Mechanisms of Cognitive Control: Active Memory, Inhibition, and the Prefrontal Cortex

Todd S. Braver
Department of Psychology, Washington University

Deanna M. Barch
Department of Psychology, Washington University

Jonathan D. Cohen
Department of Psychology, Princeton University
Department of Psychiatry, University of Pittsburgh

Running Head:
Prefrontal Mechanisms of Cognitive Control

ABSTRACT

Previous research has identified the prefrontal cortex (PFC) as a brain region that is critical for cognitive control. Currently, theorists remain divided about whether to view the PFC as primarily a coordinative, mnemonic, or inhibitory structure. A theory is presented that attempts to resolve some of the apparent conflicts between the predominant views on PFC control functions. In this theory, PFC is proposed to actively maintain representations of context information. These maintained representations provide a mechanism of control by serving as a top-down bias on the local competitive interactions that occur during processing. As such, it is suggested that PFC performs *both* mnemonic and inhibitory functions in the service of control, and that each is preferentially observable under different task situations. A series of behavioral, computational, and neuroimaging studies are presented that demonstrates how this theory can account for a wide range of data associated with performance of a simple cognitive control paradigm.

INTRODUCTION

The concept of cognitive control is central to any discussion of intelligent behavior. Phenomenologically, control is intrinsically related to notions of consciousness and agency, in that willed or intentional actions require that behavior be flexibly controlled by the agent rather than be merely fixed or reactive. High-level cognitive processes such as problem-solving, planning and decision-making appear to depend critically upon exerting control over attention, memory, and action selection. Clinical observations of patients with frontal lesions, who often exhibit a 'dysexecutive syndrome', clearly demonstrate that control has a biological basis. Thus, progress in answering questions regarding these fundamental issues in psychology will be dependent upon understanding the phenomenological reality of cognitive control in terms of the underlying psychological and biological mechanisms that give rise to it.

However, investigating cognitive control at either the psychological or biological levels has proved to be a deep and difficult problem for researchers. In the psychological literature, a number of studies have demonstrated clear distinctions that can be made between controlled processes and more automatic ones. These distinctions have been made in terms of factors such as processing speed, attentional requirements, and capacity (Posner & Snyder, 1975; Shiffrin & Schneider, 1977). More contemporary studies have acknowledged that there is actually a continuum between controlled and automatic processing (Cohen, Dunbar, & McClelland, 1990; Kahneman & Treisman, 1984). Nevertheless, the experimental work has been lacking in its ability to provide a good description of the various functions associated with cognitive control, and the mechanisms by which they arise. The problem is best exemplified in Baddeley's (1986) model of working memory. The Baddeley model has been explicit in specifying the need for a control structure as a specific subcomponent of working memory. This subcomponent, termed the central executive, is responsible for coordinating the activities of two domain-specific buffer systems and performing other necessary control functions. This model has been a very influential one, in that it provides a useful conceptual framework for understanding and interpreting the results of a wide range of

experimental data. Unfortunately, even Baddeley (1996) himself has recently admitted "that our initial specification of the central executive was so vague as to serve as little more than a ragbag (p.6)", which contained all of the phenomena that could not be readily accounted for otherwise. More recent work has addressed this problem, by attempting to more clearly specify control processes for which the executive is responsible (Baddeley, 1996; Baddeley & Della Sala, 1996). However, even with this progress, the particular mechanisms involved in mediating control still remain largely unexplained within the Baddeley model.

For researchers investigating the neurobiology of cognitive control, the prefrontal cortex (PFC) has been an area of particular focus. Over a hundred years of neuropsychological studies have provided strong evidence of the involvement of this brain region in the control of behavior. The findings are so well-accepted that in the clinical literature, the term 'frontal syndrome' refers to a particular impairment in which the normal control over social and sexual behavior is dysregulated (Hecaen & Albert, 1978; Stuss & Benson, 1986). Patients with PFC damage also often show deficits in more clearly cognitive behaviors such as planning, reasoning, problem solving, and working memory (e.g., Bianchi, 1922; Damasio, 1985; Luria, 1969; Shallice & Burgess, 1991). Lesion studies in non-human primates have paralleled the findings from the human neuropsychological literature, and neurophysiological recordings of monkey PFC neurons have shown selective activity in this brain region during simple working memory and planning tasks (Barone & Joseph, 1989; Fuster, 1989; Goldman-Rakic, 1987). A growing number of human neuroimaging studies have corroborated these findings, showing increased activity in PFC during cognitive control tasks (Baker et al., 1996; Carter, Mintun, & Cohen, 1995; D'Esposito et al., 1995; Prabhakaran, Smith, Desmond, Glover, & Gabrieli, in press), especially tasks involving working memory (Cohen et al., 1994; Grasby et al., 1993; Jonides et al., 1993; Petrides, Alivisatos, Evans, & Meyer, 1993).

Although these findings have led to general agreement that PFC is critically involved in cognitive control, the particular mechanisms surrounding its involvement remain elusive. There are three primary different perspectives regarding the role of PFC in cognition. The first is most

closely associated with Shallice's model of the Supervisory Attentional System or SAS (Norman & Shallice, 1986; Shallice, 1982; Shallice, 1988). In Shallice's model, the SAS serves as a mechanism by which complex cognitive processes are coordinated and non-routine actions are selected. Shallice has related the SAS to the functioning of the PFC, and to the particular pattern of behavioral impairments observed in frontal patients. Baddeley's work has been deeply influenced by the SAS model, and has linked it to the central executive component of working memory (Baddeley, 1986). Interestingly, in the theoretical formulations provided by Shallice and Baddeley, the SAS/central executive, and thus PFC, subserve control, but not memory functions. Indeed, Baddeley has explicitly argued that the central executive does not have any capabilities for actively maintaining information in working memory (Baddeley, 1993). Rather, the central executive serves to coordinate and organize the operation of two domain-specific buffer systems. It is these two buffer systems which are solely responsible for carrying out necessary storage functions.

This perspective stands in stark contrast with studies arising out of the animal literature on PFC function. In this literature, PFC has been found to be critical for actively maintaining information over short delay intervals in the form of sustained neuronal firing patterns (Fuster, 1989; Goldman-Rakic, 1987). From these findings, it has been argued that PFC acts primarily as a short-term memory mechanism, actively holding information "on-line" through neural activity. One of the most influential proponents of this second viewpoint is Goldman-Rakic, who has suggested that delay-period activity in PFC serves as the "cellular basis of working memory" (Goldman-Rakic, 1995). However, the functions ascribed to PFC by Goldman-Rakic seem to align more closely with the buffer system component of working memory rather than the control system, at least according to Baddeley's formulation. A third perspective on PFC function, arising out of the developmental and neuropsychological literatures, is one that views PFC primarily as an inhibitory structure, responsible for reducing interference and overriding prepotent, but situationally-inappropriate response tendencies (Dempster, 1992; Diamond, 1990; Knight, Grabowecky, & Scabini, 1995). In this formulation, the role of PFC in both active memory and other control functions have been de-emphasized.

These three influential perspectives present a confusing and often times inconsistent picture for the theorist attempting to develop a coherent theory of cognitive control and its neural substrates. Is PFC more like the central executive or memory buffer component of working memory? Does PFC play a coordinative, mnemonic, or inhibitory function in cognitive control? And, more importantly, what are the specific mechanisms by which PFC might subserve these control functions? In this article, we attempt to provide an answer to these questions, by developing an account of the mechanisms of cognitive control and their relationship to PFC function. Our account resolves some of the apparent conflicts between the predominant theories of PFC involvement in cognitive control. In particular, we suggest that PFC mediates *both* mnemonic and inhibitory functions, where each is preferentially observable under different task situations, and where each occurs in the service of cognitive control. Specifically, we argue that the control functions of PFC emerge as a direct consequence of two specific mechanisms: 1) active maintenance of task-relevant context; and 2) top-down biasing of local competitive interactions that occur during processing.

Our account builds upon and refines our previous work, which focused on the role of PFC in context processing (Cohen & Servan-Schreiber, 1992). A primary goal of this prior work was to provide an account of control-related behavioral impairments in patients with schizophrenia in terms of underlying cognitive and neurobiological disturbances. We hypothesized that the impairments in cognitive control observed in patients with schizophrenia result from deficits in the processing of context, and that these deficits are caused by an underlying neurobiological disturbance – namely, reduced effects of dopamine in PFC. These hypotheses were examined through connectionist simulations of behavioral performance in specific experimental tasks (Braver, Cohen, & Servan-Schreiber, 1995a; Cohen & Servan-Schreiber, 1992). In the present article, we refine this earlier account in a number of ways. First, we provide a more general theory of cognitive control, which provides an account of normal function as well as conditions of impairment. Second, we focus on the role of active maintenance in cognitive control, and how context information might be actively maintained within PFC in order to mediate control. Third, we demonstrate how the inhibitory

processes ascribed to PFC might emerge indirectly, through the modulatory effects of PFC activity on the competitive interactions occurring locally in task-specific pathways. Finally, we rigorously test our hypotheses through a series of behavioral, computational, and neuroimaging studies involving the same cognitive control paradigm - an "AX" version of the Continuous Performance Test (CPT, Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956). The AX-CPT paradigm provides an ideal testing ground for the theory, in that it is simple but well-controlled, and involves separable memory and inhibitory task components. The results of these studies demonstrate the power of the theory in capturing detailed aspects of both behavioral and brain imaging data collected under a variety of task conditions and within different populations.

A THEORY OF COGNITIVE CONTROL

Context and Cognitive Control

Our investigations begin by noting a basic and fundamental function of cognitive control: to flexibly adapt behavior to the demands of particular tasks, by facilitating processing of task-relevant information over other sources of competing information, and by inhibiting habitual, or otherwise prepotent responses which are inappropriate to the task. Because this control function is such a fundamental one, it occurs in even very simple task situations. As a specific example, take a situation where a speeded response is required to a particular stimulus, but only in a particular context (e.g., respond to the letter 'X' only if immediately following the letter 'A'). If the context-stimulus pairing occurs frequently, the cognitive system should begin to exploit the context to prime or facilitate processing of the stimulus. In contrast, in the rare situations where the stimulus occurs in a different context (e.g., X following the letter 'B'), the system must rely upon the information provided by the context in order to inhibit the tendency to respond. As this example makes clear, the cognitive system needs some mechanism which can exploit task-relevant context in order to properly control processing and response selection.

It is this mechanism of cognitive control that is the focus of the current article. In particular, we derive a theory that describes this mechanism in terms of both its computational properties and underlying neural substrates. The theory is composed of three central hypotheses:

1) PFC is specialized for the representation and maintenance of context information; 2) context representations act to mediate control by modulating information flow along the pathways required to support performance of a task; and 3) context information is maintained in PFC as a stable and self-sustaining pattern of neural activity. The first hypothesis was originally suggested by Cohen and Servan-Schreiber in order to capture the cognitive control impairments observed in patients with schizophrenia (Cohen & Servan-Schreiber, 1992). The second of these hypotheses was examined by Cohen and colleagues in a series of connectionist models of attentional performance in the Stroop, Eriksen and related tasks (Cohen et al., 1990; Cohen, Romero, Farah, & Servan-Schreiber, 1994; Cohen, Servan-Schreiber, & McClelland, 1992). The third hypothesis is new and will be specifically examined in the present studies. In particular, we examine how the mechanisms of active maintenance in PFC can account for the dynamics and time course of performance in the AX-CPT.

All three hypotheses presented above focus on internal representations of context information. In previous work, context was defined as prior information that is represented in such a form that it can bias selection of the appropriate behavioral response. More recently, we have attempted to refine this definition in order to better clarify its meaning and provide operational constraints (Cohen, Braver, & O'Reilly, 1996). In particular, we have suggested that representations of context can include task instructions, a specific prior stimulus, or the result of processing a sequence of prior stimuli (e.g., a discourse-level representation resulting from processing a sequence of words in a sentence). Because context representations are maintained on-line, in an active state, they are continually accessible and available to influence processing. Consequently, context can be thought of as a component of working memory, which is commonly defined as the collection of processes responsible for the on-line maintenance and manipulation of information necessary to perform a cognitive task (Baddeley & Hitch, 1994). Context can be viewed as the subset of representations within working memory which govern how other representations are used. Representations of context are particularly important for task situations in which inhibition is required, such as when there is strong competition for action selection. These response competition

situations may arise when the appropriate action is one that is relatively infrequent, or when the inappropriate action is prepotent. In this respect, context representations act in a similar fashion to goal representations within production system architectures. Maintenance of internal goal representations, or goal-related knowledge, is critical for initiating the selection of "weak"behaviors, and for coordinating their execution over temporally-extended periods, while at the same time suppressing competing, possibly more compelling behaviors.

A Connectionist Account

The theory of context and cognitive control described above was developed within the connectionist, or parallel distributed processing framework (McClelland, 1993; Rumelhart & McClelland, 1986). The connectionist framework is a natural one for concomitantly studying the computational and neural mechanisms of cognitive control, since it provides a computational architecture that is specified in neurobiological terms, and can be used to quantitatively simulate performance in cognitive tasks. In particular, the principles of this framework capture central features of computation as it is carried out in the brain. The most relevant of these are:

- *Cognition* is the result of the propagation of activation through interconnected networks of neurons. Activity is required in order to directly influence ongoing processing. Activity states are graded, and change continuously in time, rather than as discrete events.
- *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.
- *Learning* occurs by modifying the strengths of neural connections as a function of activation signals (which can also convey error and reward feedback information from the environment).
- *Variability* is an inherent characteristic of learning and processing, and results from the presence of a noise component within activation states and/or the synaptic connections between units.

- *Representations* are distributed over many neurons and brain areas, and at many different levels of abstraction and contextualization. The distributed nature of representations leads to the learning of general as well as specific information, where generalizations are produced on the basis of similarity metrics.
- *Excitation* is the primary method of communication between regions or modules. Additionally, reciprocal (recurrent, bidirectional) connections are often present between regions, allowing for interactivity in processing, and constraint satisfaction settling.
- *Inhibition* between representations exist at all levels of processing, and occur locally as lateral (within-layer) interactions. Inhibition increases as a function of the number of active representations, which has important computational benefits by enforcing relatively sparse levels of activation.
- *Memory* is achieved either by the relatively short-term persistence of activation patterns (*activity-based memory*) or longer-lasting weight (or bias) modifications (*weight-based memory*). Activity-based memory occurs through the formation of stable (attractor) activity states, arising from recurrent excitatory connections within or between modules.

The theory of cognitive control put forward here can be schematized in the form of a simple canonical model (see Figure 1). In the model, there is a main direct pathway consisting of an input module, an associative module, and an output module. This pathway is intended to capture the transformation or mapping, between raw physical input, the associations or internal representations it generates, and the output or motor response that arises as a result of processing. Although information flow between modules in this pathway is schematized as being feed-forward, in reality there are likely to be feedback connections as well, that allow for interactivity in processing. Connections change strength as a function of experience, which increase the speed and robustness with which information passes between modules. Thus, after multiple experiences with a particular stimulus-response mapping, information flow along the direct pathway becomes robust and relatively automatized (Cohen et al., 1990; Shiffrin & Schneider, 1977).

Insert Figure 1 about here

A second, indirect pathway, involving the context module, represents the functions of PFC in mediating control. There are two critical features of this pathway that provide it with the capacity for control over processing. The first is that there is strong recurrent connectivity within the context layer, which allows for the active maintenance of information. Thus, input to the context layer can be sustained through activity recirculation along mutually excitatory connections, even when the external source of input is no longer present. The second critical feature of the context pathway is its feedback connection to the direct pathway. This provides a means for activity within the context module to provide an additional source of input, which can modulate the flow of processing within the direct pathway[1]. In particular, feedback from the context layer serves to bias the local competition for representation that exists within each module, favoring one activation pathway or set of representations over their competitors. This biasing action of the context module can produce inhibitory effects on processing. These inhibitory effects are most prominent when context provides a bias that changes the outcome of the local competition, allowing a weak pathway to inhibit the more dominant one. It is important to note however, that the inhibitory effects are not direct, but rather emerge out of the interaction between the context module and the direct pathway. Thus, in the model, PFC is not thought to be primarily an inhibitory structure. Indeed, the biasing action of the context module can produce facilitory effects as well. These facilitory effects occur when context adds further support to an already strong pathway, which results in increases in the speed and robustness of information flow in that pathway. So to summarize, the theory postulates two key mechanisms regarding the role of PFC in cognitive control: 1) Context information is actively maintained over time through local recirculation of activity in PFC; and 2) Control occurs through the biasing effects of context representations on task-relevant processing along the pathway from input to response. Below, we review evidence regarding PFC function that supports these mechanisms of control in the model.

PFC and Cognitive Control

A critical component of our theory of cognitive control is that context representations are actively sustained over short periods of time within PFC. As mentioned above, the evidence that PFC is involved in active memory has primarily come from both neurophysiological studies in non-human primates and more recently, from neuroimaging findings in humans. Primate studies have typically examined active maintenance in PFC through delayed-response paradigms, in which the animal must maintain a representation of a cue stimulus over some delay, in order to respond appropriately at a later point. It is now well-established that during performance of these tasks, populations of PFC neurons exhibit sustained, stimulus-specific activity during the delay period (e.g., Fuster & Alexander, 1971; Kubota & Niki, 1971). The mnemonic properties of these neurons has been demonstrated by showing that both local and reversible lesions to PFC impair performance on these tasks, and that performance errors in intact animals are correlated with reduced delay-period activity (Bauer & Fuster, 1976; Funahashi, Bruce, & Goldman-Rakic, 1993; Fuster, 1973). Neuroimaging studies have recently begun to corroborate these findings in humans, by demonstrating that PFC activity is both modulated by active memory load (Braver et al., 1997), and sustained throughout the period over which information must be maintained (Cohen et al., 1997; Courtney, Ungerleider, Keil, & Haxby, 1997). However, the mechanisms by which activity is maintained in PFC have not yet been clearly elucidated. In the model, we have hypothesized that the mechanism by which context representations are maintained is through a recirculating pattern of activity in local circuits. This hypothesis is consistent with recent neuroanatomical evidence, which indicate a high degree of intrinsic excitatory connectivity in PFC, and a pattern of connections suggesting a discrete lattice-like arrangement of segregated modules (Levitt, Lewis, Yoshioka, & Lund, 1993; Melchitzky, Sesack, Pucak, & Lewis, in press; Pucak, Levitt, Lund, & Lewis, 1996). Moreover, this type of connectivity pattern has been shown to be computationally well-suited for active maintenance in artificial neural networks, with unit activity dynamics bearing a strong similarity to the spiking behavior of actual PFC neurons (Zipser, Kehue, Littlewort, & Fuster, 1993).

In our theory, cognitive control emerges through the biasing influence of context representations that interact with task-specific processing. In particular, the theory suggests that context can exert both inhibitory and facilitory effects on processing. The facilitory effect of PFC activity has been directly observed in single-cell recording studies of primates. For example, Miller and colleagues found enhanced activity in cortical association areas (i.e., inferotemporal cortex) during the delay period of a delayed-response task which appeared to be dependent on representations of the cue maintained in PFC (Miller & Desimone, 1994; Miller, Erickson, & Desimone, 1996). In the behavioral literature, it is well-accepted that PFC exerts an inhibitory influence over processing. Indeed, as mentioned earlier, some theorists have postulated that it is the primary functions of this brain region (Dempster, 1992; Diamond, 1990; Fuster, 1989). In terms of cognitive performance, almost all tasks which have been shown to be sensitive to PFC function involve a response competition component. As Roberts et al. (1994) have noted, "a consistent feature of the prefrontal task is that it puts a prepotent tendency in competition with an alternative response (p.375)." Examples of these types of tasks include: the Wisconsin Card Sort, the Stroop task, the A-not-B, and delayed alternation. In all of these tasks, patients with frontal damage show evidence of a failure to inhibit the prepotent response, even when the alternative is the appropriate one (Diamond & Goldman-Rakic, 1989; Freedman & Oscar-Berman, 1986; Milner, 1963; Perret, 1974).

Taken together, the empirical literature is very consistent with the idea that: 1) PFC plays a critical role in the active maintenance of information; and 2) information maintained in PFC acts to mediate control by biasing the processing of task-relevant information. However, stronger tests of this hypothesis could be achieved through more focused studies that are directly motivated by the theory. For this reason, we have conducted a series of empirical and simulation studies, presented below, which examine the mechanisms of control. These studies provide a more systematic investigation of how well the theory can predict, explain, and account for a rich set of data collected under multiple task conditions, methodologies, and population groups.

OVERVIEW OF THE AX-CPT PARADIGM AND PRESENT STUDIES

The selection of an appropriate experimental paradigm for investigation is critical when studying a process as rich and complex as cognitive control. The decision to conduct a series of studies using the AX version of the CPT was based upon a number of desirable characteristics of this task. First, the CPT has been widely studied in the clinical and neuropsychological literatures as a test of attentional control and vigilance (e.g., Nuechterlein, 1991). The more demanding versions of the paradigm are thought to rely on PFC function, as evidenced by performance deficits observed in patients with frontal lesions (Glosser & Goodglass, 1990) and other syndromes thought to involve prefrontal dysfunction, such as schizophrenia (Cornblatt & Keilp, 1994) and ADHD (Losier, McGrath, & Klein, 1996). Many neuroimaging studies have also used versions of the CPT to elicit PFC activity (Cohen et al., 1987; Rezai et al., 1993; Seidman et al., 1998; Siegel, Nuechterlein, Wu, & Buchsbaum, 1995). Second, the AX-CPT probes key aspects of cognitive control, while distilling them into a task paradigm which is as simple and tightly controlled as possible. The tightness of experimental control maximizes the interpretability of experimental results. The simplicity of the task enables it to be used with many different subject populations, and under a wide variety of task environments. Indeed, the task is very similar in structure to delayed-response tasks used in the neurophysiological literature on working memory, and thus allows easy comparison with this literature. Third, because the task is relatively simple it can be simulated in computational studies. Finally, although the task is simple, it nevertheless produces multiple performance measures which generate a rich set of data on which to base and constrain theoretical interpretations.

In the AX-CPT, subjects observe single letters presented one at a time at the center of a visual display, and are required to respond to a target probe letter (X) but only when it follows a designated cue (A). Performance is dependent on the representation and maintenance of context information, insofar as the correct response to the probe depends on knowledge of the previous cue (A or not-A). In standard versions of this task, target sequences (e.g., A-X) are typically of low frequency (e.g., 20%), and stimuli are presented at a relatively fast rate (every 1-2 seconds). In our version of the paradigm, we modifed two aspects of the task to allow for a more focused

investigation of both memory and inhibitory influences on processing. First, we selectively manipulated the delay between cue and probe in order to vary the demand on active memory. In particular, imposing a long delay between cue and probe (i.e., 5 sec or longer) produces a greater demand for context information to be actively maintained over time. Thus, active maintenance processes can be probed by measuring performance in long delay relative to short delay (i.e., 1-2 sec) trials.

Our second modification of the AX-CPT was to increase target frequency. This manipulation enabled an examination of the role of context in both biasing response competition situations, and inhibiting dominant response tendencies. Specifically, with a high frequency of target sequences (i.e., 70%), a strong association should develop between the A and X (i.e., on 87.5% of trials in which an A occurs, it is followed by an X; on 87.5% of trials in which an X occurs, it will have been preceded by an A). This association should produce two effects: 1) a prepotent tendency to make a target response to an X probe, irrespective of the preceding cue; and 2) an expectancy to make a target response following an A cue, irrespective of the subsequent probe. In such a situation, response competition effects can be examined by comparing performance in nontarget trials where the context information conflicts with the probe (BX and AY, where "B" corresponds to any non-A stimulus, and "Y" to any non-X) against performance in a third type of trial (BY) in which context agrees with the probe.

The theory of cognitive control that we have put forward provides a useful framework in which to generate insights and predictions regarding the influence of context on AX-CPT performance. There are three classes of predictions which can be made from the theory. First, context representations of the cue should facilitate appropriate responses while inhibiting inappropriate ones. Thus, following an A cue, context should act to prime the target response pathway. On AX trials, this should facilitate robust responding. However, on AY trials, the effects of context should have a negative consequence. Because context enhances the activity of the target response pathway, it should actually serve to *increase* interference in processing the probe. In other words, because of its facilitory effects on AX trials, context should have the effect of

inhibiting correct responses on AY trials, thus leading to slowing and/or increased errors. Following a non-A cue, context should act to prime the non-target response pathway by enhancing its activity. On BY trials, this has a facilitory effect by further strengthening processing. On BX trials, the context effect can be seen primarily as an inhibitory one, because it enables the non-target response to effectively compete against the otherwise dominant influence of the probe. Thus, to summarize, the model suggests that the effects of context will be most apparent in BX and AY trials, since these are the conditions involving response competition. Moreover, the differing sources of response competition on these two types of trials, allows for the generation of a counter-intuitive, and powerful prediction by the cognitive control model. Namely, the presence of strong representations of context is predicted to benefit BX performance, but to impair AY performance. Conversely, the presence of weak context representations should cause BX performance to worsen, but AY performance to improve. These effects should be observable both in terms of errors and RT measures.

A second class of predictions concerns the assumption of the theory that context is actively maintained. Thus, strong representations of context should be sustained across the delay period between cue and probe. The delay manipulation provides an opportunity to examine this prediction. In particular, when context is actively maintained without decay, its influence on performance should be just as great at the long delay as at the short. However, in cases where context representations decay over time, performance should be less influenced by these representations after a long delay. In terms of BX trials, this would cause a delay-related *decrement* in performance. In contrast, for AY trials, decaying representations should lead to a delay-related *improvement* in performance. Thus, the model predicts that the interaction of AY and BX errors with the strength of context representations should further interact with delay. The third class of predictions concerns the hypothesis that representation of maintenance of context information occurs within PFC. These predictions should be observable in both the pattern and dynamics of PFC activity during task performance. In particular, the theory predicts that PFC activity will be greater in the long delay condition relative to the short, because it is maintained across a longer

period. The theory further predicts that if context representations are strong, the activity level in PFC should be maintained over the entire delay interval. However, if context representations are decaying over the delay interval, this should also be observable in terms of decreasing PFC activity during this period.

The purpose of the present research was to test these predictions of the cognitive control model using behavioral, neuroimaging, and simulation data. In particular, through a series of seven studies we tested the following hypotheses:

- Study 1: The behavioral performance of healthy individuals performing the AX-CPT under normal conditions will reflect the presence of strong context representations, both in terms of its relationship to the various trial types, and in the effect of delay.
- Study 2: Simulations using the cognitive control model will successfully capture both the quantitative and qualitative characteristics of the normal behavioral data.
- Study 3: Simulations in which the representation and maintenance of context is weakened will make explicit the predictions of the cognitive control model regarding the consequence of this disturbance on AX-CPT behavioral performance.
- Study 4: The simulation results of Study 3 will predict the specific pattern of performance observed in patients with schizophrenia, who are hypothesized to suffer from PFC-mediated disturbances in cognitive control.
- Study 5: An experimental manipulation in healthy individuals which disturbs context processing will produce the same qualitative pattern of AX-CPT performance as that observed in schizophrenia patients.
- Study 6: Functional neuroimaging of PFC activity of AX-CPT performance will reveal an activation profile and dynamics that reflects a specific role in the representation and maintenance of context information.
- Study 7: Functional neuroimaging of AX-CPT performance under conditions when context processing is disturbed (using the experimental manipulation of Study 5) will reveal a pattern of altered PFC dynamics specifically predicted by the model.

GENERAL METHODS

All of the studies described below involved performance of the modified version of the standard AX-CPT paradigm, just described. Here we describe methods that were applied generally across studies. More specific methods that applied to particular studies are described under the appropriate section for those studies.

Participants

Participants in all studies had normal or corrected-to-normal vision and were native English speakers. Informed consent was obtained in accordance with the institutional review board, and a cash payment was given in return for participation.

Tasks and Apparatus

Participants were tested during performance a modified version of the AX-CPT. Each "trial" consisted of a cue-probe sequences. These were presented in a continuous fashion, in pseudorandom order, such that target sequences occurred with 70% frequency (A-X trials) and non-target sequences occurred with 30% frequency. Non-targets were divided evenly among the following trial types: "B-X" trials, in which an invalid cue (i.e., non-A) preceded the target; "A-Y" trials, in which a valid cue was followed by a non-target probe (i.e., non-X); and "B-Y" trials, in which an invalid cue was followed by a non-target probe. The delay between cue and probe was manipulated so that half of the trials had a short delay and half had a long delay. The exact duration of the two delay lengths varied across studies, but was always between 1-2 seconds for the short delay and 5-10 seconds for the long delay. Inter-trial interval was counterbalanced with delay, such that total trial duration was equivalent across conditions. This provided a means of controlling for general factors that might affect performance (e.g., pace of the task, response frequency, total time on task).

Letters were presented on a visual display, occurring centrally with a duration of 300 msecs (except where otherwise noted), in 24 point uppercase Helvetica font, red against a black background. The letter A was used as the cue, and the letter X was the target. The remaining letters of the alphabet served as invalid cues and non-target probes, with the exception of the letters K and

Y, which were excluded due to their similarity in appearance to the letter X. Subjects responded to every stimulus (both cue and probe), pressing one button for targets and another button for nontargets (cues were always considered non-targets). The response buttons were located on a specially constructed box connected to the computer, which recorded both response choice and reaction time with 1 millisecond accuracy. Responses were made with the middle and index fingers of the right hand. Participants had 1.3 seconds from stimulus onset in which to respond (regardless of interstimulus or intertrial interval). Responses that were slower than this limit were not recorded. Tasks were run on Apple Macintosh computers, using PsyScope software for stimulus presentation and data collection (Cohen, MacWhinney, Flatt, & Provost, 1993).

Procedure

Except where otherwise specified, participants were tested in a single testing session. Standardized instructions describing the experiment appeared on the computer, and the experimenter answered any remaining questions regarding the instructions. Subjects were asked to respond as quickly as possible to each stimulus, while maintaining accuracy. Practice trials were given until subjects felt comfortable with the task and the experimenter confirmed that the task was being performed properly. All participants were able to understand and perform the task within 20 practice trials. During the task proper, short and long delay trials were sometimes segregated into short blocks of trials. The number of blocks and trials within each block varied across studies. Subjects were allowed to rest between each block, and controlled the start of the next block by pressing a mouse button.

Data analysis

Data were analyzed using error rates (misses and false alarms), signal detection indices (d') and RTs as the dependent measures of interest. RTs were examined for correct responses only, unless otherwise noted. Analyses of non-target error rates and RTs were conducted with repeated measures ANOVAs with trial type (AY,BX,BY) and delay (short, long) as within subjects factors. Analyses of target trial error rates and RTs were conducted using paired t-tests. Target (i.e., AX) and non-target trials were analyzed separately because of their different response requirements and

their different frequencies of occurrence (i.e., 70% for AX trials, 10% for each of the non-target trials). Additional analyses, described in detail under Study 1, used response speed (fast, slow) as another within-subject factor to further examine speed-accuracy relationships. For the signal detection measures, a correction factor was applied in cases of a perfect hit rate (1.0) or false-alarm rate (0.0), to allow an unbiased estimation of d' (Nuechterlein, 1983)[2]. In addition to the traditional computation of d' (i.e., using hits and all false alarms), d' was also computed using just BX false alarms. This additional measure, hereafter referred to as d'-context, has been used in previous AX-CPT studies to provide a more specific index of sensitivity to context (Cohen, Barch, Carter, & Servan-Schreiber, 1999; Servan-Schreiber, Cohen, & Steingard, 1996). For dependent measures involving accuracy or response proportions, results were confirmed after normalizing the data using an arcsine transformation (Neter, Wasserman, & Kutner, 1990). For measures involving reaction time, traditional analyses were confirmed by using an inverse transformation of the data, to rule out spurious effects of outliers (Ratcliff, 1993). Effects that conformed to *a priori* predictions were evaluated with planned contrasts. Effects that were not predicted were evaluated in post-hoc contrasts after applying the appropriate Bonferroni corrections for multiple comparisons.

Study 1: Characteristics of Behavioral Performance

The first study was conducted in order to determine the behavioral characteristics of normal performance in the AX-CPT. The goals were twofold: a) to demonstrate both inhibitory and facilitory effects of context; and b) to demonstrate the role of context in active maintenance. Additionally, we attempted to describe the set of behavioral findings in detail, so as to tightly constrain subsequent simulation-based accounts of the processing mechanisms underlying task performance. Because of this last goal, it was critical that the findings be highly robust and reliable. However, because of its relatively simple nature, healthy adults usually find the AX-CPT to be an easy task, and are typically very fast and accurate in their responses. Thus, we expected that many of the effects present in the data would be subtle ones. As a result, we felt it necessary to maximize statistical power to detect these effects by pooling data gathered from a large sample of participants and collected across different experimental studies. Although the studies differed with respect to

some task parameters (such as delay duration, and number of trials per condition), we decided to pool the data directly, because of an interest in effects that would be robust to these changes.

Method

Participants.

A total of 209 participants recruited from the Carnegie Mellon community were included in this data set. The total data set includes a subset of data collected from 101 healthy individuals in the baseline AX-CPT that will also be discussed individually in later experiments. This subset of data comprises the 16 participants from Study 4, the 62 participants from Study 5, and the 23 participants from Study 6. Across the entire pooled data set, there were 108 males and 91 females. The mean age of participants was 23.1 years (range 18 to 51 years).

Task and Procedure

All participants performed the two delay conditions of the AX-CPT as described under General Methods. The primary variables which differed across studies were delay durations and number of trials performed. The delay duration varied between 1000-2300 msec for the short delay condition and 4900-8000 msec in the long delay condition, but in every study the difference between long and short delay durations was at least 3900 msec. A minimum of 5 trials were collected per participant for every cell of the task, but the cell size varied between 5-15 trials across studies.

Results

To organize the discussion of results, data will first be discussed in terms of main effects, and then in terms of interactions. All data below refer to responses to the probe stimulus on each trial. Although subjects responded to the cue as well as the probe on each trial, cue responses were not of theoretical interest and were only required as a means of increasing task difficulty in a non-specific manner. In general, cue responses were fast ($\underline{M}$ = 459.23, $\underline{SD}$ = 115.79) and highly accurate (less than 0.5% errors). Mean error rates and RTs for all conditions are shown in Figure 2.

Insert Figure 2 about here

Main effects of trial type

Non-target effects were examined through an analysis of trial-type (i.e., AY,BX,BY). There was a main effect of trial type on false alarm rates ($F(2,416) = 66.80$, $p<.001$), with participants making significantly more BX ($F(1,208) = 75.26$, $p<.001$) than BY false alarms (see Figure 2). There were also significantly more AY than BY false alarms ($F(1,208) = 157.28$, $p<.001$). Moreover, in directly comparing AY and BX trials, AY errors were found to be significantly more frequent than BX errors ($F(1,208) = 15.32$, $p<.001$). The ANOVA for RTs mirrored the results of the error analysis: a main effect of trial type ($F(1,208) = 271.85$, $p<.001$), with participants making significantly slower responses on AY ($F(1,208) = 980.17$, $p<.001$) and BX ($F(1,208) = 114.99$, $p<.001$) trials than on BY trials (see Figure 2). In directly comparing these two sequences, it was found that AY trials were slower than BX trials ($F(1,208) = 109.93$, $p<.001$).

Main effects of delay

In terms of d', performance was significantly worse ($F(1,208) = 24.62$, $p<.001$) in the long delay ($M = 3.30$, $SE = 0.04$) relative to the short ($M = 3.50$, $SE = 0.05$). However, the delay effect was only found for target errors ($t(208) = 5.47$, $p<.001$). False alarms did not change as a function of delay across the three non-target trial types ($F(1,208)=1.08$, $p>.1$). Subjects also had significantly slower RTs in the long delay condition relative to the short for both target ($t(208)= 8.33$, $p<.001$), and nontarget responses ($F(1,208) = 43.01$, $p<.001$).

Delay x Trial Type Interactions

Although the delay manipulation produced no main effect on non-target errors, a significant trial type x delay interaction was observed ($F(2,416) = 3.63$, $p<.05$). As illustrated in Figure 2, post-hoc contrasts suggested that the interaction was caused by a delay-related increase in AY errors with delay ($F(1,208)=3.93$, $p<.05$), combined with a non-significant tendency for BX errors to decrease with delay ($F(1,208)=1.04$,$p>.1$), and no effect of delay on BY errors ($F(1,208)=0.1$,$p>.1$). The delay manipulation had a similar effect on response speed in AY and BX trials. In particular, the main effect of delay on RT was further modulated by a delay x trial type

interaction ($F(2,416) = 6.96$, $p=.001$). Post-hoc contrasts suggested that the delay effect on was significantly greater for AY than BX trials ($F(1,208)=9.91, p<.01$).

Speed-Accuracy Relationships

The previous analyses examined the influence on accuracy and RT as two independent effects (by analyzing accuracy independent of RT, and analyzing RT after controlling for errors). Yet it is well known that tasks such as the AX-CPT, in which subjects are instructed to respond as fast as possible without sacrificing accuracy, are best analyzed in terms of a *response function* that describes the relationship between accuracy and RT (McClelland, 1993). To more specifically test speed-accuracy relationships, accuracy was conditionalized on RT across trial types.[3] Because only a small number of AY and BX trials were performed by each subject (5-15 trials per delay condition), we did not attempt to perform a full speed-accuracy analysis (conditionalizing accuracy in a parametric fashion with respect to RT as has been typically done in the literature, e.g., Gratton, Coles, Sirevaag, Eriksen, & Donchin, 1988). Instead, the data was split in a binary fashion, with each response categorized as either fast or slow, on the basis of whether it was faster or slower than the subject's median RT for that condition[4]. Response speed was then included as an additional factor in the ANOVA analysis. As shown in Table 1, a significant interaction in accuracy was observed between trial type and response speed ($F(2,414) = 118.01$, $p<.001$). Planned contrasts confirmed that the interaction was due to AY errors being greater in fast relative to slow responses ($F(1,207) = 129.34$, $p<.001$), and BX errors being greater in slow relative to fast responses ($F(1,207) = 18.62$, $p<.001$).

