6

A Systems-Level Perspective on Attention and Cognitive Control

Guided Activation, Adaptive Gating, Conflict Monitoring, and Exploitation versus Exploration

Jonathan D. Cohen, Gary Aston-Jones, and Mark S. Gilzenrat

THE SCOPE OF ATTENTION

An understanding of attention is arguably one of the most important goals of the cognitive sciences and yet also has proven to be one of the most elusive. Most attention researchers will agree that a major problem has been agreeing on a definition of the term and the scope of the phenomena to which it applies. There are no doubt as many explanations for this state of affairs as there are those who consider themselves "attention researchers." However, most will probably agree that, in large measure, this is because attention is not a unitary phenomenon—at least not in the sense that it reflects the operation of a single mechanism, or a single function of one or a set of mechanisms. Rather, attention is the emergent property of the cognitive system that allows it to successfully process some sources of information to the exclusion of others, in the service of achieving some goals to the exclusion of others. This begs an important question: If attention is so varied a phenomenon, how can we make progress in understanding it? There are two simple answers to this question: Be precise about the specific (aspects of the) phenomena to be studied, and be precise about the mechanisms thought to explain them.

In this chapter, we address a particular type of attentional phenomenon—that associated with cognitive control. Furthermore, we focus on an account that addresses not only the functional characteristics of this form of attention but also how it is implemented in neural machinery. This neurally oriented approach is attractive not only because it is intrinsically interesting to understand how the mechanisms of the brain give rise to the processes of the mind but more specifically because this exercise has proven useful in generating insights into how controlled attention operates at the systems level. By assuming that information is

represented as patterns of activity, and information processing occurs as the flow of activity, it becomes possible to understand how information represented in one part of the system can influence the processing of information in other parts of the system—that is, how attention and control operate at the systems level. The sections that follow develop this idea in greater detail, first by providing a particular example of controlled attention and how it can be modeled in terms of explicit processing mechanisms, then by showing how it can explain some of the most important observations that have been made about attention and control, and finally by reviewing recent elaborations of the basic model that have begun to address broader questions about the psychological and neural mechanisms that underlie cognitive control.

AN EXAMPLE

To see how attentional effects can be understood in terms of specific processing mechanisms, it is useful to consider an example of a model of a specific task. There are now models of a variety of tasks that could serve this purpose well. Here, we focus on a model of the Stroop task (Cohen, Dunbar, & McClelland, 1990), because this task has occupied such a central role in studies of attention (both basic and clinical), because the model illustrates in a relatively straightforward manner the principles of interest, and because this model has been used to explain a wide array of findings using the Stroop task. This model was developed within the connectionist or parallel distributed processing (PDP) framework, which has been described and elaborated in great detail elsewhere (e.g., McClelland, 1993; O'Reilly & Munakata, 2000; Rumelhart, McClelland, & PDP Research Group, 1986), and therefore we assume it is familiar (or accessible) to the reader.

In the Stroop task (Stroop, 1935), subjects must attend to one dimension of a stimulus (e.g., the color in which a word is displayed) and ignore a competing but prepotent dimension (e.g., the word itself). For example, subjects are asked to name the color of an incongruent stimulus, such as the word *green* displayed in red. In our model, units are arranged into two pathways (Figure 6.1). Stimulus units representing the color project to associative units in the color-naming pathway, which project in turn to verbal response units. The word pathway converges on the same set of verbal response units. Furthermore, connections are stronger in the word pathway, capturing the assumption that written words are more frequently and consistently mapped to their pronunciations than are visual color stimuli to the utterance of their names. As a result, with no additions to the model, it will respond to the incongruent Stroop stimulus above by "reading" the word (i.e., activating the "green" response unit). In fact, this is how human subjects respond if not instructed otherwise. That is, they produce the strongest (e.g., most familiar or salient) response to a stimulus. Critically, however, they can respond to the weaker dimension of a stimulus when asked to do so (i.e., name the color in the Stroop task). This an elementary—and perhaps the most studied—form of controlled, or voluntary, attention.

To explain this ability, we make the following set of modifications to the model. First, we assume that at-rest units have relatively low activity. This corresponds well with the properties of neurons (especially those in cortical areas), which typically exhibit relatively low firing rates at rest. This can be seen in Figure 6.2a, by noting that for an input of zero, the activity of the unit is also near zero.[1] Second, we include an additional set of task demand units,[2] each of which corresponds to one of the tasks subjects are asked to perform (color naming and word reading). We assume that each of these units is connected to all of

the associative units in the corresponding pathway. Thus, the color-naming unit is connected to the associative units in the color-naming pathway, and similarly for the word-reading unit. When one of these task demand units is activated, it sends activity to all the associative units in the corresponding pathway. This has the effect of sensitizing these units to input from the stimulus units. Because this effect sits at the core of how attention operates in this model, it is important to consider this in greater detail.

Figure 6.2a illustrates the activation function for a unit—that is, the function that determines its activity based on the summed input it receives from other units. Note that this function is nonlinear.[3] This is central to our account of attentional effects. Recall that units have low activity at rest (that is, when their input is zero). In this range the activation function is relatively flat. In other words, even if one of the stimulus units were to be activated and pass activity to the corresponding associative unit, this would have limited impact on the activity of that associative unit. Now assume that one of the task-demand units is activated. This passes activity to the associative units in that pathway. Let us assume further that the amount of activity is sufficient to move these units to the midpoint of their activation function, where this is steepest. Note that this does not provide any specific information to that pathway. That is, all the units in that pathway have been equally activated, so none drive one response more than the others. However, now any input to these units from the stimulus units will have a large impact on their activity. Even a small excitatory input to one of these associative units will quickly drive its activity up, while inhibitory input to the other will drive its activity down. In other words, the effect of activating the task-demand unit is
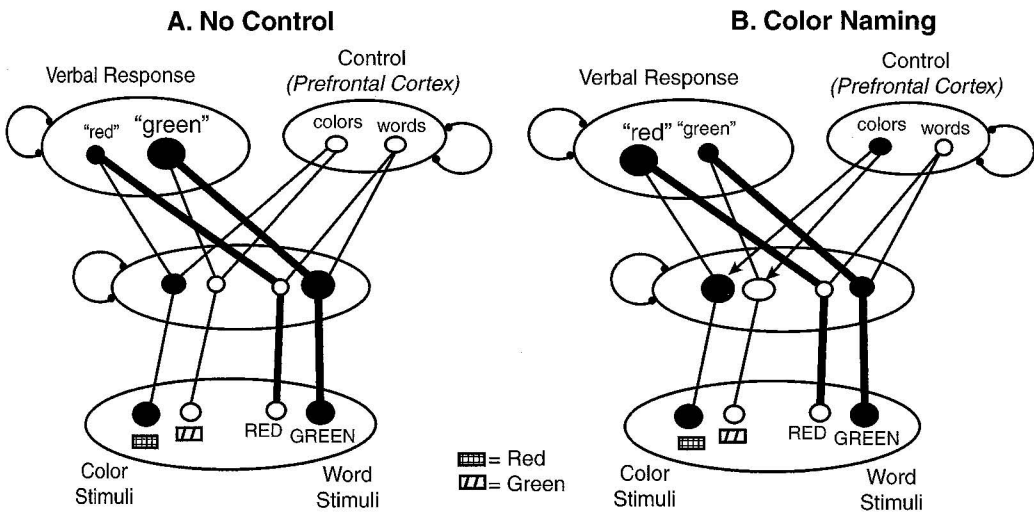

