The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition

MARCEL ADAM JUST AND SASHANK VARMA
Center for Cognitive Brain Imaging, Carnegie Mellon University, Pittsburgh, Pennsylvania

Recent findings in brain imaging, particularly in fMRI, are beginning to reveal some of the fundamental properties of the organization of the cortical systems that underpin complex cognition. We propose an emerging set of operating principles that govern this organization, characterizing the system as a set of collaborating cortical centers that operate as a large-scale cortical network. Two of the network's critical features are that it is resource constrained and dynamically configured, with resource constraints and demands dynamically shaping the network topology. The operating principles are embodied in a cognitive neuroarchitecture, 4CAPS, consisting of a number of interacting computational centers that correspond to activating cortical areas. Each 4CAPS center is a hybrid production system, possessing both symbolic and connectionist attributes. We describe 4CAPS models of sentence comprehension, spatial problem solving, and complex multitasking and compare the accounts of these models with brain activation and behavioral results. Finally, we compare 4CAPS with other proposed neuroarchitectures.

The relationship of the mind to the brain has established itself at the forefront of scientific interest in large part because of the rapid development of functional magnetic resonance imaging (fMRI). Exciting new findings have often been expressed in the form of compelling images that indicate that brain areas become activated in various tasks or fail to activate in various special populations. These brain images appear to tell a simple and straightforward story. The invited inference is that a one-to-one mapping exists between the activation of brain areas and the execution of certain psychological processes, such as the process of perceiving a face. At first glance, this story imposes some order on the results; some consistencies emerge across brain imaging studies, and some imaging results seem to be related to studies of brain lesions in neuropsychological patients. But the consistencies are limited in scope, and the mapping between the infinity of possible tasks and the finite—indeed, rather small—number of activating brain areas identified by neuroimaging techniques is greatly underdetermined, if not logically troubling. We propose that the idea of a one-to-one mapping of cortical activation to high-level cognitive processes that is suggested by the brain activation images is incorrect—a gross oversimplification of a more complex (and more interesting) many-to-many mapping, governed by more subtle organizational principles. In brief, we argue that thinking is a network phenomenon and suggest the beginnings of a theory of the organization of cortical networks.

This article explores how the various cortical areas that subserve cognition might function in conjunction with each other. The main objectives are to specify the operating principles that govern the complex, dynamic, and adaptive relationships among brain areas and to relate brain function to cognitive function. It is not our goal here to specify the cognitive specializations of each brain area, although to describe how areas work together, we necessarily have to hypothesize the cognitive endowments of some brain areas. We believe there is value in developing a theory of how various brain areas collaborate in order to realize cognitive information processing, despite the remaining uncertainty about the functional specializations of the individual areas.

As functional imaging provides progressively finer detail about brain activation, computational modeling provides a theory-building workspace in which the new pieces of information can be put together and their cofunctioning can be examined. In this workspace, the component mechanisms can be specified in detail and their ability to account for the observed phenomena can be tested, as a few initial attempts have shown (Arbib, Billard, Iacoboni, & Oztop, 2000; Fincham, Carter, van Veen, Stenger, & Anderson, 2002; Horwitz & Tagamets, 1999; Just, Carpenter, & Varma, 1999). One of the goals of this study is to instantiate the operating principles in a cognitive neuroarchitecture, 4CAPS, that has been developed expressly for this purpose. 4CAPS is a computational system for modeling cognitive performance and brain activity in a range of cognitive tasks, permitting comparisons of model accounts and empirical findings.

A central theme of our proposal is that resource constraints shape cognition. This theme has recurred in the scientific study of human thinking, from Miller's (1956) "magical number seven" article on the limitations of

153

short-term memory, to contemporary research on how working memory limitations bound cognition (Miyake & Shah, 1999). We propose that resource constraints apply not just at the level of cognitive information processing, but also at the level of cortical information processing. This is a simple and inescapable consequence of the fact that the brain, like all biological systems, is subject to hard constraints on bioenergetic and structural resources. This simple consequence, however, has profound implications: If cognition is the emergent product of computation in a network of collaborating brain areas, and if computation within each brain area and communication between brain areas consume resources, then resource constraints must shape cortical information processing, and thus cognitive information processing as well.

