan to stop by the grocery store is likely to conflict with the current plan (and its associates, like getting to work on time) more than a plan for the trip home. Thus, the latter will be favored, and survive the competition. This will activate representations of the environment associated with the trip home (e.g., dark outside, clock strikes 5:00, etc.). Thus, the system is likely to stabilize in a state comprised of the following representations: go to work; go home; go to grocery store; 5:00, and so on. As noted earlier, the hippocampus acts to bind the components of such stable neocortical states.

This leads to the question: given the coactivation of both the current plan (get to work) and the plan to stop by the grocery store, why is it that the former prevails now whereas the latter prevails at the end of the day? The answer is that at present the environment strongly favors the go to work plan, through associations in the neocortex. There are strong associations between the morning environment and the plan to go to work. It is possible to activate the plan representation of (i.e., "imagine") going to the grocery store, without actually executing it. On the other hand, at 5:00, by the same mechanism, the environment strongly supports the going home plan. However, due to hippocampally bound associations, today this environment also elicits the go-to-grocery store plan. That is, the hippocampus reinstates the set of representations bound earlier that morning. Note that this also includes the then-current plan, go-to-work. However, the current environment strongly favors the go-home

plan, and thus prevails. Finally, there is a conflict between the usual go-home plan elicited directly by the environment and the go-to-grocery store plan, elicited by the hippocampus (e.g., the usual trip home involves a left at the light, whereas the trip to the grocery involves a right). Here, both the hippocampus and PFC play a role in controlling behavior: The hippocampus helps favor the go-to-grocery store representation over the usual go-straight-home representation within the PFC. This PFC representation acts to bias behavior toward the appropriate, but less habitual response of turning left at the light. Note that this predicts that, under conditions of distraction (i.e., prefrontal cortex failure), a particular type of error will be made, the elicitation of the more habitual response (going straight home). These types of habitual errors are well-documented (Reason & Mycielska, 1982).

The astute reader will notice at least one problem with this account: Why don't you go to the grocery store on the way home tomorrow as well? That is, the end of the day today may be similar to the end of the day tomorrow, so why doesn't the environment elicit the same set of representations encoded by the hippocampus this morning? There are a number of possible responses to this question. First, we should note that, at the very least, anecdotal evidence suggests that such errors do occur. We would predict, furthermore, that the more similar the environmental cues (e.g., the weather, the activities you are engaged in at the end of the day, etc.) the more likely such errors are to occur. However, even if these occur, they are not the norm. This suggests that the representations active when associations are encoded by the hippocampus are richer than we allowed. Specifically, there are likely to be feature dimensions that encode days of the week, recency, and so on. A combination of such features may help to identify and distinguish individual days. Finally, it is likely that hippocampally encoded associations are subject to decay, so that the overall likelihood of retrieval is reduced with time.

Prefrontal Cortex-Hippocampus Interactions in Classical Learning Tasks

The previous section provides a narrative account of how prefrontal cortex and hippocampus may interact in the service of prospective memory. Such accounts are difficult to test. Here, we consider how our theory can be applied to performance in a set of laboratory tasks—blocking and latent inhibition. Evidence from animal studies suggests that these tasks are sensitive to hippocampal function, and there are at least two computational accounts of the functions of hippocampus and neocortex in these learning tasks of which we are aware (Schmajuk & DiCarlo, 1992; Gluck & Meyers, *in press*). Schmajuk and DiCarlo hypothesized that hippocampus is responsible for error-based learning in the service of predicting expected responses from sensory cues, whereas neocortex relies only on associate learning in such tasks. Gluck and Meyers

(in press) argued that the hippocampus is responsible for recoding sensory representation; these recoded representations then govern the generation of associations between stimuli and responses in the neocortex. We offer an account that is different from both of these, that describes performance in such tasks in terms of hippocampal–neocortex interactions similar to the ones we described earlier. We consider how our theory can account for findings in the blocking and latent inhibition paradigms.