FIGURE 6.1. Model of the Stroop task. Circles represent processing units and line weights the strength of connections between units. Active units are filled (larger = more active). Looped connections with small black circles indicate mutual inhibition among units within that layer. (A) No control. Activation of conflicting inputs in the two pathways (green word input and red color input) produces a response associated with the word, due to the stronger connections in the word reading pathway. (B) Control present. The color task-demand unit is activated, representing the current intent to name the color. This passes activation to the associative units in the color naming pathway, which primes those units (indicated by larger size), and biases processing in favor of activity flowing along this pathway. This biasing effect favors activation of the response unit corresponding to the color input, even though the connection weights in this pathway are weaker than those in the word pathway.
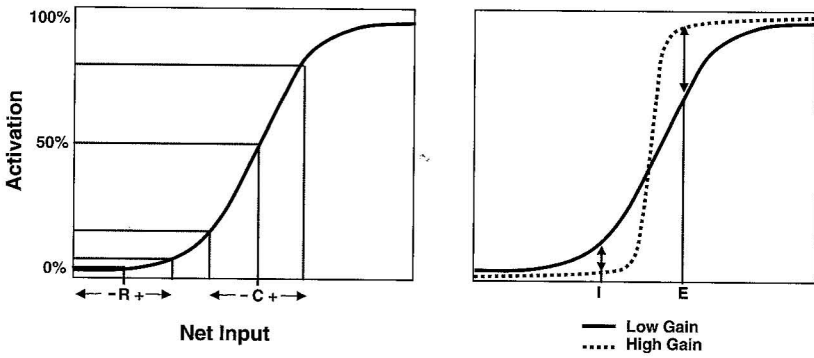
FIGURE 6.2. Activation function (see note 1 for the equation for this function). (A) Units are inhibited at rest (point labeled "R"), so that a change in the net input has relatively little effect on the unit's activity. Top-down input (from the task-demand, or control layer) places the unit near the midpoint of its activation function (point labeled "C"), where a change in the net input (+ = excitatory; – = inhibitory) has a considerably larger impact on the unit's activity. (B) An increase in gain (dotted line) increases the activity of units receiving excitatory input (E), and decreases the activity of units receiving inhibitory input (I), thus increasing the contrast between activated and inhibited units.

to bias the associative units in that pathway, placing them in the sensitive range of their activation function. This serves to modulate the responsivity of those units, making them more sensitive to the inputs.[4] This, in turn, allows the system to respond selectively to one source of information while ignoring another. For example, by activating the color-naming task-demand unit, the model can now respond to the color of the stimulus even when a conflicting word stimulus is present. That is, the model exhibits attention. This attentional effect derives from the ability of the task-demand units to guide the flow of activity along one pathway, while attenuating the flow along another. For this reason, we have come to refer to this as the guided activation theory of cognitive control (Miller & Cohen, 2001).

ATTENTION AND RELATED CONSTRUCTS

Models that implement the guided activation theory provide a quantitative account of attentional effects in a wide range of tasks (e.g., Braver, Barch, & Cohen, 1999; Braver & Cohen, 2001; Cohen, Servan-Schreiber, & McClelland, 1992; Cohen, Romero, Serven-Schreiber, & Farah, 1994; Dehaene & Changeux, 1989, 1991; Mozer, 1988; O'Reilly, Noelle, Braver, & Cohen, 2002; O'Reilly & Munakata, 2000; Phaff, van der Heijden, & Hudson, 1990; Servan-Schreiber, Bruno, Carter, & Cohen, 1998; Schneider & Detweiler, 1988). Equally important, it provides a unifying account of a constellation of processes and constructs related to attention. These are considered in the remainder of this section.

Controlled versus Automatic Processing

This is one of the oldest, and most fundamental constructs in cognitive psychology (Posner & Snyder, 1975; Shiffrin & Schneider, 1977). This distinction is cast largely in terms of the reliance on attention. Controlled processes are defined as those that rely on attention for execution, while automatic processes are defined as those that can be carried out without at-

tention. One of the earliest applications of this construct was to the Stroop task (Posner & Snyder, 1975). Color naming was considered to be controlled because it relies on attention. Without attention to the color, subjects will read the word. Furthermore, the color has no impact on word reading, even when it conflicts with the word being read. Conversely, word reading is automatic because it does not appear to rely on attention. Even when asked to name the color, if a conflicting word is present it slows the response to the color (the classic Stroop effect; MacLeod, 1991; Stroop, 1935). This is thought to reflect the fact that the word is processed even without the allocation of attention. However, there are problems with a simple dichotomous distinction between color naming and word reading in terms of controlled versus automatic processing.

First, it is not clear that any cognitive process can occur entirely independently of attention. For example, although an individual is reading the words on this page, presumably he or she is not doing so out loud. Thus word reading, at least as it is practiced in the Stroop task, is not entirely independent of attention and control. Second, it is not clear that color naming is always dependent on attention and control. In a clever experiment, MacLeod and Dunbar (1988) had participants learn associations between arbitrary shapes (displayed in black and white) and names for them that happened to be color words. At various points during training, they tested their participants' ability to name shapes that were displayed in colors that conflicted with their names. As might be guessed, they found that early in training a shape's color interfered with the ability to provide its name. In other words, in this task, color naming behaved as if it were the automatic process, contradicting the traditional suggestion that it is controlled. Kahneman and Treisman (1984) reviewed a number of other attentional findings, concluding that all processes rely on attention to some degree, and that this may vary in a graded fashion. This is consistent with MacLeod and Dunbar's findings (which also demonstrated that, as subjects became more proficient at shape naming, the color of the shape came to influence this less, while the shape's name came to interfere with color naming). Our model offers a mechanistically explicit account of these findings.

As noted earlier, in the absence of any input from the task-demand units, neither color naming nor word reading can be carried out. This is because the associative units in both pathways rest in too unresponsive a region of their activation function. Thus, even word reading requires attention. At the same time, connections in the word-reading pathway are stronger. Thus, the amount of activity of the task-demand unit needed to support word reading is less than that needed for color naming. In other words, word reading relies less on attention or control than does color naming. It is critical to note, however, that even in the absence of task-demand unit activity *some* information can flow along a pathway. Although this may not be enough to elicit an overt response, it may be enough to influence processing. Thus color naming can be influenced by information in the word pathway and therefore demands more attention when the word conflicts with the color name than when it is congruent. However, this effect is relative. For example, when color naming competes with a *weaker* process, the reverse will be true. This was the situation in MacLeod and Dunbar's (1998) shape-naming experiment, in which the association of a shape with a color word is weaker than a color with a color word. Thus, different processes vary in the degree to which they rely on attention, and this also varies for a given process based on the context in which it is carried out. From this perspective, the distinction between controlled and automatic processing is not dichotomous and absolute but, rather, graded and relative: Some processes are more automatic than others, and processes vary in their automaticity based on the context in which they occur. While this demands quantitative rather than qualitative characteri-

zation of processes, our model offers a framework in which such quantification can be carried out in terms of the connection strengths in the relevant pathways.