It is common to view resource constraints negatively, as a restriction on cognition. We offer a more nuanced, and in many ways more optimistic, perspective here. Specifically, we show that resource constraints, working in conjunction with other principles of brain organization, are fundamental to the dynamic and adaptive nature of thinking at both the cortical and cognitive levels. For example, when a task is sufficiently difficult, resource demands on the canonical large-scale network for performing the task will exceed resource supplies, and additional brain areas with spare resources and relevant functional specializations will be recruited to absorb the excess workload (Just, Carpenter, Keller, Eddy, & Thulborn, 1996). In this case, recruitment is on a just-in-time and as-needed basis. When a focal lesion drastically reduces the resources available in the canonical network, additional areas are similarly recruited on a more lasting basis. As we show below, resource constraints provide a unified account of these and other common patterns of brain activation observed in neuroimaging studies. The silver lining of resource constraints will be seen to be resource optimization—the efficient deployment of finite resources to maximize throughput—and the remarkable adaptiveness of human information processing.

The remainder of this article is composed of the following sections: first, a set of operating principles that govern the resource-constrained nature of information processing in the brain, motivated by functional neuroimaging findings; next, the instantiation of these operating principles in the 4CAPS cognitive neuroarchitecture; then a series of computational models that account for the behavioral and brain activation measures expressed within 4CAPS of several complex cognitive tasks—sentence comprehension, problem solving, spatial reasoning, and dual-tasking—in order to establish the empirical content of the operating principles and neuroarchitecture; next, a comparison of 4CAPS with other cognitive neuroarchitectures, identifying points of overlap as well as its unique claims; and to conclude, a discussion of future directions.

I. OPERATING PRINCIPLES FOR CORTICAL INFORMATION PROCESSING

One of the key challenges of cognitive neuroscience is to understand how cortical areas work together to accomplish complex cognition. Adding to this challenge is the fact that the form of the "working together" is constantly changing, as the system adaptively configures and reconfigures itself in light of changing processing needs and inherent limitations on available computational resources. We propose that the following six operating principles govern this adaptive configuration capability: An initial principle captures the current consensus in the field.

> 0. Thinking is the product of the concurrent activity of multiple brain areas that collaborate in a large-scale cortical network.

The next four principles, which constitute the theoretical core of our proposal, are relatively novel.

> 1. Each cortical area can perform multiple cognitive functions, and conversely, many cognitive functions can be performed by more than one area.
> 2. Each cortical area has a limited capacity of computational resources, constraining its activity.
> 3. The topology of a large-scale cortical network changes dynamically during cognition, adapting itself to the resource limitations of different cortical areas and to the functional demands of the task at hand.
> 4. The communications infrastructure that supports collaborative processing is also subject to resource constraints, construed here as bandwidth limitations.

Finally, we propose a measurement assumption that enables our theoretical constructs to make contact with neuroimaging data.

> 5. The activation of a cortical area, as measured by imaging techniques such as fMRI and PET, varies as a function of its cognitive workload.

In the remainder of this section, we describe the operating principles in greater detail and present examples of their operation.

0. Thinking is the product of the concurrent activity of multiple brain areas that collaborate in a large-scale cortical network. The fact that multiple areas are involved in performing any task has been confirmed by thousands of neuroimaging studies. Part of the remaining scientific challenge is to provide an account of why a *particular* set of cortical areas comes to be activated in a given task. Another part of the challenge is to characterize how the multiple areas do more than simply coactivate: They actively work together. For example, the functional connectivity between two areas, defined as the correlation or covariance between the time series of the activation of activated voxels in each, is commonly interpreted as indexing the degree of coordination between them (Friston, 1994; Horwitz, Rumsey, & Donohue, 1998). The degree of functional connectivity between two activated areas can be modulated by a number of variables, such as an increase in working memory load (Diwadkar, Carpenter, & Just, 2000), an increase in task complexity (e.g., from processing words to processing sentences; Homae, Yahata, & Sakai, 2003), or a decrease in communication band-

width due to neurological conditions such as autism (Just, Cherkassky, Keller, & Minshew, 2004). Of course, like any correlation, functional connectivity does not indicate causality, leaving open the question of whether control is symmetric, asymmetric, or coordinated by a third area. The broad conclusion, however, is that complex cognition is a product of networks that span multiple cortical areas working together in close collaboration.