Blocking. In the blocking paradigm (Kamin, 1969), subjects are first conditioned to respond to a particular stimulus (let us say, stimulus A), during a pretraining phase. During a subsequent training phase, stimulus A is paired with a second conditional stimulus (say, stimulus B). Finally, during a test phase, subjects are tested on the association between B and the unconditioned stimulus—that is, the ability of B to elicit the unconditioned response. Performance is compared with a control case in which there is no pretraining on A—that is, the subject is trained only on A and B together. The typical finding is that pretraining on A reduces conditioning to B during the A-and-B training phase. In other words, pretraining on A “blocks” the ability of B to acquire an association with the unconditioned stimulus. Lesions of the hippocampus were observed to reduce this effect, increasing conditioning to B in the face of pretraining on A (Solomon, 1977; Rickert et al., 1978).

Schmajuk & DiCarlo (1992) explained these findings by observing that, with an error-driven learning rule, pretraining on A will gradually reduce the error in predicting the unconditioned stimulus, thereby reducing the amount of error available for training during the A-and-B phase. As such, conditioning to B will be reduced. This effect is attributed to the hippocampus, presumed to use error-driven learning. Neocortex, however, is assumed to use an associative learning rule; this is influenced only by the co-occurrence of stimuli, and not by the error in predicting them. As such, during the A-and-B phase, only the co-occurrence of B and the unconditioned stimulus is relevant, and not the fact that A is already associated with the unconditioned stimulus. Thus, in the absence of the hippocampus, conditioning of B is not influenced by pretraining on A.

Our theory suggests a different account of the findings. During the A-only phase, we assume that both the hippocampus and neocortex act in parallel to encode the association between A and the unconditioned stimulus. However, there are differences in how this is encoded by the two structures. The hippocampus encodes this as an association between the stimulus, the experimental context (including a representation of the task context in the prefrontal cortex), and the unconditioned stimulus. In the neocortex, however, the association is directly between the stimulus and the unconditioned stimulus. During the A-and-B phase, however, matters change. In the absence of the hippocampus, the neocortex continues to encode associations between the stimuli pres-

ent, and would begin to form an association between B and the unconditioned stimulus. However, the hippocampus, if present, will cause the representation of B to be suppressed. This is because the presentation of A in the experimental context causes it to help reinstate the set of representations that were associated during the A-only phase, which did not include B. Although there is no evidence that the hippocampus directly inhibits representations that do not form part of the recalled set (in this case, B), we assume that the additional input provided to the representations of A and to the experimental context allow these to compete more aggressively with B, which is receiving only bottom-up support, and therefore the activation of B diminishes. This reduces the chance for new associations to form between B and the unconditioned stimulus. Again, evidence from neurophysiological studies support the existence of such competition within neocortex, and the importance of top-down influences on this competition (Desimone & Duncan, 1995).

Thus, our account, like the error-associative learning account (Schmajuk & DiCarlo, 1992), is consistent with the observation that blocking occurs in normal subjects and that it appears to be reduced in hippocampally lesioned animals (Solomon & Moore, 1975). The sensory-recoding theory (Gluck & Meyers, in press) seems to account for blocking in a parameter-dependent fashion, which may rely to greater or lesser extent upon hippocampal function. Therefore, the blocking paradigm does not provide a good test of this theory. However, it can be used to test our theory, which makes two predictions that differ from the Schmajuk & DiCarlo (1992) theory.

First, our theory predicts actual suppression of the representation of B during the A-and-B phase. The Schmajuk & DiCarlo theory does not make this prediction; it asserts that no association will occur between B and the response because A adequately predicts the response, not because B is suppressed. This could be tested. For example, activation of B could be assessed using a priming paradigm in concert with the blocking paradigm. We predict that there would be less priming of B than in an appropriate control condition, and that this reduction should correlate positively with the degree of blocking.