Insert Table 1 about here

The trial type x response speed interaction was also modulated by a further interaction with delay ($F(2,414) = 5.21$, $p<.01$). Post-hoc contrasts suggested that for early responses AY errors increased with delay ($F(1,207)=3.88, p=.05$), and BX errors significantly decreased ($F(1,207)=6.71, p=.01$). However, neither of these effects was significant for late responses (AY: $F(1,207)=0.64, p>.1$; BX: $F(1,207)=0.41, p>.1$). A similar analysis was performed on AX target

responses. There was a significant interaction between delay and response speed ($F(1,207)=17.07$, $p<.001$). As can be seen in Table 1, the delay effect was greater in slow responses than it was in fast responses. However, the delay effect was still present when only fast responses were considered ($F(1,207) = 6.01$, $p<.05$).

Discussion

The first goal of this study was to demonstrate both inhibitory and facilitory effects of context in the AX-CPT. The analyses of trial-type effects for non-target trials accomplished this goal. Inhibitory effects were measured in terms of performance on BX trials. As discussed above, the high frequency of target trials was expected to produce a prepotent tendency to make a target response whenever an X probe appears. If participants were not successfully maintaining a representation of the context provided by the cue, they would not be able to inhibit this prepotent tendency. Consequently, we would expect to observe poor BX performance. We did find evidence that the prepotent influence of the probe produced competition for response on BX trials, since errors and RT were both increased relative to the BY trials (which did not involve response competition). Nevertheless, overall BX performance was high in terms of both accuracy (less than 7% errors) and RT (~500 msec). The fact that BX performance was relatively spared suggests that context representations were used as a mechanism to inhibit target responding.

On AY trials, context effects also appeared to be evident in performance, as reflected by increased errors and RTs relative to BY trials. These effects demonstrate that the cue must have been represented as context on trials with A as well as non-A cues. In AY trials, the information provided by the cue is the only source of interference which could produce competition for response (given its association with the target probe). Both of the trial-type effects were highly significant and robust. Even though this study included data collapsed across 7 separate data sets, we observed that context effects on AY and BX trials were still significant when each of these data sets were analyzed individually[5] . Thus, the manipulation of target frequency in the AX-CPT, produced a situation in which representations of context both improve performance in some aspects of the task (i.e., BX trials), but also impair it in other aspects (i.e., AY trials). Moreover, because

context had opposing effects on these aspects of performance, its influence over processing could be determined with a high degree of experimental precision (i.e., as a statistical interaction).

The findings regarding AY performance are novel, in that they suggest that context is represented even in situations when its influence may have negative effects. This is especially interesting, since it is plausible that only invalid cues would be represented as context. In fact, at first blush, this would seem to be the most efficient performance strategy, as it requires representing context only on a small proportion of trials (20%), in which context is needed to override a prepotent response. Moreover, both faster and more accurate responses would likely be observed on AY trials if the cue were not represented as context, since here the context only serves to produce competition in responding. Nevertheless, the behavioral performance data indicates that context *is* represented on valid cue trials, which suggests two possible interpretations. First, that there is some constraint on information processing, such that it is not possible for participants to select whether a cue should be represented as context on a trial-by-trial basis. For example, it may be that access to context representation is not determined based on the identity of a stimulus, but rather only through its designation as a cue. A second possible interpretation is that representing context on valid cue trials has a facilitory effect on AX responding, which outweighs its negative effects on AY trials. Indeed, given that AX trials are 7 times more frequent than AY trials, this latter interpretation is a very plausible one. However, it was not possible to directly test this hypothesis in the current study, since it is not clear how to measure facilitation effects in AX trials. In Studies 4 and 5 we provide evidence that context representations do serve to facilitate performance on both AX and BY trials.

The second goal of the study was to demonstrate that context representations are actively maintained during performance of the AX-CPT. This goal was achieved by demonstrating that BX errors were no greater following a long delay than a short one. In other words, context was just as able to inhibit the prepotent influence of the probe following a long delay. This indicates that, in healthy individuals under normal conditions, context representations can be sustained without decay over a delay interval. Additionally, significant delay x trial type interactions were observed in both

accuracy and RT. These interactions appeared to be due to greater effects of delay on AY relative to BX trials. AY errors increased with delay while BX errors slightly decreased, and the delay-related increase in AY RT was significantly greater than the effect on BX RT. In other words, while delay slightly improved BX performance, it also produced a decrement in AY performance. Both of these effects suggest the interpretation that context actually exerted a stronger influence at the long delay than at the short, since the representation of context acts to improve BX but impair AY performance. This finding is somewhat of a surprising one, and may indicate that representations of context are not only sustained over delay periods, but that they take a relatively long time to reach full strength (i.e., > 1 sec). In Study 2, we conduct simulations of the AX-CPT which provide support for this hypothesis.

There were additional effects present in the data that helped to further characterize the processing mechanisms which contribute to AX-CPT performance. In particular, examining speed-accuracy relationships provided information regarding the dynamics of response competition. AY errors were found to be more likely to occur in the early part of the speed-accuracy curve, whereas BX errors were more likely to occur in the late part of the curve. This finding further supports the idea that the competition in AY and BX trials arises from two different patterns of interaction between context and probe. In AY trials, context primes the incorrect response, and the probe must override this tendency. Thus, the most likely cause of errors in AY trials is from making a target response (primed by the A cue) before the probe is fully evaluated as a non-X. This effect was seen in the data as a greater AY error rate on fast response trials. The opposite pattern occurred in BX trials. On these trials, context primes the correct response, but the probe interferes with this response tendency. Thus, more BX errors occurred on slow response trials, in which the probe likely had too strong of an influence over responding, something that only occurs after enough time had elapsed for it to be fully processed. We also found that these speed-accuracy relationships interacted with delay. For non-target trials, a delay x trial type x response speed interaction was observed, such that the effect of delay on AY and BX errors was only present in the early part of the speed accuracy curve. A delay x response speed interaction was also observed for

target trials, such that the delay effect was smaller for early responses than it was for late responses. Together, these findings suggest that the influence of context may be reduced for late responses. In particular, one possible interpretation is that there are non-specific task factors that may contribute to slow responding, such as fatigue or motivation. This interpretation might account for the increase in target errors for late responses, since they might reflect reduced attention to the task. Nevertheless, further research is needed in order to evaluate this possibility more fully.

The direct comparison of AY to BX performance provided further information regarding the nature of the response competition process. Specifically, we observed that performance on AY trials was both slower and more error-prone than on BX trials. Interestingly, this result was not found in two previous AX-CPT studies (Cohen et al., 1999; Servan-Schreiber et al., 1996). A primary difference between these previous studies and the current one is that in the current study participants were required to respond to both targets and nontargets, and to both cue and probe stimuli, whereas in the previous studies participants only made target responses. One interpretation of this change in the current study is that it affected both task difficulty and response threshold, and that the change in these two variables altered the balance between AY and BX performance. Under this interpretation, task difficulty increased as a consequence of the increase in response requirements (i.e., responding to all stimuli rather than just targets), while response threshold was lowered as a result of the increased response frequency (i.e., responding twice during every trial rather than once during only a portion of trials).

If the increased response frequency of the current study did, in fact, lead to participants lowering their response threshold, it might provide an explanation for the observed relationship between AY and BX errors. As discussed above, AY errors are likely to be greater when responses are made quickly, before the probe is completely processed. In contrast, BX errors are more likely when responses are made slowly, after the probe has had time to exert its full influence. Thus, a lowering of response threshold should cause an increase in AY errors but a decrease in BX errors. As a consequence, it may be that the relationship between AY and BX performance in the current study reflects the change in response threshold. Moreover, this argument suggests that even within

the current study, there should be an association between participants' response speed and the amount of AY and BX errors they make. In particular, there should be a positive association between BX RT and error rate, but a negative association between AY RT and error rate. A post-hoc regression analysis supported this prediction, demonstrating a modest positive correlation for BX responses ($r(208) = 0.256$, $p<.001$), and a modest negative correlation for AY responses ($r(208) = -0.293$, $p<.001$). Interestingly, the slope for AY responses (3.27% *increase* in errors per 100 msec decrease in RT) was twice as great as that for BX responses (1.56% *decrease* in errors per 100 msec decrease in RT). These slope differences suggest that changes in response threshold will more strongly influence AY error rates than BX error rates.

The main effect of delay on performance also provided further information regarding the processing mechanisms in the AX-CPT. In the accuracy data, there was not a main effect of delay on nontarget errors, but there was a delay-related increase in target errors. This result appears to be inconsistent with the theory, and could potentially indicate the presence of decay mechanisms affecting context representations even in healthy adults. However, this interpretation is an unlikely one given the lack of a main effect of delay on nontarget accuracy. Moreover, the finding that the target delay effect was greater for slow responses further argues against a context-based interpretation for the delay-related increase in target errors, since slow responses appear to be less influenced by context. Nevertheless, the result does appear to be statistically reliable, and should be taken into consideration in any account of AX-CPT performance. In Study 2, we show how simulations of the AX-CPT may provide a possible explanation for this effect. A main effect of delay was also observed in the RT data. The source of this effect is not clear, but previous studies have demonstrated that there number of factors which influence response latency in cued reaction time tasks. One such factor is motor priming, where one part of a sequence primes another when there is a short interval between them (Bertelson, 1961; Cleeremans & McClelland, 1991; Remington, 1969). Thus, RTs may be slower at long delays due to an absence of motor priming effects. Another factor which has been shown to cause slowing of RTs after a long delay is uncertainty about the timing of the upcoming stimulus (Shakow, 1962). Future studies will be

needed to examine this issue more closely, to determine which of these factors, or others, contribute to delay-related slowing.

Study 2: Simulating Basic Cognitive Control Effects

The results of Study 1 demonstrate the presence of response competition effects in performance of the AX-CPT. These results are consistent with the mechanistic account suggested by the cognitive control model : Response competition in the AX-CPT arises from the interaction of actively maintained context information provided by the cue with the prepotent response tendencies produced by the probe. Nevertheless, the question still remains whether the mechanisms postulated in the model are actually sufficient to capture the qualitative and quantitative aspects of the behavioral data. Additionally, there were a number of other behavioral effects that were not initially predicted by the theory. These include: 1) the delay x trial type interaction for nontarget errors and RT; 2) the main effect of delay on target errors; and 3) the delay x trial type x response speed effect on nontarget accuracy. It is not clear whether any or all of these additional effects could also be accounted for from the cognitive control mechanisms already postulated, or whether there are aspects of AX-CPT performance which remain outside of the scope of the theory. Moreover, this issue cannot be adequately addressed from examination of the theory as a purely conceptual entity. However, since the theory is specified in terms of the connectionist computational framework, it should be possible to implement it as an explicit simulation. Thus, in the current study, we conducted computer simulations to test whether the processing mechanisms we have postulated to occur in the AX-CPT can produce the particular pattern of performance observed empirically. We relied upon the rich set of behavioral data generated from this task to provide tight constraints on both the development of the model and the evaluation of its success in accounting for normal performance.

Methods

Architecture and Processing

As mentioned above, the connectionist, or PDP framework was used to constrain the development and implementation of the AX-CPT model. The architecture of the model (see Figure

3) was kept consistent with both the general principles of this framework enumerated previously, and with the canonical model of cognitive control. However, individual stimuli and responses were represented in a localist manner, to keep the model as simple and interpretable as possible. Specifically, the model consisted of a direct pathway composed of feed-forward connections between a pool of input units (4), representing the four stimulus conditions (A,B, X or Y), a pool of associative units (4), and a pool of output units (2). In addition, the cue inputs also projected to a layer of context units (2). The context layer then projected back to the pool of associative units in the direct pathway. Within each pool of units, there were lateral inhibitory connections which produced competition for representations. Additionally, within the context layer, units had strong (non-modifiable) self-excitatory connections (+6.0 weight)[6], which allowed context layer activity states to be sustained over time, in the form of a fixed-point attractor. Moreover, this pattern of connectivity reflects the assumption that local recurrent activity is the primary mechanism by which active maintenance occurs in PFC, an assumption similar to that proposed in previous active memory models of PFC (Dehaene & Changeux, 1989; Guigon, Dorizzi, Burnod, & Schultz, 1995; Moody, Wise, di Pellegrino, & Zipser, 1998; Zipser, 1991).

Insert Figure 3 about here

Processing occurred continuously over time in the model. Unit activations were governed by temporal difference equations, which included a zero-mean Gaussian noise component and a fixed negative bias on all units (-2.5)[7]. The bias parameter caused units to assume a low baseline state of activation (similar to that thought to occur in most biological neurons). The duration of relevant events within the simulation (e.g., cue and probe presentation, delay periods) were scaled to approximate the temporal relationships used in the actual task. Thus, the cue and probe were each presented for 2 time steps, the short delay lasted 7 time steps, and the long delay lasted 33 time steps. The presentation of each stimulus was simulated by adding an external source of activation (i.e., soft-clamping) to units in the input layer for a short duration. Activation states were then allowed to evolve in response to this external input. All input units were provided some external source of activation during presentation of every stimulus, in order to approximate the effects of

overlap in distributed representations, and lateral competition at the sensory stage of processing. The external input to the stimulus condition being presented was 5.0 and was 2.5 to the competing input. Following learning of lateral inhibitory weights, this translated into a peak activation of approximately 0.8 for the presented stimulus, and approximately 0.2 for the competing stimulus condition. The partial activation of competing inputs simulated the initial ambiguities that may arise during letter identification.

Training

Connection weights within the network were adjusted through training with a continuous recurrent version of the backpropagation supervised learning algorithm (Pearlmutter, 1989). The training procedure consisted of repeated presentations of each of the 8 different trial types of the AX-CPT (AX,AY,BX and BY at both short and long delays), with the presentation frequency of each type matching that of the behavioral task. This learning approach enabled optimization of weight strengths based on both the constraints of task performance and the relative frequencies of task events[8]. Before training, all modifiable connections were initialized with small random weights. Within-layer weights were constrained to be negative and took initial values in the interval [-1,0]. Between-layer weights were constrained to be positive and took initial values in the interval [0,1]. Each trial was simulated as a sequence of events occurring in the following order: cue, delay, probe, ITI. Responses to the cue were not of theoretical interest in this simulation, and as discussed in Study 1, were assumed to only have a general effect on task performance (i.e., overall task difficulty and response threshold). Consequently, the model was only trained to respond to the probe stimulus. Following presentation of the probe, activation of the output units was compared to that of the correct response for that trial type. Weights were adjusted according to the discrepancy between the network's response and the correct one. Because network responses were stochastic (due to the noise component added to each unit's input), a response was only judged to be discrepant if the output unit's activation was on the wrong side of 0.5[9]. During training, the learning rate was set at 0.01 and momentum was set at 0.9. The standard deviation of the noise term (σ) was gradually increased during training (from 0.01 to 0.5). Training continued until sum-

squared error across all trials dropped to a preset criteria (< 2.0). This was achieved after approximately 500 presentations of each trial type.

Simulations and Data Analysis

Following training, weights were fixed. Simulations of performance on each trial type were conducted by determining which of the output units was the first to surpass a prespecified threshold value (T). If neither unit reached threshold during the ITI period of the trial, that trial was counted as a non-response (the percentage of non-response trials was comparable to that observed empirically -- less than 3%). Simulations were run in groups of 200 hundred trials, with the frequency of each trial type matching that used in the empirical task (i.e., 70% targets, 10% of each non-target type; 50% short delay trials, 50% long delay trials). Each run of trials was treated as a "subject", with accuracy and RT statistics computed for the run. We collected this data across 209 runs in order to produce a data set for the model that was equivalent in terms of sample size to that collected for the behavioral study.

We attempted to satisfy the basic constraints of the behavioral data by adjusting the two free parameters (T and σ) in the model. Specifically, we optimized the model to fit the primary qualitative effects of response competition that were observed in the behavioral data. These were the main effects of trial type found for non-targets in both errors and RT. Moreover, we optimized the fit to capture the ordinal relationship observed in the behavioral data among the 3 non-target types, which was AY > BX > BY for both errors and RT. These qualitative fits of the model were determined by subjecting the simulated data to analyses that were identical to those performed on the empirical data (i.e., a repeated measures ANOVA). Additionally, we optimized the quantitative fit of the model to the accuracy data. We did not try to optimize the quantitative fit of the model to the RT data because there were a number of factors which would complicate this effort[10]. Quantitative fits were examined by conducting ANOVA analyses directly comparing the simulation to empirical data to determine whether there were any main effects of data source (simulation vs. empirical data), or any interactions between source and the two primary task factors (trial type and delay).

We then performed a set of additional analyses using these best fitting parameter values for the model. These analyses were conducted to determine whether the model captured additional properties of the behavioral data that were not specifically optimized. These included: the delay x trial type interaction in both accuracy and RT; the delay effect on target accuracy; and speed-accuracy relationships as a function of trial-type and delay. Finally, we conducted a series of analyses examining activation dynamics of network units, in order to determine the mechanisms underlying its performance. These analyses were performed by averaging activity with in context and response layer units, at every time step, across all trials. The first analysis examined correct and incorrect trials separately, collapsing across short and long delay conditions. The third analysis examined short and long delay activity separately, collapsing across correct and error trials.

Results

Parameter effects

The noise parameter was found to have a strong effect on overall model performance, increasing the confusability between the different cue and probe types. The effect of the response threshold interacted with the noise parameter. In particular, for a given moderate value of noise, the threshold appeared to affect the relative balance between AY and BX errors. This effect was similar to what we had hypothesized in Study 1. Namely, at a high threshold there were more BX than AY errors, while at a low threshold the opposite pattern was observed. Furthermore, as was also observed in the behavioral data, manipulations of response threshold appeared to have a greater effect on AY errors than on BX errors. Based on a search of this space, the parameter values that best fit the behavioral data were $T=0.69$ and $\sigma=0.95$.

Qualitative Fits

Analyses of the simulation accuracy data revealed that the main effect of trial type was highly significant for nontarget responses ($F(2,416) = 184.63$, $p<.001$). Planned contrasts confirmed that the effect was in the predicted direction, with more AY than BX errors ($F(1,208) = 10.72$, $p=.001$) and more BX than BY errors ($F(1,208) = 286.19$, $p<.001$) (see Figure 4). A similar pattern was found in the simulation RT data, with a highly significant main effect of trial type

($F(2,416) = 3457.57$, $p<.001$), which was due to slower AY than BX responses ($F(1,208) = 2619$, $p<.001$), and slower BX than BY responses ($F(1,208) = 531.64$, $p<.001$) (see Figure 4).

Insert Figure 4 about here

Quantitative Fits

There were no main effects of data source in the nontarget accuracy data ($F(1,416) = 0.0$, $p>.1$). Additionally, there was no interaction of data source with delay ($F(1,416) = 0.04$, $p>.1$). The interaction of source with trial type was marginally significant ($F(2,832) = 2.65$, $p=.07$), with the model making more AY and BY errors but less BX errors than was found empirically. However, posthoc contrasts examining each trial type separately showed that none of the differences between model and empirical data were statistically significant (all p's $> .1$). Finally, the 3-way interaction between data source, delay and trial type was also not significant ($F(2,832) = 0.22$, $p>.1$). For target accuracy, there was no main effect of data source ($F(1,416) = 1.54$, $p>.1$). However, the interaction between data source and delay was significant ($F(1,416) = 6.09$, $p= .01$), which was due to the empirical data showing a larger delay effect than the simulation.

These patterns were confirmed in a more detailed analysis comparing the simulation and empirical results separately for each of the 8 experimental conditions. These analyses confirmed that there were no significant differences between empirical and simulation data for 6 of the 8 conditions (AX long delay, AY short and long delay, BX short and long delay, and BY long delay; all p's $> .1$). There was a marginal effect for the BY short delay condition ($t(416) = 1.85$, $p=.07$), and a significant effect for the AX short delay condition ($t(416) = -2.82$, $p<.01$). For BY trials at the short delay, slightly fewer errors occurred in the simulation than were found empirically, while for AX trials at the short delay, there were fewer errors found empirically than in the simulation.

Additional Fits

We next examined the simulation results in terms of performance effects that were not specifically optimized. In particular, we tested whether the interactions between delay and trial type seen empirically were also a property of the model. In the accuracy data, the model also

demonstrated a significant interaction ($F(2,416) = 3.48$, $p<.05$). A planned contrast indicated that the nature of this interaction was identical to that found empirically, in that the difference between AY and BX accuracy was greater at the long delay than at the short delay ($F(1,208) = 4.99$, $p<.05$). A planned contrast also indicated a marginal tendency for BY errors to increase with delay ($F(1,208) = 3.67$, $p=.057$), a different finding from that observed empirically. The delay x trial type interaction in the model was not significant for RTs ($F(2,416) = 0.78$ $p> .1$). However, the absolute numerical relationships in the data were in the right direction, with AY responses being longer and BX responses being quicker in the long delay relative to the short (see Figure 4).

We then examined the effect of delay on target errors. In the empirical data, target errors significantly increased with delay, which was an effect not originally predicted in the conceptual model. However, the simulation results also revealed a significant delay effect for target accuracy ($F(1,208) = 13.65$, $p<.001$), with greater errors occurring in the long delay condition. The simulation did not perfectly capture the behavioral data, in that the magnitude of the delay effect was greater empirically (as was determined in the quantitative analysis directly comparing model to data). Nevertheless, the AX-CPT simulation was able to reproduce a qualitative pattern found in the empirical data, even though it had not been predicted *a priori* . Under ActivationDynamics we describe the mechanism which produced this effect in the simulations.

Finally, we examined speed-accuracy relationships in the model. As was done with the behavioral data, accuracy was conditionalized on response speed (determined by a median split on RT) for each run. Just as was observed empirically, a significant trial type x response speed interaction was present in the simulation ($F(1,208) = 286.39$, $p<.001$; see Table 2). Planned contrasts confirmed that the simulation results conformed to the same pattern as the behavioral data, with AY errors being greater in fast relative to slow responses ($F(1,208) = 189.75$, $p<.001$), and BX errors being greater in slow relative to fast responses ($F(1,208) = 128.55$, $p<.001$). We also examined the interaction of delay with the speed-accuracy function. The 3-way interaction for non-targets between trial type, delay and response speed which was found in the empirical data was not significant in the model ($F(2,416) = 0.25$, $p> .1$). Moreover, although the delay x response speed

interaction was significant for target accuracy ($F(1,208) = 4.17$, $p<.05$), the effect was in the opposite direction from that observed empirically, with the delay effect being greater in fast responses than in slow responses (see Table 2).

Insert Table 2 about here

Activation Dynamics

In order to better understand the mechanisms underlying performance in the model, we conducted a series of analyses examining the activation dynamics at the context layer and its effects on the output layer (results of these analyses are shown in Figures 5-7). The first analysis examined correct responses only. The most critical aspect of the activation dynamics in these trials is that maintained representations of context caused sub-threshold priming of the response layer prior to probe onset. On trials in which an "A" cue is presented, the context representation primes the target response, so that it is more strongly activated than the non-target response prior to probe onset. On trials in which a "non-A" cue is presented, the opposite pattern occurs, such that the non-target response is primed. This initial response advantage caused by the influence of context caused the nontarget RT patterns which were observed. For BY trials, because context and probe both activate the same response, RTs are quick. However, on AY and BX trials, context and probe activate different responses, which leads to interference. For AY trials, the interference occurs early after probe onset, when the initial priming of the target causes a reduction in the speed with which the probe can activate the nontarget response. For BX trials, the interference occurs later, as the influence of the probe builds up and causes a slowing in the rate at which the nontarget response reaches its asymptotic activity level.

We next examined activation dynamics in error trials. As shown in Figure 5a, for AY trials, errors were associated with greater contrast between the A and not-A representations in the context layer, such that at probe onset the A representation was more active while the not-A representation was less active. This increased contrast was likely due to the effects of noise during the delay period. The increased contrast in the context layer carried over to the output layer. For incorrect

trials, at the point of probe onset there was greater activity in the target response unit and less activity in the non-target response unit (see Figure 5b). This greater degree of target priming at the point of probe onset caused enough advantage for the target response unit that it surpassed the response threshold before the nontarget unit. Thus, in the model, AY errors are caused by having context representations which are overly strong, and which wield too much influence over the response competition process. Additionally, this effect led to the relationship observed between accuracy and response speed. AY errors were more likely to occur in fast responses, which can be explained by the greater priming of the incorrect response. Thus, when AY errors occur, it is because the incorrect response is primed strongly enough that it reaches threshold before the correct response.

Insert Figure 5 about here

For BX trials, a very different pattern was observed. As shown in Figure 6a, errors were associated with *reduced* contrast between context layer representations. More importantly on correct trials, the non-A representation was more strongly activated than the A representation, while on error trials the A representation was more strongly activated. Thus, due to noise in the encoding process, the context was incorrectly represent and maintained, albeit in a weaker fashion. This pattern of activity in the context layer dramatically affected the response layer dynamics (see Figure 6b). On correct trials, at the point of probe onset the non-target response unit was more active than the target response unit, giving it an initial advantage which allowed it to effectively compete with input coming from the probe. However, in the response layer on error trials there was equal activity in both the target and non-target units at probe onset. Because there was no advantage for the non-target unit, the response was driven by the input from the probe. Thus, in the model, BX errors appeared to be due to reduced sensitivity in the context layer, which caused responses to be primarily driven by prepotent tendencies. This finding also explains the relationship between BX accuracy and response speed that was observed. BX errors were more likely to occur on slow responses than on fast ones. This is because trials with slow responses were more likely to be

those in which the response was driven more by the probe than by context. Since the probe is more strongly associated with the incorrect response, slow responses were more error-prone.

Insert Figure 6 about here

We next compared activation dynamics between long and short delay trials (see Figure 7). The simulation and empirical data both showed that AY errors tended to increase with delay, while BX errors tended to decrease. An examination of activation dynamics reveals a mechanism which may explain how these effects arise. Specifically, in the simulations, the representation of context was slightly, but significantly, more active at the time of probe onset in the long delay than it was in the short. In other words, it appeared as if the context representation had not yet reached full strength during the short delay. This effect carried over to the response layer. In AY trials, the context effect served to more strongly prime the incorrect (target) response unit. This effect could explain how the probability of incorrect AY responding *increases* following a long delay. In BX trials, the context effect acted to more strongly prime the correct (target) response. This effect could explain how the probability of incorrect BX responding *decreases* following a long delay. Thus, the increased strength of context representation with delay may provide a common mechanism for the delay x trial type interactions also observed in the behavioral data.

Insert Figure 7 about here

Finally, we examined activation dynamics on AX target trials. On incorrect AX trials, the "non-A" representation was maintained over the delay, and thus was more strongly activated than the correct representation. The incorrect representation of context caused greater priming of the non-target than the target unit in the response layer. This initial advantage in activity was great enough to allow the incorrect response to survive competition from the probe and eventually reach threshold. We also examined the relationship between short and long delay activity. Recall that in both the behavioral data and simulation results, a significant increase in AX errors was found to occur with delay. At first blush, this finding seems difficult to explain, given the hypothesis that

context representations are actively maintained and increase in strength over the delay. If this hypothesis were true, it would seem that AX errors should decrease with delay rather than increase, since stronger representations of context should lead to greater priming of the correct target response. Indeed, this is the exact interpretation that appeared to explain the delay-related decrease in BX errors. Further, an initial examination confirmed that in AX trials the average activity of the A context unit was, in fact, greater at the long delay than it was at the short. However, a more detailed analysis of the activation trajectories helped to provide an explanation.

This analysis was conducted by determining the relative frequency of trials in which: 1) the correct context representation was strongly activated (i.e., greater than 0.5); 2) the incorrect representation was strongly activated; and 3) there was an uncommitted context representation (i.e., both units activated less than 0.5). For the short delay condition, there were a small percentage of trials (2.5%) that fell into this last category at the time of probe onset. Further analyses suggested that even with an uncommitted context representation, the correct target response was eventually made in almost all of these trials. This is not surprising given that the probe had a strong association with a target response. In contrast, in the long delay condition, by the point of probe onset there were almost no trials in which the context representation remained uncommitted. Moreover, there was an equal increase in the frequency of trials of both of the other two categories. In other words, because of the noise in the system, it appeared that on half of the trials that were initially uncommitted, the context layer eventually committed to the correct representation, while on the other half it committed to the incorrect representation. The critical point is that those trials in which the context committed to the incorrect representation should eventually result in an error. This pattern suggests that the net error should increase in the long delay relative to the short, since a proportion of trials that led to an error in the long delay condition were originally trials in which a correct response was made in the short condition. Indeed, the value of this increase, 1.25%, was exactly the amount in which errors increased between the short and long delay in the simulation results.

Discussion

The simulation results demonstrate that the cognitive control model can account for detailed characteristics of performance in the AX-CPT. With two free parameters, the simulation captured the inhibitory and facilitory effects of context that were evident in the main effects of trial type in errors and RT, as well as the particular relationship between the 3 non-target sequences. Additionally, the model captured the quantitative pattern of the accuracy data as well, with only 1 of the 8 experimental conditions (AX short delay trials) producing a reliable difference between the simulation and empirical results. Further, analyses of activation dynamics in the simulation provided insights into the mechanisms which produced these patterns of performance. For example, the trial-type effects observed in the simulations appeared to be a direct consequence of the biasing function of context, which produced sub-threshold priming of the response layer prior to probe onset. In AY trials, errors appeared to be caused by overly strong target priming on trials with high contrast in context layer representations, while in BX trials, errors result insufficient non-target priming on trials where context representations were weak and/or inaccurate. The dynamics also explained the relationship between accuracy and response speed, since overly strong priming on AY trials leads both to errors and fast RTs (since the activation level is close to threshold), whereas a lack of priming on BX trials would lead both to errors and slow RTs (since activation is farther from threshold).

Moreover, the results also demonstrate that the model is sufficient to account for additional effects in the data that were not predicted prior to conducting simulations. In particular, the active memory component of the model was sufficient to capture the surprising effects of delay on performance (i.e., the delay x trial type interaction in non-target errors, and the delay effect on target errors). The analysis of activation dynamics provided confirmation that these delay effects were, in fact, due to active maintenance of context. In particular, at the long delay, there was a slight, but significant increase in the strength of context representations relative to the short delay. This stabilization of context layer representations translated into greater priming at the response layer, which lead to a greater opportunity for AY errors, and less opportunity for BX errors. The analysis was able to provide an explanation for the counterintuitive finding of increased AX errors with

delay, suggesting that it was also due to the stabilization of context representations. In particular, at the short delay, there were still a proportion of trials in which the context representation has not been committed, while at the long delay the representation had resolved, with half of these previously uncommitted trials actually representing the incorrect context. When context is uncommitted, the response was driven by the influence of the probe, which led to correct responses on those short delay trials. However, on the proportion of trials in which the incorrect context was eventually represented, an error was made. Since the frequency of trials in which the incorrect context is represented increases at the long delay, this led to a net increase in errors.

The closeness of the fit between the simulations and the empirical data suggest that the cognitive control model may be capturing important features of the processing mechanisms underlying behavioral performance in the AX-CPT. Moreover, there are a number of predictions that arise from the model regarding internal processing dynamics. In particular, the model suggests that: 1) AY error trials will be associated with strong representations of context and overactive target response priming; 2) BX error trials will be associated with weak, inaccurate context representations and an absence of non-target response priming; 3) Long delay trials of all types will be associated with stronger context representations and greater response priming than short delay trials; 4) AX trials at the long delay will have a greater proportion of committed context representations, and priming of the incorrect response (since half of the committed context representations will commit to the incorrect context). Although it might be difficult to directly examine the strength of context representations through purely behavioral measures (but see Studies 6 and 7 for evidence of how they might be examined with functional neuroimaging methods), it might be possible to examine response priming through the use of a force dynamometer as a response indicator. An approach frequently adopted in the psychophysiological literature (Gehring, Goss, Coles, Meyer, & Donchin, 1993; Gratton et al., 1988; Kutas & Donchin, 1980) is to use the dynamometer as a continuous index of sub-threshold muscle tension prior to actual responding (since a response is delivered by squeezing the dynamometer to some prespecified fraction of maximal force). With this method, muscle tension could be measured in

both the correct and incorrect response channels during the delay period of AX-CPT trials. Muscle tension activity which correspond to the predicted patterns of response priming predicted would help to further support the claims of the model with respect to internal processing mechanisms.

Despite its ability to capture many detailed aspects of performance in the AX-CPT, the simulation is not perfect in its present form. In particular, there was a clear mismatch between the model and behavioral data in terms of the effect of delay on speed-accuracy relationships. This mismatch was most notable in the differences observed at the late part of this speed-accuracy curve. A detailed comparison of the two sets of results revealed that, in the behavioral data, the late part of the speed-accuracy curve appeared to be less sensitive to the effects of context, whereas this was not the case in the simulation. One plausible explanation of this discrepancy is that it is the result of general task factors that were outside the scope of the model. For example, it is possible that subjects' motivational or arousal states during performance preferentially affect the late part of the speed-accuracy curve, since these would be additional aspects of behavior that could likely to contribute to response slowing. Since these general factors were not simulated, they would not contribute to the speed-accuracy pattern in the results. However, it is also possible that the mismatch between model and data reflects relevant aspects of context representational dynamics that were not properly captured in the simulation. Specifically, in the simulation, context representations were maintained throughout the entire response period (e.g., see Figure 7). Yet it is possible that instead context representations are either updated or inhibited at some point following probe presentation, such that they are less active late in the response period. This explanation would also produce a speed-accuracy curve which showed less influence of context for responses which were slow, thus occurring after the period when the representations were not as strong. Moreover, it would seem that this possibility warrants further study, as it appears equally plausible, and may have important consequences for our understanding of the mechanisms of context representation.

Study 3: Simulating Impairments in Cognitive Control

The cognitive control model has been successful in accounting for the patterns of normal behavioral performance in the AX-CPT. Additionally, the model makes claims about the

consequences of cognitive control impairments on task performance. The model postulates that disturbances in the representation and maintenance of context information should lead to a very specific pattern of performance deficits in the AX-CPT, both in terms of response competition effects for AY and BX errors, and in the further interaction of delay with these effects. Specifically, if context *representation* is impaired, the model predicts that BX performance will deteriorate, since correct responding is dependent on the influence of context to override the influence of the probe. Additionally, if context *maintenance* is impaired, the model predicts that BX performance will show a further decrement at the long delay, since the representation will be at its weakest. In contrast, the model predicts AY performance will improve if context *representations* are weakened, since the priming influence of context creates competition in making the correct response. Moreover, if context *maintenance* is impaired, the model predicts that AY performance will be further improved at the long delay.

In this study, we attempted to make these predictions of the model explicit, by simulating disturbances in context processing. In these simulations, both context representation and context maintenance were weakened, in order to examine their effect on task performance. We simulated this weakening of context by reducing the gain parameter of context units in the model. The gain parameter governs the input-output function of units, such that with reduced gain, units are less sensitive to their net input. In stochastic networks (i.e., those in which noise is an intrinsic component of processing), gain reductions have been shown to produce impairments in signal detection capabilities (Servan-Schreiber, Printz, & Cohen, 1990). Furthermore, reductions of gain in units representing context have been shown to produce performance impairments and delay effects in tasks (e.g., the Stroop and lexical disambiguation) that rely on cognitive control (Cohen & Servan-Schreiber, 1992). Thus, we hypothesized that reducing gain at the context layer would lead to impairments in both context representation and context maintenance. Moreover, these impairments would result in a specific pattern of performance deficits in the AX-CPT that are consistent with a failure of cognitive control. Finally, to test the specificity of the cognitive control model, we also conducted simulations in which gain was reduced at the input layer, rather than at the

context layer. This manipulation may correspond to degrading perceptual representations of the input. The cognitive control model predicts that although gain reductions at the input layer may cause certain types of performance impairments, these will be behaviorally distinct from those produced by gain reductions at the context layer.

Methods

Architecture and Processing

Simulations were conducted using the AX-CPT model employed in Study 2, but with reduced gain either on the context units or on the input units. We reduced gain for context units to the same degree as was done in previous simulations examining the effects of context on cognitive task performance (Cohen & Servan-Schreiber, 1992). Thus, for the context gain reduction (CGR) model, the context gain parameter was reduced to a value of 0.6, from a baseline value of 1.0. This same value was chosen for the input gain reduction (IGR) model, although the gain reduction was applied to the input rather than context units. In order to reduce the degrees of freedom in the simulation, all other aspects of the models were held constant, including all of the connection weights and the noise level. However, the threshold parameter (T) was retained as a free parameter, since it has been shown that the optimal response threshold changes when gain is decreased (Servan-Schreiber et al., 1990).

Simulations and Data Analysis

Simulations were conducted in exactly the same manner as in Study 2. To produce data of comparable sample size in the reduced gain models we simulated the same number of trials as was done in Study 2 (i.e., 209 runs of 200 trials each). Accuracy and RT statistics were collected for each run separately and each run was treated as a "subject". Simulation data were subjected to a mixed ANOVA with model-type (intact vs. CGR) treated as a between-subjects factor, and the other task conditions treated as repeated measures. The data were examined for effects of reduced context gain on performance, specifically in terms of interactions between gain and the primary task factors (e.g., delay and trial type). Additionally, performance in the CGR model was compared against the IGR model in order to determine the specificity of findings.

Results

Parameter effects

The primary effects of the gain manipulation were found to be robust to the value of response threshold. However, when context gain was reduced and threshold was unchanged, RT is increased, and the number of trials in which activation does not reach threshold (i.e., "no response" trials) is also increased. Additionally, we observed that the overall false alarm rate showed a non-monotonic pattern, such that as with decreasing threshold there were decreasing number of false alarms until T = 0.55. With T < 0.55, false alarms began increasing again. Therefore, we took T=0.55 to be the optimal threshold when context gain=0.6, and fixed this value for the simulations described below. The same threshold was used (T=0.55) in the IGR simulations, in order to provide as close of a comparison as possible of the two reduced gain models.