1. Each cortical area can perform multiple cognitive functions, and conversely, many cognitive functions can be performed by more than one area. This operating principle has two aspects. The first is that each area performs multiple functions. This multiplicity of functional specializations is well established in the neuropsychological literature—for instance, see Mesulam's (1990) reference to Broca's area as both the syntactic pole and the articulatory pole of processing. More recently, reviews of functional neuroimaging studies have shown that, over and over again, a given cortical area is activated in a number of different types of tasks (Cabeza & Nyberg, 1997, 2000). In particular, Duncan and Miller (2002) have proposed a multifunction view of prefrontal cortex, based in part on the findings of functional imaging studies. Cortical assignment for more than one function holds at even smaller spatial scales than typical imaging studies have afforded. At the spatial resolution allowed by BOLD fMRI in typical cognitive studies, individual voxels (volume elements typically of 30–50 mm^3) are usually sensitive to multiple task variables. For example, in sentence comprehension, a fair proportion of the activating voxels in Wernicke's area (left posterior superior temporal gyrus [STG]) respond to both lexical and syntactic manipulations (Keller, Carpenter, & Just, 2001).[1] The theory of multiple specializations is also compatible with seminal single-unit studies that have demonstrated that individual neurons are active during more than one function (Georgopoulos, Lurito, Petrides, Schwartz, & Massey, 1989; Nicolelis, 1997; Rao, Rainer, & Miller, 1997; Sanes & Donoghue, 1997). Sensory events and motor actions are coded as patterns of neural activity, akin to waves of activation, and therefore individual neurons can participate in multiple codes.

The second, and more novel, aspect of this principle is that some functions are implemented by more than one area. We propose that there is some redundancy of function across cortical areas, such that some functions can be performed by more than one brain area, although the nature of the implementations themselves can (and likely does) vary. This redundancy is revealed when focal damage to one cortical area results in the recruitment of another area to perform a similar function. For example, damage to areas involved in language processing, such as Broca's area, sometimes leads to the activation of homologous right-hemisphere areas soon after stroke (Thulborn, Carpenter, & Just, 1999). Such redundancy may generally be the basis of certain kinds of cortical plasticity. Redundancy of function is also observed in healthy brains, especially when a difficult task overwhelms the resource supplies (i.e., computational capabilities) of the canonical brain areas for the task. For example, when normal young adults process sentences that are syntactically more complex, there is increasing activation of the right homologue of Broca's area (Just, Carpenter, Keller, et al., 1996). These types of findings suggest an overlap of language functions between Broca's area and its right homologue, with Broca's area being primary (i.e., more efficient) for some tasks, such as syntactic analysis, and its right homologue being primary for others, such as prosodic processing.

2. Each cortical area has a limited capacity of computational resources, constraining its activity. This operating principle is also relatively novel. Thinking is biological work, with an upper limit on resource availability. The finite resource capacities of various brain systems not only slow and degrade performance on difficult tasks, but also fundamentally shape cortical and cognitive information processing, just as the shape of a riverbed constrains the water flow of a river. Resource constraints force the system to dynamically configure and reconfigure itself to adapt to limitations. In this role, resource constraints act not just as a stifling restriction on thinking, but also as a function that forces much of the adaptivity of human cognition, a theme that will emerge throughout this article.

One imaging study that addressed capacity constraints by attempting to breach the cognitive system's limit (Callicott et al., 1999) used the *N*-back task to examine how increases in the number of items to be encoded, retained, and recalled affected the slope of the activation in several cortical areas. Prefrontal areas (which are associated with the working memory functioning required by this task) showed a monotonic increase in activation from $N = 0$, to $N = 1$, to $N = 2$, as shown in Figure 1. (There was also a decrease from $N = 2$ to $N = 3$ in this area, as most participants could no longer accurately perform the task and essentially gave up. Other areas showed different patterns, because their workloads did not necessarily increase in this way.) This is a familiar pattern of increasing activation in response to an increase in cognitive workload within some range, with acceptable performance accuracy.