Types of Attention

We began the previous section by defining controlled processes as those that rely on attention. But how do we define attention? In our model, attention can be defined as the influence that activity in the task-demand units has on processing in the color and word pathways. Note that this does not rely on any qualitatively distinct mechanisms. Attention arises from the flow of activity between units and over connections that are qualitatively identical to those used to actually perform the task. This suggests that we can define attention, in the most general way, as the modulatory influence that representations of one type have on selecting which (or to what degree) representations of other types are processed, that is, how representations of one type guide the flow of activity among other types. Thus, the representation of an object may influence which sensory features are processed, just as in our model the representation of a task demand influenced which dimensions of the stimulus were processed. The representation of a strong stimulus may even have a "bottom-up" influence on representations of task demands.[5] This idea accords well with a theory of attention that has emerged from the neurophysiological literature—the biased competition theory proposed by Desimone and Duncan (1995). This theory assumes that in the brain, different representations compete for expression, and that the role of attention is to bias this competition in favor of some competitors over others. The source of the bias can be bottom up (e.g., driven by a stimulus) or top down (driven by a higher level "template"). In our model, the top-down flow of activity from the task-demand units literally biased the associative units in each pathway, modulating their responsivity and thereby influencing the competition between information in each pathway and guiding its flow along one relative to the other.

As appealing as the generality of this perspective on attention is, it raises an important issue. On the one hand, the guided activation theory and the biased competition theory emphasize the general nature of attention and the broad range of circumstances in which attentional interactions can occur. This provides a general framework for thinking about attention and, in our case, explicitly modeling attentional interactions. On the other hand, the very breadth of this range invites the following, as yet unanswered question: Are there meaningful distinctions to be made between attentional interactions that occur in different domains, at different levels of processing, or in different "directions" (e.g., top down vs. bottom up). For example, are there systematic differences in the dynamics or scope of modulatory interactions? One answer to this question was implicit in the introduction to this chapter, in the choice to focus on "controlled attention," suggesting that attentional interactions related to "control" exhibit a cohesive set of properties that distinguish them from other types of attentional interactions. This choice is motivated both by functional and neurobiological considerations and bears a close similarity to a highly influential taxonomy proposed by Posner and Petersen (1990; see also Posner, 1980). These authors distinguished between three attentional systems in the brain: an anterior attentional system (housed in the frontal lobes) associated with cognitive control and action selection, a posterior attentional system (housed in the parietal and occiptal lobes) associated with orienting and perceptual attention, and an arousal system (subserved by brainstem neuromodulatory systems) associated with sustained attention and vigilance. Work building on our model of attention and its relationship to neural mechanisms has reached a largely convergent perspective that makes

complementary contributions to our understanding of how attention and cognitive control operate and are implemented in the brain.

Functional Requirements for Cognitive Control

The modeling efforts previously discussed have focused on a specific set of mechanisms that explain how cognitive control gives rise to attentional effects. However, a more general consideration of cognitive control suggests that additional mechanisms are required for its operation. To see this, let us first consider the function of the task-demand units in the Stroop model. So far, we have focused on their attentional effects—that is, their ability to select one source of information for processing over another. However, more generally, they can be viewed as implementing a mapping from a particular set of inputs to a particular set of outputs. For example, the color task-demand unit represents the relationship between color stimuli and their names. From this vantage point, the task-demand units can be seen as carrying out the function of rules, intentions, or goal representations in other theoretical frameworks. All these specify a relationship between existing states (determined by external sensory inputs or internal influences such as memories, emotions, etc.) and desired outcomes that demand particular behaviors in order to be achieved.

Two critical requirements for the representation of task demands, rules, intentions, or goals are that these be actively maintained while the relevant behaviors are performed and then adaptively updated when behavior has achieved the desired outcome or is no longer appropriate: the task is complete, the rule has changed, or the intention or goal has been achieved. This suggests that the apparatus responsible for cognitive control include mechanisms responsible for active maintenance and adaptive updating. Interestingly, these functions are just those for which the prefrontal cortex (PFC) appears to be specialized. This has led us to propose that the PFC subserves the function carried out by the task-demand units in our model: The active and sustained representation of task demands, such as rules and goals—what we have sometimes referred to more broadly as internal representations of context (Cohen, Braver, & O'Reilly, 1996; Cohen & Servan-Schrieber, 1992)—that bias, or guide, the flow of activity along task-relevant pathways, in accord with the guided activation theory. However, the original model lacks critical features, such as the ability to determine on its own which task-demand representation should be active, just how active this should be, for how long, and how this should be updated when a new one is required—that is, it lacks mechanisms for adaptive updating.

Work over the past decade has directly addressed these issues by augmenting the basic model, constrained by neuroscientific data. Although a detailed consideration of these developments is beyond the scope of this chapter, a brief review illustrates how cognitive control can be implemented in a neural system that is self-organized and self-regulated, without recourse to unexplained mechanisms or intelligence (i.e., without the need for a "homunculus").

A NEURAL SYSTEM FOR CONTROLLED ATTENTION

Active Maintenance

The first requirement for a system of control is that it be able to actively maintain representations of task demand, rules, or goals over temporally extended periods (e.g., during performance of the task), in the absence of external support. For example, subjects do not need

to be reminded, trial after trial, to name the color of the Stroop stimulus in a block of trials. Models of sustained activity in PFC have implemented this property using recurrent connectivity (Braver, Cohen, & Servan-Schreiber, 1995; Durstewizt, Seamans, & Sejnowski, 2000; Wang, 1999; Zipser, 1991). This gives rise to attractor dynamics, allowing sets of units with mutually excitatory connections to actively maintain themselves in the absence of input (i.e., as an "attractor state"; Hopfield & Tank, 1986). Recently, biophysically more detailed explorations have suggested that intracellular mechanisms may also contribute to active maintenance, allowing individual units to be "latched" into an on or an off state (Frank, Louhry, & O'Reilly, 2001). However, the importance that this has for models at the systems level remains to be explored.

Adaptive Updating

A second critical property is that the system must be able not only to maintain task representations but also to update these appropriately. This requires that representations in PFC resist perturbation by task-irrelevant inputs (i.e., avoid distraction), while responding to inputs that signal the need for a change (i.e., avoid perseveration). The ability for appropriate updating is central to the flexibility of cognitive control, and disturbances of PFC are known to be associated with distractibility, perseveration, or both. There is growing evidence to support the hypothesis that updating relies on a dopamine-mediated adaptive gating mechanism subserved by the ventral tegmental area (VTA), a dopaminergic nucleus in the brainstem that projects widely to prefrontal areas. In initial work (Braver & Cohen, 2000; O'Reilly, Braver, & Cohen, 1999), we evaluated the plausibility of a simple version of this hypothesis, by implementing a transient gating signal in the task-demand (PFC) layer, that rendered these units temporarily responsive to input from posterior structures (see Figure 6.3). In the absence of this gating signal, representations in the task-demand layer were insensitive to exogenous input, allowing them to maintain the current task demand representation against impinging sources of interference. However, when a gating signal occurred, inputs from other parts of the system could drive activity in the task-demand layer, activating a new representation in that layer. More recently, we have begun to explore the possibility that a more powerful system involving dopaminergic projections to the basal ganglia subserves this gating function (e.g., Frank et al., 2001). This elaboration allows more focused forms of gating that can support hierarchical updating of goals, subgoals, and so on. Nevertheless, the proposal that a dopaminergic, brainstem-mediated gating mechanism regulates the updating of goal representations in PFC raises a critical question: How does this system know when to produce a gating signal, and what new state to produce in the PFC?