Trial type effects

A significant interaction was found between gain and trial type ($F(2,832) = 289.59$, $p<.001$) in the CGR model. Planned contrasts indicated that the CGR model made less AY errors ($F(1,416) = 6.46$, $p<.05$) and more BX errors ($F(1,416) = 418.51$, $p<.001$) than the intact model (see Figure 8). Additionally, the difference between BX and AY error rates was greater in the CGR model than in the intact model ($F(1,416) = 323.45$, $p<.001$). In contrast, BY errors were found to significantly decrease with reduced context gain ($F(1,416) = 11.05$, $p=.001$). D'-context was significantly reduced in the CGR model ($F(1,416) = 1096.13$, $p<.001$) (see Figure 8). The RT effects mirrored those of accuracy. Although there was a main effect of reduced context gain such that responses were speeded overall ($F(1,416) = 3626.04$, $p<.001$), there was also a significant interaction with trial type ($F(2,832) = 901.66$, $p<.001$). As shown in Figure 9, planned contrasts indicated that the speeding of AY responses observed in the CGR model was greater than that observed for BX responses ($F(1,416) = 977.4$, $p<.001$). The speeding of RT was also found to be greater for AY relative to BY responses ($F(1,416) = 1909.19$, $p<.001$).

Insert Figures 8 and 9 about here

Delay x Trial Type effects

A significant 3-way interaction was found in accuracy between gain, trial type and delay ($F(2,832)=38.04$, $p<.001$) in the CGR model. Planned contrasts confirmed that the difference in performance between the intact and CGR model was greatest at the long delay, with BX performance being most impaired ($F(1,416) = 42.31$, $p<.001$) and AY performance being most enhanced ($F(1,416) = 8.30$, $p<.01$) (see Figure 8). However, there was no effect of delay on BY errors in the CGR model ($F(1,416) = 0.94$, $p>.10$). The simulation results also revealed a significant interaction between gain and delay on d'-context ($F(1,416) = 60.75$, $p<.001$), with context sensitivity being significantly lower at the long delay relative to the short in the CGR model ($F(1,416) = 153.64$, $p<.001$) (see Figure 8) . The simulation also revealed a gain x trial type x delay interaction in RT ($F(2,832) = 7.315$, $p=.001$). Planned contrasts indicated that for AY trials, RT decreased with delay in the CGR model, but increased with delay in the intact model ($F(1,416) = 3.57$, $p=.06$). For BX trials, RT increased with delay in the CGR model, but decreased with delay in the intact model ($F(1,416) = 6.88$, $p<.01$). Interestingly, posthoc contrasts also revealed that BY trials showed the same effect as BX, in that there was increase in RT at the long delay in the CGR model but not in the intact model ($F(1,416) = 10.1$, $p<.01$). A similar pattern was also observed in AX target RTs: a gain x delay interaction ($F(1,416) = 42.9$, $p<.001$), caused by a delay-related slowing of responses in the CGR model.

Speed-Accuracy Relationships

The simulation results revealed a significant gain x trial type x speed interaction in the CGR model ($F(2,832) = 62.37$, $p<.001$; see Table 3). Planned contrasts demonstrated that the reduction in AY errors in the reduced gain model was greatest for fast responses ($F(1,416) = 34.95$, $p<.001$), while the increase in BX errors in the reduced gain model was greatest for slow responses ($F(1,416) = 126.98$, $p<.001$). However, there appeared to be no further interaction of delay with this error pattern, as the 4-way interaction was not significant ($F(2,832) = 2.15$, $p>.1$).

Insert Table 3 about here

Comparison with IGR Model

The simulations revealed some similarities between the two reduced gain models (see Figures 8 and 9). Relative to the intact model, the IGR also showed a large increase in BX errors ($F(1,416) = 266.8$, $p<.001$), and a large decrease in d'-context ($F(1,416) = 777.9$, $p<.001$). However, there were a number of important differences between the models as well, primarily in the interaction of delay on performance (see Figures 8 and 9). Specifically, in comparing non-target performance in the IGR and CGR models, a significant 3-way interaction between model-type, trial-type, and delay was found in both accuracy ($F(2,832) = 65.8$, $p<.001$) and RT ($F(2,832) = 8.9$, $p<.001$). Planned contrasts revealed that in the IGR model AY errors increased ($F(1,416)=21.5$, $p<.001$) and BX errors decreased with delay ($F(1,416)=17.2$, $p<.001$), although the opposite pattern was found in the CGR model. Additionally, contrasts revealed that in the IGR model both BX ($F(1,416)=22.7$,$p<.001$) and BY RT ($F(1,416)=30.0$,$p<.001$) decreased with delay, whereas the opposite pattern occurred in the CGR model. Similarly, target RTs showed a significant delay-related decrease in the IGR model ($F(1,416)=679.6$, $p<.001$), although they increased with delay in the CGR model. Finally, a significant model-type x delay interaction was found in d'-context ($F(1,416) = 45.2$, $p<.001$). Although there was a slight, but significant delay-related decrease in d'-context in the IGR model ($F(1,416)=5.3$, $p<.05$), this decrease was significantly less than that observed for the CGR model.

Activation Dynamics

The analysis of activation dynamics made explicit how reducing gain at the context layer caused the contrast between context representations to be decreased, such that they exert less influence over responding at the output layer. Figure 10 makes this point clearly, by comparing activity in the intact and CGR models across the context and output layers. At the context layer, for both BX and AY trials, the maintained representation is less active in the CGR model at probe onset, while the representation which is not supposed to be maintained shows greater activity. The reduced contrast in context representations is directly due to the reduction in context gain, which causes the maintained representation to be less sensitive to its self-excitatory input and the

competing representation to show less effects of lateral inhibition. Moreover, this context-layer effect is the source of the change in BX and AY responding. For BX trials, weakened context causes less priming of the non-target response prior to probe onset. Thus, following probe onset, there is greater activity in the target response unit, which leads to increased response competition. The increased competition is manifest both as greater errors and more RT interference. For AY trials, weakened context causes less priming of the target response. Thus, following probe onset, there is less activity in the target response unit, leading to decreased response competition. The decreased competition is manifest both as reduced errors and less RT interference. However, for both types of trials, errors occur in the CGR model for the same reasons as they do in the intact model. Specifically, for BX error trials, context is weaker than on correct trials, while on AY error trials, context is stronger than on correct trials.

Insert Figure 10 about here

We next compared activity dynamics as a function of delay. The effects of delay most strongly differentiated the CGR model from both the intact model and the IGR model. In the CGR model, the reduction of gain in the context layer causes context representations to be susceptible to the accumulating effects of noise. Thus, as delay increases, the maintained context decays in activation. As a result, the longer the delay the less influence of context will have over responding. In the IGR model, the input to the context layer is degraded, as a result of decreased gain over the input layer. However, the strong recurrent connections and lateral inhibition act to "clean up" this degraded input, causing the context representations to increase in strength with time. Consequently, in the IGR model, the longer the delay, the greater the effect of context. As can be seen in Figure 11, the maintained context representation decreases with delay in the CGR model (averaged across all trial types), but increases with delay in the IGR model, and stays relatively constant in the intact. Thus, at the time of probe onset for the long delay condition, the context representation is at its least active in the CGR model but at its most active in the intact and IGR models. This impairment in active maintenance in the CGR directly leads to altered response competition at the output layer.

For BX trials, there is less priming of the non-target response at the long delay, and consequently more competition from the target response. For AY trials, there is less priming of the target response at the long delay, which in turn, produces less competition.

Insert Figure 11 about here

Discussion

The primary goal of this study was to explicitly validate the claims of the theory regarding the effect of impairments in the representation and maintenance of context on AX-CPT performance. To produce these impairments, we simulated the effects of reduced gain in the context layer, since previous work has demonstrated that this manipulation: a) produces *representation* deficits, by increasing susceptibility to noise (i.e., reducing SNR); and b) produces *maintenance* deficits, since the noise effects accumulate over time (Cohen & Servan-Schreiber, 1992). The simulations results demonstrate how both of these impairments affect AX-CPT performance. Analyses of activity layer dynamics showed that the representation of context was weaker in the reduced context gain simulations, with less contrast observed between the two different context representations. This resulted in increased competition for response in BX trials, and decreased competition in AY trials. Additionally, in the CGR model, the maintenance of context was impaired. The contrast between context representations further decreased with delay, such that contrast was lower at the long delay than at the short. Consequently, at the long delay, context exerted even less of an influence over response competition.

The impact of reduced gain on context layer activity dynamics directly translated into changes in model performance. Representation deficits arising from reduced context gain were observed as significant interactions between gain and trial type. In terms of accuracy, d'-context was significantly worse and BX errors were significantly increased, while AY errors were actually reduced. In terms of RT, BX responses were relatively slowed, while AY responses were relatively speeded. Context representation impairments also significantly interacted with the speed-accuracy function. The increase in BX errors was greatest in the late part of the function, while the decrease

in AY errors was greatest in the early part of the function. Maintenance deficits were also observed, in terms of significant interactions with delay. D'-context further decreased at the long delay, and BX errors further increased, while AY errors further decreased.

Deficits in active maintenance also affected RT. BX responses were slower, while AY responses were faster at the long delay, which is the opposite pattern from that found with normal gain. Interestingly, reducing gain also produced delay-related slowing in AX and BY trials, even though these conditions do not involve response competition. This finding was not predicted prior to conducting simulations, but is consistent with the general claim of the model that, in the absence of response competition, the primary effects of context will be facilitory ones. Moreover, the findings regarding AX and BY RTs are analogous to, and consistent with, findings from computational modeling of the Stroop task (Cohen et al., 1990). In that study, it was observed that the strength of context representations affected RTs for both word reading and neutral color naming, even though these tasks appear not to involve response competition. Thus, taken together, the simulation results make clear that a single disturbance involving the representation and maintenance of context can produce three different types of performance deficits related to cognitive control: inhibitory failures (BX errors), memory failures (delay effects in AY and BX errors and in d'-context), and reduced facilitation effects (AX and BY slowing).

The simulation results also make clear the specificity of effects attributed to the representation and maintenance of context, due to the inclusion of the IGR model. In the IGR model, reduced gain over the input layer produced changes in performance that were clearly distinct from the changes observed when context representations were disturbed in the CGR model. The differences between the two models were most clearly observable in the interaction with delay. In the IGR model, performance showed effects of delay that were similar to the intact model, with BX errors decreasing and AY errors increasing. This pattern of performance lends further support to the idea that impairments in the active maintenance of context information are specifically and selectively responsible for the delay effects observed in the CGR model. A very similar disturbance (i.e., reducing gain in a different layer of the model) which does not affect the active maintenance of

context will not produce such delay-related effects on performance. Consequently, the simulation results indicate that delay-related effects on performance in the AX-CPT serve as a measure of the integrity of context maintenance. Interestingly, however, the simulation results also indicate that disturbances to other parts of system can result in changes to the representation of context. Specifically, in the IGR model, like the CGR model, there was a large increase in BX errors. This result suggests that disturbing perceptual representations can indirectly affect the strength of context representations (which are required for correct performance on BX trials), by affecting how well the cue information gets encoded as context.

Taken together, these results provide further support for the role that cognitive control plays in AX-CPT performance. The simulations indicate that impairments in both the representation and maintenance of context will produce should produce a distinct pattern of AX-CPT performance that is easily discernible in terms of its behavioral signature. Although certain aspects of performance should show deficits (e.g., d'-context, BX error rate and RT), others should be unaffected (BY error rate), and still others should actually be improved (AY error rate and RT). Moreover, the influence of delay on these effects should most clearly and most selectively reveal the integrity with which context representations are actively maintained. Additionally, the simulation results make predictions regarding the pattern of activity dynamics associated with impairments in the maintenance of context. In the CGR, but not the IGR model, context layer activity was found to decay over the course of the delay period. These results directly translate into predictions regarding the effects of context impairments on PFC activity. Specifically, PFC activity should be lower with weakened context representation, and this pattern should further interact with delay. Moreover, delay-related activity in PFC should be found to decay over the course of a delay period under conditions when context maintenance is disturbed. This prediction is directly tested in Study 7. Conversely, the model also suggests that the AX-CPT might be well-suited as a behavioral probe of cognitive control impairments that result from PFC dysfunction. This last prediction forms the basis for the next study.

Study 4: Testing Model Predictions in Patients with Schizophrenia

The simulation results of Study 3 provide an explicit demonstration of the consequences of context impairments on AX-CPT performance. However, these results would be more compelling if the model could be used in a truly predictive fashion, to generate testable hypotheses about AX-CPT performance in a population hypothesized to have cognitive control impairments. One such population is patients with schizophrenia. Schizophrenia has commonly been associated with a failure of cognitive control (Callaway & Naghdi, 1982; Nuechterlein & Dawson, 1984). Moreover, the use of reduced gain as a means of simulating context impairments may actually provide a plausible model of the physiological abnormalities associated with schizophrenia. Many investigators have argued that some of the cognitive deficits present in schizophrenia may result from a disturbance of dopamine neurotransmitter activity in PFC (Davis, Kahn, Ko, & Davidson, 1991; Goldman-Rakic, 1991). Phsyiologically, dopamine is thought to act on target neurons by potentiating their responsivity to both inhibitory and excitatory input (Chiodo & Berger, 1986; Penit-Soria, Audinat, & Crepel, 1987). In previous work, we have argued that this physiological action of dopamine can be simulated as a change in gain, and that reductions in context layer gain may capture the effects of reduced prefrontal dopaminergic tone, which is thought to be present in schizophrenia (Cohen & Servan-Schreiber, 1993; Servan-Schreiber et al., 1990). More importantly, we have previously shown that reducing gain in the context layer can simulate the pattern of performance deficits of schizophrenia patients on a number of cognitive tasks (Braver, Cohen, & Servan-Schreiber, 1995b; Cohen & Servan-Schreiber, 1992). However, in this previous work, the simulations captured only general aspects of behavioral performance.

Thus, in the current study, we attempted to test the model's predictions regarding cognitive control by examining more detailed and specific aspects of behavior in schizophrenia patients performing the AX-CPT. Moreover, a number of steps were taken in this study to address methodological issues which often plague studies of clinical populations, and particularly studies of patients with psychiatric diseases. First, close attention was paid to matching patient and control groups on multiple demographic variables, since patients often show a different demographic profile than college undergraduates, who are the typical participants in cognitive studies. Second,

only patients who had never received neuroleptic medication were tested, since neuroleptics may both mask underlying deficits as well also produce additional cognitive impairments (Blanchard & Neale, 1992; Spohn & Strauss, 1989). Third, only patients who had never before been hospitalized for psychosis were tested, since both hospitalization and the developmental progression of the illness may have important consequences for cognitive functioning in schizophrenia. Together, addressing these three issues provides a firmer basis on which to draw conclusions regarding the nature of cognitive control impairments in patients with schizophrenia.

Methods

Participants

Participants were 16 healthy controls and 16 patients with schizophrenia or schizoaffective disorder. Healthy controls were recruited through local advertisements and were evaluated using the Structured Clinical Interview for DSM-III-R. All patients were neuroleptic-naive, and were recruited into the study if they were experiencing any type of psychotic symptom and it was their first episode of psychiatric hospitalization or contact with outpatient psychiatric services. Patients were followed longitudinally, and were confirmed to have a diagnosis of schizophrenia 6 months after their participation in this study[11]. Diagnoses were determined through a diagnostic and disposition conference which included information from the Structured Clinical Interview for DSM-III-R (SCID-PD, Spitzer, Williams, Gibbon, & First, 1990), administered by trained research personnel and a thorough chart review. In addition, the Brief Psychiatric Rating Scale (BPRS; Overall, 1974) was used to evaluate symptom severity. Ratings were completed by trained research team members who regularly participated in evaluation sessions to insure reliability. All BPRS ratings were made within one week of testing and all raters were blind to the performance of participants in the tasks.

In both groups participants were excluded for: a) age greater than 50 or less than 14; b) WAIS-R Full Scale IQ lower than 70; c) non-English native language; d) a lifetime diagnosis of substance dependence or any substance use disorder within six months of testing; e) neurologic disorders or any family history of hereditary neurologic disorder; or f) pregnancy. Potential patients were also excluded for: a) history of psychosis longer than 5 years; or b) head injury, if

symptom onset occurred within 3 years of injury. Potential controls were excluded if they had: a) any lifetime history of Axis I disorder or any first order family history of a psychotic disorder; or b) treatment with any psychotropic medication within 6 months prior to testing. The healthy controls were matched with patients for age, gender, race, and years of father's education (to match approximately for socioeconomic status). Focused contrasts indicated that the healthy controls did not differ from the schizophrenic patients on any of these variables. The demographic and clinical characteristics of both participant groups are shown in Table 4.

Insert Table 4 about here

Tasks and Procedure

Delay durations for this study were 1000 msec for short delay trials and 4900 msec for long delay trials. 13 of the participants (7 controls, 6 patients) performed the task with short and long delay trials segregated into short blocks of 10 trials each. In the blocked version, blocks were presented in pseudorandom order, with the constraint that both delay conditions appear in every set of 2 blocks. The other 19 participants (9 controls, 10 patients) performed the task in a mixed design, with short and long trials randomly interleaved. All participants performed 300 trials total (15 cycles of the two delay blocks in the blocked version, and 3 blocks of 100 trials each in the mixed version)[12], with 150 trials in each of the short and long delay conditions.

Data analysis

Analyses on all dependent measures of interest were conducted by including diagnostic group as a between subjects factor. Data were examined for both main effects of group and any interactions between group and the other task factors (e.g., delay and trial type)[13].

Results

Trial type effects

There was a significant reduction in d'-context for patients ($F(1,30)=11.6$, $p<.01$). There was also significant group x trial type interaction for nontarget errors ($F(2,60)=4.87$, $p=.01$). As can be seen in Figure 12, planned contrasts indicated that patients made significantly more errors on

BX trials than controls ($F(1,30)=6.95$, $p<.05$). In contrast, there was no significant difference between patients and controls in AY errors ($F(1,30)=2.40$, $p>.10$). A planned contrast also revealed that patients made more BY errors than controls ($F(1,30)=7.48$, $p<.05$). An analysis of nontarget RTs revealed a trend-level interaction between group and trial type ($F(2,60)=2.85$, $p=.07$; see Figure 13). Planned contrasts indicated that patients showed greater slowing of RTs, compared to controls, in BX trials than in AY trials ($F(1,30) = 4.34$, $p<.05$), as well as greater slowing in BY trials relative to AY ($F(1,30) = 5.62$, $p<.05$). Moreover, in AY trials, patient RTs were not significantly slower than controls $F(1,30) = 0.34$ $p>.10$).

Insert Figures 12 and 13 about here

Delay x Trial Type effects

A significant group x delay interaction was observed for d'-context ($F(1,30) = 6.07$, $p<.05$), with patients showing a significant decrease at the long delay relative to the short ($F(1,30) = 12.70$, $p<.01$), and controls showing no delay effect ($F(1,30) = .001$, $p>.10$). A three-way interaction between group, delay and trial-type was also observed for nontarget errors ($F(2,60)=3.97$, $p<.05$). Planned contrasts further confirmed that the difference in performance between patients and controls increased at long delays for BX trials, but decreased at long delays for AY trials ($F(1,30)=5.18$, $p<.05$). Additionally, BY errors did not increase with delay, and instead showed a marginal trend towards a delay-related decrease ($F(1,31)=2.68$, $p=.10$). However, focused contrasts also revealed that patients made significantly more AY errors than controls at the short delay ($F(1,30) = 4.85$, $p<.05$), and did not increase BX errors at the long delay ($F(1,15)=.03$, $p>.1$). The three-way interaction in the RT data was not significant ($F(2,60)=0.35$, $p>.1$). Nevertheless, it was the case that numerically, the results were in the direction predicted by the simulation. Patients showed less delay-related slowing than controls for AY trials, but more slowing for BX and BY trials. For target responses, the group x delay interaction was significant at the trend level ($F(1,30) = 2.79$, $p=.1$), with delay-related slowing greater for patients. Finally, the group x trial type x response speed (i.e., fast vs. slow responses) interaction was not significant ($F(2,60)=.23$, $p>.1$).

Discussion

The results of this study provide some confirmation of the model's ability to predict the behavioral consequences of a specific disturbance in cognitive control. In particular, 4 of the model's 8 predictions regarding patient performance were confirmed: the main effect of group and group x delay interaction in d'-context; and the group x trial type and group x trial type x delay interactions in non-target accuracy. Two additional predictions of the model were also partially confirmed: the trend-level group x trial type interaction in non-target RT and the trend-level group x delay interaction in target RT. Note also that a number of these predictions involved a pattern of performance that was both highly specific and counterintuitive. Patient performance conformed to predicted patterns in demonstrating reduced context sensitivity and increased BX errors, but more importantly also demonstrated that these deficits increased with delay. Moreover, the simulations predicted that performance on AY trials would show a very different pattern, and this prediction was also confirmed. Patients were not significantly different from controls in either AY error-rate or RT, and their performance actually improved with delay. It is also worth noting that the findings with regard to d'-context and BX errors replicate earlier results (Cohen et al., 1999; Servan-Schreiber et al., 1996), which strengthens their validity. However, the results with respect to AY trials are new and add further weight to the hypothesis that the effects are specific to the representation and maintenance of context. Finally, the simulations were employed in a truly predictive fashion, in that they were published (for an earlier incarnation of the model) prior to ever testing patients in this particular version of the paradigm (Braver et al., 1995a). Taken together, these results are consistent with two primary hypotheses of the model: 1) that AX-CPT performance critically depends upon cognitive control; and 2) that control is achieved by the representation and maintenance of context which acts to bias response competition.

Conversely, by supporting the model's hypotheses regarding the influence of cognitive control on AX-CPT performance, the results also provide further confirmation of a specific cognitive control deficit in patients with schizophrenia. In order to demonstrate a differential deficit in a particular cognitive process, it is necessary to show that performance deficits in patients are not

greatest in the conditions which are also most difficult for controls (Chapman & Chapman, 1978; Chapman & Chapman, 1989). The findings with respect to AY performance satisfy this requirement. In the long delay condition, controls make the most false alarms and have the slowest RTs on AY trials, suggesting that it is the most difficult condition. Yet on these trials, the error-rate and RT of patients is statistically no different from controls. In contrast, at the long delay, patients made the greatest number of false alarms and had the slowest RTs for BX trials. It seems clear that these patterns of performance could not have arisen solely from a generalized performance impairment. Rather, it is more likely that the effects are due to a specific deficit in context representation and maintenance, since both were explicitly predicted by the model.

However, it is still possible that patients do suffer from additional impairments that are superimposed on a specific disturbance in cognitive control. The findings provided support for this interpretation as well, in that patients exhibited an overall slowing in RT, and increased errors even in the task condition that was least difficult for controls (i.e., BY trials). Although these findings were not predicted by the model, similar results have been commonly observed in the literature, and may reflect a generalized "cognitive slowing" or some type of motor-related impairment (Nuechterlein, 1977; Walker, 1981). This interpretation might also explain the reduced success of the model in confirming predictions involving RT. Specifically, the results failed to confirm a group x trial type x delay interaction or a group x trial type x response speed interaction, and the other two RT effects (the group x trial type interaction and the group x delay interaction in target RT) were only significant at the trend level. It is likely that all of these RT effects would be influenced by motor-related impairments or slowed information processing. Finally, these factors might even provide an explanation for the mismatch between simulation and patient data regarding AY and BX errors. This mismatch primarily arose from patients making a greater number of errors for both trial-types at the short delay than was predicted by the simulation. Responses at the short delay would be most likely to be affected by motor-deficits or slowing, since these rely upon making a second response (to the probe) quickly after a first (to the cue). Thus, it may be that the model could more fully capture the patterns found in the patient data by simulating an additional

impairment that affects processing speed or response execution. For example, it has been previously suggested that motor effects on response speed might be captured by changing the gain on units in the output layer of the model (Servan-Schreiber, 1990). Additionally, changes in processing speed have been examined by manipulating the rate at which information accumulates at each processing stage (Cohen & Servan-Schreiber, 1992). Thus, an important direction for future research might be to examine the effects of these sorts of manipulations on performance of the model.

Another possible explanation for the failure to confirm the RT predictions of the model is that the current experiment did not have sufficient statistical power to detect these effects. In fact, there is support for this interpretation from the simulation results themselves. The two primary predictions of the model are the 3-way interactions between gain, trial type and delay in both accuracy and RT. However, the effect size for accuracy interaction is much larger than for the RT interaction (0.6 vs. 0.24) Moreover, for this level of effect size, detecting a significant effect (at the .05 level) with 50% power would require a sample of 100 participants in each group. However, the current study involved 16 participants in each group, which only provides a 10% chance of actually detecting the effect (calculated from the conversion tables provided by Cohen (1988)). The RT data from the current study also support this hypothesis. As noted above, all of the RT effects were in the correct numerical direction as predicted by the model, although the magnitude was too low enough to reach statistical significance. Thus, to provide a fairer test of the model, it would be necessary to increase the size of the sample. Indeed, the recruitment of first-episode patients for this study is still on-going, so it may be possible to detect significant RT effects in future analyses making use of larger samples.

The limitations of the current study notwithstanding, the results do provide validation of the utility of the AX-CPT paradigm as a means to probe for suspected cognitive control impairments in clinical populations. These investigations could be further examined through the testing of other populations suspected to also have cognitive control impairments, such as healthy older adults and those with Alzheimer's disease, patients with ADHD or obsessive-compulsive disorder, and those

suffering from Parkinson's disease. Of course, a major component of the theory proposed here is that these cognitive control impairments reflect a disturbance in the ability to maintain and represent context. Moreover, the theory suggests that PFC is centrally involved in context representation and maintenance. Thus, a prediction of the model is that those populations which show delay-related impairments in AX-CPT performance will be those whose pathology involves PFC dysfunction. Conversely, another obvious prediction of the model is that patients who are known to have PFC lesions should also show performance impairments in the AX-CPT[14]. In Study 6, we provide direct evidence that suggests the involvement of PFC in AX-CPT performance, and further suggests that it is dorsolateral regions of PFC that are most critical for active maintenance of context. In the following study, we show how cognitive control impairments might also be observed in healthy individuals under specific task situations.

Study 5: Inducing Cognitive Control Impairments in Healthy Individuals

In the preceding study, very specific behavioral impairments in AX-CPT performance were observed in patients with schizophrenia. In the cognitive control model, these impairments were simulated by reducing gain in the context module, which weakened the representation and maintenance of context information. The gain manipulation in the model was hypothesized to correspond to a core physiological disturbance which may be present in schizophrenia, namely, reduced dopamine effects in PFC. However, it would be desirable if the model could be further validated by testing its predictions in a population other than patients with schizophrenia. For example, it would be particularly useful if performance deficits related to cognitive control could be induced in healthy individuals, using appropriate experimental manipulations.

There are a number of reasons why studying healthy individuals is desirable. First, testing model predictions in healthy individuals would help to address issues regarding its generalizability. For example, it is possible that the findings from the previous study (Study 4) are specific only to patients with schizophrenia. Establishing that the same effects can be observed in healthy individuals would demonstrate that impairments in cognitive control need not be produced solely through physiological disturbances. Second, testing healthy individuals would help to address

issues of experimental control. The findings from the schizophrenia study failed to confirm a number of model predictions. One possible explanation for this failure is that in addition to cognitive control impairments, patients with schizophrenia also suffered from additional impairments which could not be captured in the model. This problem can be addressed by testing healthy individuals, since all performance deficits would be induced through experimental manipulations. Conversely, this approach would also help address the specificity of the model's predictions, since it would be possible to determine if the predicted performance deficits arise from experimental manipulations that are thought to involve cognitive control, but not from manipulations that do not involve control mechanisms. Finally, testing healthy individuals would help to address issues of statistical power. When studying clinical populations, such as patients with schizophrenia, it is often difficult to obtain a large enough sample size to ensure adequate power to detect the effects of interest. This problem may have also contributed to the failure to confirm a number of model predictions. With healthy individuals, it is much easier to obtain a sample size that increases power to sufficient levels. Moreover, if performance deficits are experimentally induced, they can be studied in a within-subjects design, which would further serve to increase power by reducing between-subjects error variance.

In order to appropriately test the model's predictions in healthy individuals, it is necessary to first experimentally induce a state that corresponds to the one found in the reduced gain model. Interestingly, there exist formal analyses of the effects of gain which provide insight into how this could be achieved. In particular, it has been previously demonstrated that gain serves to modulate information processing by altering the signal-to-noise ratio (SNR) across a network of units (Servan-Schreiber et al., 1990). Specifically, with reduced gain, SNR is lower for a given noise level. Conversely, increasing noise will also decrease SNR for a given level of gain. Thus, manipulating noise and manipulating gain should have formally equivalent effects on information processing within a network. As a consequence, increasing the noise level during performance of the AX-CPT should have the same effects on behavior for healthy individuals (i.e., individuals with

normal gain) that a physiological disturbance in prefrontal dopamine has for patients with schizophrenia (i.e., reduced gain).

Additional studies of gain effects on information processing have demonstrated that changes in network SNR depend on the particular point in the network in which gain is manipulated (Servan-Schreiber, 1990; Servan-Schreiber, Bruno, Carter, & Cohen, in press). Thus, in the cognitive control model, in order to achieve the same effect as a selective gain reduction at the context layer, it is necessary to selectively increase noise at this layer as well. To produce this state experimentally, interfering stimuli were visually presented during the delay period, in the form of irrelevant distractor letters. This manipulation was hypothesized to selectively affect context representations, by increasing noise during periods in which context information must be actively maintained but no other processing is required. The specificity of this manipulation was also tested, by determining the effects of increasing noise at other stages of processing. Thus, a separate manipulation was also included, in which both cue and probe stimuli were visually degraded. This latter manipulation was intended to selectively increase noise during the perceptual stage of processing, thus providing a control condition in which to examine the effects of interference. In particular, the degraded condition might correspond well to the simulations of AX-CPT performance conducted with the IGR model in Study 3.

Method

Subjects

Participants in this study were 62 individuals (31 males and 31 females) recruited from the Carnegie Mellon community, with a mean age of 20.69 years (SD = 3.47, Range 18 to 37 years).

Tasks

Participants performed the AX-CPT at both short and long delays, under three conditions: baseline, degraded, and interference. Figure 14 shows a schematic of each of the three conditions. Delay durations for were 2300 msec for short delay trials and 6200 msec for long delay trials. The baseline condition was performed as described under General Methods. In the degraded condition, 85% of the pixels were randomly removed from each stimulus. In the interference condition,

distractor letters appearing in a different color (white) were presented in addition to the cue and probe letters. Participants were required to respond to the distractors to ensure encoding (by pressing the non-target button), but were instructed to otherwise ignore them when determining whether each probe (presented in red) was a target. In short-delay trials, 1 distractor appeared during the delay period, and in long-delay trials 4 distractors appeared during the delay. Distractors also appeared during the ITI in a counterbalanced relationship (i.e., 4 distractors during the ITI of short-delay trials, and 1 distractor in the ITI of long-delay trials). This resulted in an equal number of total distractors (5) presented during each trial, and provided a control against the general effects of distractor presentation. Distractors were presented in a sequential manner, for the same duration as cue and probe stimuli (i.e., 300 msec), with a 1000 msec ISI (regardless of whether the next stimulus was another distractor, a cue, or a probe).

Insert Figure 14 about here

Procedure

Thirty-nine of the participants performed the AX-CPT across 3 testing sessions lasting approximately an hour each. A single task condition was performed in each session (both delays), and the order of sessions was counterbalanced across subjects. Long and short delay trials were segregated into blocks of 10 trials each, and a total of 10 blocks of each delay type in each condition were performed, yielding 100 trials in each condition for both delays. Blocks were presented in pseudorandom order, with the constraint that both delay conditions appear in every set of 2 blocks. The other 23 participants performed the task in two testing sessions, with the first lasting about an hour and the second lasting about an hour and a half. Long and short delay trials were segregated into blocks of 20 trials each, and a total 6 blocks of each delay type in each condition were performed, yielding 120 trials in each condition for both delays. In both sessions all 3 task conditions (baseline, interference, degraded) were performed, 2 blocks of each in the first session and 4 blocks of each in the second session. Although each block consisted of a single task condition, the three conditions were performed in an interleaved fashion across blocks. Block order

was pseudorandom, with the constraint that all 6 block types (3 conditions x 2 delays) appeared in every set of 6 blocks[15].

Data analysis

Analyses on all dependent measures of interest were conducted by including task condition as a within subjects factor. Data were examined for both main effects of task condition and any interactions between task condition and the other experimental factors (e.g., delay and trial type)[16].

Results

Trial-type effects

Performance of the interference condition produced a significant reduction in context sensitivity, as indexed by d'-context ($F(1,61) = 77.79$, $p<.001$; see Figure 15). For non-target trials, a marginal condition x trial type interaction was found ($F(2,122) = 2.39$, $p = 0.09$; see Figure 15). Similar to what was observed in patients with schizophrenia, planned contrasts indicated a trend toward more errors on BX trials ($F(1,61) = 3.58$, $p = 0.06$), but no significant difference on AY trials ($F(1,61) = 0.84$, $p >.10$). Additionally, there were no significant differences between conditions in terms of BY trials ($F(1,61) = 0.00$, $p >.10$). A significant task condition x trial type interaction was also observed in the RT data ($F(2,122) = 8.12$, $p<.001$; see Figure 16). Under interference, participants showed a generalized speeding of RTs (main effect: $F(1,61) = 214.10$, $p<.001$), which was likely related to the increased response frequency of this condition since participants responded to distractors as well as the cue and probe). Planned contrasts indicated that the interaction occurred because speeding of RT under interference was significantly greater for AY trials than for BX trials ($F(1,61) = 13.11$, $p<.001$).

Insert Figures 15 and 16 about here

Delay x Trial Type effects

A trend-level interaction was found between condition and delay in d'-context ($F(1,61) = 2.96$, $p=.09$). Planned contrasts indicated that d'-context significantly decreased with delay in the interference condition ($F(1,61) = 13.44$, $p<.01$), but not at baseline ($F(1,61) = 1.81$, $p>.10$) (see

Figure 15). For non-target errors, the interference manipulation produced a significant task condition x delay x trial-type interaction ($F(2,122) = 5.19$, $p<.01$), which replicated the findings observed with schizophrenia patients. Planned contrasts confirmed that in the interference condition, AY errors decreased from the short to long delay ($F(1,61) = 4.32$, $p<.05$), while BX errors increased from the short to long delay ($F(1,61) = 9.88$, $p<.01$). In contrast, delay did not influence either AY errors ($F(1,61) = 0.01$, $p>.10$) or BX errors ($F(1,61) = 0.19$, $p>.10$) in the baseline condition. There was no effect of delay on BY errors in either the baseline condition ($F(1,61) = 0.42$, $p>.10$) or the interference condition ($F(1,61) = 0.43$, $p>.10$). An analysis of nontarget RT also revealed the presence of a significant 3-way interaction ($F(2,122) = 4.28$, $p<.05$; see Figure 16). Planned contrasts indicated that the speeding of RT in the interference condition relative to baseline was reduced for both BX ($F(1,61) = 11.61$, $p=.001$) and BY trials ($F(1,61) = 7.03$, $p=.01$) at the long delay, but was unchanged for AY trials ($F(1,61) = 0.00$, $p>.10$). Finally, there were also delay-related effects on target RTs. In particular, there was a significant condition x delay interaction ($F(1,61) = 22.19$, $p<.001$), such that RT speeding on AX trials under interference was less at the long delay than at the short (see Figure 16).

Speed-Accuracy Relationships

A significant 3-way interaction between task condition, trial type and response speed was found when comparing baseline to interference ($F(2,122) = 3.45$, $p<.05$) (see Table 5). Planned contrasts confirmed that the interference-related decrease in AY errors was greater in fast responses than slow ones ($F(1,61) = 4.45$, $p<.05$), while the interference-related increase in BX errors was greater in slow responses ($F(1,61) = 10.51$, $p<.01$). The task condition x trial type x response speed interaction observed in the interference condition was further modulated by an interaction with delay ($F(2,122)=6.03$, $p<.01$). This 4-way interaction appeared to be due to greater effects of interference coming at the long delay relative to the short. Posthoc contrasts suggested that for AY trials the interference-related decrease in errors was only found for *fast* responses at the long delay ($F(1,61)=6.26$, $p<.05$). For BX trials, the greatest interference-related increase in errors was for *slow* responses at the long delay ($F(1,61)= 9.87$, $p<.01$).

Insert Table 5 about here

Comparison to Degraded Condition

There were some similarities in the effect of degradation on AX-CPT performance. In comparing the degradation to baseline condition, participants showed a significant reduction in d'-context ($F(1,61) = 126.13$, $p<.001$) and task condition x trial type interactions in both errors ($F(2,122) = 9.35$, $p <.001$; see Figure 16) and RT ($F(2,122) = 7.71$, $p =.001$; see Figure 15. Planned contrasts revealed significantly greater errors for BX trials in the degraded condition relative to baseline ($F(1,61) = 12.25$, $p =.001$), but no significant differences for AY trials ($F(1,61) = 0.49$, $p>.10$). However, there were a number of important differences between interference and degradation as well, primarily in the interaction of delay on performance. Specifically, in comparing nontarget performance in the degraded and interference conditions, a significant task condition x trial type x delay interaction was observed ($F(2,122) = 5.13$, $p<.01$). Planned contrasts indicated that while BX errors increased with delay in the interference condition, they showed a tendency to decrease in the degraded condition ($F(1,61) = 2.54$, $p=.11$). Similarly, while AY errors decreased with delay in the interference condition, they showed no change in the degraded condition ($F(1,61) = 0.85$, $p>.10$).

A similar 3-way interaction was also found in the nontarget RT data ($F(2,122) =3.55$, $p<.05$; see Figure 16). Planned contrasts indicated that while RTs for both BX ($F(1,61) = 11.74$, $p=.001$) and BY trials ($F(1,61) = 19.08$, $p<.001$) increased with delay in the interference condition, there was no delay effect for either BX ($F(1,61) = 1.59$, $p>.10$) or BY ($F(1,61) = 0.58$, $p>.10$) in the degraded condition. In examining target RTs a significant task-condition x delay interaction was found in comparing the interference and degraded conditions $F(1,61) = 4.17$, $p<.05$). Planned contrasts indicated that while there was significant delay-related slowing of AX RTs under interference ($F(1,61) = 38.71$, $p<.001$), there was no such change under degradation ($F(1,61) = 1.91$, $p>.10$). Finally, when response speed was included as an additional factor, it was found that this factor significantly modulated the 3-way interaction between the interference and degraded conditions

($F(2,122) = 3.09$, $p<.05$). Planned contrasts revealed that the 3-way interaction between task condition, trial type and delay was significant for fast responses ($F(2,122) = 5.58$, $p<.01$), but was not significant for slow responses ($F(2,122) = 0.48$, $p>.10$).