3. The topology of a large-scale cortical network changes dynamically during cognition, adapting itself to the resource limitations of different cortical areas and to the functional demands of the task at hand. This novel operating principle is central to our proposal. The composition of a large-scale cortical network and its pattern of inter-area connectivity are not static but can vary dynamically during task performance. This proposal stands in contrast to the widespread conception that each cognitive task is performed by a fixed set of brain areas—a fixed neural substrate. According to the dynamic alternative we advocate, the "underlying neural substrate" is a moving target, changing not only from trial to trial (i.e., from stimulus item to stimulus item), but also from moment to moment within a trial.

This dynamic configuration and reconfiguration of a large-scale network is a response to the fluctuating availability of resources within brain areas and the changing functional demands of a task. The membership and connectivity of a network—its *topology*—first of all can change as the functional demands of a task increase *quan-*
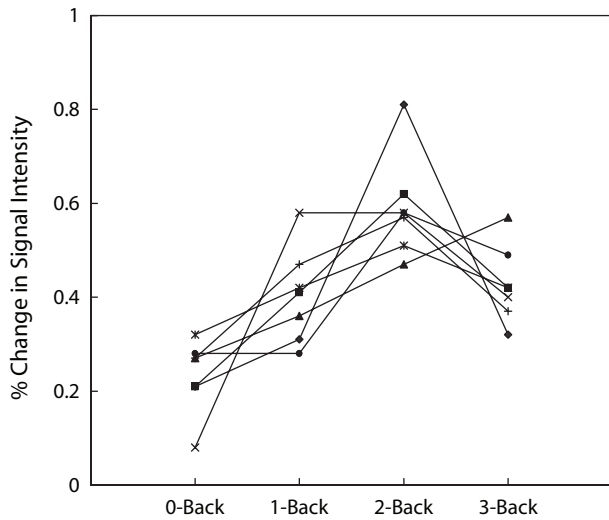
Figure 1. Capacity constraints in a single area (left dorsolateral prefrontal cortex) in an *N*-back task, in which activation increases with workload up to a capacity limit (from Callicott et al., 1999). The seven curves represent the data from 7 subjects.

titatively, so that additional areas become active and are recruited into the network. For example, when a sentence comprehension task is made more difficult by progressively increasing the structural complexity of the sentence, activation in the right-hemisphere homologue of Wernicke's area systematically increases from a negligible to a substantial level (Just, Carpenter, Keller, et al., 1996), as shown in Figure 2. We interpret this example of dynamic network reconfiguration as follows: When the resources of an area well-specialized for performing task-relevant functions (e.g., Wernicke's area) are exhausted, additional areas that are less specialized for the same functions (e.g., its right homologue) are recruited into the network to absorb the excess workload. In addition, the topology of a network can change when the functional demands of performing a task change *qualitatively*, as when an increase in difficulty requires strategic planning. For example, if a sentence comprehension task requires considerable problem solving in addition to language processing, it will dynamically activate and recruit left dorsolateral prefrontal cortex (DLPFC), an area associated with executive function and working memory, into the language network (Newman, Just, & Carpenter, 2002).

The dynamic entry of new cortical areas into a large-scale network is incremental and continuous, not all-or-none. For example, in the Newman et al. (2002) study, left DLPFC activation in the difficult comprehension task occurred earlier or later, depending on where in the sentence the maximal reasoning load occurred. In this regard, dynamic cortical reconfiguration provides just-in-time and as-needed neural support for the changing demands of task performance.