The answer to this question comes from important new discoveries regarding the effects of dopamine. Once thought to mediate the hedonic value of a reward, recent work suggests that dopamine release may function as a learning signal, reinforcing associations that provide better predictions of reward (Shultz, Dayan, & Montague, 1997). Importantly, the parameter used to implement this function in computational models (Montague, Dayan, & Sejnowski, 1996) bears a remarkable similarity to the parameter we have used in models to implement a dopamine-based gating signal (Braver & Cohen, 2000). In a series of models, we have illustrated that implementing concurrent effects of the dopamine signal on reinforcement learning and gating allows the system to associate stimuli with the gating signal that predict reward, and thus learn how to update representations in the PFC appropriately (e.g., Braver & Cohen, 2000; Rougier & O'Reilly, 2002). We have used these mechanisms to account for detailed behavioral and neurobiological data regarding the function of the
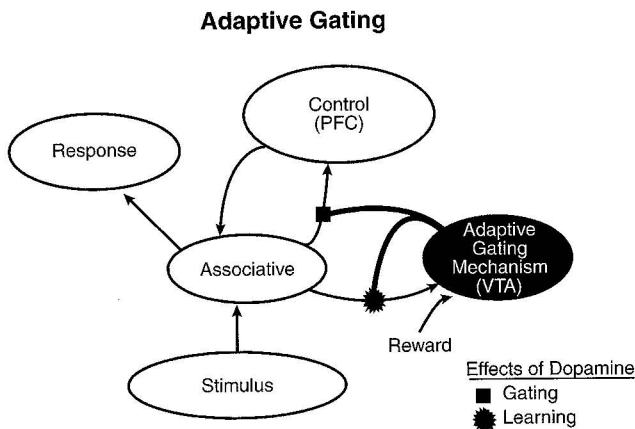
Adaptive Gating

FIGURE 6.3. Model with an adaptive gating mechanism (VTA). Activity of the VTA regulates input to the PFC layer from the associative layer, at the same time training connections from the associative layer back to the VTA. Training occurs according to the temporal differences (predictive Hebbian) learning algorithm (Sutton, 1988). This compares inputs from the associative layer with reward signals, and strengthens connections from the associative layer to the VTA for cues that successfully predict reward. This interacts synergistically with the gating mechanism, because cues that are associated with PFC representations that lead to the procurement of reward thereby predict reward. This strengthens their connections to the VTA, so that in the future they will be more likely to generate a gating signal, activate their associated PFC representation, better predict reward, and therefore receive greater strengthening, and so on.

PFC (O'Reilly et al., 2002; Reynolds & Braver, 2002) in tasks that rely on the flexible deployment of control.

Conflict Monitoring and Regulation of the Degree of Control

The mechanisms described thus far address the ability to represent task-demand information, maintain it, and update it as needed. However, they do not address a different set of questions, which is how the system knows when top-down control is needed in the first place, and just how much is needed to achieve a goal? The need for such mechanisms becomes apparent if we assume that, as in the rest of the brain, representations within the PFC compete for expression. Such competition might provide a partial explanation for the capacity limitations of cognitive control. Such capacity limitations are readily apparent to anyone who has tried to attend to e-mail and talk on the phone at the same time or to talk to a passenger while driving under adverse conditions. These limitations are also fundamental to the distinction between controlled and automatic processing, which assumes that controlled processing relies on a capacity-limited attentional system (e.g., the use of dual-task designs to identify reliance on controlled processing rests on this assumption). Given competition within the PFC, it becomes important for the system to determine whether PFC representation is needed to perform a given task, and if so, how active the representation must be to support adequate performance. Insofar as a weakly active PFC representation leaves room for others to also be active (and thereby other goals to be simultaneously pursued), it is advantageous for the system to titrate the activity of PFC representations to current task needs. In our work with the original Stroop model, we examined the requirements that different processes had for top-down control. As discussed earlier, weaker pathways (e.g., for color

naming) required stronger activation of the corresponding task-demand unit to achieve the same level of performance, and this was especially true when conflicting information was present in competing pathways. Conversely, stronger pathways needed less (but always some) top-down support. Recently we have begun to consider what mechanisms might support the adaptive regulation of task demand activity to meet current task needs.

Our approach to this problem was inspired by the observations about attention with which we began: The primary function of attention is to support the processing of task-appropriate sources of information against competition from interfering sources. Put another way, the role of attention is to reduce *conflicts* in processing. Therefore, the occurrence of conflict provides a natural signal of the need for attentional control We have proposed that monitoring for conflicts in processing is subserved by a specific neural system: the anterior cingulate cortex (ACC). This hypothesis is more fully elaborated in Botvinick, Braver, Yeung, Ullsperger, Carter, and Cohen (Chapter 7, this volume). In brief, the conflict monitoring hypothesis argues that the ACC is responsive to conflict in processing pathways—in particular those that are subject to attentional control by the PFC (i.e., mappings from inputs to outputs)—and that activity in the ACC signals the need to more strongly activate representations in the PFC, in order to better support processing in the task-relevant pathway(s). For example, if the color-naming unit is not sufficiently activated in the Stroop model, then information flowing along this pathway will not compete effectively with interfering information arriving from the word pathway (in the case of an incongruent stimulus). Both response alternatives will become activated, and conflict will ensue. The conflict-monitoring hypothesis asserts that such conflict will engage the ACC, signaling the need to increase activity of the color-naming unit (see Figure 6.4). This hypothesis, and predictions that derive from it, have now received considerable support from behavioral and neuroscientific findings (for a review, see Botvinick et al., Chapter 7, this volume).[6]

The conflict-monitoring hypothesis makes the strong claim that while the ACC plays a role in conflict monitoring and the recruitment of attentional control, it is not responsible for the *allocation* of control. This function is ascribed to the PFC. This assertion contrasts with specific claims about the role of the ACC made by Posner and Petersen (1990) in their original formulation of the anterior attentional system (see also Posner & Dehaene, 1994; Posner & DiGirolamo, 1998). However, it does not violate the more general spirit of their proposal that top-down attentional control is subserved by a frontal system involving the ACC. The primary thrust of the conflict-monitoring hypothesis is to add further specification to one component of this system and suggest a modified set of structure–function relationships. One limitation of the conflict-monitoring hypothesis, however, is that in its present form it does not precisely characterize the mechanisms by which ACC conflict monitoring engages PFC control. The models that have been developed to date implement this as a direct influence of the ACC on the PFC. More recently, we have begun to explore alternative mechanisms that may mediate this influence. One hypothesis is that this involves the locus coeruleus.