Discussion

The results of this study provide an important measure of validation for the model, by demonstrating how failures of cognitive control can be observed in healthy individuals through appropriate experimental manipulations. The interference manipulation produced a pattern of AX-CPT performance that was consistent with a weakening of the representation and maintenance of context. In particular, the changes in performance under interference (relative to baseline) were consistent with 6 out of the 8 predictions of the reduced gain model (condition effect in d'-context, condition x trial type interaction in RT, condition x trial type x delay interaction in accuracy, condition x trial type x delay interaction in RT, condition x delay interaction in target RT, and condition x trial type x response speed interaction in accuracy), and partially consistent with 2 other predictions (the trend-level condition x delay interaction in d'-context and the trend-level condition x trial type interaction in accuracy). Indeed, the qualitative pattern of performance of healthy individuals under interference was strikingly similar to that of patients with schizophrenia. In both groups, context sensitivity decreased and BX errors increased, while AY errors decreased. Furthermore, in both groups these effects interacted with delay, such that they were stronger at the long delay than at the short. The behavioral similarity of the interference condition to the performance of schizophrenia patients in the AX-CPT was specifically predicted by the cognitive control model, in that both conditions are hypothesized to result in impaired representation and maintenance of context. Additionally, the use of a healthy population provided a means of addressing the two major limitations of the schizophrenia study, which were small sample size and reduced experimental control. Moreover, the current study helps establish the generalizability of the model, by demonstrating that its predictions can be tested both in healthy and clinical populations.

Interestingly, however, there were some differences between the performance of patients with schizophrenia and healthy individuals under interference. Although patients made more BY

errors than controls, there was no increase in BY errors under interference. Further, although patients had slower overall RTs than controls, under interference RTs were significantly faster than at baseline. These two differences provide further support for the hypothesis that the performance deficits seen in schizophrenia patients may involve other impairments in addition to ones involving cognitive control, rather than a failure of the model to predict the effects of weakened context representations. There were also subtle differences between the two groups in measures that did reflect cognitive control function. In particular, the context-related increase in BX errors was greater in patients, but the context-related decrease in AY errors was greater under interference. Additionally, the context-related increase in AX errors was greater under interference (11.2% increase relative to baseline) than that found in patients (7.8% increase relative to controls). It is unlikely that the degree of impairment was simply greater in patients, since in two of the conditions the interference manipulation produced greater effects. Nevertheless, this finding could indicate that the context-related impairments produced by interference differ in subtle ways from those observed in patients with schizophrenia. However, further study is needed to answer this question more clearly.

The addition of the degraded condition in this study helped to address the specificity of the model's predictions regarding the role of cognitive control in the AX-CPT. The studies of gain effects within neural networks conducted by Servan-Schreiber and colleagues (Servan-Schreiber, 1990; Servan-Schreiber et al., in press) suggest that reducing gain selectively at the context layer should have different effects from selective gain reductions at other layers of the model. This hypothesis was further confirmed in the simulation of the AX-CPT, in which reductions of gain at the input layer produced qualitative difference in behavioral performance in comparison to gain reductions at the context layer. We hypothesized that the experimental manipulation of stimulus degradation would provide a means of testing this assumption empirically, as it should correspond to the effects of selectively reducing gain at the input layer of the model (by increasing noise at the perceptual stage of processing). The effects of degradation on performance support this hypothesis. In the two strongest predictions of the model – the condition x trial type x delay

interactions in both accuracy and RT – we observed significant effects of interference not only compared to baseline, but also compared to degradation. Whereas AY errors decreased and BX errors increased with delay in the interference condition, under degradation AY errors did not significantly change and BX errors actually decreased. In RT, BX responses got slower with delay under interference, while they were faster under degradation. Moreover, these effects mimic the simulation results comparing the CGR and IGR models. Thus, the results are highly consistent with the hypothesis that interference, but not degradation, produced an impairment in the maintenance of context information.

In contrast, the two manipulations appeared to have much more similar effects with respect to the representation of context. In particular, the degraded condition, when compared to baseline, showed the same condition x trial type interactions in accuracy and RT that were found under interference. Under both conditions context sensitivity decreased, BX errors and RT interference tended to increase while AY errors and interference tended to decrease. These similarities were also consistent with the simulation results from the IGR and CGR models, and further support the hypothesis that both manipulations resulted in impairments to the representation of context. Under degradation, context representations are weaker due to inaccuracies in the encoding process which arise from having noisier perceptual representations of the cue. The inclusion of the degraded condition in this study thus provides a means of addressing the specificity of the model, by demonstrating that only manipulations which affect context information directly produce impairments in both the representation *and* maintenance of context. Conversely, the experimental results under degradation also suggest that cognitive control can appear to be impaired following disturbances that only indirectly affect context representation. However, impairments in context maintenance only occurred following direct disturbances. Consequently, the model suggests that experimental tests of the integrity of cognitive control will be most selective if they include task conditions that probe the maintenance as well as the representation of context information.

The experimental approach taken in this study is similar to that of other recent studies in the neuropsychology and psychopathology literatures. In these studies, healthy individuals were tested

under various task manipulations as a means of examining hypotheses regarding the mechanisms of cognitive impairment in disorders such as visual agnosia (Vecera & Gilds, 1998), aphasia (Miyake, Carpenter, & Just, 1994), schizophrenia (Barch & Berenbaum, 1994), and PFC lesions (Duncan & Sussman, 1995; Roberts, Hager, & Heron, 1994). In particular, the experimental manipulations in these studies were chosen to increase load in the specific cognitive processes thought to be disturbed in a given disorder. The logic is that if the pattern of performance deficits observed in patients can also be elicited in healthy individuals under high-load conditions, it support the involvement of that particular cognitive process in the disorder. Additionally, by demonstrating that similar patterns of impairments can be elicited from healthy individuals under high-load conditions, the studies also help establish that normal and disordered functioning exist on a continuum rather than as qualitatively distinct categories. The current study is analogous to these others, in that the interference manipulation can be seen as increasing the load on a specific cognitive control process, namely, the representation and maintenance of context. Furthermore, by demonstrating that the interference manipulation produced a similar pattern of performance deficits to the one seen in patients with schizophrenia, the study provides support for the hypothesis that cognitive control impairments can be observed in the absence of physiological disturbance, under appropriate task conditions.

However, the current study also differs from the previous ones in terms of the particular nature of the load manipulation. The most commonly accepted approach in the cognitive literature for increasing load is to use a dual-task manipulation. Dual-task manipulations provide a means of decreasing the resources available for a particular cognitive process or operation (Hitch & Baddeley, 1976; Logie, Zucco, & Baddeley, 1990; Pashler, 1994; Wickens, Kramer, Vanasse, & Emanuel, 1983). In contrast, the manipulation used in the current study cannot truly be considered to be a dual-task one. Although the interference condition has a greater task demand overall (since stimulus presentation rate and response frequency increase), the task instructions do not require any further processing of distractor stimuli beyond an early encoding stage (since responding is only dependent upon encoding the presence of distractors and not their identity). Consequently,

interference does not specifically decrease the resources available for context processing, since the *amount* of information that must be represented and maintained as context is unchanged. On the other hand, the interference manipulation can be seen as a high-load condition in that the context must be represented and maintained in the face of increased noise. The decision to use interference over the more standard dual-task manipulation was motivated by the cognitive control model. In the model, reducing gain over the context layer produced a particular cognitive control impairment which captured the pattern of performance deficits seen in patients with schizophrenia. In order to elicit these same impairments in healthy individuals, we employed a task manipulation hypothesized to affect information processing in a manner formally equivalent to a reduction in context layer gain. The results support this hypothesis, by demonstrating that the manipulation was successful in eliciting the anticipated effects. However, it is still possible that equivalent effects on performance could also be achieved through some type of dual-task manipulation that specifically increases the amount of information that must be simultaneously maintained as context. Indeed, this hypothesis is one that should be explicitly tested in future studies.

Study 6: Active Maintenance of Context in Prefrontal Cortex

The results of the previous studies provide an important measure of support for the cognitive control model. Specifically, the computational mechanisms postulated to underlie successful control over behavior in the AX-CPT -- namely, the representation and maintenance of context -- were explicitly implemented in computer simulations and found to successfully account for and predict a rich and complex set of behavioral data in both healthy and clinical populations during performance of the task. However, the model also makes claims about the neural substrates of these mechanisms of cognitive control. In particular, the model proposes that context information is represented in PFC, and is maintained there across delays as a sustained pattern of neural activity. In this study, we attempted to test these claims directly using functional neuroimaging methods, which provide a means of examining brain activity during AX-CPT performance.

Although previous neuroimaging findings have already demonstrated PFC activation during the performance of CPT tasks, in the current study we attempted to ask more specific questions about the role of PFC in actively maintaining context information. To ask these questions, we exploited the fact that task requirements in the AX-CPT are identical in both long and short delay conditions, such that they are only differentiated in terms of the duration over which context must be maintained. Thus, if context is maintained in PFC, greater PFC activity should be observed in blocks of long delay trials than in blocks of short delay trials. Moreover, if context is maintained across the delay period, then analyses of PFC activity dynamics should reveal a sustained pattern of activation across the delay period. The study was conducted using functional magnetic resonance imaging (fMRI), which provides both high spatial resolution and temporal resolution that is sufficient to answer simple questions regarding the dynamics of activity within PFC.

Methods

Participants.

Participants in this study were 23 neurologically normal right-handed individuals (13 males and 10 females), with a mean age of 23.32 (SD = 9.27, Range 18 to 51 years).

Task.

All participants performed the two delay conditions of the AX-CPT. The short delay lasted 1 sec and the long delay was 8 sec. Delay and ITI were counterbalanced across the two conditions, and cue and probe durations were 500 msec (see Figure 18). Thus, total trial duration was 10 sec in both conditions. Eleven participants performed only these two primary delay conditions and the other 12 participants also performed a third delay condition. In this third condition both the delay and ITI lasted 9.5 sec, yielding a total trial duration of 20 sec (referred to as "extra-long"; see Figure 17). This third condition provided a means by which to dissociate cue-related activity from probe-related activity, discussed further below. Responses were made using a hand-held response box with fiber-optic connections to a Macintosh computer in the scanner control room running PsyScope software (Cohen et al., 1993).

Insert Figure 17 about here

Trials were blocked by delay condition, and an equal number of blocks were run in each of the conditions. Scanning occurred during performance of all blocks. For the 12 participants performing all 3 delay conditions, block duration was 4 minutes, and 4 blocks of each condition were performed. In the short and long delay conditions, each block contained 24 trials, while in the "extra long" delay condition each block contained 12 trials. For the other 11 participants, 6 blocks of each condition were performed, in 2 minute blocks containing 12 trials. Blocks were run in a pseudorandom order, such that all conditions were sampled once before starting the next cycle.

MRI Scanning Procedures.

Images were acquired with a 1.5T GE Signa whole body scanner. Twenty-four slices (3.75 mm^3 isotropic voxels) were acquired parallel to the AC-PC line. Functional scans were acquired with a 4-interleave spiral-scan pulse sequence (TR = 640 ms, TE = 35 ms, FOV = 24 cm, flip = 40°) (Noll, Cohen, Meyer, & Schneider, 1995). This pulse sequence allowed 8 slices to be acquired every 2.5s. Scanning was synchronized with stimulus presentation so that 4 scans of 8 slice locations were acquired during each 10 s trial, and 8 scans of these locations acquired on every 20s trial for the additional 12 subjects (see Figure 17). A first set of 8 locations was scanned for three consecutive trials, followed by two additional sets of 8 different locations, each scanned during three consecutive trials. Following this change of location, a fourth trial was presented in the absence of scanning, in order to allow the MRI signal to achieve steady state. This scanning procedure enabled 3 scans of all 24 slice locations to be acquired every 12 trials. The order in which slice locations were acquired was counterbalanced within subjects across blocks, as well as across subjects. This same scanning technique has been used in previous studies (e.g., Cohen et al., 1997), to track the temporal dynamics of activation within each task trial. Anatomical scans were acquired at the same locations as the functional images, using a standard T1-weighted pulse sequence.

Image Analysis Procedures.

Images were co-registered and pooled across subjects using a procedure similar to one used in PET studies (Woods, Mazziotta, & Cherry, 1993). This procedure has been used successfully in

previous fMRI studies to increase statistical power and permit direct quantitative identification of regions that change activity reliably across subjects (Barch et al., 1997; Braver et al., 1997; Cohen et al., 1997). Participants' structural images were aligned to a reference brain using an automated algorithm (Woods, Cherry, & Mazziotta, 1992). All functional images were corrected for movement and scaled to a common mean (to reduce the effect of scanner drift or instability). The functional images were then registered to the reference brain using the alignment parameters derived for the structural scans, and smoothed using an 8 mm FWHM Gaussian filter (to reduce the effects of anatomic variability across subjects). The imaging data, pooled across subjects, were analyzed in a voxelwise manner using a focused contrast procedure testing for greater activity in the long delay relative to the short. The statistical map generated by this procedure was then thresholded for significance using a cluster-size algorithm (Forman et al., 1995), which takes account of the spatial extent of activation to correct for multiple comparisons. A cluster-size threshold of 8 voxels and a per-pixel alpha of 0.005 was chosen, thus ensuring that the image-wise false positive rate was 0.005. The anatomical location of each active region was then determined by reference to the Talairach atlas (Talairach & Tournoux, 1988). Finally, active regions were subjected to an ANOVA analysis to determine the time course of activation by examining the effect of scan-within-trial (1-4 or 1-8).

Results

Delay Effects

Only three brain regions showed significantly greater activation in the long delay relative to the short delay condition. All of these regions were located within frontal cortex: left middle frontal gyrus, right middle frontal gyrus, and left inferior frontal gyrus. As shown in Table 6, both middle frontal gyrus regions corresponded to dorsolateral PFC (DLPFC; BA 46/9), and the left inferior frontal gyrus region corresponded to Broca's area (BA 44/6). These findings regarding PFC are consistent with earlier reported results using a subset of 11 subjects from this group (Barch et al., 1997)[17]. Together, the results support the hypothesis that PFC is involved in actively maintaining context information provided by the cue during the delay period.

Insert Table 6 about here

Time Course Effects

We next examined whether the delay effects observed in PFC were modulated during the time course of the trial. An ANOVA conducted on these 3 regions-of-interest (ROIs) revealed that none showed any interaction between delay and scan-within-trial (all ps > .1; see Figure 18), suggesting that the delay effect was sustained across the entire trial. A similar analysis was also conducted on the subset of subjects who performed the extra-long condition, in order to confirm the effect and to determine whether the sustained activity effect was specific to the delay period. Consequently, in the extra-long condition scans were subdivided into those corresponding to the delay period (scans 2-5) and those corresponding to the ITI (scans 6-8 and 1).[18] For the DLPFC ROIs (averaged together to increase statistical power), a trend-level main effect was observed for greater activity during the delay period than during the ITI period ($F(1,11)=4.41$, $p=.06$) (see Figure 19). In addition, the delay effect again did not interact with scan ($F(3,33) = 0.05$, $p>.1$). The main effect findings were further examined by computing "area under the curve" measures for each of the two events (i.e., the summed increase in signal across the 4 scans). This analysis revealed significantly greater area under the curve for delay-related vs. ITI-related activity ($F(1,11)=5.09$, $p<.05$). The Broca's area ROI did not show the same pattern. Delay-related activity was not significantly greater than ITI-related activity ($F(1,11)=1.35$, $p>.1$), and the interaction with scan was marginally significant ($F(3,33) = 2.33$, $p=.09$).

Insert Figures 18 & 19 about here

Discussion

This study was conducted to validate the model's predictions regarding the role of PFC in the representation and maintenance of context information. Using fMRI, we examined PFC activity during performance of the AX-CPT, which is postulated to rely on actively maintained context. Many previous neuroimaging studies have already demonstrated PFC involvement in both CPT tasks and others which rely upon the active maintenance of context (Baker et al., 1996; Cohen et al.,

1994; Cohen et al., 1987; Fiez et al., 1996; Seidman et al., 1998; Smith, Jonides, & Koeppe, 1996). However, the current findings go beyond these, by establishing the specific characteristics of PFC activity during task performance. First, we established that PFC activity increases when the duration of active maintenance increases. By comparing activity between short delay and long delay trials, we were able to hold all aspects of task processing constant except for the proportion of time over which context must be maintained. We identified PFC regions whose activity reliably and significantly increased in the long delay condition relative to the short. Second, we established that maintenance-related activity was selective to PFC. In the comparison of short and long delay performance across 23 subjects, only 3 regions were found whose activity significantly increased with delay. These 3 regions were all located within PFC – bilaterally in dorsolateral PFC, and in left inferior frontal cortex in a region corresponding to Broca's area. Third, we determined that activity in the dorsolateral PFC (but not in Broca's area) was selectively increased during the delay period relative to the ITI. Finally, we established that the dynamics of PFC activity were consistent with a sustained increase in activity across the entire delay interval. The effect of delay was found to be constant across the entire trial when comparing the long delay condition against the short, and across the entire delay period when comparing delay against ITI. Thus, taken as a whole, the findings from this study strongly support the hypothesis that context information is selectively represented within PFC, and actively maintained there over delay periods.

The finding of a lack of delay-related activation outside of PFC in the AX-CPT has interesting implications for the nature of active memory mechanisms. In particular, a common hypothesis regarding the mechanism of active memory is that information is sustained over delays through recirculation of activity between widely spread cortical (and possibly subcortical) regions. This hypothesis is supported by single-cell recording studies which have found sustained delay-related activity in other cortical regions outside of PFC, such as posterior parietal cortex and inferotemporal cortex (Constantinidis & Steinmetz, 1996; Fuster & Jervey, 1981; Miller, Li, & Desimone, 1993), as well as in subcortical regions, such as the basal ganglia and thalamus (Fuster & Alexander, 1973; Schultz & Romo, 1988). Additionally, neuroimaging studies often show

coactivation of PFC with posterior regions during performance of working memory tasks (Courtney, Ungerleider, Keil, & Haxby, 1996; Smith et al., 1995; Swartz et al., 1995). However, the current study did not result in such a pattern of coactivation between PFC and posterior cortical (or subcortical) regions. This finding suggests that recirculation of activity between widespread areas may not be the primary mechanism underlying active memory for context, or at least not the only mechanism by which active memory might occur. An alternative possibility, is that active maintenance occurs as recirculation of activity that is intrinsic to PFC, either through local networks, or through cross-hemispheric connections. Either of these possibilities is supported by the current results, since the delay-related activity in PFC was bilateral. A third possibility is that active memory within PFC is due to intracellular mechanisms, rather than to network properties. In particular, it is possible that activity patterns in PFC neurons show bistability, with one stable state corresponding to baseline and other corresponding to the maintenance of information (Guigon et al., 1995; Kirillov, Myre, & Woodward, in press). Unfortunately, the methods used in the current study cannot provide any evidence as to whether active memory occurs through intracellular or local circuit mechanisms.

The PFC regions that satisfied all of the criteria for active maintenance of context were found in the dorsolateral portion of PFC, in the left and right middle frontal gyrus. The specific location of these regions corresponded to Brodmann's area (BA) 46/9, which has been an area long associated with active maintenance in both animal neurophysiological and human neuroimaging studies. The most reliable sites for detecting maintenance-related activity in primate single-cell recordings have been in the region around the principal sulcus (Goldman-Rakic, 1987), which appears to be directly homologous to BA 46 (Rajkowska & Goldman-Rakic, 1995). Moreover, this region has been observed to be critical for performance in delayed-response tasks, as impairment can be observed even following highly circumscribed lesions (Butters, Pandya, Sanders, & Dye, 1971). In humans, BA 46/9 has also been commonly observed in neuroimaging studies involving active maintenance and working memory (for reviews see, Fiez et al., 1996; McCarthy, 1995). In particular, this region has been found to be engaged by tasks requiring maintenance of a wide

range of stimuli including, faces (Courtney et al., 1996), letters (Cohen et al., 1994), digits (Grasby et al., 1994; Petrides, Alivisatos, Meyer, & Evans, 1993), locations (McCarthy et al., 1994), and abstract shapes (Petrides et al., 1993; Swartz et al., 1995). More recently, a few neuroimaging studies which have attempted to isolate the maintenance component of working memory have also demonstrated activity in BA 46/9 (Cohen et al., 1997; Courtney et al., 1997; Fiez et al., 1996). As such, the current results are consistent with previous findings, and provide more specific support for the hypothesis that dorsolateral PFC is responsible for actively maintaining representations of context information through sustained patterns of neural activity.

It is interesting that the left inferior frontal cortex was found to exhibit increased activity as a function of delay in the condition-based analysis, but in the event-related analysis exhibited a pattern of activity dynamics that was qualitatively different from that observed for DLPFC. This dissociation between inferior and dorsolateral PFC is consistent with our previous findings using the n-back task (Cohen et al., 1997). In that study, we observed that DLPFC showed sustained activity that was sensitive to working memory load. In the left inferior frontal cortex, activity was also load sensitive, but appeared to interact with scan, such that the activation was more sustained at higher levels of load. This potential functional dissociation between left inferior and dorsolateral PFC has also been observed by other investigators (e.g., Awh et al., 1996). Many other studies have observe left inferior frontal cortex activation during language and verbal working memory tasks (Fiez et al., 1996; Frackowiak, 1994; Zatorre, Meyer, Gjedde, & Evans, 1996). In these studies, activity in this region has been typically interpreted as reflecting articulatory planning and rehearsal (e.g., Paulesu, Frith, & Frackowiak, 1993). This interpretation would also be consistent with our delay effect, since it is plausible that participants engaged in rehearsal of the cue to improve maintenance over the delay. However, if this were the case we would expect to observe a more sustained pattern of activity dynamics in the event-based analysis, since rehearsal is presumably an ongoing, repetitive process. On the other hand, it is possible that the minimal load imposed by the cue (i.e., a single letter), did not require rehearsal to occur throughout the delay period. This latter interpretation is also consistent with the findings from the n-back study, in which left inferior

frontal cortex showed a somewhat transient pattern of activity dynamics during the 1-back condition (which also only required maintenance of a single letter).

Although our results point to a selective role for dorsolateral PFC in active maintenance during the AX-CPT, previous findings have suggested that other regions within PFC may also subserve active maintenance functions. Some of these findings have suggested that other PFC regions are involved in active maintenance of different types of information, For example, in a single-cell recording study, Wilson et al. (1993) have observed more ventrolateral regions of PFC (around the arcuate sulcus) that also show delay-related activity in delayed-response tasks. Interestingly, Wilson et al. (1993), found a dissociation in the delay-related activity between neurons in dorsal and ventral PFC. Neurons in the principal sulcus showed increased activity under conditions where location information had to be maintained, but not under conditions which required maintenance for identity information. Neurons in the arcuate sulcus showed the opposite pattern of activity, thus establishing a double dissociation based on the content of information in active memory. There have been reports of similar sorts of dissociations in the neuroimaging literature, distinguishing between verbal, spatial or object working memory (Smith & Jonides, in press). However, these types of dissociations have often failed to replicate under more rigorous conditions (Cohen et al., 1998). The results from the neuroimaging literature have led to speculation by some authors that PFC is organized not by the content of information being maintained, but rather by functional distinctions, such as between the maintenance and manipulation of information (D'Esposito et al., 1998; Owen, Evans, & Petrides, 1996; Petrides, 1996). Under this interpretation, ventral and inferior regions of PFC (i.e., BA 44/45/47) subserve pure maintenance functions, whereas dorsolateral regions of PFC (i.e., BA 46/9) are primarily involved in situations where the contents of working memory must be manipulated or monitored.

Given the current controversy between the organization-by-content and the organization-by-process views, it is not possible to say definitively whether the activation of BA 46/9 in the current study reflects the particular content of the information being maintained (i.e., verbal material) or rather something about the functional processes being engaged (e.g., pure maintenance vs.

"manipulation"). However, it is possible that maintenance of context information is qualitatively different from the more domain-specific "buffer-type" of maintenance typically discussed in standard theories of working memory (Baddeley, 1992). For example, in traditional verbal working memory tasks (e.g., immediate serial recall), it may be sufficient to maintain stimulus information as an articulatory or phonologically-based representation. Maintaining information using this type of representational code might involve ventral or inferior regions of PFC. In contrast, such articulatory or phonologically based representations may be necessary, but not *sufficient*, to drive performance in the AX-CPT. Instead, AX-CPT performance is critically dependent upon transforming the cue stimulus into a context-based representation, which carries information about the consequences of the cue for future stimulus evaluation and response. For example, following presentation of a "C" cue, it is necessary to transform this information such that the cue is represented as "non-A", and that this representation is used to bias the system to make a non-target response to a subsequent X probe. It is the representation and maintenance of information in this context-based code that we feel best characterizes the functional role of dorsolateral PFC and may distinguish it from other PFC regions.

Study 7: Effects of Cognitive Control Impairment on PFC Activity

The results of the previous study (Study 6) support the hypothesis of the cognitive control model that representations of context information are actively maintained in PFC. This last study tests a corollary of the hypothesis, that when the representation and maintenance of context is impaired, this impairment should be observable in terms of altered activity dynamics within PFC. To examine this hypothesis we examined brain activity during the performance of the AX-CPT under both baseline and interference conditions. The behavioral data of Study 5 suggest that the representation and maintenance of context is weakened under interference conditions. Consequently, we predicted that PFC activity during this condition would show changes in delay-related dynamics that would be similar to those observed in the simulations of Study 3 carried out with the CGR model. In particular, we expected to observe that activity in PFC would decay over the delay period during performance of the interference AX-CPT. To test the specificity of these

predicted effects, a subset of participants also performed the AX-CPT under degraded conditions, as was done in Study 5.

Methods

Participants.

Participants in this study were 21 neurologically normal right-handed individuals (11 males and 10 females), with a mean age of 24.62 (SD = 5.08, Range 20 to 37 years).

Task.

Participants performed the "extra-long" delay condition of the AX-CPT described in Study 6. Duration parameters were as in that study (500 msec stimulus duration, 9.5 sec delay and ITI). The task was performed under both baseline and interference conditions. In the interference condition, a total of 8 distractors were presented during each trial, 4 during the delay period and 4 during the ITI. Distractors were presented in a sequential manner, for the same duration as cue and probe stimuli (i.e., 500 msec), with a 1500 msec ISI (regardless of whether the next stimulus was another distractor, a cue, or a probe). In addition to the two primary task conditions, 13 participants also performed the AX-CPT under degraded conditions. Stimulus degradation was manipulated exactly as in the previous behavioral study (Study 5), with 85% of the pixels randomly removed from each stimulus. This subset of 13 participants performed 5 blocks each of these 3 task conditions, and the remaining 8 participants performed 8 blocks each of the baseline and interference conditions. Blocks were run in a pseudorandom order, such that each condition was sampled once before starting the next cycle. Each block contained 10 trials, and scanning occurred during performance of each trial.

MRI Scanning Procedures.

In this study, a 2-interleave spiral-scan pulse sequence (TR = 1250ms, TE = 35ms, FOV = 24cm, flip=60°) was used which allowed us to simplify the acquisition procedure. The sequence enabled the acquisition of 16 axial locations ($3.75mm^3$) every 2.5s. Scans were synchronized to the start of every trial, and eight scans were acquired during the course of the trial.

Image Analysis Procedures.

The preprocessing of the functional imaging data included movement correction, mean normalization, smoothing, and co-registration of subjects using the procedures described in Study 6. Data were analyzed through an ANOVA with task condition (baseline vs. interference) and scan-within-trial (scans 1-8) as factors. Because the study was hypothesis-driven, the analysis was conducted in a confirmatory manner, using the 3 PFC ROIs derived from Study 6. The confirmatory analysis was carried out by co-registering the subjects of the current study to those of the previous study (Study 6). The ROIs were then laid over each brain. The voxels corresponding to each ROI were averaged together and analyzed as a composite. For regions showing significant effects, post-hoc analyses were conducted on individual voxels within the ROI. The primary effect of interest was the presence of a task x scan interaction.

Results

Behavioral Performance

Analyses of the behavioral data indicated that one subject was not performing the interference task appropriately (the subject made 100% errors on AX trials). Consequently, this subject's behavioral and imaging data were excluded from all further analyses. For the remaining 20 subjects, the interference effect appeared to replicate the results of Study 6. A significant decrease in d'-context was observed under interference (baseline = 2.89, interference=2.32; $t(19) = 3.39$, $p<.003$). The task condition x trial type interaction for nontarget accuracy was not significant ($F(2,38)=1.43$, $p>.1$). However, numerically, there was both an increase in BX errors (baseline = 8.5%; interference = 16.1%) and a decrease in AY errors (baseline = 7.7%; interference = 5.2%). For nontarget RT, the interaction was marginally significant ($F(2,38)=2.5$,$p=.09$). Post-hoc contrasts indicated that although AY responses were significantly slower than BX responses at baseline ($t(19) = 2.2$, $p<.05$), there were no differences under interference ($t(19) = 0.1$,$p>.1$).

Imaging Data

In the baseline condition, all 3 ROIs showed highly significant main effects of scan (all ps $< .001$), suggesting that these regions were sensitive to task events. Moreover, when comparing baseline to interference, a significant task x scan interaction was observed for the left DLPFC

region ($F(7,133)=2.19$, $p<.05$). This interaction was not significant for either the right DLPFC region ($F(7,133)=1.62$, $p>.1$) or the left inferior region ($F(7,133)=0.847$, $p>.1$). Interestingly, however, the left inferior region did show a main effect of task, with greater activity under interference ($F(1,19)=4.53$, $p<.05$). Neither of the two DLPFC regions showed such effects (both ps > .1). The task x scan interaction in the left DLPFC was also followed up by post-hoc voxelwise analyses. These revealed that the effect was highly consistent, with 8 of the 9 voxels showing interactions significant at the $p < .10$ level. The nature of this interaction can be seen in Figure 20, which shows the pattern of activity dynamics in both task conditions for the two most significant voxels (averaged together). Finally, to test for specificity of this interference effect in the left DLPFC region, we examined the effects of degradation in the subset of participants performing all 3 task conditions. There was no task x scan interaction when comparing baseline to degraded conditions ($F(7,77) = 0.9$, $p>.1$), and the pattern of activity dynamics under degradation looked qualitatively similar to baseline. In contrast, the comparison of degraded to interference revealed a trend-level task x scan interaction ($F(7,77) = 1.84$, $p=.09$).

Insert Figure 20 about here

Discussion

The goal of this study was to demonstrate that the cognitive control model can account for the pattern of PFC activity dynamics under both normal and impaired conditions. The model suggested that under interference conditions the representation and maintenance of context would be weakened, and that this would be reflected in terms of a decay in PFC activity during the delay period of the AX-CPT. We conducted a second fMRI experiment that built upon the results of the first one. In particular, we conducted a confirmatory analysis which enabled us to identify the exact regions of PFC that were previously shown to demonstrate delay-related effects in the AX-CPT. Moreover, this type of analysis provided the strongest test of our hypotheses, both anatomically and conceptually, by forcing us to demonstrate that these previously identified regions would: 1) replicate the pattern of activity dynamics observed in the baseline AX-CPT in a

completely different group of participants; and 2) also show effects of interference which were indicative of a delay-related decay in activity. These hypotheses were both confirmed by the data. We observed a significant task condition x scan interaction in the left DLPFC region. As Figure 20 shows, the activity dynamics in the baseline condition are similar to that observed in the extra-long condition of Study 6 (cf. Figure 19). However, under interference, the cue-related activity decayed more rapidly during the delay period. It is significant that the interference condition did not reduce the peak amplitude of the cue-related response but rather affected its latency and duration. This pattern further supports that the interference effect in left DLPFC does not reflect an overall attenuation of activity, but rather a decay of activity. As such, the results are fully consistent with the simulations performed under conditions when context representation and maintenance is disturbed (i.e., the CGR model of Study 3). In those simulations, context-related activity was also found to decay over the delay period.

The finding that performance of the degraded condition of the AX-CPT did not produce this change in PFC activity dynamics is consistent with the behavioral results (e.g., Study 5), and further supports the specficity of the effect. Specifically, PFC activity dynamics were not altered by task manipulations which increase task difficulty by increasing demand on perceptual processing stages. It is only when task manipulations are specifically targeted towards increasing the difficulty of context processing that these alterations could be observed. It is also interesting to note that without benefit of the model, one might have predicted that the interference condition would have a different effect on PFC activity. For example, one plausible hypothesis is that under interference the PFC response would be one of enhancement rather decay. The logic of this hypothesis is that the interference condition increases active memory load, and that the response to this increased load would be increased PFC activity, as has been observed in other working memory tasks, such as the n-back (Braver et al., 1997; Jonides et al., 1997). The model makes clear that interference does not increase memory load per se, because no more information must be maintained as context. Rather, the model suggests that the presence of interfering items should serve to weaken the representation and maintenance of context, and that this effect should accumulate over the delay period.

Although the results are not consistent with the hypothesis that DLPFC activity is enhanced under interference, they do not rule out the possibility that compensatory strategies are employed by subjects to counteract the effects of interference. Indeed, the data provide some tantalizing hints in this regard. First, in the left DLPFC region it is interesting to note that although activity following the cue decays more quickly under interference, it rises more quickly as well. In particular, at scan 2 the cue-related response is appears to be greater under interference than baseline, although this effect was not statistically significant. One intriguing hypothesis is that under interference, the initial representation of context is amplified, possibly due to enhanced attentional allocation. This early enhancement may help to counteract the disruption and decay of context representation due to the subsequent presentation of distractor stimuli. A second indication of potential compensatory strategies in play during interference is the finding of increased activity in the left inferior frontal cortex. If activity in this region represents the presence of articulatory planning and rehearsal, the increased activity under interference could reflect the increased use of this processing strategy during the interference condition. Indeed, it is well-known and well-documented that subjects use rehearsal as a primary strategy for maintaining information in the face of interference (e.g., Baddeley, 1986). However, these interpretations are still somewhat speculative and premature, given that they weren't the primary focus of this experiment. Nevertheless, they do make clear the need for a more systematic study of the relationship between subjects' strategies and brain activity patterns.

Another issue worth noting is the lack of a task x scan interaction in the right DLPFC region. In the previous study (Study 6), this region had shown a pattern of effects which were almost identical to that observed for the left DLPFC region. Moreover, upon inspection, they appeared to be located in homologous regions of cortex. However, in the current study, these two regions showed very different patterns. Not only was the task x scan interaction not significant on the right, but it also appeared qualitatively as if the interference condition resulted in no change in activity dynamics, thus arguing against a lack of power. Although we have no explanation for this lack of an effect, it does suggest that the two regions might in fact be functionally dissociable.

Obviously, further work is needed to provide support for this assumption, but the results do make clear the point that demonstration of dissociability between very similar cortical regions may require the use of systematic testing and subtle task manipulations.

The results of this study also have interesting implications for neuroimaging studies of patients with schizophrenia. In the present work we have suggested that the interference condition of the AX-CPT produces disturbances in the representation and maintenance of context information. This disturbance is reflected both in terms of behavioral performance and PFC activity dynamics. As discussed above, we have also suggested that performing the AX-CPT under interference puts healthy subjects in an experimental state that is somewhat analogous to that experienced by schizophrenia patients. Thus, based on the current results, the model makes strong predictions about the pattern of brain activity that would be expected if functional neuroimaging studies were conducted with schizophrenia patients performing the AX-CPT. First, the model predicts that patients should show reduced delay-related activity in PFC during AX-CPT performance. Second, the model predicts that patients should show a pattern of PFC activity dynamics similar to that observed by healthy conditions in the interference condition; that is, activation which appears to decay over the delay period. We have begun to test these predictions empirically, and preliminary results are consistent with the predictions of the model (Barch et al., in preparation).

GENERAL DISCUSSION

The seven studies presented in this paper have provided convergent support for a theory that addresses a fundamental aspect of cognitive control – the ability to facilitate the processing of task-relevant information and the selection of task-appropriate responses, while inhibiting competing sources information and prepotent, but inappropriate, response tendencies. These control functions were studied in the context of performance in the AX-CPT, a well-controlled and easily interpretable task for examining response competition processes. Furthermore, the theory was implemented as an explicit model developed within the connectionist framework, such that simulations of task performance could be conducted. The findings from these studies demonstrate that the theory and

model can account for: a) a wide range of behavioral effects observed in healthy individuals under normal task conditions; b) the pattern of deficits observed in patients with schizophrenia; c) the similarity of these deficits to those observed in healthy individuals under interference conditions; d) the selectivity and dynamics of PFC activity during task performance under baseline and interference conditions.

The theory provides an explanation for these findings that is specified in terms of both psychological and neural mechanisms. At the psychological level, the theory postulates that cognitive control is mediated through representations of prior context, which serve to bias the local competitions that occur at many stages of processing in task-relevant pathways. At the biological level, the theory claims that context information is represented within PFC (specifically in dorsolateral regions), and that these representations can be maintained for short periods of time as sustained patterns of neural activity. The connectionist framework provides a means of bridging these two levels of explanation, by demonstrating how information processing dynamics reflect neurobiological constraints, and how specific behavioral effects can arise from manipulations of neurobiological variables. Moreover, an important contribution of the theory is that it integrates the three different perspectives on the role of PFC in cognitive control that were described in the Introduction: coordination, active memory, and inhibition. Specifically, we argue that it is precisely because PFC is able to actively maintain a specific type of information – context – that it can serve to coordinate task-appropriate responses and inhibit competing, but task-inappropriate responses. As such, the theory presented here provides new insights into cognitive control and related concepts such as working memory, inhibition, attention, and PFC function. Below, we discuss each of these concepts in turn, and the relationship of the theory to other computational accounts addressing similar issues. Finally, we discuss additional issues in cognitive control that are not addressed by the current model, and that provide directions for future research and development.

Cognitive Control

A primary component of the theory is that control is exerted by activation of a mechanism which serves to bias on-going processing by both facilitating task-appropriate information and

responses, and by inhibiting competing information and responses which are task-inappropriate. In this respect, the theory makes claims similar to that of Posner & Snyder (1975) and Shiffrin & Schneider (Shiffrin & Schneider, 1977), who also argue that the conscious control of cognition results in both costs and benefits in task performance. However, the mechanisms by which these costs and benefits accrue are largely left unspecified in traditional theories. In the theory presented here, the mechanism of control is that of biasing activation states through sustained representations of context. This mechanism is in agreement with Norman & Shallice (1986), who argue that the SAS mediates control in a purely modulatory fashion, that is "neither sufficient nor necessary to cause selection (p.8)".