4. The communications infrastructure that supports collaborative processing is also subject to resource constraints, construed here as bandwidth limitations. This operating principle is perhaps the most novel of our proposals. In addition to the constraints on the activation observed in individual areas, there also appear to be constraints on the conjoint functioning of multiple activated cortical areas. These systemwide resource constraints are particularly manifest when participants perform two complex cognitive tasks concurrently. If there were no systemwide resource constraints, one would expect the activation observed during dual-tasking to simply be the sum of the activations observed when performing each single task in isolation, or perhaps the activation could be slightly overadditive because of additional overhead (coordination) costs. However, when auditory sentence comprehension and mental rotation are performed concurrently, activation during dual-tasking is far less than the sum of the activations when performing comprehension or rotation alone (Just, Carpenter, Keller, et al., 2001), as shown in Figure 3. This is consistent with the proposal here that some resource constraints apply to sets of cortical areas, complementing the constraints that hold within cortical areas. As described below, we interpret these interarea resource constraints as bandwidth limitations on the communications infrastructure that supports collaborative processing within the task-specific large-scale networks (Just et al., 2004).

5. The activation of a cortical area, as measured by imaging techniques such as fMRI and PET, varies as a function of its cognitive workload. The final operating principle is a measurement assumption. Cognition is biological work; this is the key to interpreting fMRI data. In fact, an alternative name that was once proposed for *functional* neuroimaging was "*brain work* imaging" (Haxby, Grady, Ungerleider, & Horwitz, 1991), a labeling of the measured property that we aim to make precise. In particular, cortical areas engage in order to perform neural computation, and this computation consumes biological resources. The *capacity utilization* of a cortical area is the
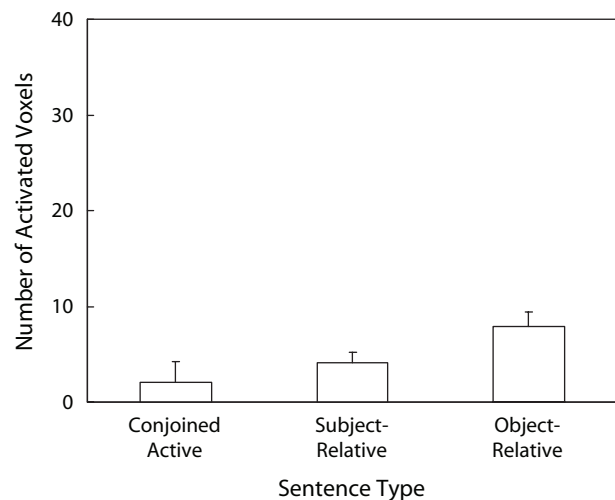

Figure 2. Progressive activation of the right homologue of Wernicke's area (right superior temporal gyrus) as a function of sentence complexity (from Just, Carpenter, Keller, et al., 1996).
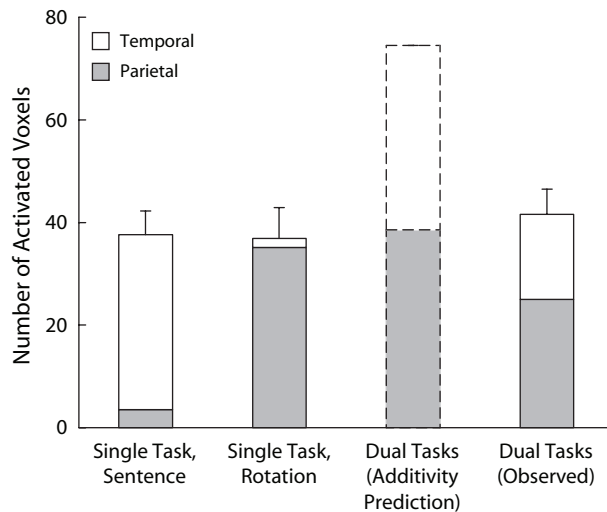
Figure 3. Activation in two areas during a language task and during a spatial task, predicted dual-task activation under the assumption of additivity, and the observed dual-task activation (from Just, Carpenter, Keller, et al., 2001).

proportion of its resources currently being consumed, relative to its total supply. We claim that the capacity utilization of a cortical area is indexed by activation as measured by fMRI or PET. Capacity utilization is an instantaneous and dynamic property of a cortical area. It can be measured as the average activation over a relatively long period of time spanning several trials, as in studies that employ block designs, or it can be measured multiple times during a single trial, as in event-related fMRI studies. Although the precise physiological mechanisms that couple neural computations with the fMRI BOLD response are currently unclear, most accounts ultimately rest on some bioenergetic principle related to resource consumption and regeneration (e.g., Logothetis, 2003). Our reference to a specific biological process is intentionally at a more molar level because, for our purposes, the capacity utilization of a cortical area can be conceptualized as the aggregate of many resource-bounded neurobiological mechanisms.