The locus coeruleus (LC) is the brainstem neuromodulatory nucleus responsible for most of the norepinephrine (NE) released in the brain (Berridge & Waterhouse, 2003; Foote, Bloom, & Aston-Jones, 1983). It has widespread projections throughout the neocortex. Previous work (Servan-Schreiber, Printz, & Cohen, 1990) has suggested that the effects of NE release can be modeled as a change in the gain (steepness) of the activation function of connectionist units (see Figure 6.2b). This has the effect of augmenting the activity of
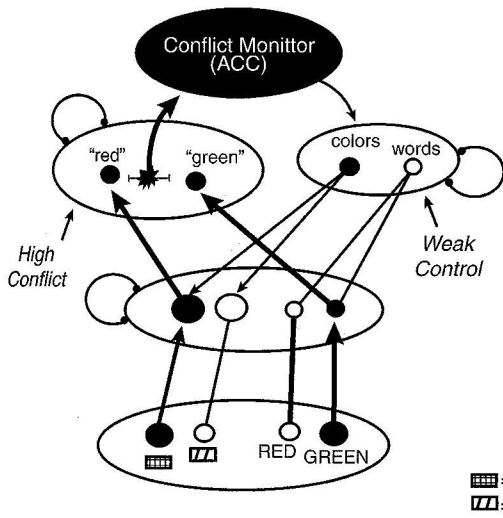
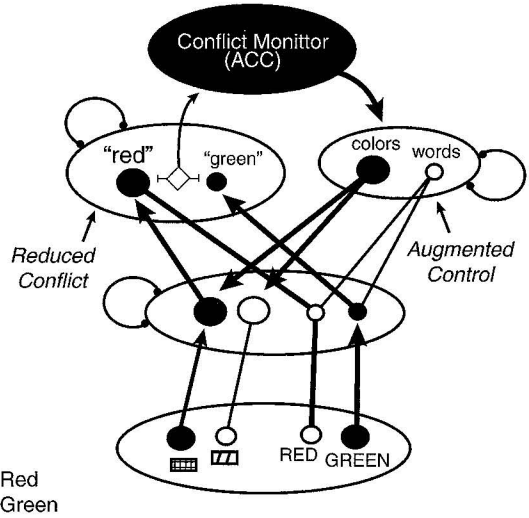## A. Detection of Conflict   B. Recruitment of Control

FIGURE 6.4. Model with conflict-monitoring mechanism (ACC). (A) Weak control allows conflict to develop at the response layer as a result of competing inputs from different pathways. This is detected by the ACC. (B) Detection of conflict within the ACC augments activated units in the PFC layer, producing additional top-down control and a reduction of conflict.

units that are already activated and further suppressing the activity units that are already being inhibited. That is, an increase in gain increases the contrast of the current pattern of activity. This produces precisely the effect within the PFC that is needed in response to conflict. In fact, such contrast enhancement of PFC representations is how adjustments in control have been implemented in our simulations of the conflict-monitoring hypothesis to date (Botvinick, Braver, Carter, Barch, & Cohen, 2001; Yeung, Botvinick, & Cohen, in press). However, in these models, contrast enhancement in the PFC was assumed to be produced directly by ACC activity. While this remains a possibility, we have begun to consider the possibility that this effect is actually mediated by NE release from the LC.

LC-mediated modulation of the PFC is consistent with several lines of evidence, including modeling work that specifies a role for the LC in attentional modulation (e.g., Gilzenrat et al., 2002; Holmes, Nieuwenhuis, & Gilzenrat, & Cohen, 2002; Robertson, Mattingley, Rorden, & Driver, 1998; Usher, Cohen, Servan-Schreiber, Rajkowski, & Aston-Jones, 1999; Yu & Dayan, 2002), as well as recent neuroanatomic evidence suggesting that the ACC is a primary source of cortical projections to the LC (Rajkowski, Lu, Zhu, Cohen, & Aston-Jones, 2000). The appeal of this hypothesis is that it provides a mechanism by which conflict detection within the ACC can augment control without specific knowledge about which particular representations in PFC require augmentation: A global signal can have specific effects. Although details concerning the dynamics of LC-mediated modulation of the PFC are beyond what can be considered here, it is important to note that our hypothesis is that transient (phasic) rather than sustained (tonic) release of NE mediates the modulation of PFC. This forms part of a more general theory about the role of the LC in regulating attention, that we briefly review next.

Exploration versus Exploitation and Regulation of the Focus of Attention and Control

The control mechanisms described previously explain how behavior can be guided by representations of task demands, rules, intentions, or goals in the PFC, adaptively modulating their activity as needed to support task performance. According to the conflict-monitoring hypothesis, when performance degrades and conflict increases, control should increase. However, what about situations in which behavior continues to fall short, despite compensatory adjustments in control? For example, in a Stroop experiment, what if the color stimulus is progressively degraded? The gradual increase in conflict should lead to concomitant increases in control. However, at some point, if the color is degraded beyond recognizability, it makes no sense to further augment control. Rather, control should be withdrawn from this task and some other goal should be pursued. Or, consider the following situation that is perhaps more ecologically valid: An animal is picking berries from a tree. At first berries are everywhere, and the task may not require much effort or attention, but as berries become more scarce, more attention is needed. After some point, however, increasing attention will not help; there are just too few berries left to make the effort worthwhile. At this point, some other behavior should be pursued. These situations suggest that as conflict increases and reward diminishes, at some point the relationship between conflict and control should reverse. This tension between optimizing control to reap the benefits of the current behavioral program and abandoning the current program when it may be more advantageous to sample alternative behavioral programs is well recognized in machine learning, where it is referred to as the trade-off between exploitation and exploration.

We have hypothesized that interactions between the ACC and the LC may regulate the balance in this trade-off (Usher et al., 1999). This builds on observations made by Aston-Jones, Rajkowski, Kubiak, and Alexinsky (1994), indicating that in the awake behaving monkey, the LC shifts between two operating modes that correspond closely with behavioral performance in a simple target detection task. In the "phasic mode," when the animal is performing optimally, the LC shows only moderate levels of tonic discharge, while it exhibits a phasic response selectively to target stimuli but not to distractors. In the "tonic mode," the tonic level of discharge is higher, but there are no phasic responses to target stimuli, reaction time to targets is slower, and the animal commits a greater number of false alarms to distractors.[7] Usher and colleagues (1999) developed a biophysically plausible model of LC function that accounted for transitions between phasic and tonic modes in terms of a single physiological variable (coupling between LC cells) and at the same time explained the impact of these shifts on task performance.[8] In brief, the model suggests that in the phasic mode, the strong LC response to target stimuli facilitates their processing by transiently increasing gain which, at the response layer, has the effect of lowering response threshold. Because this effect is selective to targets, it improves both target detection performance and reaction time. In contrast, in the tonic mode, the absence of a phasic response accounts for the increase in response time to target stimuli, while the increased but indiscriminant release of NE lowers the system's response threshold to other stimuli (e.g., distractors), accounting for the increase in false alarms.