The current theory shares many other commonalities with the Norman & Shallice model, including a focus on response competition and selection and the assignment of control functions to PFC. However, this account goes beyond Norman & Shallice, by additionally beginning to specify the properties of PFC representations, and by implementing these ideas within simulation models. In particular, we have argued that PFC representations encode context information, which can be in a abstract form, such as that provided by task instructions, but may also be much more specific, such as that provided by prior stimuli (as in the AX-CPT). Furthermore, we have suggested that a critical property of PFC representations is that they can be actively maintained for short-periods of time, which enables them to modulate processing of subsequent inputs or selection of subsequent responses. Most importantly, through computer simulations, we have explicitly demonstrated that these ideas regarding PFC function can account for detailed aspects of behavioral and neuroimaging data from both healthy individuals and those with disorders affecting PFC (i.e., patients with schizophrenia).

The theory presented in this paper also addresses the relationship between controlled and automatic processes. Processes requiring cognitive control have previously been argued to be slow, effortful, affected by cognitive load, and subject to interference effects, whereas automatic processes are fast, effortless, and involuntary (Posner & Snyder, 1975; Shiffrin & Schneider, 1977). In the model, these distinctions can be seen in terms of the relationship between the direct and indirect

pathway. Stimulus-response mappings that have strong connections within the direct pathway exhibit many of the characteristics of automaticity. Processing is fast and robust, and therefore less subject to load and interference effects (e.g., AX and BY trials in the AX-CPT). In contrast, stimulus-response mappings that are dependent upon support from the indirect pathway require a greater degree of cognitive control (e.g., BX trials). Processing is: a) slower, because of the indirect flow of information; b) more effortful, because it involves resolving competition between representations; c) subject to interference effects from the competing pathway; and d) modulated by load, because active maintenance can be disrupted (for example, when noise levels are high, as in Study 5). However, the current theory builds on previous work by our group (Cohen et al., 1990), by postulating that the controlled-automatic distinction is not categorical, but rather one that lies on a continuum governed by the frequency of experience.

In particular, in the model of the AX-CPT, dominant response tendencies emerge directly from the relative frequencies with which the various stimulus-response mappings are experienced. For example, because the mapping between the "X" stimulus-target response mapping is experienced most frequently, its strength is greatest. Thus, the mediating influence of context representations are more important for BX trials than for AX trials. However, the prepotency of the X-target response mapping was not absolute, since performance on AX trials was also found to be impaired when context representations were weakened. Moreover, these results illustrate the continuous nature of interference effects, since prepotencies were induced after only a limited number of experiences (i.e., participants only learned the stimulus-response mappings during the experimental session). Consequently, it is likely that the response prepotencies in the AX-CPT are highly malleable, and would quickly change if the relative frequencies of the stimulus conditions were manipulated (e.g., making BX trials more frequent than AX trials). Indeed, this type of manipulation is one that could fruitfully be studied in future research.

Working Memory

The theory presented in this paper provides new insights regarding the relationship between cognitive control and working memory. The most prominent model of working memory in the

cognitive psychology literature has been the one put forward by Baddeley (1986). In Baddeley's framework, working memory is a triarchic system composed of two domain-specific buffers, the phonological loop and the visuospatial scratchpad, and a central executive that oversees and coordinates their operation. However, in Baddeley's model, the central executive largely serves as a catch-all designation for a number of control functions that remain for the most part undifferentiated in terms of their underlying mechanisms (but for recent progress in this area, see Baddeley, 1996). In contrast, the account that we have put forward has focused on a critical control function, the ability to facilitate task-relevant information and responses while inhibiting competing information and responses. In the studies presented, we have attempted to specify the nature and mechanisms of this function in terms of the: a) the task situations in which it is most critical (e.g., when there is competition due to response prepotencies); b) the dynamics of its normal operation and its influence on behavior (e.g., facilitation and inhibition effects on target and non-target responses; interactions with delay; speed-accuracy relationships); c) its neural substrate (e.g., dorsolateral PFC) and d) the causes and consequences of its breakdown (e.g., the effects of interference manipulations).

Additionally, we have argued that a fundamental mechanism of cognitive control is the active maintenance of a particular type of task-relevant information, which we've referred to as context. Moreover, because context is actively maintained, it can act to influence how incoming information is represented and processed. In many ways, this account is consistent with the common definition of working memory as a system that involves both the temporary storage and manipulation of task-relevant information (Baddeley, 1986; Just & Carpenter, 1992). However, at least in the formulation of working memory advocated by Baddeley, storage and control are strictly segregated, with storage occurring only in the two buffers, and control operating through the central executive. The account we have proposed rejects the strong form of this distinction, by suggesting that storage and control can be intrinsically related. In particular, a primary feature of the model which provides the capacity for cognitive control is the active maintenance of context representations. Furthermore, active maintenance is a central aspect of control function, in that it

allows a means for task-relevant information to be utilized in a manner that allows it to directly bias on-going processing. Indeed, this feature of the model leads to a number of new predictions regarding working memory. For example, the theory suggests that storage in working memory does not necessarily require the operation of the two buffer systems. In particular, the studies we have presented in this paper argue that maintenance of cue-related information in the AX-CPT occurs within the context layer, which in Baddeley's terminology, could be thought of as a component of the central executive. Moreover, in healthy individuals under normal task conditions, we observed that maintenance was not subject to decay (at least over the delay intervals tested). In the model, this characteristic is due to the strong recurrent connectivity within the context layer that enables representations to be self-sustaining and robust to normal levels of noise. This property of context maintenance may differentiate it from maintenance within the two buffer systems. Specifically, it has been argued that information within the phonological loop decays in about 2 seconds, if not actively refreshed by a rehearsal process (Baddeley, Thomson, & Buchanan, 1975).

On the other hand, it is possible to formulate an alternative interpretation of the results which argues that cue-related information in the AX-CPT was actually maintained within the buffer systems (most likely the phonological loop, given the nature of the stimuli). In particular, it could be argued that decay did not occur in the AX-CPT because participants were engaging in rehearsal of the cue over the delay. Moreover, this interpretation could also provide a coherent account of both the effects of interference (Study 5) and the neuroimaging results (Study 6), by attributing interference effects to be similar to the irrelevant speech effects found in the verbal working memory literature (Salame & Baddeley, 1982), and the delay-related activation of left inferior frontal cortex as evidence of articulatory-based rehearsal (Paulesu et al., 1993). However, there a number of problems with this interpretation. First, in a previous study of the AX-CPT, it was found that articulatory suppression, the standard method for preventing verbal rehearsal, produced no greater decrements in task performance than another manipulation which did not affect rehearsal (Gupta, 1995). Second, the delay-related activity of left inferior frontal cortex was increased rather than decreased by the interference manipulation (Study 7). Thus, it is hard to argue that the

interference manipulation prevented the use of rehearsal processes. Third, the interference effects observed in Study 6 are likely caused by a mechanism different than the one causing irrelevant speech effects, since irrelevant speech effects are thought to only occur when the distracting information is presented auditorily (and thus has obligatory access to phonological processing mechanisms, e.g. Jones & Morris, 1992; Salame & Baddeley, 1989). Since the distractors in the interference study were presented visually, it does not seem as if they would produce obligatory phonological recoding. Nevertheless, further studies are needed to more systematically demonstrate that active maintenance occurs in the AX-CPT without the involvement of the two buffer systems hypothesized by the Baddeley model. This area of research is a critical one, because the discovery of distinctions between maintenance of context information and other forms of short-term maintenance could have wide-spread implications for the study of working memory. In particular, these distinctions have the potential for resolving conflicting findings within the experimental literature regarding such issues as domain-specificity, decay rates, interference effects, neuroanatomical substrates, and capacity limits (Miyake & Shah, in press).

Inhibition

The role of inhibition in cognition is an area garnering much recent attention (for reviews see Dagenbach & Carr, 1994; Dempster & Brainerd, 1995). In much of this work, it has been demonstrated that certain effects on processing appear not to be easily accounted for by activation of processing pathways, but can only be explained if one postulates a mechanism that serves to attenuate or suppress activation. This recognition that inhibition as well as facilitation effects occur in cognition has prompted some theorists to argue that these two processes may be mediated by two fundamentally different mechanisms (Posner & Snyder, 1975; Rafal & Henik, 1994). In the cognitive control model, we provide a demonstration of how a single mechanism can produce both inhibitory and facilitory effects. In the model, facilitation and inhibition both occur when maintained representations in the context module bias local competitive interactions in the task-specific modules of the direct pathway. When this bias favors the representation which also receives strong bottom-up support, facilitation in processing should result, as exemplified in the

case of AX and BY trials in the AX-CPT task. When context representation was disturbed (in the simulations of Study 3 and the interference task of Study 5), AX and BY responses were slowed and AX errors increased. The model also shows how inhibitory effects can occur when the context-mediated bias favors the weaker pathway, such as occurs for BX trials. In this case, context is required to overcome the bottom-up support given to the task-inappropriate representation. When context representation and maintenance is disturbed, performance on BX trials greatly diminishes, leading to the appearance of an inhibitory deficit. Thus, in the model facilitation and inhibition are just flip sides of the same coin, determined solely by where the context-mediated bias occurs. Consequently, the model provides a demonstration of how the apparent behavioral independence of facilitation and inhibition need not necessarily indicate two separate underlying mechanisms.

A second aspect of the model which bears on discussions of inhibition is in the nature of the inhibitory mechanism itself. In the model, inhibition is an ubiquitous aspect of processing, and one that occurs at a local level throughout the system. Much as others have postulated (e.g., McClelland, 1993), we assume that inhibition occurs in the form of lateral competitive interactions within each processing module. The role of context is to serve as top-down input which can act to bias the outcome of these competitions. However, we postulate that context representations can only have excitatory effects on processing. This idea is consistent with general principles of brain organization in which long-range connectivity between cortical regions is mediated solely by excitatory contacts (i.e. pyramidal neurons). Studies of PFC neuroanatomy are also consistent with this principle, in that the projections from PFC to posterior systems (e.g., parietal and temporal cortex) are excitatory ones (Fuster, 1989). Thus, in the model, inhibitory effects of context occur only indirectly, by preferentially activating one competitor over another. This feature of the model is also in accord with other biologically-based computational models of inhibition, most notably that of Houghton and Tipper (1996). In particular, the Houghton and Tipper model suggests that inhibition is not due to a "central" mechanism, but instead occurs as a purely local, and distributed phenomenon.

It is important to note, however, that there may be other forms of inhibition that are not yet accounted for in the model. For example, in the Houghton and Tipper model, certain experimental effects such as negative priming (Tipper & Cranston, 1985) and inhibition of return (Posner & Cohen, 1984) effects are captured by an "inhibitory rebound" in activation that occurs after distractor inputs are removed. This type of mechanism is not present in the cognitive control model, and may represent an important component of many inhibitory effects in cognition.

Attention

The theory we have advocated links the mechanisms of cognitive control with attentional function. In the model, context representations in the indirect pathway exert a modulatory influence over processing in the direct pathway. Thus, context representations can be seen as an attentional template (Duncan & Humphreys, 1989), in that they preferentially select or facilitate task-relevant processing streams, while suppressing or filtering out irrelevant ones. In the current work, we have focused on a task, the AX-CPT, in which attention is required for selection of the task-appropriate response rather than for selecting particular dimensions or features of perceptual input. Context serves to direct attention to the appropriate response for a potentially ambiguous stimulus (e.g., X) by priming the activation level of that response during the delay period. This role of cognitive control in directing attention to action is very similar to the role ascribed to the SAS in the Norman & Shallice model. However, in other work, we have also argued that context representations are not restricted to influencing processing at the action selection stage and can bias competition at multiple stages of processing (Cohen et al., 1990; Cohen et al., 1994; Cohen et al., 1992). This has important implications, because it suggests that the same control mechanism can be engaged in a common fashion across tasks which appear to make different attentional demands.

The alignment of attentional functions to the PFC in the cognitive control model is consistent with a number of accounts put forward regarding the neurobiology of attention (e.g., Posner & Petersen, 1990). In particular, the particular role that we have ascribed to PFC in attention is highly compatible with the biased competition model put forward by Desimone & Duncan (1995). In their model, competition for representation occurs at many stages of processing,

and attention serves to bias this competition, by providing an additional source of activation which favors task-relevant inputs and task-appropriate responses. As in the cognitive control model, the competition for representations occurs locally as a within-layer phenomenon, and attentional bias may occur in a "top-down" fashion. Additionally, the biased competition model suggests that PFC representations will not be equally involved in all attentional tasks, since bias can also occur in a bottom-up, stimulus-driven fashion. This latter constraint is also compatible with the cognitive control model. Moreover, our theory suggests that PFC involvement will be most critical when tasks require that the attentional bias derived from internal representations of prior context (e.g., task instructions, prior stimuli, the result of processing multiple stimuli).

PFC function

A central premise of the cognitive control model is that context information is represented and maintained within PFC. This idea is consistent with a wide range of data from neurophysiology, neuropsychology, and most recently, neuroimaging. Moreover, the model provides a reinterpretation of the literature on active memory in PFC. Goldman-Rakic (1987) has argued that PFC is specialized for actively maintaining sensory information in working memory, by demonstrating that neuronal activity in PFC shows sensitivity to specific sensory features, and that impairing this activity (e.g., through lesions or pharmacological manipulations) produces selective impairments in the maintenance of this information. The theory presented here is consistent with these findings, but further suggests that in at least dorsolateral PFC, active maintenance occurs for particular type of information, namely, context representations. Because only the behaviorally relevant aspects of prior information are stored in context representations, active maintenance in dorsolateral PFC is not likely to be equivalent to a short-term memory trace. As such, the representations may be insensitive to certain featural or response distinctions. Moreover, because context may be represented as a goal state or the end result of integrating a series of inputs, actively maintained representations in PFC might be very complex and not easily analyzed by the standard methods of sensory physiology (i.e., receptive fields mapping, post-stimulus histograms). Indeed, initial support for these hypotheses comes from the recent work of

Miller and colleagues, who have demonstrated that neuronal activity in PFC: a) is responsive to only the behaviorally-relevant subset of information in a visual cluttered array (Rao, Rainer, & Miller, 1997b); b) is insensitive to intervening, but irrelevant inputs, presented during the delay period between cue and probe (Miller et al., 1996); and c) represents both spatial and object identity information in an integrated form when task demands require that both domains of information be maintained to respond appropriately (Rao, Rainer, & Miller, 1997a). Future studies along these lines will help to shed more light on the distinctions between active memory as it occurs in PFC, and the active memory and short-term storage functions of other regions (e.g., inferotemporal cortex and posterior parietal cortex).

The theory that we have proposed also suggests a common processing framework by which to integrate the active memory functions of PFC with its putative role in control and inhibition. In particular, the theory suggests that active maintenance in PFC occurs in the service of control, by sustaining representations of context that provide a modulatory influence over processing. Thus, it is precisely the capability of PFC representations to be sustained over time that provides them with the capacity to exert control over subsequent behaviors. Moreover, because the information maintained in PFC is of a specific type (i.e., context), the representations are flexible enough to modulate processing of many different types of inputs, across many different brain regions. The model also suggests an integration of inhibition and active memory, by demonstrating how the suppression of competing responses is dependent upon actively maintaining a representation of the task-relevant context. In the AX-CPT model, the maintained context serves to prepare the task-appropriate response over the delay period, so that it may successfully compete with a stronger, but inappropriate one. This view is similar to the one espoused by Fuster (1989), who has postulated that at least three behavioral functions may be performed within PFC: provisional (short-term) memory, preparatory set, and interference control. However, Fuster has also suggested that these three functions may be both functionally and anatomically dissociable (e.g., inhibitory functions occur within orbitofrontal regions whereas memory functions occur with dorsolateral PFC). The neuroimaging studies conducted here are consistent with the idea that dorsolateral PFC is involved

with active maintenance. However, the cognitive control model also suggests that dorsolateral PFC might be involved in inhibitory functions as well. Specifically, the work here demonstrates that in the AX-CPT task, both inhibitory failures (e.g., increased BX errors) and memory failures (e.g., delay-related decrements in d'-context) can occur as the result of a disturbance to a single underlying mechanism. The work of Roberts et al. (1994) also provides support for this idea. In the Roberts et al. (1994) study, changes in working memory load were found to interact with inhibitory control in the antisaccade task. Nevertheless, it is still possible that there are either functional or anatomical dissociations within PFC between mechanisms specialized for inhibition and those for active maintenance. Additional studies will be needed to help resolve this issue.

Relationship to Other Computational Models

The theory presented in this paper suggests a number of key computational principles that relate to cognitive control, including active maintenance, context representation, and mechanisms of impairment. In this section we discuss how each of this principles relate to other computational models in the literature.

Active Maintenance. A central feature of the model is that context representations are actively maintained. The computational mechanism which enables active maintenance in the model is the presence of recurrent excitatory connections within the context layer, which allows activation states to sustain themselves in the absence of afferent input. This proposal is similar to other ones that have been put forward regarding active memory in PFC. Zipser and colleagues (1993) have shown that the representations that develop in a recurrent neural network model trained to perform short-term memory tasks show a striking representational similarity to PFC neurons, in terms of the dynamics and selectivity of activity patterns. Dehaene and Changeux have also used self-excitatory connectivity in their models of PFC function to simulate performance in both the A-not-B task (Dehaene & Changeux, 1989) and the Wisconsin Card Sort (Dehaene & Changeux, 1992). Guigon et al. (1995) have also used a similar mechanism to examine delayed-response tasks, but argue that sustained activity states in PFC are the product of intrinsic bistability within single neurons rather than a result of recirculating activity among neural assemblies. Unfortunately, both

their model and the cognitive control model abstract away from many neurophysiological variables that would be relevant to this issue. As such, a more detailed understanding of the neural mechanisms of active maintenance will be dependent upon additional data from neuroanatomical and neurophysiological studies.

Context Representations. The model suggests that a central control function played by PFC is the modulation of processing pathways by representations of context information. We have postulated that context representations are similar to behavioral goals. In this manner they behave similarly to the goal states in production system architectures, such as SOAR (Newell, 1991) or ACT-R (Anderson, 1993), which modulate what actions are taken (i.e., which productions are fired) in a particular context. However, we have also argued that more specific, or fine-grained information may be represented as context (such as the status of the cue in the AX-CPT model). Moreover, the means by which context representations modulate processing in the rest of the system is identical to that of other units. In other words, context units are just like all the other units in the model in that they send their activation states along weighted connections to influence the net input of the units to which they connect. This differentiates the role of context representations in our models from production system models, in which there is a categorical distinction between goal representations and other types of representations in terms of information content and influence on the rest of the system. This feature of the cognitive control model also differentiates it from that of Schneider & Detweiler (1987), who have simulated a wide-range of working memory phenomena using a model that is for the most part based on connectionist principles. However, their model, like ACT-R and SOAR, also makes a categorical distinction between the mechanisms of control and those used for other aspects of processing.

Mechanisms of Impairment. We have argued that impairments in cognitive control can arise when the representation and/or maintenance of context information within PFC is weakened. In the AX-CPT model, these impairments were simulated by reducing the gain on units in the context layer. Other computational models have suggested that prefrontal impairments in cognitive control can be simulated through different mechanisms. Kimberg and Farah (1993) have shown

that weakening the associative links between goal states and activated elements of declarative memory results in performance deficits in tasks sensitive to PFC damage. Although the models of Kimberg & Farah were simulated within a production system architecture, an analogous lesion could be achieved in the AX-CPT model by reducing the feedback weights from the context layer to the associative layer. This type of lesion would likely produce deficits in task performance that were similar to weakened representations of context, since activity within the context layer would exert less influence over processing. However, it is not clear if this type of damage would also produce behavioral impairments that interact with delay. Since frontal patients have not yet been tested on the AX-CPT paradigm presented here, it is not clear whether damage to PFC does actually produce delay-related deficits in task performance (see Footnote 14).

A different mechanism of cognitive control impairment has been suggested by both Dehaene and Changeux (1992) and Levine and Pruiett (1989). These investigators have observed that prefrontal-like deficits can also arise from weakening the influence of error-related feedback on performance. Given that explicit feedback is not provided in the AX-CPT following the initial practice period, it is unlikely that this type of impairment is responsible for the task performance deficits observed in patients with schizophrenia, or in healthy individuals under interference. However, further research is needed to determine if weakening context representations would produce behavioral deficits similar to those modeled by Dehaene and Changeux (1992) and Levine and Pruiett (1989) in tasks that do rely upon explicit error-related feedback, such as the WCST.

Unresolved Issues in Cognitive Control

The theory presented in this paper provides an explicit computational theory of cognitive control. Although the theory shows promise, its implementation in terms of specific processing mechanisms is still highly simplified, and it lacks critical components necessary for a more general account of control. In the following section, we discuss important aspects of cognitive control that have been incorporated in the current model, and that must be addressed by any comprehensive theory.

Capacity Limitations. Possibly the most critical aspect of cognitive control that is unaddressed in the model is that of capacity limitations. One of the most robust findings in cognitive psychology is that there appear to be strict limits in the amount of information that can be attended to, actively maintained, or manipulated at any one time (Baddeley, 1994; Miller, 1956). Indeed, capacity limits play a central role in the working memory model of Carpenter and Just (1992). They and other investigators have argued that these limitations are the fundamental variable that determine individual differences in a wide-range of cognitive activities, and even general intelligence (Carpenter, Just, & Shell, 1990; Daneman & Merikle, 1996; Duncan, 1995; Engle, Cantor, & Carullo, 1992; Hasher & Zacks, 1988; Kyllonen & Christal, 1990). However, the nature and source of these limitations still remains largely a mystery. One plausible argument is that capacity limits reflect an innate structural or physiological constraint, such as the total amount of activation that can be achieved (i.e., "the red line"). Another hypothesis is that capacity limits arise from computational considerations, such as the need to reduce interference, or cross-talk, between activated items.

The latter hypothesis is one that could be fruitfully explored through computational analyses. One intriguing possibility is that limitations in capacity are somehow related to the representation of context. This possibility is a plausible one, given the central role we have argued for context representations in attention, working memory, and cognitive control. However, an investigation of this hypothesis requires a further elaboration of the mechanisms that shape the development and organization of these representations. In particular, the definition of context that we have put forward in this paper is largely an operational one, and is relatively unconstrained. For example, it still remains to be specified how context representations can emerge that provide the flexibility characteristic of human behavior, without requiring infinite capacity. In recent work, we have begun to lay out a representational scheme that we believe can address these issues (O'Reilly, Braver, & Cohen, in press). Briefly, we have proposed that the requirement that context be both self-sustaining and capable of biasing processing in task-appropriate ways works synergistically to produce representations that are independent and categorical. Moreover, because context

representations are independent, they can be used in arbitrary combinations, which enables a high degree of flexibility. However, the independence requirement also could lead to an increased tendency for spurious representations to become activated, and incorrectly bias processing in the rest of the system. Consequently, it may be the case that greater levels of inhibition are necessary to avoid this problem. If this hypothesis were correct, it could lead to a limit in the number of context representations that could be activated at any one time, thus providing a source of capacity constraint. Computational studies that examine this hypothesis in greater detail provide an important direction for future research.

Other brain systems. Another important issue that has not been addressed by the current model is the interaction between context representations in PFC and other brain systems specialized for action or memory. With regard to action, cognitive control appears to be necessary for the coordination and organization of temporally extended sequences of well-learned actions. Moreover, it has been observed that this capability seems to breakdown following PFC damage (Duncan, 1986; Schwartz, 1995). With regard to memory, it seems likely that there are neural systems that are particularly specialized for the short-term maintenance of both temporal and spatial relationships. These systems may correspond to the STM buffers frequently described in both the verbal and spatial working memory literatures (Logie, 1995; Vallar & Shallice, 1990). Additionally, it is known that there are at least two neural systems responsible for maintaining information over longer term periods. One system is specialized for the rapid formation of arbitrary memory traces, and appears to be located within the hippocampal formation, while the other is slower, more structured, and involves the neocortex (McClelland, McNaughton, & O'Reilly, 1995).

It would seem that the coordination of these systems in a task-appropriate manner is a control function that is necessary in many behavioral situations, and is one that has been noted by a number of investigators (Baddeley, 1996; Ericsson & Kintsch, 1995; Gupta & MacWhinney, 1997; Shallice, 1988). This type of coordination and organization of other brain systems may occur through the sustained influence of context representations. In terms of action, context representations may act to bias sequence selection, to ensure that the task-appropriate sequences are

produced over competing sequences that might be prepotent in a given situation. In short-term memory systems, it appears that representations decay quickly if not continually reactivated. Maintained representations of context within PFC may serve this function, by providing a sustained source of top-down input based on the current behavioral goal. Moreover, for both short and long-term memories, context representations may provide a mechanism for search and retrieval of particular subsets of information. However, the computational dynamics of these processes have not yet been well-worked out (Burgess & Hitch, 1992; Houghton, 1990; Kawamoto & Anderson, 1985). Nevertheless, relevant data on all of these issues can already be found in the neuroimaging literature. Specifically, both verbal and spatial working memory tasks appear to involve the coactivation of PFC with other brain regions thought to be specialized for verbal or spatial STM (inferior frontal cortex, posterior parietal cortex, e.g., Smith et al., 1996). Additionally, PFC activation is now a robust finding in tasks involving search and retrieval from episodic memory (Buckner, 1996; Tulving, Kapur, Craik, Moscovitch, & Houle, 1994). PFC activation has also been observed in studies that require participants to imagine a particular sequence of actions or make judgments regarding sequence order (Partiot, Grafman, Sadato, Flitman, & Wild, 1996). More detailed information on the relationship between PFC and these other brain systems could be obtained from future studies which make use of time-course or correlational information about activation patterns between brain regions.

The control of control. It is necessary that the mechanisms involved in control have some way of being controlled themselves. For example, context representations should be more actively engaged to bias processing when task demands increase. Task difficulty information could be conveyed by explicit feedback about error rates, but this sort of feedback is often not available in many task situations. Thus, the cognitive system may require some sort of "on-line monitor" that provides an index of the demand for control. A specific brain region that may play this particular role in processing is the anterior cingulate cortex. This hypothesis is consistent with the growing literature documenting the reliable presence of anterior cingulate activity in various cognitive control tasks (Buckner et al., 1995; Carter et al., 1995; Casey et al., 1997; Paus, Petrides, Evans, & Meyer,

1993; Petersen, Fox, Posner, Mintun, & Raichle, 1989; Raichle et al., 1994). Our own previous work has contributed to this literature by demonstrating anterior cingulate activity in a functional neuroimaging study comparing the baseline AX-CPT with the more difficult degraded condition (Barch et al., 1997). More recently, we have found that activity in this region is also sensitive to errors in the AX-CPT, and to the increased representational conflict found in AY and BX trials, relative to AX and BY (Carter et al., 1998). Based on these findings we have begun to hypothesize about the particular computational mechanisms underlying anterior cingulate activity dynamics, and their relationship to cognitive control. Specifically, we have hypothesized that the anterior cingulate may be monitoring representational "energy", which is an index that should be correlated with increased errors, task difficulty, and more specifically, conflict between representations. Monitoring changes in representational energy might be particularly useful for controlled processing, because they provide a way for the system to gauge the difficulty of a task, without requiring explicit error-related feedback. More importantly, this hypothesis enables the cognitive system to have access to global knowledge regarding its own internal state without requiring a homunculus, which has been a fundamental problem for theories of cognitive control. Through a series of computational simulations, we have recently demonstrated the plausibility of our hypothesis, by showing how it can account for the pattern and dynamics of anterior cingulate activity across a wide range of task domains (Botvinick, Braver, Carter, Barch, & Cohen, submitted).

It is likely that changes in representational energy in the anterior cingulate could play a modulatory function during learning. In computational terms, knowledge of energy could be highly useful for changing learning rate, which governs how quickly connections are strengthened. In a stochastic system, energy level could also be used to increase variability (temperature), which helps to ensure that the system reaches the optimal energy state. Interestingly, these two parameters (learning rate and temperature) have been shown to be formally equivalent (Thimm, Moerland, & Fusler, 1996). By using both explicit feedback regarding errors as well as more continuous information regarding the current energy state, the cognitive system would have tremendous flexibility in controlling itself. However, a learning system of this sort would also have to act

relatively quickly, in order to respond to a dynamic and rapidly changing environment. For example, in the AX-CPT, each new cue can be thought of as changing the environment, which, in turn, should alter the behavioral goal. In effect, an A cue should set up a behavioral goal to make a target response if the next stimulus is an X. The presentation of a non-A cue should change this goal, so that a non-target response is always made to the subsequent stimulus. The ability to rapidly switch behavioral goals has been previously noted to be an important aspect of cognitive control. Many investigators have discussed this function in relation to attention switching, memory updating, or dual-task situations (Baddeley, 1996; Engle, Kane, & Tuholski, in press; Morris & Jones, 1990; Rogers & Monsell, 1995). Moreover, impairments in these types of tasks have been observed following PFC damage (Baddeley, Della Sala, Papagno, & Spinnler, 1997; Duncan, Emslie, Williams, Johnson, & Freer, 1996; Milner, 1963; Owen, Roberts, Polkey, Sahakian, & Robbins, 1991; Rubinstein, Evans, & Meyer, 1994). Thus, it may be the case that both error-related learning and goal switching involve the modulation of context representations by some other system.

A hypothesis that we have recently been exploring is that these two modulatory functions may be mediated by the dopamine system. For many years, dopamine has been known to be involved in the processing of reinforcement (Willner & Scheel-Kruger, 1991; Wise & Rompre, 1989). More recently, it has been suggested that phasic activity in dopamine neurons conveys an error signal related to the temporal prediction of reward, which can be used to drive learning in the rest of the brain (Montague, Dayan, & Sejnowski, 1996). In a different line of research, dopamine activity has been interpreted as producing a switching signal that alters the current motor pattern or behavioral set (Oades, 1985). Moreover, there is mounting evidence that dopamine influences both maintenance-related activity in PFC and behavioral performance in cognitive control tasks (Kimberg, D'Esposito, & Farah, 1997; Luciana, Depue, Arbisi, & Leon, 1992; Sawaguchi & Goldman-Rakic, 1994; Servan-Schreiber, Carter, Bruno, & Cohen, in press). However, the specific mechanisms by which dopamine may interact with PFC to modulate both learning and goal switching have not previously been elucidated. In other recent work, we have conducted simulation

studies which specifically examine the interactions between dopamine and PFC and their implications for learning, goal switching, and cognitive control (Braver & Cohen, in press). In particular, we have suggested that the role of dopamine in reward-prediction learning provide it with a pattern of activity dynamics that can simultaneously be exploited to "gate" information into PFC. We have argued that this gating function of dopamine serves to both update context representations at task appropriate junctures while simultaneously protecting these representations from being disrupted by task-irrelevant inputs (i.e., interference). Our future research will aim to incorporate these recent accounts of dopamine and anterior cingulate function into the cognitive control model in order to address a much wider range of issues relevant to cognitive control. In particular, our goal will be to account for behavioral performance phenomena in more complex and naturalistic cognitive control tasks in terms of a unified and neurobiologically-plausible set of control mechanisms.

CONCLUSIONS

The theory presented in this paper provides a detailed account of a wide-range of empirical phenomena associated with performance of the AX-CPT. More importantly, it provides a useful framework in which to understand the central mechanisms and key information processing principles underlying a fundamental cognitive control function. In particular, the cognitive control model suggests that flexible adaptation of behavior to particular task demands is achieved through the active representation and maintenance of context information that occurs within PFC. We have discussed how this model can provide new insights that help to resolve debates within the literatures on controlled processing, working memory, attention and PFC function. However, there are a number of critical issues in these literatures that remain unaddressed by the model. Future research on these issues will hopefully get us closer to a more comprehensive understanding of how we, as biological organisms, exert control over own thoughts and behavior.

FOOTNOTES

1 Context representations are likely to influence response selection processes, consistent with "selection for action" theories of attentional control (Allport, 1989; Neumann, 1987). However, in Figure 1, we have schematized feedback from the context module also occurring at earlier points in the processing stream. This illustrates our assumption that context may exert an effect on interpretative or perceptual processes as well. This assumption is consistent with evidence that PFC activity can modulate perceptual as well as response-related processes (e.g., Knight, Hillyard, Woods, & Neville, 1980).

2 The correction factor is based on both the number of target and non-target trials. As a result, in cases of perfect hit or false-alarm rates, d' is reduced when less trials are used to generate this measure. Although the correction procedure normalizes results across conditions within an analysis, it may artifactually change d' values across analyses and studies. This fact should be noted when examining results.

3 This prediction could also be examining by conditionalizing RT based on accuracy. Specifically, the model would predict that AY errors should be faster than correct responses, while BX errors should be slower. Indeed, the predictions were confirmed for both AY (error responses: M=582 msec, SE=6; correct responses: M=377 msec, SE=10) and BX RTs (error responses: M=618 msec, SE=15; correct responses: M=505, SE=8). However, this result was not analyzed statistically, given that over half of the subjects made no errors on either condition.

4 One subject was thrown out from this analyses as a result of having too few responses to categorize in one condition (due to many trials in which no response was made).

5 However, an analysis which included experiment as an additional factor did reveal a number of experiment-related effects. In particular, for the accuracy data there was a significant experiment x trial type interaction ($F(12,406)=3.13, p<.001$), and for the RT data there was a main effect of experiment ($F(6,202)=21.31$, $p<.001$), plus experiment x trial type ($F(12,404)=2.01$, $p<.05$), and experiment x delay ($F(6,202)=6.05, p<.001$) interactions. Most of these effects appeared to be related to differences found in the data sets obtained from Studies 3 and 5. This

finding was not completely unsurprising since these two studies were the most different from the others in terms of the demographic profile of participants. In most of the studies the participants were Carnegie Mellon Undergraduates. In contrast, Studies 3 and 5 included older participants, who often did not have a college-level education. These factors may have contributed to differences in performance, and thus should be examined more carefully in future studies.

6 This "hardwiring" of self-excitatory connections reflects our assumption that the active maintenance properties of PFC arise by either maturational or learning mechanisms that are beyond the scope of current consideration.

7 The following equation was used:

$$I_j(t+1) = \left(\gamma \sum w_{ij} y_i + B + I_j(t) \right) dt + \sigma Z_i(t) \, \mathrm{sqrt}(dt)$$

where

$$y_j = 1 / (1 + e^{-I_j})$$

is the activation of unit j, I_j is the total input to j, dt is the time-step of integration, γ is the gain, B is the bias, $Z_i(t)$ is a standard independent Gaussian random variable, and σ is the variance of the distribution. See Braver et al.(1995a) for further details.

8 This approach to learning has been termed "neural system identification" by Zipser (1992), and can be distinguished from other approaches which try to more closely simulate the developmental or neurobiological mechanisms of learning (McClelland, 1994; Miller, Keller, & Stryker, 1989; Munakata, McClelland, Johnson, & Siegler, in press; O'Reilly, 1996).

9 Interestingly, the use of noise during learning appeared to affect the final weights which developed. Specifically, a comparison of training runs with and without noise tended to show that noise during training led to larger lateral inhibitory weights, and smaller input weights to the context layer. Informal analyses suggested that these changes arose in order to compensate for overly high activity levels in the competing inputs that occurred as a result of noise-related variability on a small fraction of trials. Moreover, the reduced input weight to the context layer may have contributed to the delay effects on context representational strength that are described below.

[10] These included the need to account for: 1) components of processing which were not simulated (e.g., early visual processing, response execution); 2) the temporal scaling used in the model; 3) other RT effects such as motor priming or stimulus onset uncertainty (discussed in Study 1). Accounting for these effects would necessitate including additional parameters to the model, which would greatly increase its degrees of freedom.

[11] Four patients originally tested were later diagnosed with another psychiatric illness, and their results were excluded from further study.

[12] Analyses examining the effect of blocking in the two sub-groups of participants did not reveal any significant interactions of this variable with the primary task factors (all p's > .1). There was a main effect of blocking on RT, such that participants in the mixed design were significantly faster ($F(1,28) = 7.22$, $p<.05$) than those in the blocked design. There were no main effects of blocking on accuracy ($F(1,28) = 0.08$, $p>.1$)

[13] One patient did not make any correct responses for BX trials in the long delay condition. To permit completion of the ANOVA table for correct response RTs, this cell was replaced by taking the group mean RT for that condition.

[14] It is important to note that the model does not predict that patients who suffer from PFC damage will necessarily show decrements in AX-CPT performance that interact with delay. In particular, if the lesion is such that it disturbs the representation of context, there may be as much of a control impairment observed at short delays as there is at long delays. Indeed, the model predicts that delay-related decrements in performance will only occur when the disturbance is one that causes context representations to weaken over time.

[15] Analyses examining the effects of blocking in the two sub-groups of participants revealed no main effects of this variable or any interactions with the primary task factors for either accuracy or RT (all p's > .1).

[16] One participant did not make any correct responses for BX delay trials in the degraded condition (both long and short delays). To permit completion of the ANOVA table for correct

response RTs, these cells were replaced by taking the mean RT for the degraded condition in long and short delay BX trials.

17 In contrast to the findings with PFC, an additional delay effect that was previously observed in parietal cortex failed to replicate. However, in the original observation, the delay effect was found to interact with scan-within-trial. Thus, it may be the case that the parietal region was either less reliably affected by the delay manipulation or functionally dissociable from the PFC regions.

18 The coding of scans corresponding to delay and ITI periods was shifted forward by 1 scan, in order to take into account the approximately 3 second hemodynamic lag present in the fMRI response (Kwong et al., 1992; Savoy et al., 1995).

REFERENCES

Allport, A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 631-682). Cambridge, MA: The MIT Press.

Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Baddeley, A. (1994). The magical number seven: Still magic after all these years. *Psychological Review, 101*(2), 353-356.

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49(A)*, 5-28.

Baddeley, A., & Della Sala, S. (1996). Working memory and executive control. *Phil. Trans. R. Soc. Lond. B, 351*, 1397-1404.

Baddeley, A., Della Sala, S., Papagno, C., & Spinnler, H. (1997). Dual-task performance in dysexecutive and nondysexecutive patients with frontal lesion. *Neuropsychology, 11(2)*, 187-194.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556-559.