Second, because we assume that the neocortex has the capacity for error-based learning, but over a much longer time frame, we expect that even hippocampally lesioned subjects should show a blocking effect with sufficient training in the A-only phase. Over time, a cortical association would develop between A and the experimental context, obviating the need for the hippocampus to produce the top-down activation of A (during A-and-B phase) that inhibits B and produces the blocking effect. In contrast, the entire premise of the error-vs.-associative learning account is that neocortex is capable only of associative learning, and therefore should never show blocking on its own.

Latent Inhibition. This refers to any reduction of conditioning to a stimulus that can be attributed to pre-exposure to that stimulus (Lubow, 1989).

Thus, in latent inhibition paradigms, subjects are exposed to a stimulus in a pretraining, or pre-exposure phase, without presentation of the unconditioned stimulus. This is followed by a conditioning phase, during which the stimulus is paired with the unconditioned stimulus. The rate of conditioning is slower when compared with that of subjects who have not been pre-exposed to the stimulus, using a wide variety of different stimuli and tasks (Lubow, 1989). Furthermore, lesions of the hippocampus appear to reduce the effect of stimulus pre-exposure, such that the rate of conditioning during the training phase is increased.

The traditional account of this phenomenon considers it in terms of selective attention (Lubow, 1989): It is assumed that subjects become familiar with and learn to ignore the uninformative stimulus during the pre-exposure phase, reducing its salience during subsequent conditioning. Schmajuk and DiCarlo (1992) and Gluck and Meyers (in press) offered closely related accounts in which associative resources are preferentially allocated to informative stimuli, and reduced for uninformative ones.

Our account is somewhat different from traditional ones. During the phase of pre-exposure to stimulus A, we assume that the hippocampus forms an association between this stimulus, the environmental context, and the representation active within prefrontal cortex. Although typically there is no experimentally determined task that the subject performs during the pre-exposure phase, we assume that subjects are engaged in some set of activities (such as exploration, grooming, etc.) that are associated with corresponding "state" (i.e., internal context) representations within prefrontal cortex. Thus, during the test phase, presentation of A causes the hippocampus to reactivate this set of representations. This set contrasts with the set of representations activated by presentation of the unconditioned stimulus. This stimulus is strongly associated with a specific behavior, the unconditioned response, that is different than the behaviors engaged in by the subject during the pre-exposure phase. Furthermore, we assume that the unconditioned stimulus and response activate a representation within prefrontal cortex (e.g., a state of alertness or anticipation), that differs from the representation within prefrontal cortex during the pre-exposure phase. Thus, stimulus A and the unconditioned stimulus favor the activation of different and conflicting sets of representations, or states. As we discussed in the orange juice example, we assume that constraint satisfaction, operating within the neocortex, favors one or the other of these two states, or an alternation between them. Conditioning of an association between two stimuli (in this case, stimulus A and the unconditioned stimulus) requires that these be coactivated. Thus, competition between incompatible states accounts for the latent inhibition effect. Finally, without pre-exposure to stimulus A, it will not be associated with any hippocampally mediated state, and therefore no competition will arise within the neocortex upon presentation of the unconditioned stimulus.

This account shares something in common with our account of the blocking phenomenon: Both assume that hippocampus is responsible for binding together representations during the initial phase, and that recall of this set of representations is mediated by hippocampus during subsequent phases of the experiment. In the blocking paradigm, this recall leads to the inhibition of the novel stimulus, interfering with conditioning on that stimulus. In the latent inhibition paradigm, the novel stimulus is salient, and thus is unlikely to be inhibited. However, as we suggested, it has strong associations with representations that conflict with those of the pre-exposure stimulus, and it is the competition between these sets of representations that interferes with conditioning.

As described, an important feature of our account of blocking and latent inhibition that distinguishes it from others is that our suggests that the prefrontal cortex plays a significant role in these phenomena in addition to that played by the hippocampus. Another important difference between our account of latent inhibition and more traditional ones is that we do not assume that pre-exposure to a stimulus reduces its representation, or makes this less distinct. More fundamentally, we do not assume that subjects actively ignore the familiar stimulus, but simply that its representation—within the experimental context—has become incompatible with that of the unconditioned stimulus, and that this incompatibility interferes with conditioning.