Neuroimaging studies that use *graded* designs generally find that, within some range of acceptable performance accuracy, activation in relevant areas increases monotonically with the hypothesized computational demand imposed. This finding has been observed in a number of contexts: in word span tasks that vary the number of words to be stored and recalled (Grasby et al., 1994); in working memory *N*-back tasks that vary *N* (Braver, Cohen, Jonides, Smith, & Noll, 1997; Rypma, Prabhakaran, Desmond, Glover, & Gabrieli, 1999); in sentence comprehension tasks that vary the amount and nature of clause embedding (Just, Carpenter, Keller, et al., 1996; Röder, Stock, Neville, Bien, & Rösler, 2002); in mental rotation tasks that vary the rotation angle (Carpenter, Just, Keller, Eddy, & Thulborn, 1999; Just, Carpenter, Maguire, Diwadkar, & McMains, 2001); and in problem-solving tasks that vary the depth of search (Baker et al., 1996; Newman, Carpenter, Varma, & Just, 2003). Figure 4 shows how the

volume and intensity of brain activation increase with the computational demand placed on one of the cortical areas central to performance of three different tasks—sentence comprehension, mental rotation, and Tower of London problem solving.

II. THE 4CAPS ARCHITECTURE

The operating principles introduced in the previous section have been embodied in a new cognitive neuroarchitecture, 4CAPS. Cognitive architectures are unified theories of cognition (Newell, 1990) that take the form of computational formalisms. As unified theories, they claim to support accounts of all forms of cognition. 4CAPS focuses in particular on complex forms of cognition such as language comprehension, problem solving, spatial reasoning, and dual-tasking. As computational formalisms, cognitive architectures claim to embody the basic information-processing capabilities of the mind. That is, they define the available representational formats, the transformational operations on these representations, and control structures that organize the application of operations to representations over time. As a cognitive neuroarchitecture, 4CAPS makes the additional claim of admitting a plausible neural implementation. As we will see below, this neural implementation is not of the conventional sort: Its cognitive information-processing mechanisms are not directly reduced to neural information-processing mechanisms. Rather, it supports a natural definition of resource utilization in brain areas that can be compared with functional neuroimaging measures.

4CAPS is the most recent member of an architectural family that includes CAPS (Thibadeau, Just, & Carpenter, 1982) and 3CAPS (Just & Carpenter, 1992). These architectures have been the basis of models of behavioral performance in many domains: mental rotation (Just & Carpenter, 1985); analogical problem solving (Carpenter, Just, & Shell, 1990); discourse comprehension (Goldman & Varma, 1995; Thibadeau et al., 1982); human–computer interaction (Byrne & Bovair, 1997; Huguenard, Lerch, Junker, Patz, & Kass, 1997); Tower of Hanoi problem solving in normal subjects (Just, Carpenter, & Hemphill, 1996) and in patients with frontal lobe lesions (Goel, Pullara, & Grafman, 2001); and sentence comprehension in normal readers (Just & Carpenter, 1992) and in aphasic readers (Haarmann, Just, & Carpenter, 1997). 4CAPS inherits many of the computational mechanisms of its predecessors, and thus their ability to account for the behavioral data (response times and error rates) on complex cognition; in this sense, it is a *cognitive* architecture. The new feature of 4CAPS is that it relates cognitive information to cortical information processing by instantiating the operating principles described above. As a result, 4CAPS models can also account for the neuroimaging data on complex cognition. In this sense, it is a cognitive neuroarchitecture.