Baddeley, A. D. (1993). Working memory or working attention? In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control: A tribute to Donald Broadbent* (pp. 152-170). Oxford: Clarendon Press.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology, 8*(4), 485-493.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575-589.

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*, 515-526.

Barch, D. M., & Berenbaum, H. (1994). The relationship between information processing and language production. *Journal of Abnormal Psychology, 103*, 241-250.

Barch, D. M., Braver, T. S., Nystrom, L., Forman, S. D., Noll, D. C., & Cohen, J. D. (1997). Dissociating working memory from task difficulty in human prefrontal cortex. *Neuropsychologia, 35*, 1373-1380.

Barch, D. M., Carter, C. S., Braver, T. S., Sabb, F. W., McDonald, A., Noll, D. C., & Cohen, J. D. (in preparation). Selective deficits in prefrontal cortex regions in medication naive schizophrenia patients. .

Barone, P., & Joseph, J. P. (1989). Prefrontal cortex and spatial sequencing in macaque monkey. *Experimental Brain Research, 78*, 447-464.

Bauer, R. H., & Fuster, J. M. (1976). Delayed-matching and delayed-response deficit from cooling dorsolateral prefrontal cortex in monkeys. *Journal of Comparative and Physiological Psychology, 90*(3), 293-302.

Bertelson, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology, 13*, 90-102.

Bianchi, L. (1922). *The mechanism of the brain and the function of the frontal lobes*. Edinburgh: Livingstone.

Blanchard, J. J., & Neale, J. M. (1992). Medication effects: Conceptual and methodological issues in schizophrenia research. *Clinical Psychology Review, 12*, 345-361.

Botvinick, M. M., Braver, T. S., Carter, C. S., Barch, D. M., & Cohen, J. C. (submitted). Evaluating the demand for control: Anterior cingulate cortex and crosstalk monitoring. .

Braver, T. S., & Cohen, J. D. (in press). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In S. Monsell & J. Driver (Eds.), *Attention and Performance XVIII* . Cambridge, MA: MIT Press.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage, 5*(1), 49-62.

Braver, T. S., Cohen, J. D., & Servan-Schreiber, D. (1995a). A computational model of prefrontal cortex function. In D. S. Touretzky, G. Tesauro, & T. K. Leen (Eds.), *Advances in Neural Information Processing Systems* (Vol. 7, pp. 141-148). Cambridge, MA: MIT Press.

Braver, T. S., Cohen, J. D., & Servan-Schreiber, D. (1995b). Neural network simulations of schizophrenic performance in a variant of the CPT-AX: A predicted double dissociation. *Schizophrenia Research, 15*(1-2), 110.

Buckner, R. L. (1996). Beyond HERA: Contributions of specific prefrontal brain areas to long-term memory retrieval. *Psychonomic Bulletin & Review, 3*(2), 149-158.

Buckner, R. L., Petersen, S. E., Ojemann, J. G., Miezin, F. M., Squire, L. R., & Raichle, M. E. (1995). Functional anatomical studies of explicit and implicit memory retrieval tasks. *Journal of Neuroscience, 15*(1), 12-29.

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31*, 429-460.

Butters, N., Pandya, D., Sanders, K., & Dye, P. (1971). Behavioral deficits in monkeys after selective lesions to the middle third of the sulcus principalis. *Journal of Comparative Physiological Psychology, 76*, 8-14.

Callaway, E., & Naghdi, S. (1982). An information processing model for schizophrenia. *Archives of General Psychiatry, 39*(March), 339-347.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the raven progressive matrices test. *Psychological Review, 97*(3), 404-431.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D. C., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science, 280*, 747-749.

Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: An [^{15}0]-$H_2$0 PET study of Stroop task performance. *Neuroimage, 2*, 264-272.

Casey, B. J., Trainor, R. J., Orendi, J. L., Schubert, A. B., Nystrom, L. E., Cohen, J. D., Noll, D. C., Giedd, J., Castellanos, X., Haxby, J., Dahl, R. E., & Rapoport, J. L. (1997). A pediatric functional MRI study of prefrontal activation during performance of a Go-No-Go task. *Journal of Cognitive Neuroscience, 9*, 835-847.

Chapman, L. J., & Chapman, J. P. (1978). The measurement of differential deficit. *Journal of Psychiatric Research, 14*, 303-311.

Chapman, L. J., & Chapman, J. P. (1989). Strategies for resolving the heterogeneity of schizophrenics and their relatives using cognitive measures. *Journal of Abnormal Psychology, 98*, 357-366.

Chiodo, L., & Berger, T. (1986). Interactions between dopamine and amino-acid induced excitation and inhibition in the striatum. *Brain Research, 375*, 198-203.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120(3)*, 235-253.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: L. Erlbaum Associates.

Cohen, J. D., Barch, D. M., Carter, C. S., & Servan-Schreiber, D. (1999). Schizophrenic deficits in the processing of context: Converging evidence from three theoretically motivated cognitive tasks. *Journal of Abnormal Psychology, 108*, 120-133.

Cohen, J. D., Braver, T. S., & O'Reilly, R. (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, 351*(1346), 1515-1527.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account 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 nonspatial working memory task with functional MRI. *Human Brain Mapping, 1*, 293-304.

Cohen, J. D., MacWhinney, B., Flatt, M. R., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers, 25*(2), 257-271.

Cohen, J. D., Nystrom, L. E., Braver, T. S., Sabb, F. W., Delgado, M. R., & Noll, D. C. (1998). *fMRI studies of the topographic organization of working memory representations in prefrontal cortex.* Paper presented at the Cognitive Neuroscience Society, Fifth Annual Meeting, San Francisco, CA.

Cohen, J. D., Perstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature, 386*, 604-608.

Cohen, J. D., Romero, R. D., Farah, M. J., & Servan-Schreiber, D. (1994). Mechanisms of spatial attention: The relation of macrostructure to microstructure in parietal neglect. *Journal of Cognitive Neuroscience, 6*(4), 377-387.

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., & Servan-Schreiber, D. (1993). A theory of dopamine function and cognitive deficits in schizophrenia. *Schizophrenia Bulletin, 19*(1), 85-104.

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, R. M., Semple, W. E., Gross, M., Nordahl, T. E., Delisi, L. E., Holcomb, H. H., King, A. C., Morihisa, J. M., & Pickar, D. (1987). Dysfunction in a prefrontal substrate of sustained attention in schizophrenia. *Life Sciences, 40,* 2031-2039.

Constantinidis, C., & Steinmetz, M. A. (1996). Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *Journal of Neurophysiology, 76,* 1352-1355.

Cornblatt, B. A., & Keilp, J. G. (1994). Impaired attention, genetics, and the pathophysiology of schizophrenia. *Schizophrenia Bulletin, 20*(1), 31-62.

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cerebral Cortex, 6,* 39-49.

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature, 386,* 608-612.

D'Esposito, M., Aguirre, G. K., Zarahn, E., Ballard, D., Shin, R. K., & Lease, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Cognitive Brain Research, 7,* 1-13.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature, 378*(November 16), 279-281.

Dagenbach, D., & Carr, T. H. (Eds.). (1994). *Inhibitory processes in attention, memory, and language.* San Diego: Academic Press.

Damasio, A. R. (1985). The frontal lobes. In K. M. Heilman & E. Valenstein (Eds.), *Clinical Neuropsychology* (pp. 339-375). New York: Oxford University Press.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review, 3*, 422-433.

Davis, K., Kahn, R., Ko, G., & Davidson, M. (1991). Dopamine in schizophrenia: A review and reconceptualization. *American Journal of Psychiatry, 148*(11), 1474-1486.

Dehaene, S., & Changeux, J. P. (1989). A simple model of prefrontal cortex function in delayed-response tasks. *Journal of Cognitive Neuroscience, 1*, 244-261.

Dehaene, S., & Changeux, J. P. (1992). The Wisconsin card sorting test: Theoretical analysis and modeling in a neuronal network. *Cerebral Cortex, 1*, 62-79.

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Towards a unified theory of cognitive development and aging. *Developmental Review, 12*, 45-75.

Dempster, F. N., & Brainerd, C. J. (Eds.). (1995). *Interference and inhibition in cognition*. San Diego: Academic Press.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193.

Diamond, A. (1990). The development and neural bases of memory functions as indexed by the A-not-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.

Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's A-not-B task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research, 74*, 24-40.

Duncan, J. (1986). Disorganization of behaviour after frontal lobe damage. *Cognitive Neuropsychology, 3*(3), 271-290.

Duncan, J. (1995). Attention, intelligence, and the frontal lobes. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 721-733). Cambridge, MA: MIT Press.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30,* 257-303.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433-458.

Duncan, K., & Sussman, D. (1995). Towards a cognitive account of frontal lobe function: Simulating frontal lobe deficits in normal subjects. *Annals of the New York Academy of Sciences, 769,* 289-304.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 972-992.

Engle, R. W., Kane, M., & Tuholski, S. (in press). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* . New York, NY: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211-245.

Fiez, J. A., Raife, E. A., Balota, D. A., Schwarz, J. P., Raichle, M. E., & Peterson, S. E. (1996). A positron emission tomography study of the short-term maintenance of verbal information. *The Journal of Neuroscience, 16(2)*(January 15), 808-822.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine, 33,* 636-647.

Frackowiak, R. S. J. (1994). Functional mapping of verbal memory and language. *Trends in Neuroscience, 17*(3), 109-115.

Freedman, M., & Oscar-Berman, M. (1986). Bilateral frontal lobe disease and selective delayed response deficits in humans. *Behavioral Neuroscience, 100,* 337-342.

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. (1973). Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory. *Journal of Neurophysiology, 36,* 61-78.

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.

Fuster, J. M., & Alexander, G. E. (1973). Firing changes in cells of the nucleus medialis dorsaalis associated with delayed response behavior. *Brain Research, 61,* 79-91.

Fuster, J. M., & Jervey, J. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science, 212,* 952-955.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4,* 385-390.

Glosser, G., & Goodglass, H. (1990). Disorders in executive control functions among aphasic and other brain-damages patients. *Journal of Clinical and Experimental Neuropsychology, 12(4),* 485-501.

Goldman-Rakic. (1995). Cellular basis of working memory. *Neuron, 14,* 477-485.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In F. Plum & V. Mountcastle (Eds.), *Handbook of Physiology - The Nervous System V* (Vol. 5, pp. 373-417). Bethesda, MD: American Physiological Society.

Goldman-Rakic, P. S. (1991). Prefrontal Cortical Dysfunction in Schizophrenia: The Relevance of Working Memory. *Psychopathology and the Brain*, 1-23.

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.

Grasby, P. M., Frith, C. D., Friston, K. J., Simpson, J., Fletcher, P. C., Frackowiak, R. S. J., & Dolan, R. J. (1994). A graded task approach to the functional mapping of brain areas implicated in auditory-verbal memory. *Brain, 117*, 1271-1282.

Gratton, G., Coles, M. H., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 331-344.

Guigon, E., Dorizzi, B., Burnod, Y., & Schultz, w. (1995). Neural correlates of learning in the prefrontal cortex of the monkey: A predictive model. *Cerebral Cortex, 5*, 135-147.

Gupta, P. (1995). *Word learning and immediate serial recall: Toward an integrated account.* Unpublished Ph.D., Carnegie Mellon University.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59*, 267-333.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension and aging: A review and a new view. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 22, pp. 193-225). New York: Academic Press.

Hecaen, H., & Albert, M. L. (1978). *Human neuropsychology*. New York: Wiley.

Hitch, G. J., & Baddeley, A. D. (1976). Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology, 28*, 603-621.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, M. C., & M. Zock (Eds.), *Current Research in Natural Language Generation* .

Houghton, G., & Tipper, S. P. (1996). Inhibitory mechanisms of neural and cognitive control: Applications to selective attention and sequential action. *Brain and Cognition, 30*, 20-43.

Jones, D., & Morris, N. (1992). Irrelevant speech and serial recall: Implications for theories of attention and working memory. *Scandinavian Journal of Psychology, 33*, 212-229.

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. A. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience, 9*, 462-475.

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.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity, *Varieties of Attention* (pp. 29--59): Academic Press, Inc.

Kawamoto, A., & Anderson, J. A. (1985). A neural network model of multistable perception. *Acta Psychologica, 59*, 35-65.

Kimberg, D. Y., D'Esposito, M., & Farah, M. J. (1997). Effects of bromocriptine on human subjects depend on working memory capacity. *Neuroreport, 8*, 381-385.

Kimberg, D. Y., & Farah, M. J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior. *Journal of Experimental Pscyhology: General, 122*(4), 411-428.

Kirillov, A. B., Myre, C. D., & Woodward, D. J. (in press). Bistability, switches and working memory in a two-neuron inhibitory-feedback model. *Biological Cybernetics*.

Knight, R. T., Grabowecky, M. F., & Scabini, D. (1995). Role of human prefrontal cortex in attentional control. *Advances in Neurology, 66*, 21-34.

Knight, R. T., Hillyard, S. A., Woods, D. L., & Neville, H. J. (1980). The effects of frontal and temporal-parietal lesions on the auditory evoked potential in man. *Electroencephalography and Clinical Neurophysiology, 50*(1-2), 112-124.

Kubota, K., & Niki, H. (1971). Prefrontal cortical unit activity and delayed alternation performance in monkeys. *Journal of Neurophysiology, 34*, 337-347.

Kutas, M., & Donchin, E. (1980). Preparation to respond as manifested by movement-related brain potentials. *Brain Research, 202(1),* 95-115.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, P., Kennedy, D. N., Hoppel, B. E., S., C. M., Turner, R., Cheng, H. M., Brady, T. J., & Rosen, B. R. (1992). Dynamic magnetic resonance of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Science, 89*(June), 5675-5679.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity? *Intelligence, 14*, 389-433.

Levine, D. S., & Prueitt, P. S. (1989). Modeling some effects of frontal lobe damage-novelty and perseveration. *Neural Networks, 2*, 103-116.

Levitt, J. B., Lewis, D. A., Yoshioka, T., & Lund, J. S. (1993). Topography of pyramidal neuron intrinsic connections in Macaque monkey prefrontal cortex (areas 9 and 46). *The Journal of Comparative Neurology, 338*, 360-376.

Logie, R. H. (1995). *Visuo-spatial working memory.* Hove, UK: Lawrence Erlbaum Associates.

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica, 75*(55-74).

Losier, B. J., McGrath, P. J., & Klein, R. M. (1996). Error patterns of the continuous performance test in non-medicated and medicated samples of children with and without ADHD: A meta-analytic review. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 37,* 971-987.

Luciana, M., Depue, R. A., Arbisi, P., & Leon, A. (1992). Facilitation of working memory in humans by a D_2 dopamine receptor agonist. *Journal of Cognitive Neuroscience, 4*(1), 58-68.

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.

McCarthy, G. (1995). Functional neuroimaging of memory. *The Neuroscientist, 1*(3), 155-163.

McCarthy, G., Blamire, A. M., Puce, A., Nobre, A. C., Bloch, G., Hyder, F., Goldman-Rakic, P., & Shulman, R. G. (1994). Functional magnetic resonance imaging of human prefrontal cortex during a spatial working memory task. *Proceedings of the National Academy of Sciences, 91*, 8690-8694.

McClelland, J. L. (1993). Toward a theory of information processing in graded, random, and interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV: Synergy's in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 655-688). Cambridge, MA: MIT Press.

McClelland, J. L. (1994). The interaction of nature and nurture in development: A parallel distributed processing perspective. In P. Bertelson, P. Eelen, & G. d'Ydewalle (Eds.), *International Perspectives on Psychological Science, Volume I: Leading themes* (pp. 57-88). Hillsdale, NJ: Erlbaum.

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.

Melchitzky, D. S., Sesack, F. R., Pucak, M. L., & Lewis, D. A. (in press). Synaptic targets of pyramidal neurons providing intrinsic horizontal connections in monkey prefrontal cortex. *Journal of Comparative Neurology*.

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*(0), 5154-5167.

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *Journal of Neuroscience, 13*, 1460-1478.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Miller, K. D., Keller, J. B., & Stryker, M. P. (1989). Ocular dominance column development: Analysis and simulation. *Science, 245*(August), 605-615.

Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology, 9*, 90-100.

Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology, 11*(6), 671-717.

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

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.

Moody, S. L., Wise, S. P., di Pellegrino, G., & Zipser, D. (1998). A model that accounts for activity in primate frontal cortex during a delayed-match-to-sample task. *Journal of Neuroscience, 18*, 399-410.

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology, 81*, 111-121.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (in press). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in objects permanence tasks. *Psychological Review*.

Neter, J., Wasserman, W., & Kutner, M. H. (1990). *Applied linear statistical models*. Boston: Irwin Press.

Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on Perception and Action* . Hillsdale, NJ: Erlbaum.

Newell, A. (1991). *Unified Theories of cogntion*. Cambridge, MA: Cambridge University Press.

Noll, D. C., Cohen, J. D., Meyer, C. H., & Schneider, W. (1995). Spiral K-space MR Imaging of cortical activation. *Journal of Magnetic Resonance Imaging, 5*(1), 49-56.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In T. Shallice & D. A. Norman (Eds.), *Consciousness and Self-regulation* : Plenum Press.

Nuechterlein, K. H. (1977). Reaction time and attention in schizophrenia: A critical evaluation of the data and theory. *Schizophrenia Bulletin, 3*, 45-57.

Nuechterlein, K. H. (1983). Signal detection in vigilance tasks and behavioral attributes among offspring of schizophrenic mothers and among hyperactive children. *Journal of Abnormal Psychology, 92*, 4-8.

Nuechterlein, K. H. (1991). Vigilance in schizophrenia and related disorders. In S. R. Steinhauer, J. H. Gruzelier, & J. Zubin (Eds.), *Handbook of Schizophrenia Vol. 5: Neuropsychology, psychophysiology, and information processing* (pp. 397-433). Amsterdam: Elsevier.

Nuechterlein, K. H., & Dawson, M. E. (1984). Information processing and attentional functioning in the developmental course of schizophrenia disorders. *Schizophrenia Bulletin, 10*(2), 160-203.

O'Reilly, R. C. (1996). *The LEABRA model of neural interactions and learning in the neocortex.* Unpublished Ph.D. thesis, Carnegie-Mellon University, Pittsburgh.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (in press). A biologically-based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* . New York: Cambridge University Press.

Oades, R. D. (1985). The role of noradrenaline in tuning and dopamine in switching between signals in the central nervous system. *Neuroscience and biobehavioral reviews., 9,* 261-282.

Owen, A. A., Evans, A. C., & Petrides, M. (1996). Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: A positron emission tomography study. *Cerebral Cortex, 6,* 31-38.

Owen, A. M., Roberts, A. C., Polkey, C. E., Sahakian, B. J., & Robbins, T. W. (1991). Extra-dimensional versus intra-dimensional set shifting performance following frontal lobe excisions, temporal lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia, 29,* 993-1006.

Partiot, A., Grafman, J., Sadato, N., Flitman, S., & Wild, K. (1996). Brain activation during script event processing. *Neuroreport, 7,* 761-766.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116(2),* 220-244.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature, 362,* 342-345.

Paus, t., Petrides, M., Evans, A. C., & Meyer, E. (1993). Role of the human anterior cingulate cortex in the control of oculomotor, manual, and speech responses: A positron emission tomography study. *Journal of Neurophysiology, 70*(2), 453-469.

Pearlmutter, B. (1989). Learning state space trajectories in recurrent neural networks. *Neural Computation, 1,* 263-269.

Penit-Soria, J., Audinat, E., & Crepel, F. (1987). Excitation of rat prefrontal cortical neurons by dopamine: An in vitro electrophysiological study. *Brain Research, 425,* 263-274.

Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behaviour. *Neuropsychologia, 12,* 323-330.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience, 2*(1), 153-170.

Petrides, M. (1996). Lateral frontal cortical contribution to memory. *Seminars in the Neurosciences, 8,* 57-63.

Petrides, M. E., Alivisatos, B., Evans, A. C., & Meyer, E. (1993). Dissociation of human mid-dorsolateral from posterior dorsolateral frontal cortex in memory processing. *Proceedings of the National Academy of Science, U.S.A., 90,* 873-877.

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., & Cohen, Y. A. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance X: Control of language processes* . Hillsdale, NJ: Lawrence Erlbaum.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annu. Rev. Neurosci., 13,* 25-42.

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: Lawrence Erlbaum Associates.

Prabhakaran, J., Smith, J. A. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (in press). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance and Raven's progressive matrices test. *Cognitive Psychology*.

Pucak, M. L., Levitt, J. B., Lund, J. S., & Lewis, D. A. (1996). Patterns of intrinsic and associational circuitry in monkey prefrontal cortex. *Journal of Comparative Neurology, 376*, 614-630.

Rafal, R., & Henik, A. (1994). The neurology of inhibition: Integrating controlled and automatic processes. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 1-51). San Diego: Academic Press.

Raichle, M. E., Fiez, J. A., Videen, T. O., MacCleod, A. K., Pardo, J. V., Fox, P. T., & Petersen, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex, 4*, 8-26.

Rajkowska, G., & Goldman-Rakic, P. S. (1995). Cytoarchitectonic definition of prefrontal areas in the normal human cortex: II. Variability in locations of areas 9 and 46 and relationship to the talairach coordinate system. *Cerebral Cortex, 5(4)*, 323-337.

Rao, S. C., Rainer, G., & Miller, E. K. (1997a). Integration of what and where in the primate prefrontal cortex. *Science, 276*(May), 821-824.

Rao, S. C., Rainer, G., & Miller, E. K. (1997b). Selective attention gates working memory for objects in monkey prefrontal (PF) cortex. *Society for Neuroscience Abstracts, 23*, 1615.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*(3), 510-532.

Remington, R. J. (1969). Analysis of sequential effects in choice reaction times. *Journal of Experimental Psychology, 82*, 250-257.

Rezai, K., Andreasen, N. C., Alliger, R., Cohen, G., Swayze, V., & O'Leary, D. S. (1993). The neuropsychology of the prefrontal cortex. *Archives of Neurology, 50*(6), 636-642.

Roberts, J., R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General, 123*(4), 374-393.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*, 207-231.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology, 20*(5), 343-350.

Rubinstein, J., Evans, J. E., & Meyer, D. E. (1994). Task-switching in patients with prefrontal cortex damage. *Society for Cognitive Neuroscience Abstracts, 1*, 37.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. (Vol. 1 and 2). Cambridge, MA: MIT Press.

Salame, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 150-164.

Salame, P., & Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology, 41A*, 107-122.

Savoy, R. L., Bandettini, P. A., O'Craven, K. M., Kwong, K. K., Davis, T. L., Baker, J. R., Weisskoff, R. M., & Rosen, B. R. (1995). *Pushing the temporal resolution of fMRI: Studies of very brief visual stimuli, onset variability and asynchrony, and stimulus-correlated changes in noise*. Paper presented at the Society of Magnetic Resonance, 3rd Meeting.

Sawaguchi, T., & Goldman-Rakic, P. (1994). The role of d1-dopamine receptors in working memory: Local injections of dopamine antagonists into the prefrontal cortex of rhesus monkeys performing an oculomotor delayed-response task. *Journal of Neurophysiology, 63(6),* 1401-1412.

Schneider, W., & Detweiler, M. (1987). A connectionist/control architecture for working memory. In Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 21, pp. 54-114).

Schultz, W., & Romo, R. (1988). Neuronal activity in the monkey striatum during the initiation of movements. *Experimental Brain Research, 71,* 431-436.

Schwartz, M. F. (1995). Re-examining the role of executive functions in routine action production. *Annals of the new York Academy of Sciences, 15,* 321-335.

Seidman, L. J., Breiter, H. J., Goodman, J. M., Goldstein, J. M., Woodruff, P. W., O'Craven, K., Savoy, R., Tsuang, M. T., & Rosen, B. R. (1998). A functional magnetic resonance imaging study of auditory vigilance with low and high information processing demands. *Neuropsychology, 12,* 505-518.

Servan-Schreiber, D. (1990). *From physiology to behavior: Computational models of catecholamine modulation of information processing.* Unpublished Ph.D., Carnegie Mellon University.

Servan-Schreiber, D., Bruno, R., Carter, C. S., & Cohen, J. D. (in press). Dopamine and the mechanisms of cognition. Part I: A neural network model predicting dopamine effects on selective attention. *Biological Psychiatry*.

Servan-Schreiber, D., Carter, C. S., Bruno, R., & Cohen, J. D. (in press). Dopamine and the mechanisms of cognition. Part II: D-amphetamine effects in human subjects performing a selective attention task. *Biological Psychiatry*.

Servan-Schreiber, D., Cohen, J. D., & Steingard, S. (1996). Schizophrenic deficits in the processing of context: A test of a theoretical model. *Archives of General Psychiatry, 53,* 1105-1113.

Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A network model of catecholamine effects: Gain, signal-to-noise ratio, and behavior. *Science, 249*, 892-895.

Shakow, D. (1962). Segmental set: A theory of the formal psychological deficit in schizophrenia. *Archives of General Psychiatry, 6*, 1-17.

Shallice, T. (1982). Specific impairments of planning. *Phil. Trans. R. Soc. Lond., 298*, 199-209.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge: Cambridge University Press.

Shallice, T., & Burgess, P. (1991). Higher-order cognitive impairments and frontal lobe lesions in man. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Frontal lobe function and dysfunction* . New York: Oxford University.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning automaticity, attending, and a general theory. *Psychological Review, 84*, 127-190.

Siegel, B. V., Nuechterlein, K. H., Wu, J. C., & Buchsbaum, M. S. (1995). Glucose metabolic correlates of continuous performance test performance in adults with a history of infantile autism, schizophrenics, and controls. *Schizophrenia Research, 17*, 85-94.

Smith, E. E., & Jonides, J. (in press). Working memory: A view from neuroimaging. *Cognitive Psychology*.

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6*, 11-20.

Smith, E. E., Jonides, J., Koeppe, R. A., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial vs. object working memory: PET investigations. *Journal of Cognitive Neuroscience, 7*(3), 337-356.

Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1990). *Structured clinical interview for DSM-III-R--patient edition (SCID-P, version 1.0)*. Washington, D.C.: American Psychiatric Press.

Spohn, H. E., & Strauss, M. E. (1989). Relation of neuroleptic and anticholinergic medication to cognitive functions in schizophrenia. *Journal of Abnormal Psychology, 98*, 367-380.

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes*. New York: Raven Press.

Swartz, B. E., Halgren, E., Fuster, J. M., Simpkins, E., Gee, M., & Mandelkern, M. (1995). Cortical metabolic activation in humans during a visual memory task. *Cerebral Cortex, 5*, 205-214.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.

Thimm, G., Moerland, P., & Fusler, E. (1996). The interchangeability of learning rate and gain in backpropagation networks. *Neural Computation, 8*, 451-469.

Tipper, S. P., & Cranston, M. (1985). The negative priming effect: Inhibitory effects of ignored primes. *The Quarterly Journal of Experimental Psychology, 37A*, 571-590.

Tulving, E., Kapur, S., Craik, F. I., 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.

Vallar, G., & Shallice, T. (1990). *Neuropsychological impairments of short term memory*. Cambridge, England: Cambridge University Press.

Vecera, S. P., & Gilds, K. S. (1998). What processing is impaired in apperceptive agnosia? Evidence from normal subjects. *Journal of Cognitive Neuroscience, 10*, 568-580.

Walker, E. (1981). Attentional and neuromotor functions of schizophrenics, schizoaffectives, and patients with other affective disorders. *Archives of General Psychiatry, 38*(Dec), 1355-1358.

Wickens, C., Kramer, A., Vanasse, L., & Emanuel, D. (1983). Performance of Concurrent Tasks: A Psychophysiological Analysis of the Reciprocity of Information-Processing Resources. *Science, 221*, 1080-1083.

Willner, P., & Scheel-Kruger, J. (1991). *The mesolimbic dopamine system: From motivation to action.* New York: Wiley.

Wise, R. A., & Rompre, P.-P. (1989). Brain dopamine and reward. *Annual Review of Psychology, 40*, 191-225.

Woods, R. P., Cherry, S. R., & Mazziotta, J. C. (1992). Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography, 16*, 620-633.

Woods, R. P., Mazziotta, J. C., & Cherry, S. R. (1993). MRI - PET registration with automated algorithm. *Journal of Computer Assisted Tomography, 17*(4), 536-546.

Zatorre, R. J., Meyer, E., Gjedde, A., & Evans, A. C. (1996). PET studies of phonetic processes in speech perception: Review, replication, and re-analysis. *Cerebral Cortex, 6*, 21-30.

Zipser, D. (1991). Recurrent network model of the neural mechanism of short-term active memory. *Neural Computation, 3*, 179-193.

Zipser, D. (1992). Identification models of the nervous system. *Neuroscience, 47*(4), 853-862.

Zipser, D., Kehue, B., Littlewort, G., & Fuster, J. (1993). A Spiking Network Model of Short-Term Active Memory. *The Journal of Neuroscience, 13*, 3406-3420.

AUTHOR NOTES

This research represents a portion of the dissertation completed by Todd S. Braver in partial fulfillment of the requirements for a doctoral degree at Carnegie Mellon University. The research was supported by an APA Dissertation Research Grant (T.S.B.) and NIMH Grants MH47073, MH52864, MH45156, MH47566 (J.D.C.). Preliminary versions of the work were presented at the Annual Meeting of the Society for Research in Psychopathology in 1996, the Annual Meeting of the Psychonomic Society in 1997, and the Third and Fourth International Conferences on Functional Mapping of the Human Brain in 1997 and 1998.

We gratefully acknowledge the support and assistance of many individuals in conducting this work: Sandra Banks and Sarah Shomstein helped in data collection for Study 1 and Study 5; Amy Sanders, Leigh E. Nystrom, Fred W. Sabb, and Douglas C. Noll provided invaluable technical support for Study 6 and Study 7; Randy O'Reilly and Chad Dawson, are thanked for providing the PDP++ simulation environment used in Study 2 and Study 3; Charles Hachten aided with recruitment for Study 4; Cameron S. Carter provided helpful suggestions for the design and implementation of Study 4 and Study 7. We wish to thank the staff and patients of Western Psychiatric Institute and Clinic for their kind cooperation and support. We also thank James L. McClelland, David Plaut, Marcel A. Just, David A. Lewis, and David A. Balota for helpful comments on earlier drafts of this article.

Table 1

Study 1: Accuracy as a Function of Response Speed

			Errors (percent)	
Response Speed	Condition	Delay	M	SE
Fast	AX	Short	4.1	0.5
		Long	5.2	0.5
	AY	Short	18.2	1.7
		Long	22.1	2.0
	BX	Short	6.0	0.9
		Long	3.4	0.7
	BY	Short	0.5	0.2
		Long	0.9	0.4
Slow	AX	Short	4.3	0.4
		Long	7.7	0.6
	AY	Short	2.4	0.6
		Long	3.1	0.7
	BX	Short	7.9	0.9
		Long	8.6	1.2
	BY	Short	1.1	0.4
		Long	0.5	0.3

Table 2

Study 2: Accuracy as a Function of Response Speed

			Errors (percent)	
Response Speed	Condition	Delay	M	SE
Fast	AX	Short	6.3	0.3
		Long	7.8	0.3
	AY	Short	14.6	1.0
		Long	17.0	1.2
	BX	Short	3.0	0.6
		Long	3.4	0.6
	BY	Short	0.1	0.1
		Long	0.4	0.2
Slow	AX	Short	3.4	0.2
		Long	3.8	0.2
	AY	Short	3.8	0.6
		Long	4.7	0.6
	BX	Short	14.2	1.2
		Long	11.8	1.1
	BY	Short	0.60	0.2
		Long	1.2	0.3

Table 3

Study 3: Accuracy as a Function of Response Speed

			Errors (percent)	
Response Speed	Condition	Delay	M	SE
Fast	AX	Short	11.0	0.4
		Long	16.8	0.4
	AY	Short	11.9	1.0
		Long	9.3	1.0
	BX	Short	6.2	0.8
		Long	13.7	1.2
	BY	Short	0	0
		Long	0.1	0.1
Slow	AX	Short	10.3	0.4
		Long	13.5	0.4
	AY	Short	6.9	0.8
		Long	5.5	0.7
	BX	Short	31.9	1.5
		Long	43.3	1.7
	BY	Short	0.1	0.1
		Long	0.4	0.2

Table 4

Study 4: Clinical and Demographic Characteristics

	Group			
	Normal Controls		Schizophrenia Patients	
	M	SD	M	SD
Age (in years)	25.13	5.5	24.88	6.78
Sex (% male)	50		56	
Parent's Education (in years)	14.1	2.9	13.61	3.2
Education (in years)	16.25	2.67	13.40	1.36
Global Assessment Scale	------		31.63	7.82
Total BPRS	------		61.31	10.95

Table 5

Study 5: Accuracy as a Function of Response Speed

			Task Condition					
			Baseline		Interference		Degraded	
Response Speed	Condition	Delay	M	SE	M	SE	M	SE
Fast	AX	Short	6.7	1.2	27.1	2.9	15.7	1.9
		Long	6.4	1.1	28.8	2.6	20.9	2.3
	AY	Short	19.9	3.1	18.8	3.2	16.5	2.8
		Long	21.9	3.7	11.3	2.4	13.7	2.8
	BX	Short	7.4	1.7	1.5	0.9	13.6	2.5
		Long	3.5	1.0	5.0	1.7	7.5	2.4
	BY	Short	1.3	0.6	0.4	0.4	3.0	1.1
		Long	0.5	0.5	1.5	0.9	1.5	0.9
Slow	AX	Short	5.2	0.7	7.7	1.4	22.3	1.7
		Long	8.1	1.0	8.1	0.8	26.5	2.1
	AY	Short	3.4	0.9	3.8	1.8	5.7	1.5
		Long	1.6	0.7	4.2	1.3	5.1	1.4
	BX	Short	7.2	1.6	13.2	2.4	15.9	2.7
		Long	8.1	1.8	16.4	3.0	16.8	2.9
	BY	Short	0.6	0.5	1.1	0.8	1.6	0.9
		Long	0.3	0.3	0.6	0.4	2.4	0.9

Table 6

Study 6: Significant Activation as a Function of Delay

Regions of Interest	Brodmann Area(s)	X[a]	Y[a]	Z[a]	Volume (mm^3)[b]	Maximum Z-score[c]
Left DLPFC	46/9	-30	47	26	701	3.06
Right DLPFC	46/9	32	41	32	2026	3.06
Left Inferior Frontal Cortex	44/6	-54	9	17	1658	3.88

[a] X, Y, and Z are coordinates in a standard stereotactic space [Talairach, 1988 #31] in which positive values refer to regions right of (X), anterior to (Y), and superior to (Z) the anterior commissure (AC).

[b] Volume refers to the number of voxels (converted to mm^3) which reached statistical significance in each region of interest.

[c] F values were converted to Z-scores to provide a measure of effect size independent of sample size.

FIGURE CAPTIONS

Figure 1. Simple canonical model of cognitive control

Figure 2. Data from Study 1. Errors and reaction times in the AX-CPT task. Error bars for all figures represent 1 Standard Error of the Mean (S.E.M.). Short = Short delay condition; Long = Long delay condition.

Figure 3. Diagram of the AX-CPT model used in Study 2.

Figure 4. Data from Study 2. Errors and reaction times in the AX-CPT model. For all model data, reaction times are in simulated activation time steps.

Figure 5. Data from Study 2. Activation dynamics in AY correct vs. incorrect trials.

Figure 6. Data from Study 2. Activation dynamics in BX correct vs. incorrect trials.

Figure 7. Data from Study 2. Activation dynamics in long vs. short delay.

Figure 8. Data from Study 3. Errors and d-'context in the reduced gain AX-CPT models

Figure 9. Data from Study 3. Reaction times in the reduced gain AX-CPT models.

Figure 10. Data from Study 3. Activation dynamics in reduced vs. normal gain models.

Figure 11. Data from Study 3: Activation dynamics as a function of delay in the reduced gain models.

Figure 12. Data from Study 4: Errors and d'-context for patients with schizophrenia and matched controls.

Figure 13. Data from Study 4: Reaction times for patients with schizophrenia and matched controls.

Figure 14. Schematic of the three task conditions used in Study 5.

Figure 15. Data from Study 5: Errors and d'-context across the three task conditions

Figure 16. Data from Study 5: Reaction times across the three task conditions

Figure 17. Schematic of the experimental and scanning design used in Study 6. In the short and long delay conditions, four scans are acquired per trial. In the "extra long" delay condition, eight scans were acquired per trial.

Figure 18. Data from Study 6. Delay related activity as a function of scan within trial for the three

active regions within PFC (shown in an axial slice 24mm superior to the AC-PC plane).

Figure 19. Data from Study 6. The time course of event-related activity in the PFC regions (the dorsolateral PFC regions are averaged together).