Relationship of Blocking and LI to Prospective Memory Tasks

The specific laboratory tasks we discussed (blocking and LI) illustrate task dimensions, and the operation of processing mechanisms that are relevant for understanding the kinds of errors likely to occur under conditions in which prospective memory operates. We predict that pre-exposure to task variables and responses will affect subsequent planning and memory. For example, consider the possibility that, in the "missing orange juice" scenario, one discovers a second item during the day that needs to be purchased at the grocery store on the way home. We predict that this second item is more likely to be forgotten than if it had been the only one needed, due to a blocking-like effect. When the attempt is made to add the second item to the mental shopping list, we expect that reinstatement of the hippocampal "go to the grocery store on the way home" plan, and its prior association with orange juice will actually tend to interfere with the association of the new item with the plan.

We can also imagine a case in which latent inhibition will affect prospective memory performance. As was suggested (Mark McDaniel, personal communication, July, 1994), we can consider the case where the grocery store on the way home is actually an expensive convenience store that was avoided in the past. In this case, the plan to stop at the convenience store will have to overcome this prior "no response" association, and we predict that more failures

to "stop and shop" would occur in this situation, as compared to one in which the store is a regular venue.

CONCLUSION

In this chapter, we review findings concerning the functions of prefrontal cortex and hippocampus, and their roles in information processing. We also describe computational mechanisms that were proposed to account for these, and describe how these may interact to give rise to the coordinated execution of planful behavior and prospective memory, in particular. We discuss how these ideas relate to more traditional concepts such as automaticity, and how they may apply to specific laboratory tasks that can be used to test them. At this stage, however, the interaction between the computational mechanisms we propose is conjectural. Our claims need to be substantiated, both in functioning simulations that implement these mechanisms, and in empirical validation of the predictions they make. These suggest future activities that, with time, we hope to be able to look back on as an example of prospective memory at work.

REFERENCES

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Barone, P., & Joseph, J. P. (1989). Role of the dorsolateral prefrontal cortex in organizing visually guided behavior. *Brain Behavior and Evolution*, 33, 132-135.
- Becker, J. T., Mintun, M. A., Diehl, D., DeKosky, S. T., & Dobkin, J. (1993). Functional neuroanatomy of verbal memory as revealed by word list recall during PET scanning. *Society for Neuroscience Abstracts*, 19(2), 1079.
- Bianchi, L. (1922). *The mechanism of the brain and the function of the frontal lobes*. Edinburgh, Scotland: Livingstone.
- Blamire, A. M., McCarthy, G., Bloch, G., Rothman, D. L., & Shulman, R. G. (1993). Functional magnetic resonance imaging of human pre-frontal cortex during a spatial working memory task. In *Proceedings of the Society for Magnetic Resonance in Medicine, 12th Annual Scientific Meeting*, 3, 1412.
- 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., 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., Forman, S. D., Braver, T. S., Casey, B. J., Servan-Schreiber, D., & Noll, D. C. (1994). Activation of prefrontal cortex in a non-spatial working memory task with functional MRI. *Human Brain Mapping*, 1, 293-304.
- Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *American Journal of Psychology*, 105, 239-269.
- Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, 210, 207-209.