The specificity of the LC phasic response to the target (in the phasic mode) has been demonstrated in a number of additional studies, including reversal experiments in which LC has been shown to reliably acquire the new target before the animal's overt behavior has done so (e.g., Aston-Jones, Rajkowski, & Kubiak, 1997). These characteristics suggest that

in the phasic mode, the LC functions as an attentional filter that selects for the occurrence (i.e., *timing*) of task-relevant stimuli, much as cortical attentional systems filter the *content* of a stimulus. The existence of such a temporal filter is consistent with several recent psychophysical studies (e.g., Coull & Nobre, 1998). At the same time, by increasing the gain of cortical representations, the LC phasic response can also enhance the effects of cortical selection by content (e.g., the top-down effects of the PFC). Together, these effects allow the LC phasic response to selectively facilitate responses to task-relevant stimuli when they occur. One immediate question, however, is, What is the adaptive value of the tonic mode, which seems to deteriorate performance? We have argued that the tonic mode may support exploratory behavior. By reducing phasic responses to target stimuli, and increasing tonic NE release, the system is more effectively driven by task-irrelevant stimuli. Such responsiveness is not adaptive with respect to the specified experimental task, because it permits the processing of task-irrelevant stimuli and the sampling of other behavioral programs. However, it may be highly adaptive if either the current task is no longer remunerative, or if the environment changes and more valuable opportunities for reward or new behavioral imperatives have appeared. From this perspective, a shift from the LC phasic to the tonic mode may shift the behavioral strategy from exploitation (when this is no longer adaptive) to exploration (when a new goal should be sought). Loosely speaking, this "throws the ball up in the air, so another team can take it." Viewed from the perspective of attentional control, the LC phasic mode supports the current control state (exploitation), while the LC tonic mode provokes a withdrawal of control from the current task, favoring the sampling of other behavioral goals (exploration), which raises one more important question: What information can the system use to determine whether it should exploit (LC phasic mode) or explore (LC tonic mode)? The answer to this question closes the loop, both in the control system and in our discussion.

We hypothesize that ACC conflict monitoring can provide the necessary information. This hypothesis requires only one additional assumption: that the ACC is able to integrate conflict over two time frames—a short one (seconds) and a long one (minutes). Consider the following two circumstances. In one, performance is good and there are still rewards to be accrued from the current task, but there are occasional lapses in performance producing transient increases in conflict (e.g., on single trials). Under these conditions, control should be increased each time there is a lapse, to restore performance. That is, control should be increased when long-term conflict is low but there is a momentary increase in short-term conflict. In contrast, consider a different situation in which performance has been poor and conflict has been *persistently* high. At some point, this situation should encourage the withdrawal of control, irrespective of short-term changes in conflict. A similar situation should arise when, irrespective of performance, opportunities for reward diminish. A relatively simple equation can capture these relationships,[9] which indexes the need for control as a function of short-term conflict and reward, discounted by the accumulation of long-term conflict and dimunition of reward (see Figure 6.5). This computation can be used to drive shifts between LC phasic and tonic modes by influencing simple physiological parameters (Brown et al., 2004; Usher et al., 1999). Taken together, these mechanisms would constitute a self-regulating system that is responsive to demands for control over different time scales and is sensitive to the current value of exploration versus exploitation (Figure 6.6). These mechanisms are consistent with known properties of the LC, ACC, and their anatomic connectivity; however, their validation presents a challenge to further neurophysiological investigation.[10]
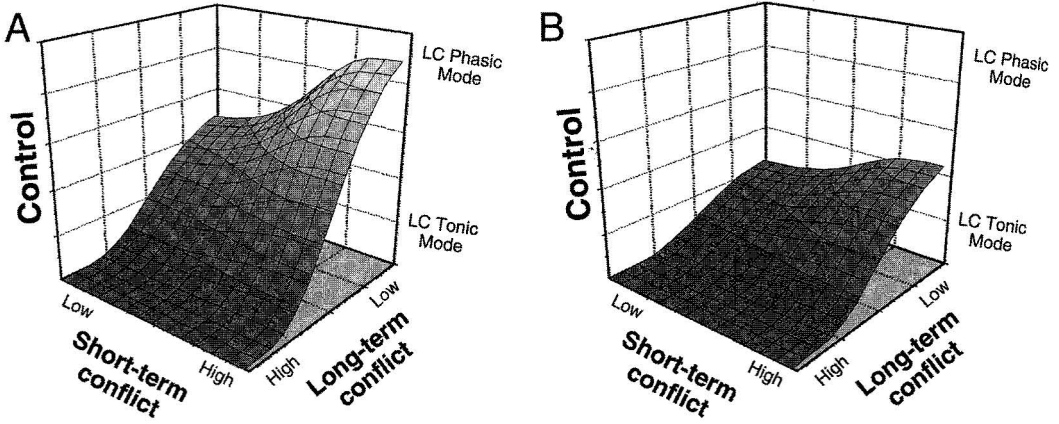
FIGURE 6.5. Relationship of control to conflict and reward (see note 9 for the equation defining this relationship). (A) Under conditions of high reward and low long-term conflict, transient increases in conflict (short-term conflict) produce greater control, by driving LC into the phasic mode. This effect is diminished as long-term conflict increases, driving LC into tonic mode. (B) Low reward damps the effect of short-term conflict, driving the LC into the tonic mode irrespective of conflict.
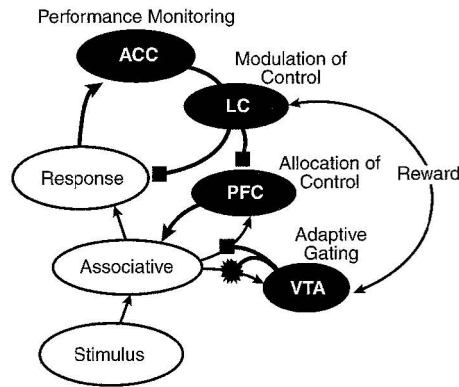
Neural System for Adaptive Regulation of Control

FIGURE 6.6. Integrated neural system for the adaptive regulation of control. This system is able to adaptively update control representations in PFC, learn how to do so, and modulate the strength of top-down control in response to prevailing balance between reward and short-term and long-term degradations in performance (as indexed by conflict).

CONCLUSIONS AND FUTURE CHALLENGES

This chapter has reviewed recent theoretical work, pursued within the connectionist framework, that addresses the nature of controlled attention and its neural implementation. The goal of this work is to provide a mechanistically explicit account of how processes that give rise to the phenomena of attention and control are implemented in the brain. The hypotheses reviewed suggest that control relies on the activation of appropriate representations in the PFC. These representations can be thought of as task demands, rules, intentions, or goals that direct behavior to produce desired outcomes by biasing processing and guiding the flow of activity along pathways responsible for mapping inputs to desired outputs. As summarized in Figure 6.6, PFC representations are regulated by several mechanisms, including a dopamine-mediated system for updating PFC representations in specific task contexts, and learning how to do so; an ACC-mediated system for assessing the demand for control; and an LC-mediated system for modulating PFC representations in response to these demands. These hypotheses define a mechanistically explicit, self-regulating system of control that is responsive to adaptive needs at different time scales and of fundamentally different types (e.g., exploration vs. exploitation). One important feature of these hypotheses is their suggestion that brainstem neuromodulatory systems—once thought to be responsible for the regulation of nonspecific aspects of psychological function, such as motivation (dopamine) and arousal (NE)—may play a significantly more central and specific role in information processing.