Figure 20. Data from Study 7. The time course of event-related activity in left dorsolateral PFC

Figure 1

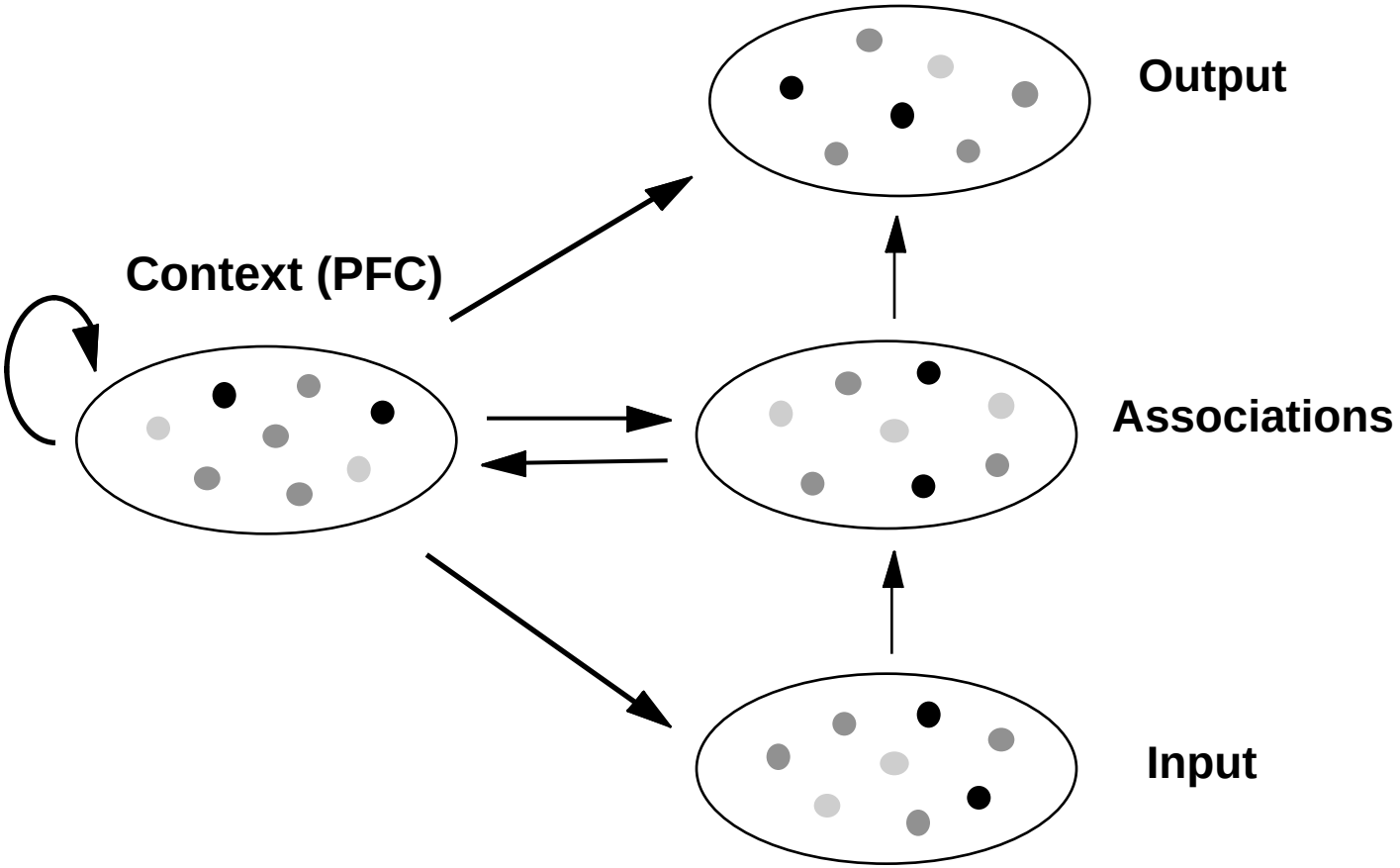


Figure 2

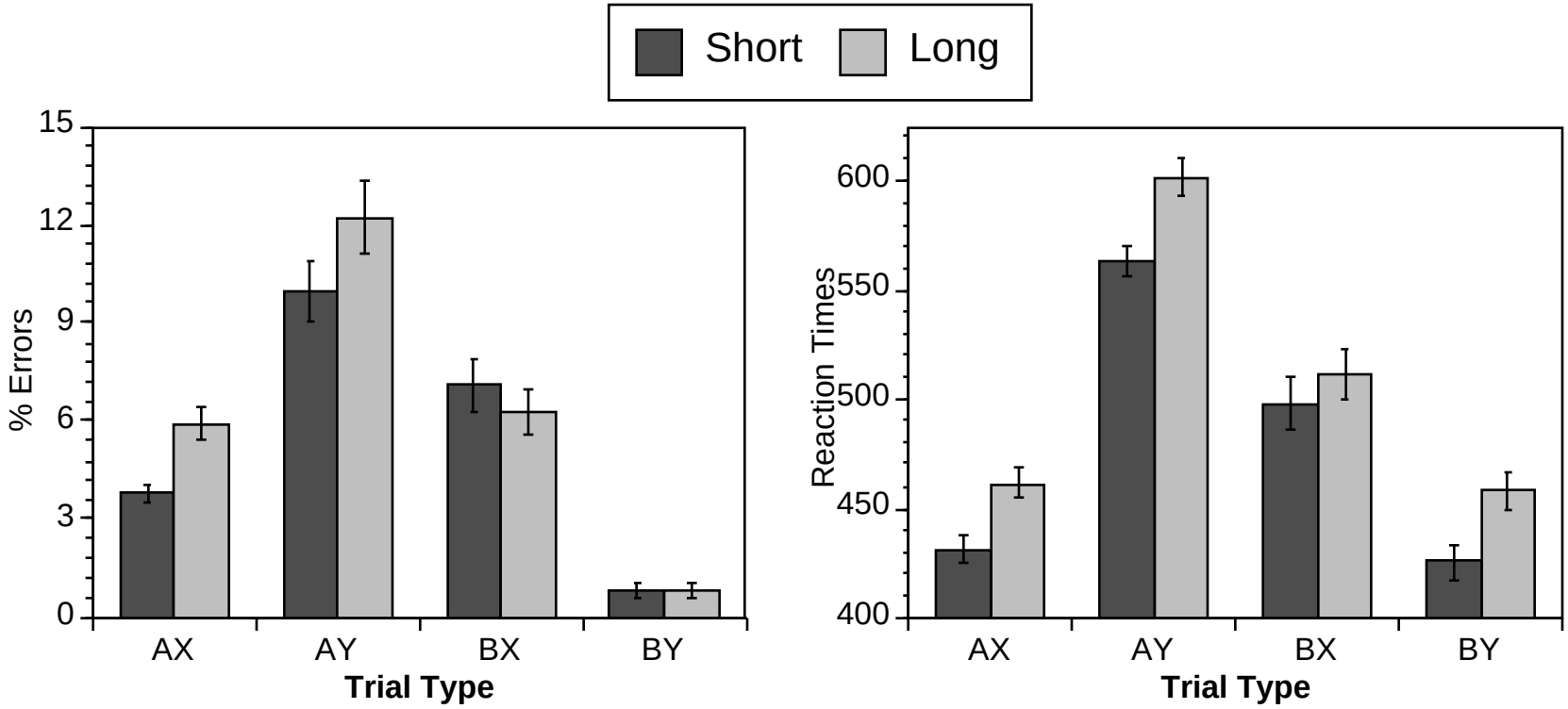

Short
Long
% Errors
15
12
9
6
3
0
AX
AY
BX
BY
Trial Type
Reaction Times
600
550
500
450
400
AX
AY
BX
BY
Trial Type

Figure 3

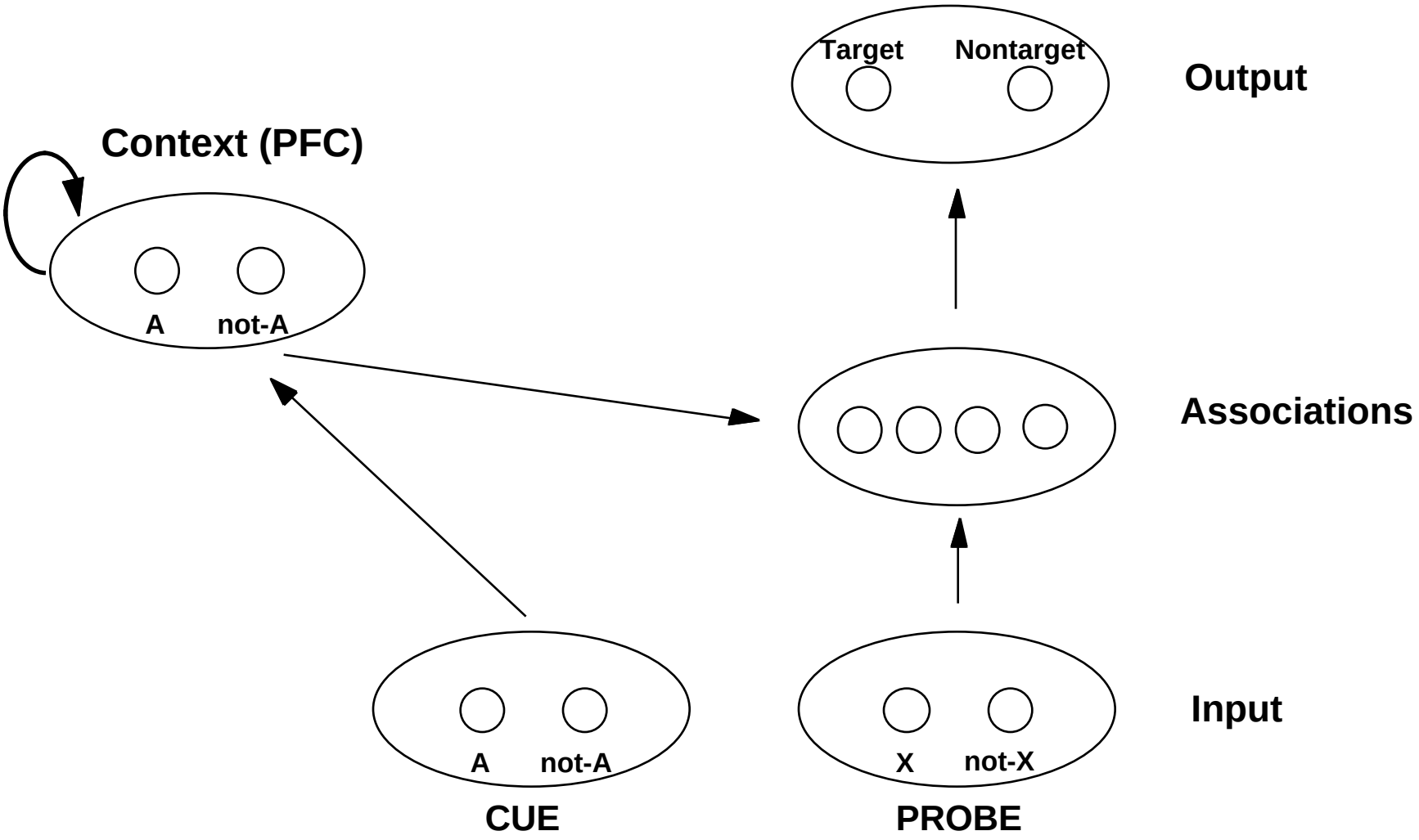

Target
Nontarget
Output
Context (PFC)
A
not-A
Associations
A
not-A
CUE
X
not-X
PROBE
Input

Figure 4.

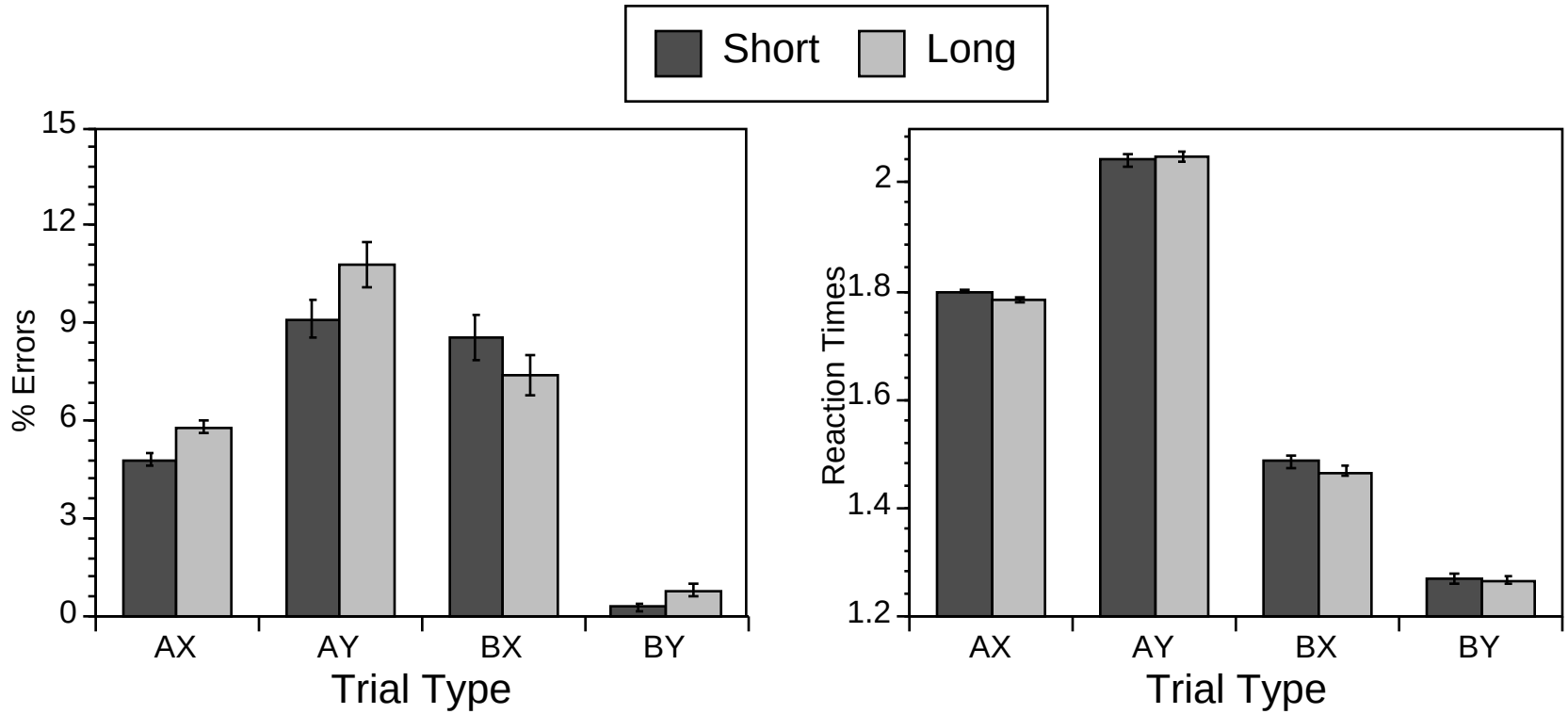
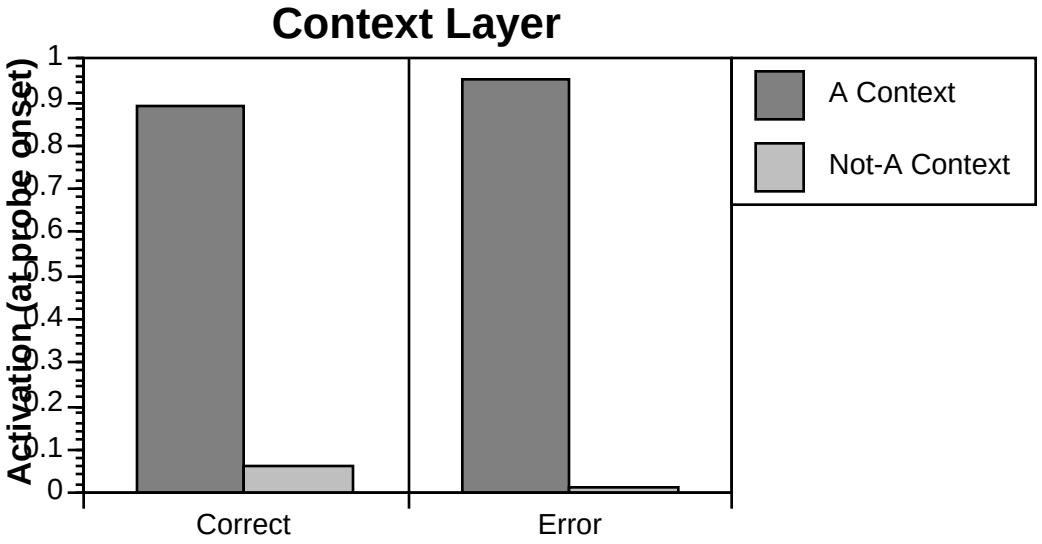

Short
Long
% Errors
15
12
9
6
3
0
AX
AY
BX
BY
Trial Type
Reaction Times
2
1.8
1.6
1.4
1.2
AX
AY
BX
BY
Trial Type

Figure 5.

A.


Context Layer
Activation (at probe onset)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Correct
Error
A Context
Not-A Context

B.

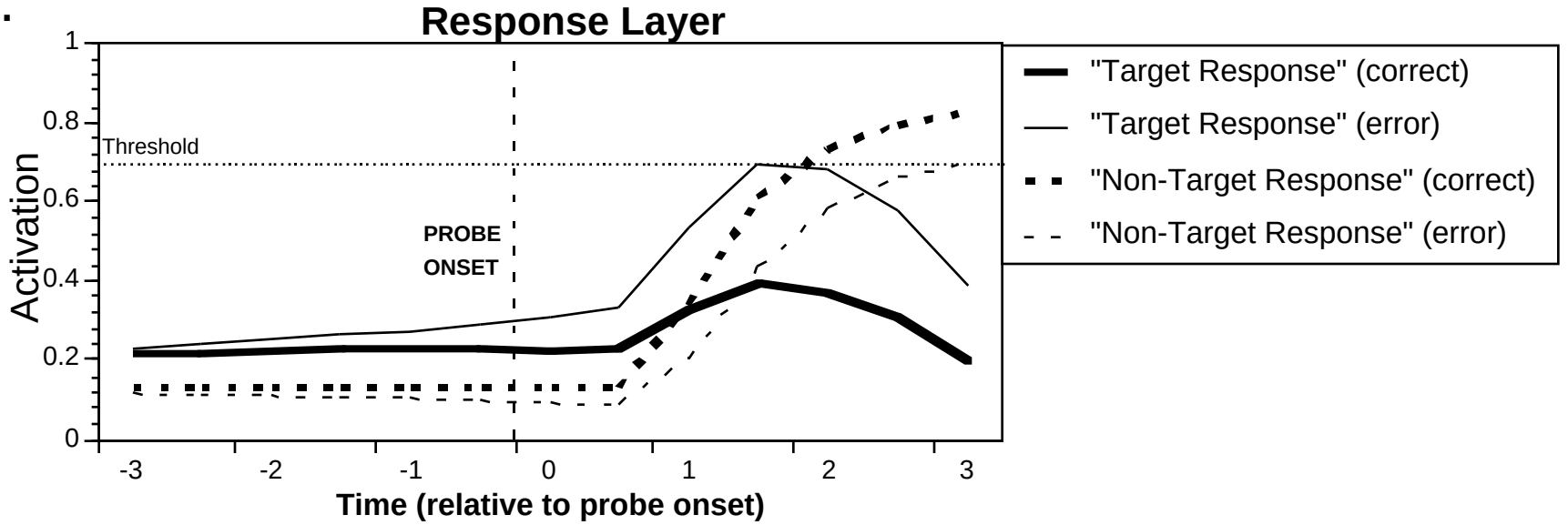
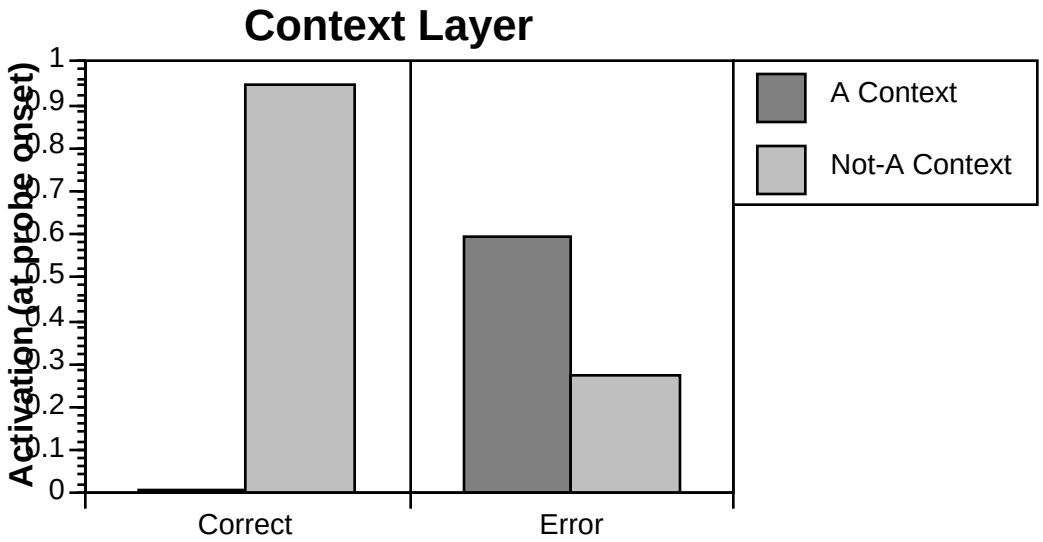

Response Layer
Activation
1
0.8
0.6
0.4
0.2
0
Threshold
PROBE
ONSET
-3
-2
-1
0
1
2
3
Time (relative to probe onset)
"Target Response" (correct)
"Target Response" (error)
"Non-Target Response" (correct)
"Non-Target Response" (error)

Figure 6.

A.


Context Layer
Activation (at probe onset)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Correct
Error
A Context
Not-A Context

B.

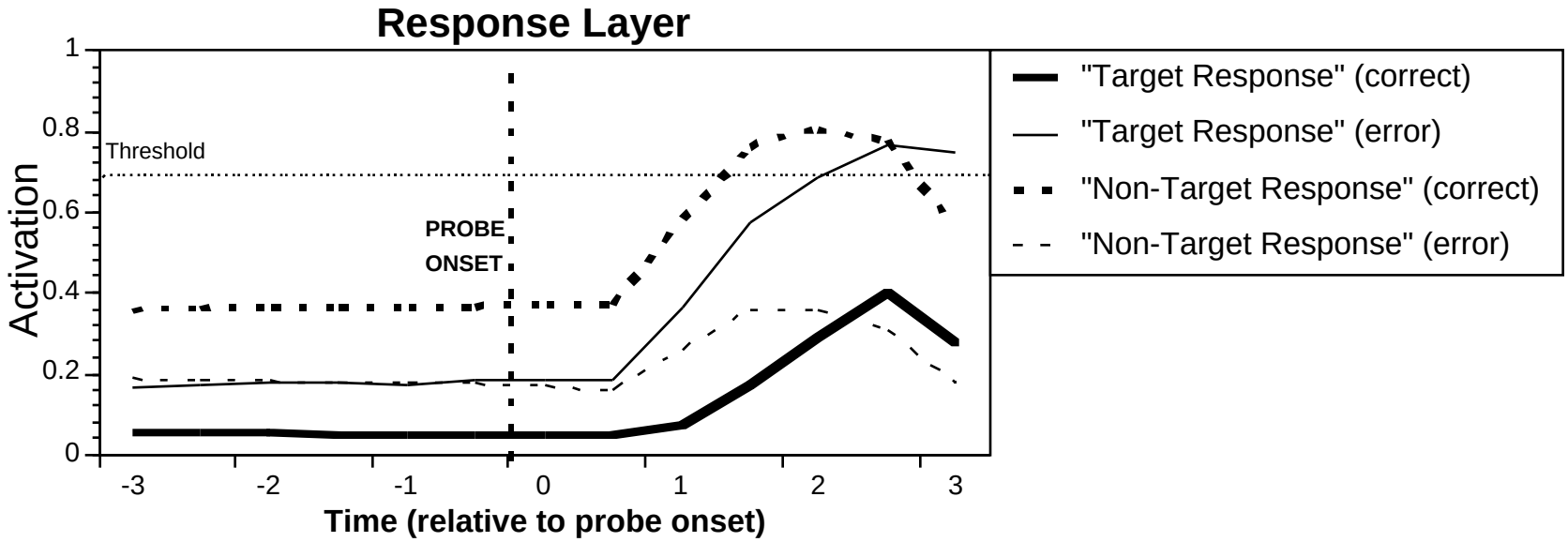

Response Layer
Activation
1
0.8
0.6
0.4
0.2
0
Threshold
PROBE
ONSET
-3
-2
-1
0
1
2
3
Time (relative to probe onset)
"Target Response" (correct)
"Target Response" (error)
"Non-Target Response" (correct)
"Non-Target Response" (error)

Figure 7

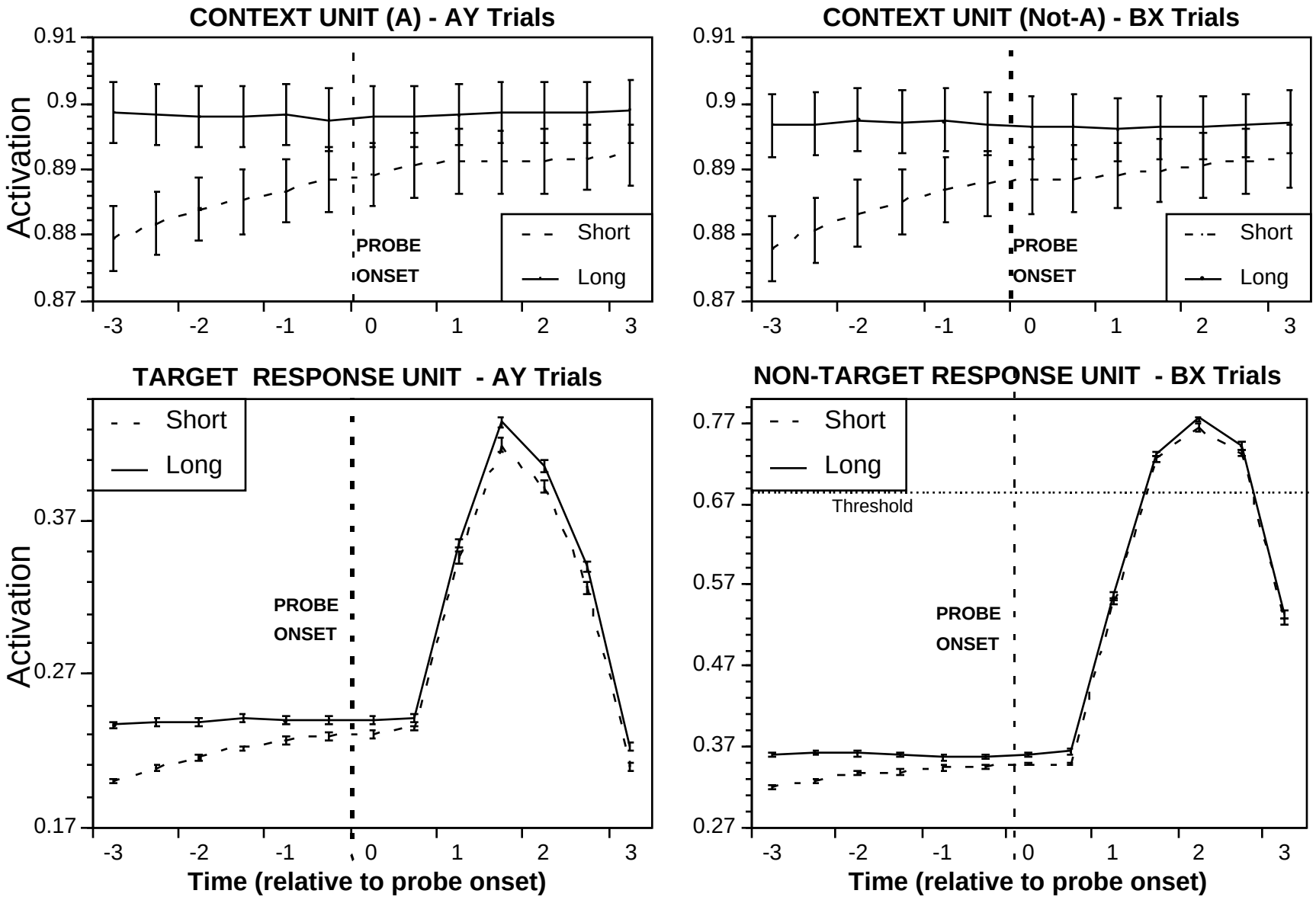

CONTEXT UNIT (A) - AY Trials
CONTEXT UNIT (Not-A) - BX Trials
TARGET RESPONSE UNIT - AY Trials
NON-TARGET RESPONSE UNIT - BX Trials
Activation
Short
Long
PROBE
ONSET
Threshold
Time (relative to probe onset)
0.91
0.9
0.89
0.88
0.87
0.77
0.67
0.57
0.47
0.37
0.27
0.17
-3
-2
-1
0
1
2
3

Figure 8

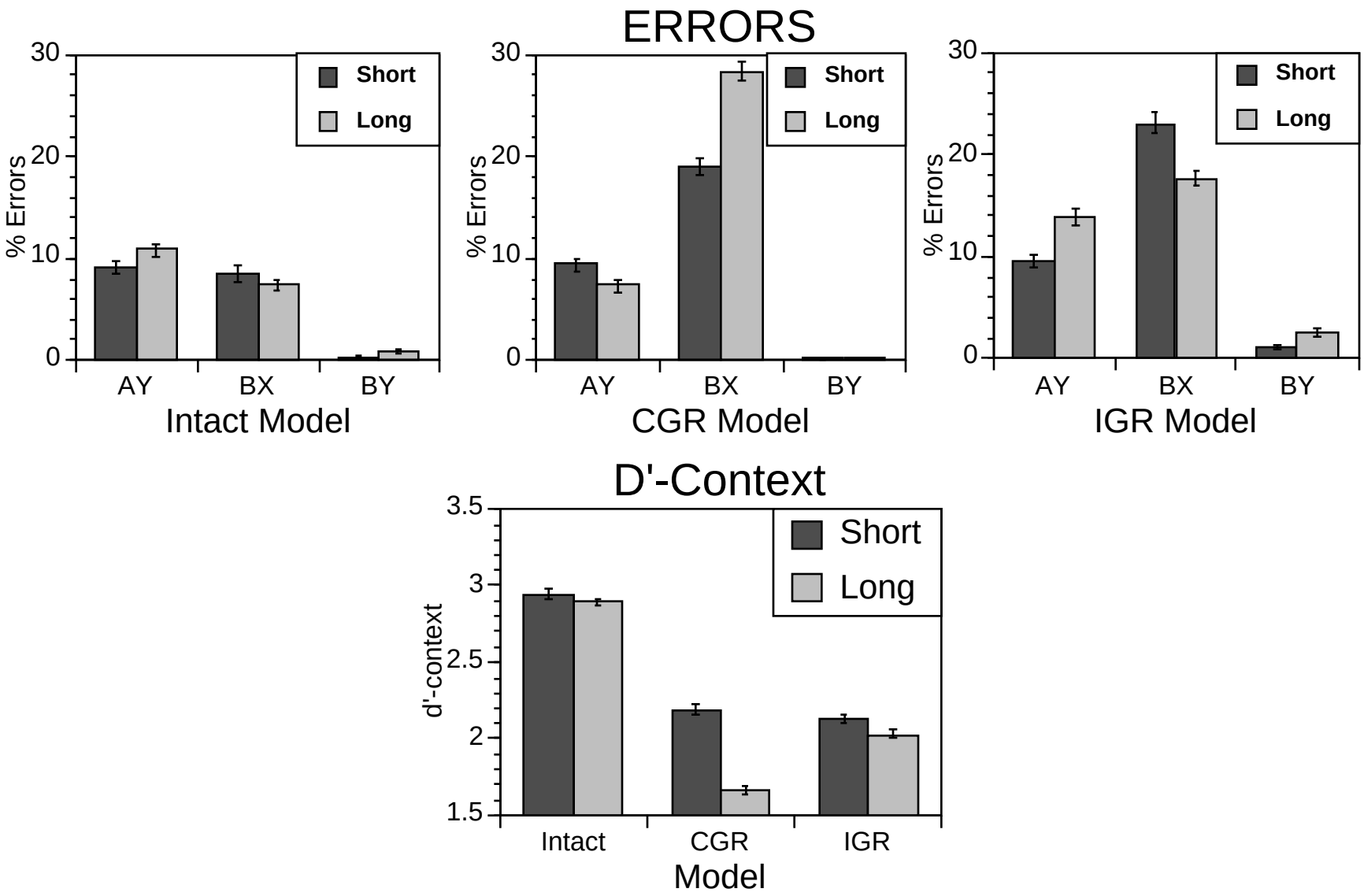

ERRORS
% Errors
30
20
10
0
Short
Long
AY
BX
BY
Intact Model
CGR Model
IGR Model
D'-Context
d'-context
3.5
3
2.5
2
1.5
Intact
CGR
IGR
Model

Figure 9

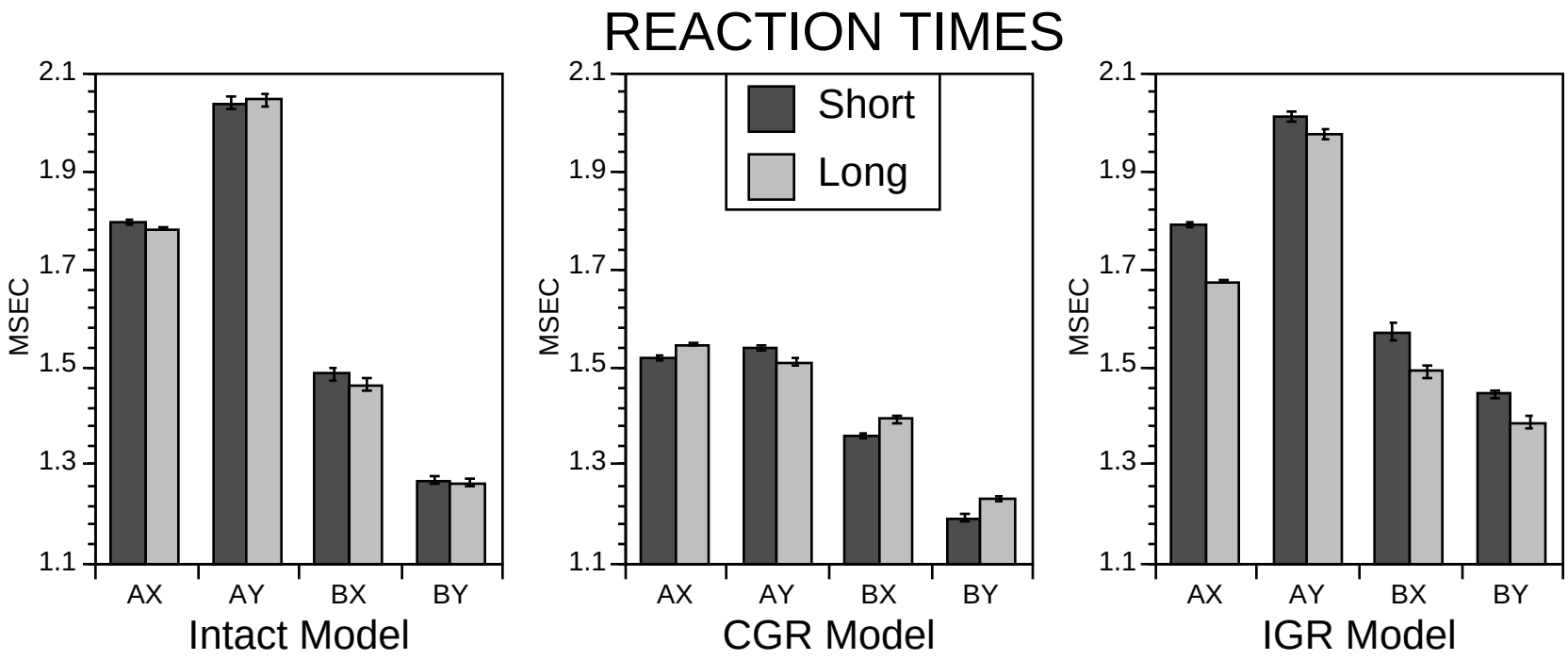

REACTION TIMES
MSEC
2.1
1.9
1.7
1.5
1.3
1.1
AX
AY
BX
BY
Intact Model
Short
Long
CGR Model
IGR Model

Figure 10

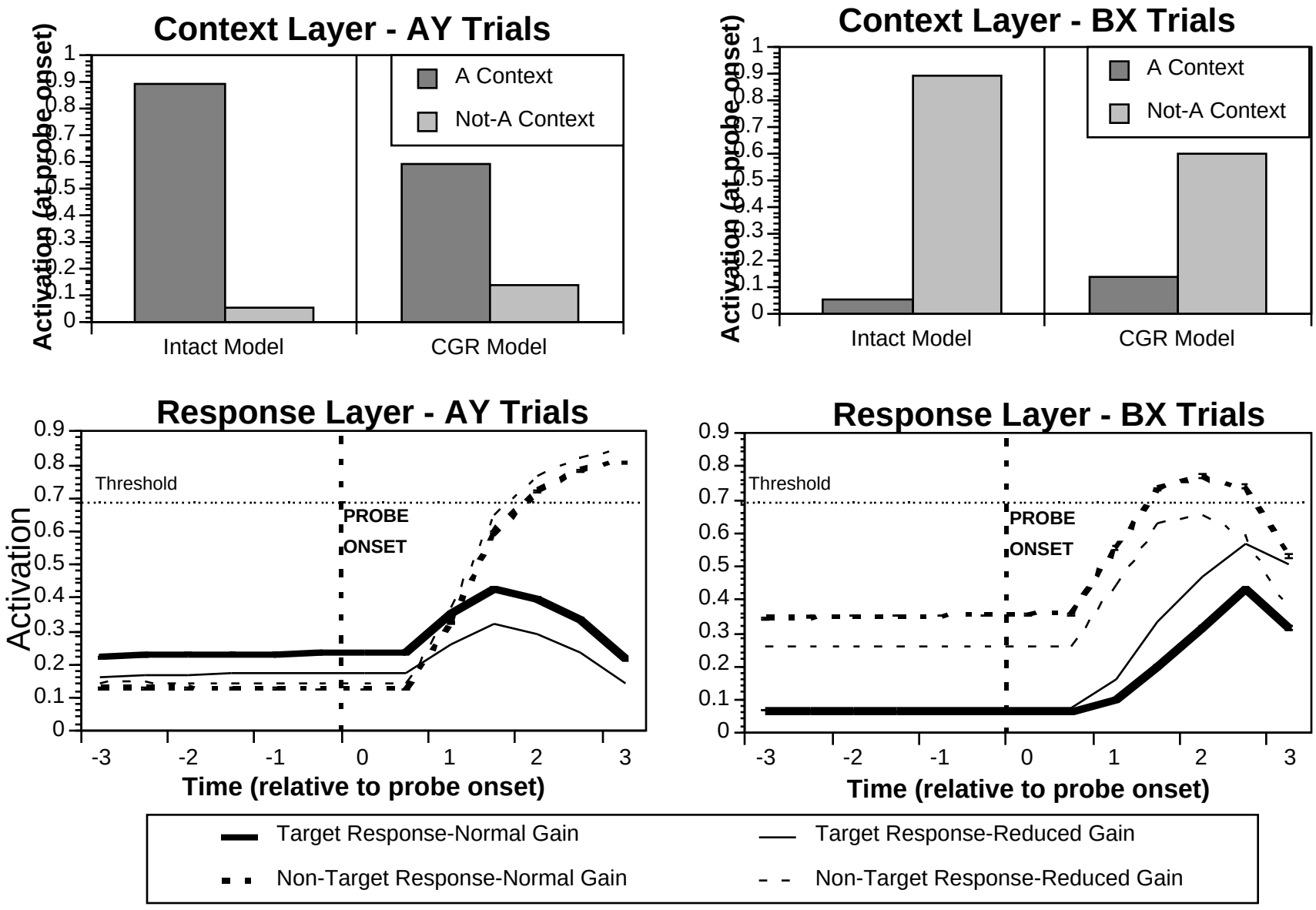

Context Layer - AY Trials
Activation (at probe onset)
A Context
Not-A Context
Intact Model
CGR Model
Context Layer - BX Trials
Activation (at probe onset)
A Context
Not-A Context
Intact Model
CGR Model
Response Layer - AY Trials
Activation
Threshold
PROBE ONSET
Time (relative to probe onset)
Response Layer - BX Trials
Threshold
PROBE ONSET
Time (relative to probe onset)
Target Response-Normal Gain
Target Response-Reduced Gain
Non-Target Response-Normal Gain
Non-Target Response-Reduced Gain