- Damasio, A. R. (1979). The frontal lobes. In K.M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology* (pp. 3360-3411). New York: Oxford University Press.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193.
- Diamond, A., & Doar, B. (1989). The performance of human infants on a measure of frontal cortex function, the delayed response task. *Developmental Psychobiology*, 22(3), 271-294.
- Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's A B task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research*, 74, 24-40.
- Diamond, A. (1990). The development and neural bases of memory functions as indexed by the A B and delayed response tasks in human infants and infant monkeys. In A. Diamond (Ed.), *The development and neural bases of higher cognitive functions* (pp. 267-317). New York: New York Academy of Science Press.
- Duncan, J. (1986). Disorganization of behavior after frontal lobe damage. *Cognitive neuropsychology*, 3(3), 271-290.
- Einstein, G. O., Holland, L. J., McDaniel, M. A., & Guynn, M. J. (1992). Age-related deficits in prospective memory: The influence of task complexity. *Psychology and Aging*, 7, 471-478.
- Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 717-726.
- Ferrier D. (1886). *Functions of the Brain*. London: Smith and Elder.
- Fuster, J. M. (1980). *The prefrontal cortex*. New York: Raven.
- Fuster, J. M. (1989). *The prefrontal cortex. Anatomy, physiology and neuropsychology of the frontal lobe*. New York: Raven.
- Fuster, J. M. (1985a). The prefrontal cortex and temporal integration. In A. Peters & E. G. Jones (Eds.), *Cerebral cortex* (pp. 151-177). New York: Plenum.
- Fuster, J. M. (1985b). The prefrontal cortex, mediator of cross-temporal contingencies. *Human Neurobiology*, 4, 169-179.
- Gluck, M. A., & Meyers, C. E. (in press). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus*.
- Goldberg, T. E., Weinberger, D. R., Berman, K. F., Pliskin, N. H., & Podd, M. H. (1987). Further evidence of dementia of the prefrontal type in schizophrenia? *Archives of General Psychiatry*, 44, 1008-1014.
- 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.
- Gorenstein, E. E., Mammato, C. A., Sandy, J. M. (1989). Performance of inattentive-overactive children on selected measures of prefrontal-type function. *Journal of Clinical Psychology*, 45, 619-632.
- Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning Memory and Cognition*, 10, 164-178.
- Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl type card sorting problem. *Journal of Experimental Psychology*, 38, 404-411.
- Grasby, P. M., Frith, C. D., Friston, K. J., Bench, C., Frackowiak, R. S. J., & Dolan, R. J. (1993). Functional mapping of brain areas implicated in auditory-verbal memory function. *Brain*, 116, 1-20.
- Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J., Allen, D. G., Kurkijan, M., Schapiro, M. B., Rapoport, S. I., & Grady, C. L. (1993). Lateralization of frontal lobe activity associated with working memory for faces changes with retention interval: A parametric PET-rCBF study. *Society for Neurosciences Abstract*, 19, 1284.
- Hecaen, H., & Albert, M. L. (1978). *Human neuropsychology*. New York: Wiley.
- Hitzig, E. (1874). *Untersuchungen uber das Gehirn*. Berlin: Hirschwald.

- Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S., & Mintun, M. A. (1993). Spatial working memory in humans as revealed by PET. *Nature*, 363, 623–625.
- Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, R. Davies, & J. Beatty (Eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.
- Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and Aversive Behavior* (pp. 279–296). New York: Appleton-Century-Crofts.
- Lubow, R. E. (1989). *Latent inhibition and conditioned attention theory*. Cambridge, England: Cambridge University Press.
- Luria, A. R. (1969). Frontal lobe syndromes. In P. J. Vinken & G. W. Bruyn (Eds.), *Handbook of clinical neurology* (Vol. 2, pp. 725–757). New York: Elsevier.
- MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology*, 14, 126–135.
- Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London, B*, 262, 23–81.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (in press). 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*.
- McClelland, J. L., McNaughton, B. L., O'Reilly, R. C., & Nadel, L. (1992). Complementary roles of hippocampus and neocortex in learning and memory. *Society for Neuroscience Abstracts*, 18(2), 1216.
- McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 24, pp. 109–164). San Diego: Academic Press.
- McNaughton, B. L., & Morris, R. G. M. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neurosciences*, 10(10), 408–415.
- Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology*, 9, 90–100.
- Milner, B. (1966). Amnesia following operation on the temporal lobe. In C. W. M. Whitty & O. L. Zangwill (Eds.), *Amnesia* (pp. 109–133). London: Butterworth & Co.
- Mishkin, M., & Pribram, K. H. (1955). Analysis of the effects of frontal lesions in monkeys. I. Variations on delayed alternation. *Journal of Comparative Physiological Psychology*, 388, 492–495.
- Mishkin, M. (1964). Perseveration of central sets after frontal lesions in monkeys. In J. M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior* (pp. 219–241). New York: McGraw-Hill.
- Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, 12, 313–324.
- O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a tradeoff. *Hippocampus*, 6, 661–682.
- Passingham, R. E. (1985). Memory of monkeys (*Macaca mulatta*) with lesions in prefrontal cortex. *Behavioral Neuroscience*, 99, 3–21.
- Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behavior. *Neuropsychologia*, 12, 323–330.
- Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia*, 20, 249–262.
- Petrides, M. E., Alivisatos, B., Meyer, E., & Evans, A. C. (1993). Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Proceedings of the National Academy of Science, U.S.A.*, 90, 878–882.

- Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Reason, J. R., & Mycielska, K. (1982). *Absentminded? The psychology of mental lapses and everyday errors*. Englewood Cliffs, NJ: Prentice-Hall.
- Rickert, E. J., Lorden, J. F., Smyly, E., & Callahan, M. (1978). Hippocampectomy and the attenuation of blocking. *Behavioral Biology*, 22, 147-160.
- Robinson, A. L., Heaton, R. K., Lehman, R. A. W., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. *Journal of Consulting Clinical Psychology*, 48, 605-614.
- Rosendkild, K. E. (1979). Functional heterogeneity of the prefrontal cortex in the monkey: A review. *Behavioral Neural Biology*, 25, 301-345.
- Rosvold, K. E., Szwed, M. K., Mirsky, A. F., & Mishkin, M. (1961). The effect of frontal-lobe damage on delayed response performance in chimpanzees. *Journal of Comparative Physiological Psychology*, 54, 368-374.
- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the microstructure of cognition* (Vols. 1 and 2). Cambridge, MA: MIT Press.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & PDP Research Group (Eds.), *Parallel Distributed Processing. Volume 1: Foundations* (pp. 318-362). Cambridge, MA: MIT Press.
- Schacter, D. L., Delaney, S. M., & Merikle, E. P. (1990). Priming of nonverbal information and the nature of implicit memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 83-123). San Diego: Academic Press.
- Schmajuk, N. A., & DiCarlo, J. J. (1992). Stimulus configuration, classical conditioning, and hippocampus function. *Psychological Review*, 2, 268-305.
- Scoville, W., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20, 11-21.
- Servan-Schreiber, D., Cohen, J. D., & Steingard, S. (1994). Schizophrenic performance in a variant of the CPT-AX: A test of theoretical predictions concerning the processing of context. Unpublished manuscript.
- 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.
- Shimamura, A. P., & Squire, L. R. (1989). Impaired priming of new associations in amnesia. *Journal of Experimental Psychology: Learning Memory and Cognition*, 15, 721-728.
- Solomon, P. R. (1977). Role of the hippocampus in blocking and conditioned inhibition of the rabbit's nictitating membrane response. *Journal of Comparative and Physiological Psychology*, 91, 407-417.
- Solomon, P. R., & Moore, J. W. (1975). Latent inhibition and stimulus generalization of the classically conditioned nictitating membrane response in rabbits (*Oryctolagus cuniculus*) following dorsal hippocampal ablation. *Journal of Comparative and Physiological Psychology*, 89, 1192-1203.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 195-231.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Stuss, D. T., & Benson, D. F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin*, 95, 3-28.
- Stuss, D. T., & Benson, D. F. (1986). *The Frontal Lobes*. New York: Raven Press.

- Swedo, S. E., Shapiro, M. B., Grady, C. L., Cheslow, D. L., Leonard, H. L., Kumar, A., Friedland, R., Rapoport, S. I., & Rapoport, J. L. (1989). Cerebral glucose metabolism in childhood-onset OCD. *Archives of General Psychiatry*, 46, 518-523
- Tulving, E. (1983). *Elements of Episodic Memory*. Oxford: Clarendon Press.
- White, H. (1989). Learning in artificial neural networks: A statistical perspective. *Neural Computation*, 1, 425-464.
- 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.