Of course, these mechanisms represent only a first step toward a more complete understanding of the neural mechanisms that underlie attention and control. First, they address only controlled attention, and not the many other forms and levels of attentional effects. However, even within the scope of controlled attention, many challenges remain. For example, it is possible to register a goal or intention for the future and then to dispatch this for pursuit sometime (hours, days or even years) in the future. These forms of control cannot be explained by active maintenance of representations in PFC alone, but they are likely to involve interactions between the PFC and medial temporal lobe structures that subserve episodic memory (Cohen & O'Reilly, 1996). There are also critical interactions between control and motivation, most likely involving interactions between the PFC and limbic structures. Perhaps the most perplexing puzzle that remains concerns the nature of representations in the PFC. Our models to date have stipulated the presence of representations in the PFC required to perform a given task. However, it seems unlikely that without infinite capacity, the PFC can house all the possible representations needed to meet the arbitrarily large set of potential task demands. A characteristic feature of human behavior is the flexibility of control, manifest as the ability to perform novel tasks, or to creatively structure new forms of behavior in a novel task environment. How then, with a large but finite set of resources, can the system exhibit the flexibility we witness in our everyday behavior? This question is closely related to an equally important one: How do representations develop in PFC?

Recent work has begun to address many of the questions raised here (for reviews, see Miller & Cohen, 2001; O'Reilly et al., 1999, 2002; Rougier, Noelle, Braver, Cohen, & O'Reilly, 2004). Nevertheless, the human ability to flexibly deploy attention and control, navigating the vast repositories of information available both from the environment and the system's stored knowledge, to respond appropriately under familiar circumstances and creatively under unfamiliar ones, remains one of the most fundamental and interesting myster-

ies of science. We hope to have illustrated in this chapter that this mystery need not remain intractable to theoretical analysis.

ACKNOWLEDGMENTS

NOTES

1. This reflects the assumption that units are negatively biased. This can be seen from the expression for the activation function: activity $= 1/(1 + e^{-(\text{net input})})$. For a net input of 0, this evaluates to an activity of 0.5. However, by assuming that a constant negative bias value is added to the net input (e.g., –4 as was used in Cohen et al., 1990), the unit will have a low activity value for a zero net input.

2. Task-demand units have, in different papers—both our own and those of others—variously been referred to as attention units, context units (designating the "task context"), mapping units, and goal units, which reflects the various perspectives that can be taken on their function (to which we return at the end of the chapter). It also highlights the perils of pursuing scientific research using natural language and the value of explicit, formal models. Ultimately, our theory of attention and its relationship to cognitive control is expressed in the form of the model and *not* in the words that are used to describe it. Thus, where any ambiguity or disagreement may arise about what is claimed by our theory, the model itself (and whatever elaboration of it is required to address the issue) is the final arbiter and not any words used to describe it.

3. The assumption of nonlinearity is justified on both theoretical and empirical grounds. Theoretically, it has been shown that using nonlinear rather than linear units confers considerably greater computational power on the system (e.g., Rumelhart, Hinton, & Williams, 1986). Empirically, it is clear that neurons (whether considered individually or in populations) have finite upper and lower bounds on their activity levels.

4. Two points are worth noting about this effect. First, the term "modulation" often implies a multiplicative effect. Here, however, modulation is produced by adding (biasing) activity from the task-demand unit to the associative units. Nevertheless, because this addition occurs to the net input, which appears in the exponent of the activation function, the effect is in fact multiplicative. Second, note that the effect of this modulation is to place units in the linear range of their activation function, which provides a justification for mathematical models that assume linear response properties (e.g., Brown & Holmes, 2001), by suggesting that those units primarily responsible for processing (i.e., those that are in the "focus of attention") reside in the linear range of an otherwise nonlinear response function.

5. This assumes that there are bottom-up connections to the task-demand units in the model. Although these were not included in the original Cohen et al. (1990) model, variants have included such connections (e.g., Cohen & Huston, 1994; O'Reilly & Munakata, 2000).

6. While this hypothesis proposes that conflict monitoring is subserved by the ACC, it does not claim that this is the *only* function of this structure. Rather, we view conflict monitoring as one of a family of functions subserved by the ACC, that monitor internal processing states for breakdowns in performance, much as the amygdala is thought to monitor the environment for external signs of threat.

7. At present, it is not clear whether these represent dichotomous modes, or ends of a continuum of states that the LC can occupy.
8. Recently, Brown and colleagues (2004) have proposed that changes in baseline firing rate may also serve to drive transitions between LC phasic and tonic modes. Which of these mechanisms (electrotonic coupling, baseline firing rate, or both) is actually operative in the LC remains an area of inquiry. However, what is relevant for present purposes is that LC mode can be determined by one or two easily regulated physiological parameters.
9. Control = reward / [f(conflict$_{long-term}$)*(1 – f(conflict$_{short-term}$))], where f(conflict) = $1/(1 + e^{-conflict})$
10. The LC receives extensive projections from both the ACC and the orbitofrontal cortex (Aston-Jones et al., 2002; Rajkowski et al., 2000), which may provide evaluative information regarding both performance and rewards (e.g., Bush et al., 2002; Rolls, 2000). We should also note that the LC has extensive projections throughout the brain (except the hypothalamus and striatum, which it does not innervate). We propose that while NE release in PFC directly modulates control representations, simultaneous release of NE in areas outside the PFC serves to reinforce this effect in other sensory, motor, and associative areas.

REFERENCES

Aston-Jones, G., Rajkowski, J., & Kubiak, P. (1997). Conditioned responses of monkey locus coeruleus neurons anticipate acquisition of discriminative behavior in a vigilance task. *Neuroscience, 80,* 697–715.

Aston-Jones, G., Rajkowski, J., Kubiak, P., & Alexinsky, T. (1994). Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. *Journal of Neuroscience, 14*(7), 4467–4480.

Aston-Jones, G., Rajkowski, J., Lu, W., Zhu, Y., Cohen, J. D., & Morecraft, R.J. (2002). Prominent projections from the orbital prefrontal cortex to the locus coeruleus in monkey. *Society for Neuroscience Abstracts, 28,* 86–89.

Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus-noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Research: Brain Research Review, 42,* 33–84.

Botvinick, M. M., Braver, T. S., Carter, C. S., Barch, D. M., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624–652.

Braver, T. S., Barch, D. M., & Cohen, J. D. (1999). Cognition and control in schizophrenia: A computational model of dopamine and prefrontal function. *Biological Psychiatry, 46*(3), 312–328.

Braver, T. S., & Cohen, J. D. (2000). 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: Control of cognitive processes* (pp. 713–737). Cambridge, MA: MIT Press.