Figure 11

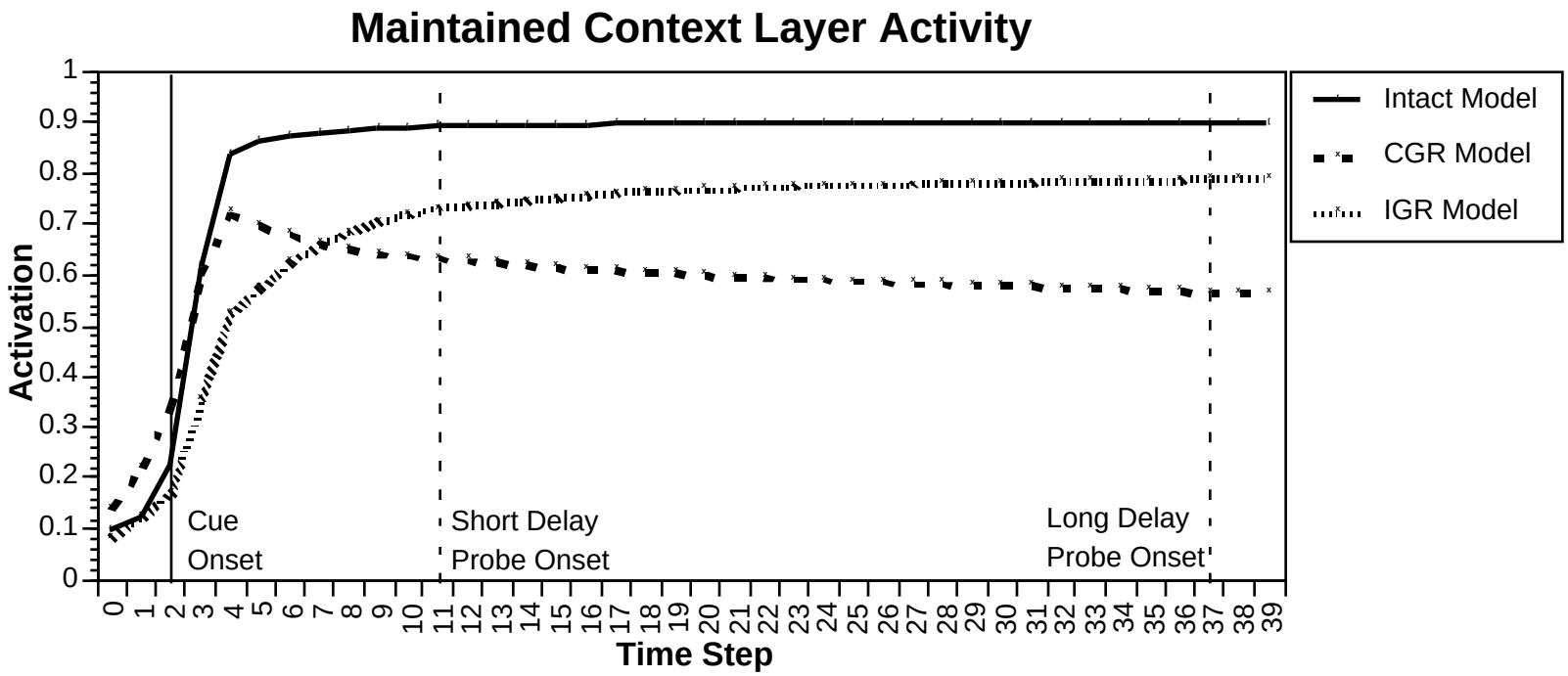

Maintained Context Layer Activity
Activation
Time Step
Intact Model
CGR Model
IGR Model
Cue
Onset
Short Delay
Probe Onset
Long Delay
Probe Onset

Figure 12

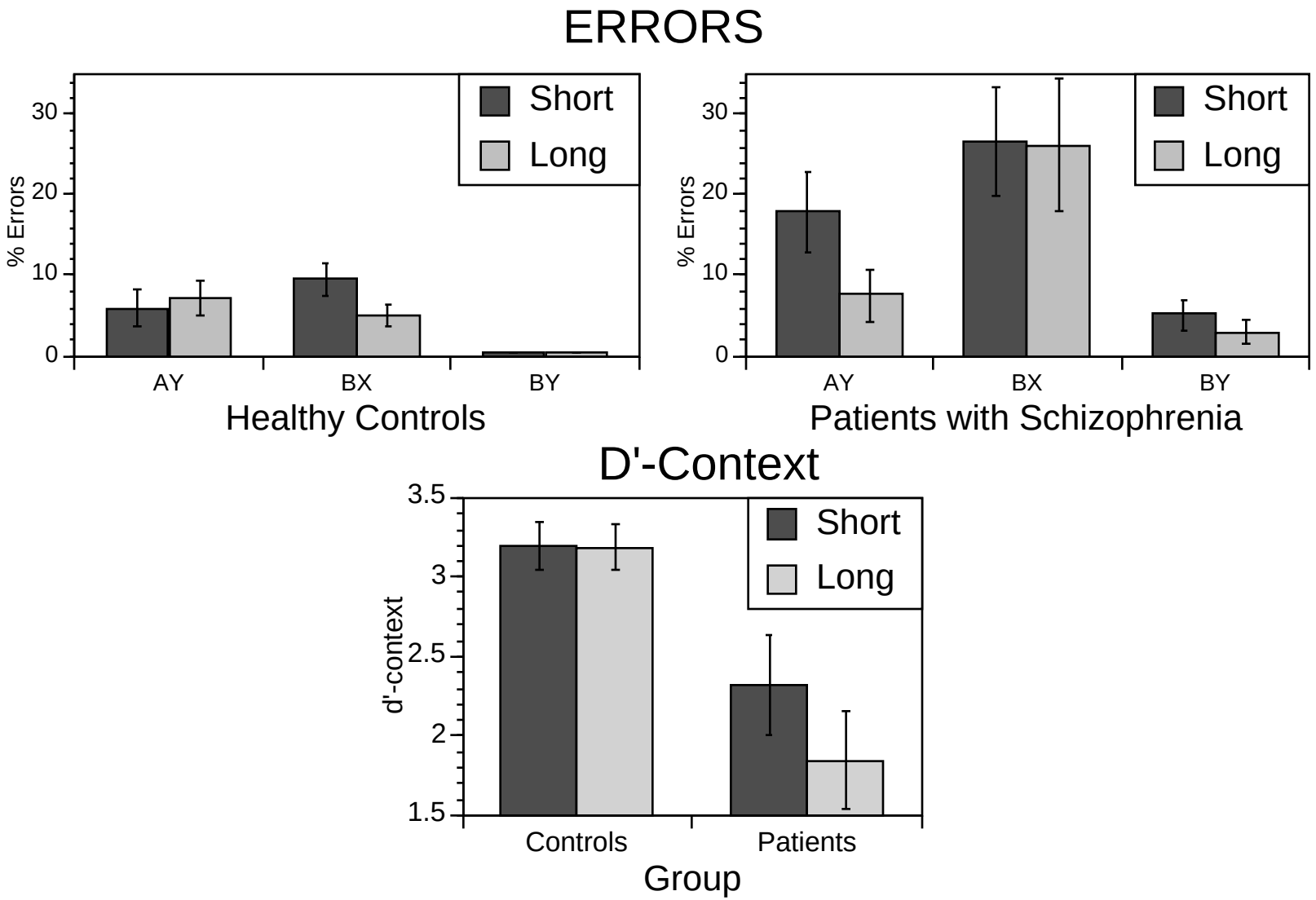

ERRORS
Short
Long
% Errors
30
20
10
0
AY
BX
BY
Healthy Controls
Patients with Schizophrenia
D'-Context
d'-context
3.5
3
2.5
2
1.5
Controls
Patients
Group

Figure 13

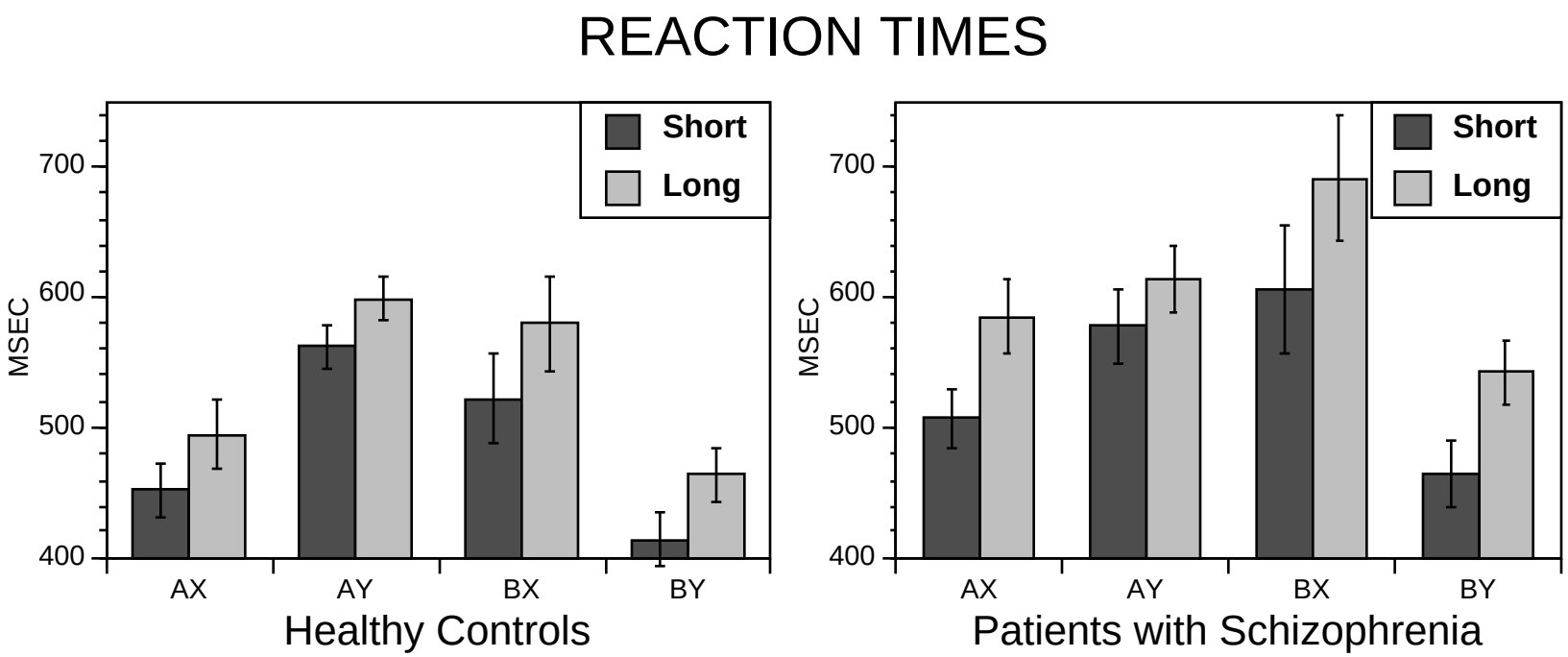

REACTION TIMES
Short
Long
MSEC
700
600
500
400
AX
AY
BX
BY
Healthy Controls
Short
Long
MSEC
700
600
500
400
AX
AY
BX
BY
Patients with Schizophrenia

Figure 14

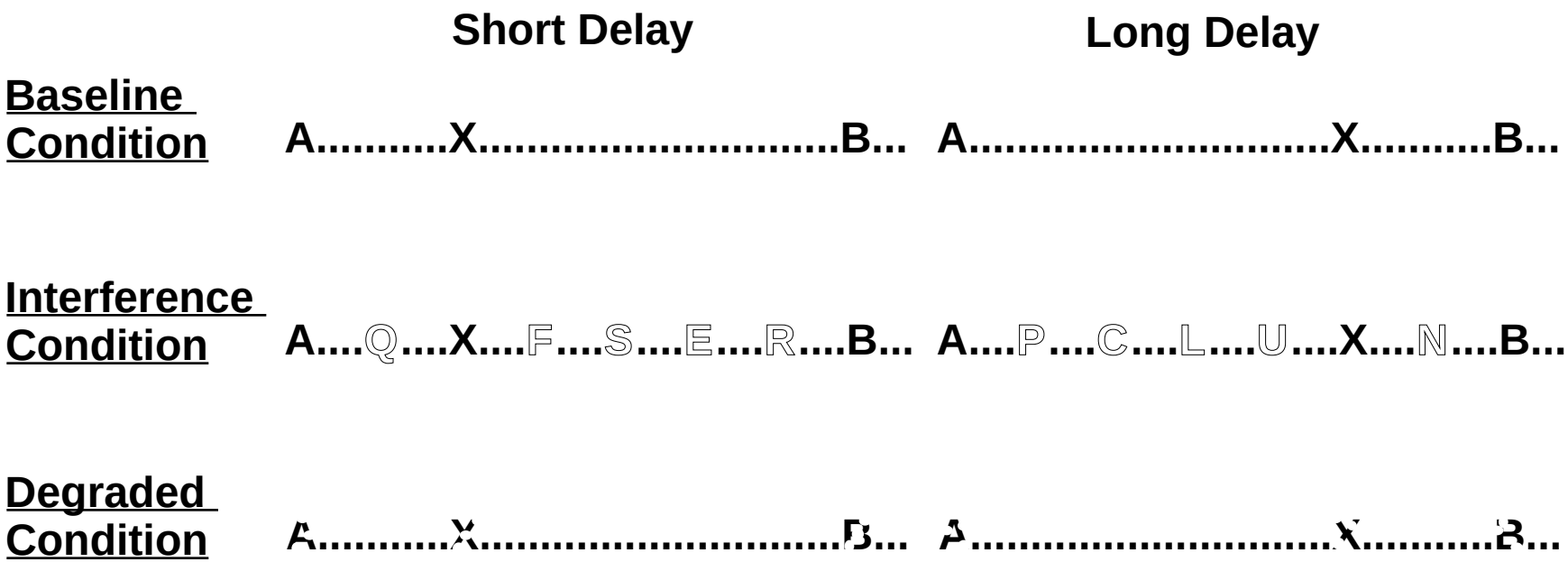

Short Delay
Long Delay
Baseline Condition
A.....X.....B...
A.....X.....B...
Interference Condition
A....Q....X....F....S....E....R....B...
A....P....C....L....U....X....N....B...
Degraded Condition
A.....X.....B...
A.....X.....B...

Figure 15

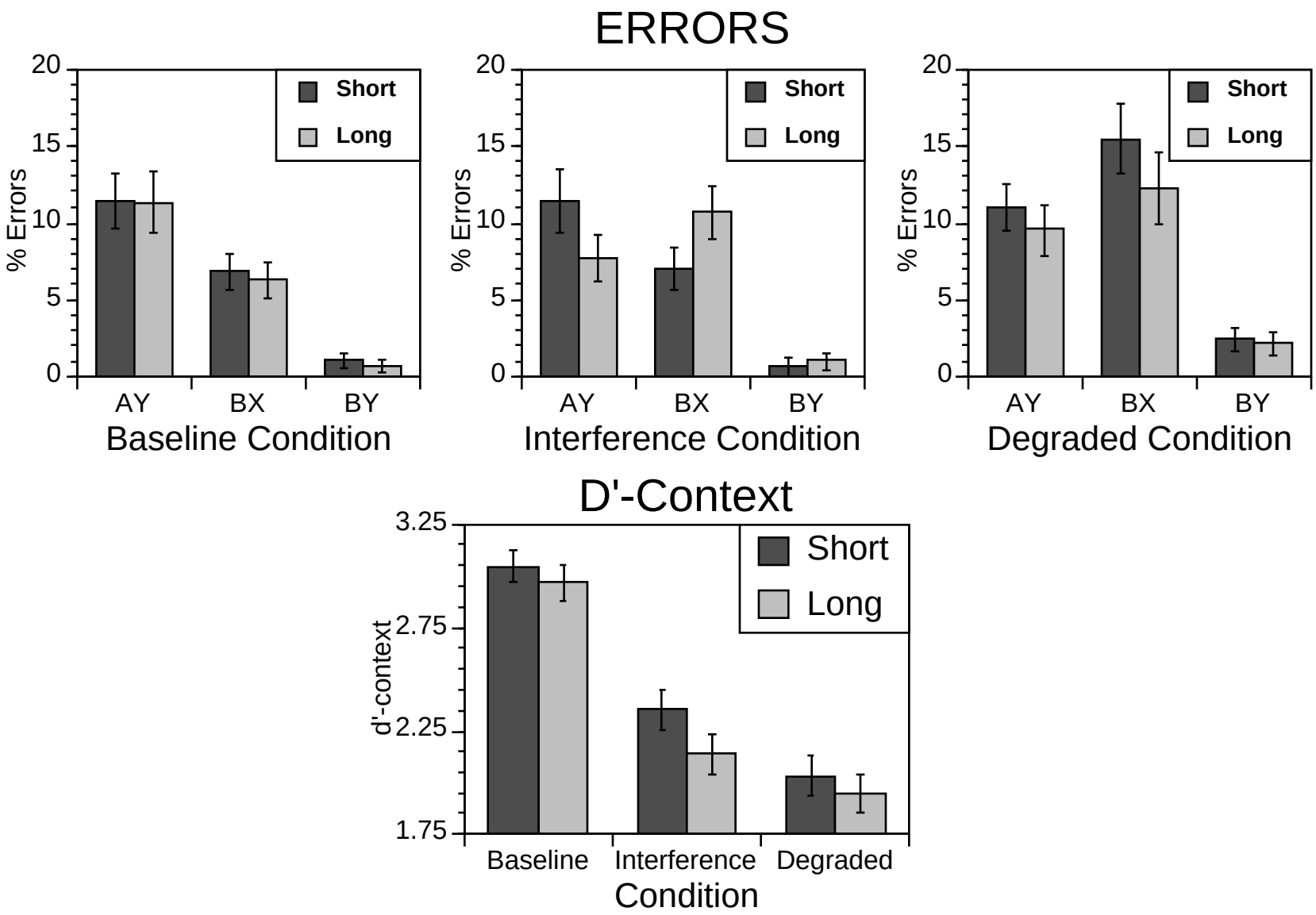

ERRORS
% Errors
20
15
10
5
0
Short
Long
AY
BX
BY
Baseline Condition
Interference Condition
Degraded Condition
D'-Context
d'-context
3.25
2.75
2.25
1.75
Baseline
Interference
Degraded
Condition

Figure 16

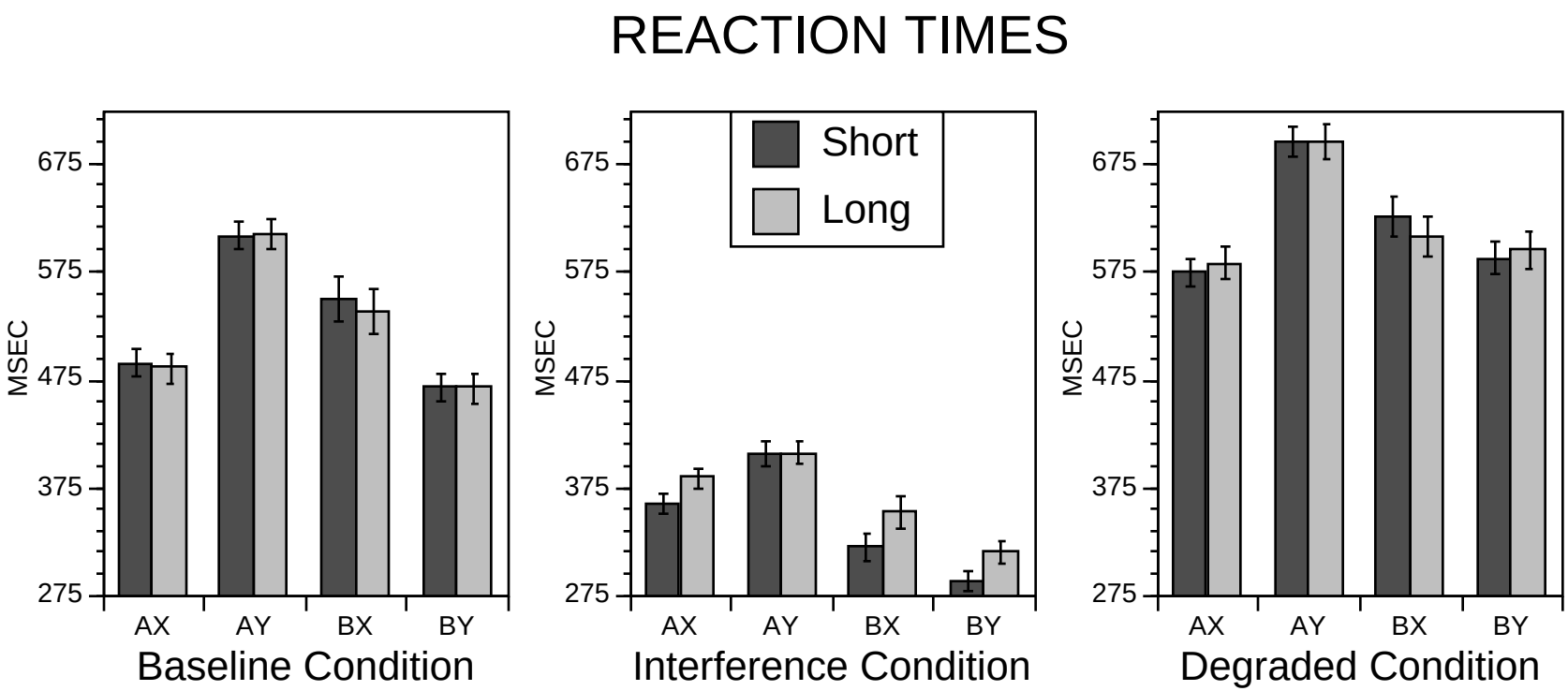

REACTION TIMES
MSEC
675
575
475
375
275
AX
AY
BX
BY
Baseline Condition
Short
Long
Interference Condition
Degraded Condition

Figure 17

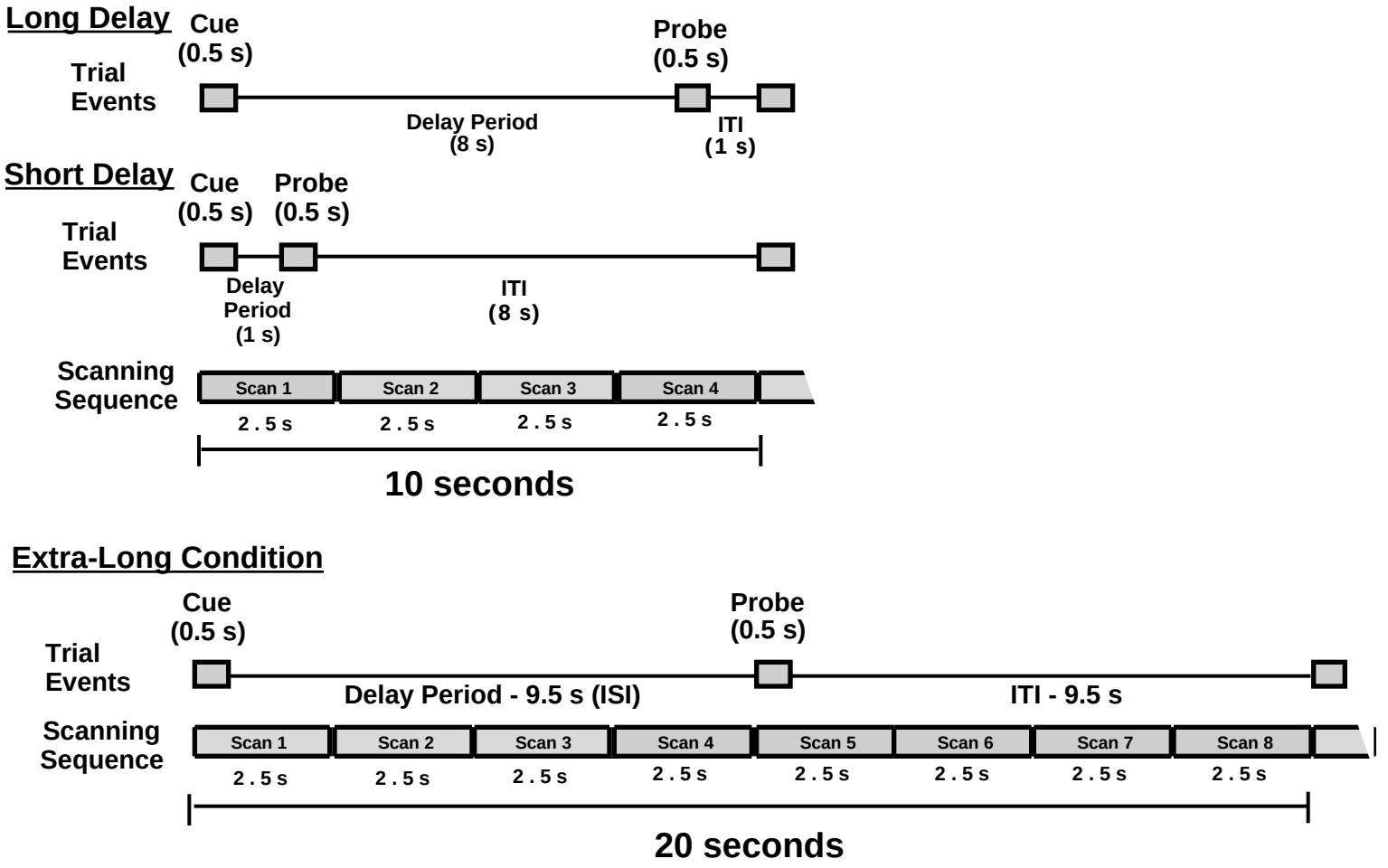

Long Delay
Cue
(0.5 s)
Probe
(0.5 s)
Trial
Events
Delay Period
(8 s)
ITI
(1 s)
Short Delay
Cue
(0.5 s)
Probe
(0.5 s)
Trial
Events
Delay
Period
(1 s)
ITI
(8 s)
Scanning
Sequence
Scan 1
Scan 2
Scan 3
Scan 4
2.5 s
2.5 s
2.5 s
2.5 s
10 seconds
Extra-Long Condition
Cue
(0.5 s)
Probe
(0.5 s)
Trial
Events
Delay Period - 9.5 s (ISI)
ITI - 9.5 s
Scanning
Sequence
Scan 1
Scan 2
Scan 3
Scan 4
Scan 5
Scan 6
Scan 7
Scan 8
2.5 s
2.5 s
2.5 s
2.5 s
2.5 s
2.5 s
2.5 s
2.5 s
20 seconds

Figure 18

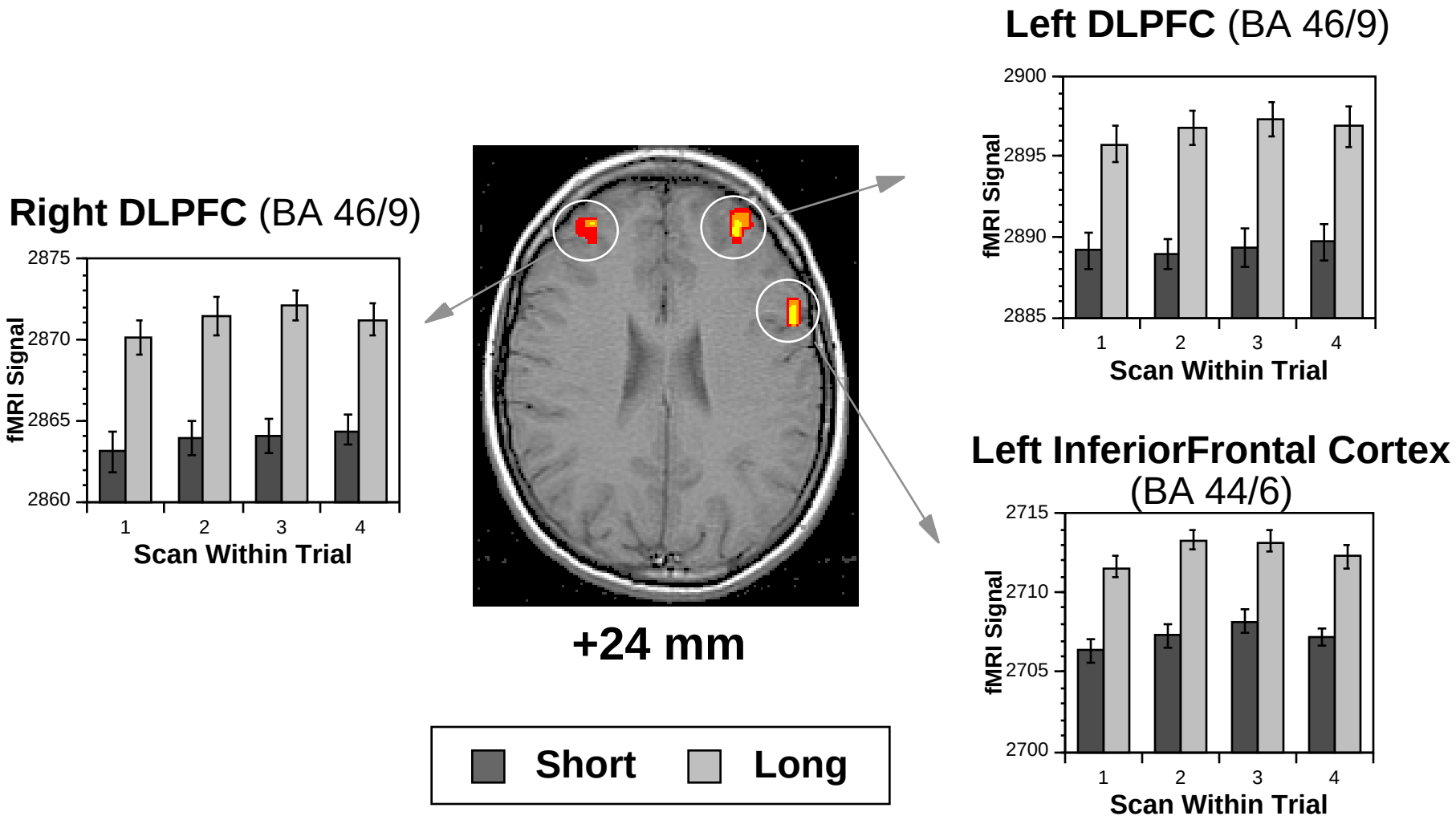

Left DLPFC (BA 46/9)
2900
2895
2890
2885
fMRI Signal
1
2
3
4
Scan Within Trial
Right DLPFC (BA 46/9)
2875
2870
2865
2860
fMRI Signal
1
2
3
4
Scan Within Trial
+24 mm
Left InferiorFrontal Cortex
(BA 44/6)
2715
2710
2705
2700
fMRI Signal
1
2
3
4
Scan Within Trial
Short
Long

Figure 19

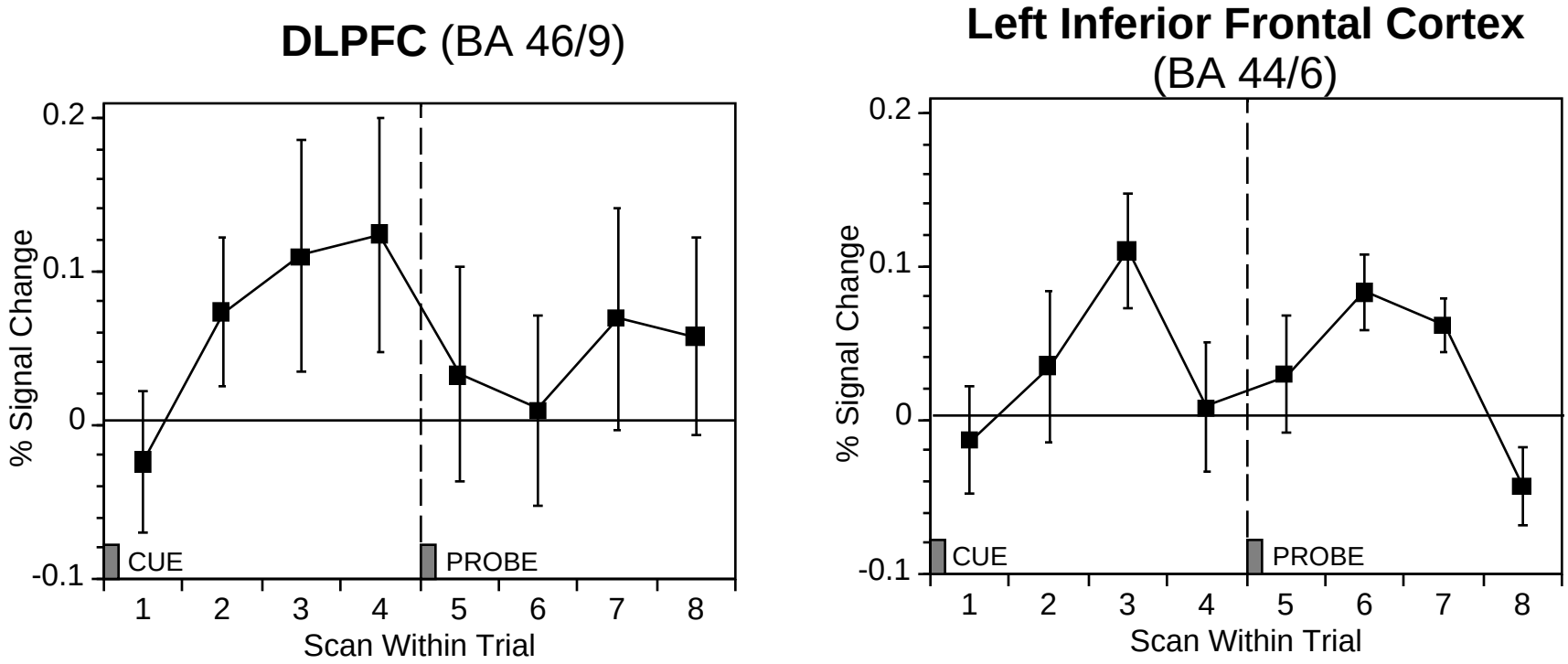
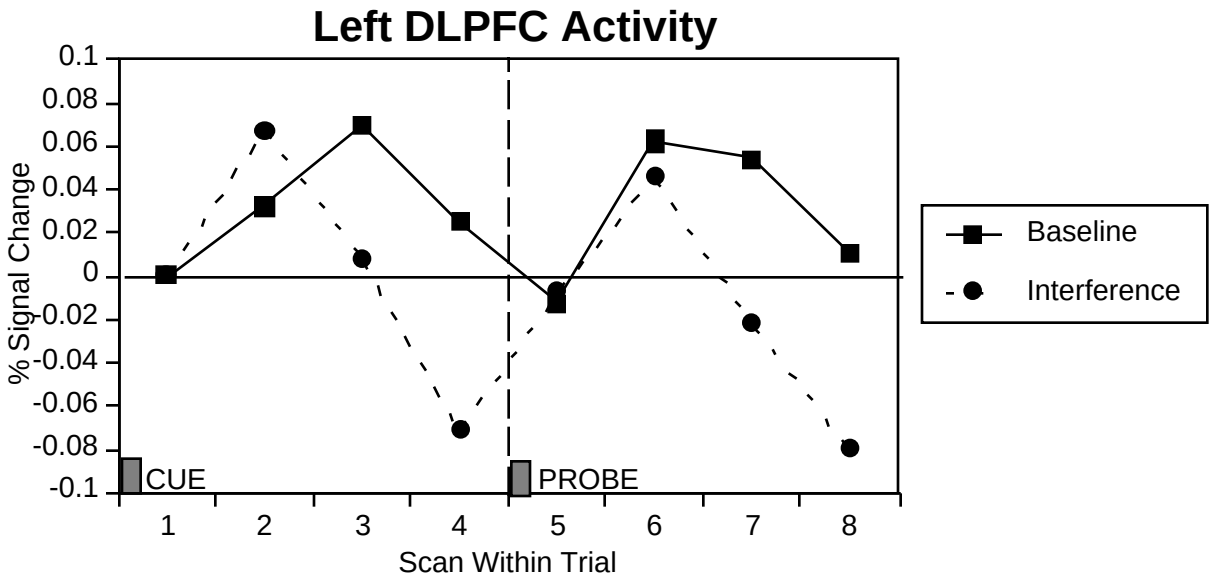

DLPFC (BA 46/9)
0.2
0.1
0
-0.1
% Signal Change
CUE
PROBE
1
2
3
4
5
6
7
8
Scan Within Trial
Left Inferior Frontal Cortex
(BA 44/6)
0.2
0.1
0
-0.1
% Signal Change
CUE
PROBE
1
2
3
4
5
6
7
8
Scan Within Trial

Figure 20


Left DLPFC Activity
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
% Signal Change
CUE
PROBE
1
2
3
4
5
6
7
8
Scan Within Trial
Baseline
Interference