Braver, T. S., & Cohen, J. D. (2001). Working memory, cognitive control, and the prefrontal cortex: Computational and empirical studies. *Cognitive Processing, 2,* 25–55.

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

Brown, E., & Holmes, P. (2001). Modeling a simple choice task: Stochastic dynamics of mutually inhibitory neural groups. *Stochastics and Dynamics, 1*(2), 159–191.

Brown, E., Moehlis, J., Holmes, P., Clayton, E., Rajkowski, J., & Aston-Jones, G. (2004). *The influence of spike rate and stimulus duration on noradrenergic neurons.* Manuscript submitted for publication.

Bush, G., Vogt, B. A., Holmes, J., Dale, A. M., Greve, D., Jenike, M. A., et al. (2002). Dorsal anterior cingulate cortex: a role in reward-based decision making. *Proceedings of the National Academy Science USA, 99*(1), 523–528.

Cohen, J. D., Braver, T. S., & O'Reilly, R. C. (1996). A computational approach to prefrontal cortex, cognitive control, and schizophrenia: Recent developments and current challenges. *Philosophical Transactions of the Royal Society of London B, (Biological Sciences), 351,* 1515–1527.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing model of the Stroop effect. *Psychological Review, 97*(3), 332–361.

Cohen, J. D., & Huston, T. A. (1994). Progress in the use of parallel distributed processing models for understanding attention and performance. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 453–476). Cambridge, MA: MIT Press.

Cohen, J. D., & O'Reilly, R. C. (1996). A preliminary theory of the interactions between prefrontal cortex and hippocampus that contribute to planning and prospective memory. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications* (pp. 267–295). Hillsdale, NJ: Erlbaum.

Cohen, J. D., Romero, R. D., Servan-Schreiber, D., & Farah, M. J. (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., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *American Journal of Psychology, 105*, 239–269.

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience, 18*, 7426–7435.

Dehaene, S., & Changeux, J. P. (1989). A simple model of prefrontal cortex function in delayed-response tasks. *Journal of Cognitive Neuroscience, 1*(3), 244–261.

Dehaene, S., & Changeux, J. P. (1991). The Wisconsin card sorting test: Theoretical analysis and modeling in a neuronal network. *Cerebral Cortex, 1*(1), 62–79.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193–222.

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *Journal of Neurophysiology, 83*(3), 1733–1750.

Foote, S. L., Bloom, F. E., & Aston-Jones, G. (1983). Nucleus locus ceruleus: New evidence of anatomical and physiological specificity. *Physiology Review, 63*, 844–914.

Frank, M. J., Loughry, B., & O'Reilly, C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, and Behavioral Neuroscience, 1*(2), 137–160.

Gilzenrat, M. S., Holmes, B. D., Holmes, P. J., Rajkowski, J., Aston-Jones, G., & Cohen, J. D. (2002). A modified Fitzhugh-Nagumo system simulates locus coeruleus-mediated regulation of cognitive performance. *Neural Networks, 15*, 647–663.

Holmes, B. D., Nieuwenhuis, S., Gilzenrat, M. S., & Cohen, J. D. (2002). *The role of locus coeruleus in mediating the attentional blink: A neurocomputational model*. Poster presented at the annual meeting of the Psychonomic Society, Kansas City, MO.

Hopfield, J. J., & Tank, D. W. (1986). Computing with neural circuits: A model. *Science, 233*, 625–633.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuranam & D. R. Davies (Eds.), *Varieties of attention* (pp. 29–59). New York: Academic Press.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163–203.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*(1), 126–135.

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: Synergies*

in experimental psychology, artificial intelligence, and cognitive neuroscience (pp. 655–688). Cambridge, MA, MIT Press.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24,* 167–202.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience, 16*(5), 1936–1947.

Mozer, M. (1988). A connectionist model of selective attention in visual perception. In *Proceedings of the tenth annual conference of the Cognitive Science Society* (pp. 195–201). Hillsdale, NJ: Erlbaum.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically-based neural network model of working memory. In P. Shah & A. Miyake (Eds.), *Models of working memory* (pp. 375–411). Cambridge, UK: Cambridge University Press.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain.* Cambridge, MA: MIT Press.

O'Reilly, R. C., Noelle, D. C., Braver, T. S., & Cohen, J. D. (2002). Prefrontal cortex in dynamic categorization tasks: Representational organization and neuromodulatory control. *Cerebral Cortex, 12,* 246–257.

Phaf, R. H., Van der Heijden, A. H. C., & Hudson, P. T. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22*(3), 273–341.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3–25.

Posner, M. I., & Dehaene, S. (1994). Attentional networks. *Trends in Neuroscience, 17,* 75–79.

Posner, M. I., & DiGirolamo, G. J. (1998). Executive Attention: Conflict, target detection and cognitive control. In R. Parasuraman (Ed.), *The attentive brain* (pp. 401–423). Cambridge: MIT Press.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 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, NJ: Erlbaum.

Rajkowski, J., Lu, W., Zhu, Y., Cohen, J., & Aston-Jones, G. (2000). Prominent projections from the anterior cingulate cortex to the locus coeruleus (LC) in rhesus monkey. *Society of Neuroscience Abstracts, 26,* 838.15.

Reynolds, J. R., & Braver, T. S. (2002, April). *Computational and neural mechanisms of task-switching.* Paper presented at the ninth annual meeting of the Cognitive Neuroscience Society, San Francisco.

Robertson, I. H., Mattingley, J. B., Rorden, C., & Driver, J. (1998). Phasic alerting of neglect patients overcomes their spatial deficit in visual awareness. *Nature, 395,* 169–172.

Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex, 10*(3), 284–294.

Rougier, N. P., Noelle, D. C., Braver, T. S., Cohen, J. D., & O'Reilly, R. C. (2004). *Prefrontal cortex and the flexibility of cognitive control: Rules without symbols.* Manuscript in submission.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* Cambridge, MA: MIT Press.

Schneider, W., & Detweiler, M. (1988). The role of practice in dual-task performance: toward workload modeling in a connectionist/control architecture. *Human Factors, 30*(5), 539–566.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275,* 1593–1599.

Servan-Schreiber, D., Bruno, R., Carter, C., & Cohen, J. D. (1998). Dopamine and the mechanisms of cognition. Part I: A neural network model predicting dopamine effects on selective attention. *Biological Psychiatry, 43,* 713–722.

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.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning automaticity, attending and a general theory. *Psychological Review, 84*(2), 127–190.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Sutton, R. (1988). Learning to predict by the method of temporal difference. *Machine Learning, 3, 9.*

Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowski, J., & Aston-Jones, G. (1999). The role of locus coeruleus in the regulation of cognitive performance. *Science, 283*, 549–554.

Wang, X. J. (1999). Synaptic basis of cortical persistent activity: The importance of NMDA receptors to working memory. *Journal of Neuroscience, 19*, 9587–9603.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (in press). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychologial Review.*

Yu, A. J., & Dayan, P. (2002). Acetylcholie in cortical inference. *Neural Networks, 15*, 719–730.

Zipser, D. (1991). Recurrent network model of the neural mechanism of short-term active memory. *Neural Computation, 3*(2), 179–193.