Below, we describe 4CAPS, focusing on how it realizes the operating principles, and in particular on how it incorporates resource constraints on information processing. The details of the computational mechanisms it inherits from
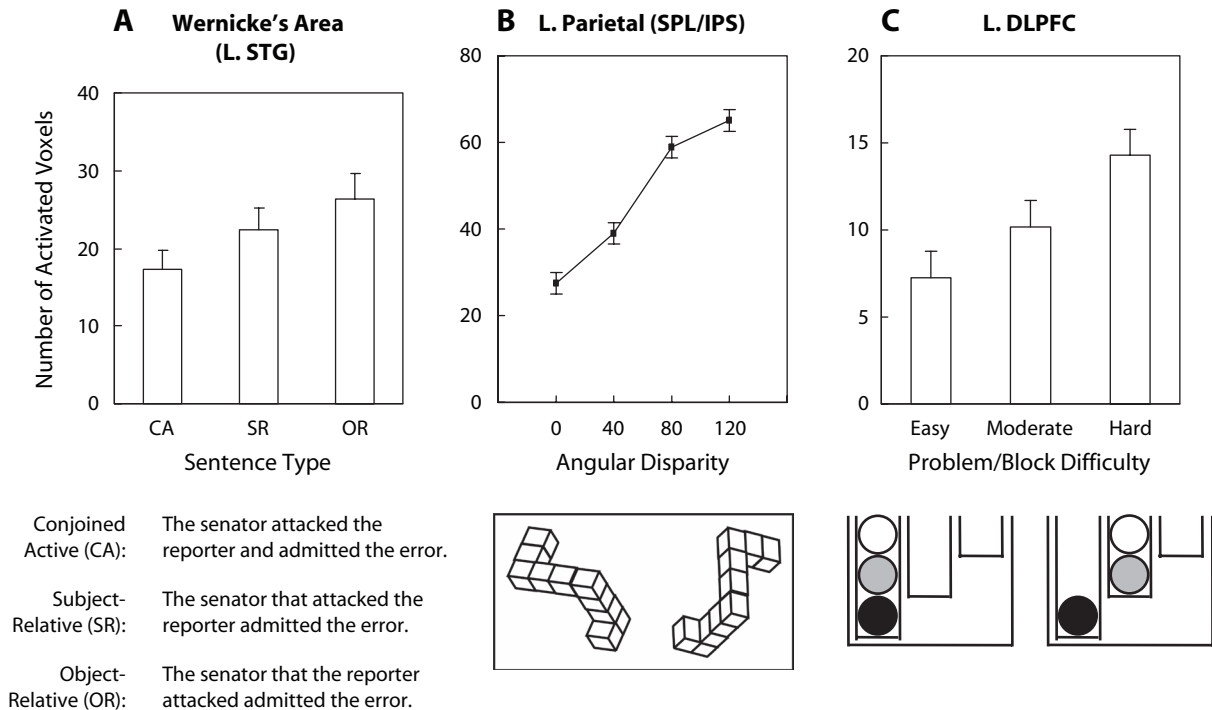
Figure 4. Modulation of activation in left-hemisphere areas as a function of (A) sentence complexity (from Just, Carpenter, Keller, et al., 1996), (B) angular disparity in mental rotation (from Carpenter et al., 1999), and (C) problem difficulty (i.e., number of moves required for solution) in the Tower of London task (from Newman et al., 2003).

CAPS and 3CAPS are available in articles on those architectures (Just & Carpenter, 1987, 1992; Just & Varma, 2002; Thibadeau et al., 1982). The details of 4CAPS and of the particular models described in subsequent sections are available at the 4CAPS Web site, www.ccbi.cmu.edu/4CAPS, which provides both documentation and source code.

A 4CAPS model consists of a set of *centers* intended to correspond to cortical areas that activate in a given task. Each center is a hybrid symbolic/connectionist system with a resource supply of fixed capacity. Centers are symbolic in that they are production systems. Declarative knowledge is represented by declarative memory elements, each possessing a number of attributes (or features), and each attribute has a symbolic or numeric value. Procedural knowledge is encoded by *condition–action* production rules. The condition aspect of a production rule specifies the enabling conditions (a pattern of declarative elements whose attribute values satisfy the specified constraints) for the rule to match. The action aspect is an action to be taken when the production rule matches and the production fires. (Actions are described below.) The characteristic ability of production systems to model high-level cognition derives from their ability to bind variables that relate declarative elements to one another, which allows them to transparently process constituent-structured (i.e., embedded) representations (Fodor & Pylyshyn, 1988).

4CAPS models also have connectionist properties: They make use of activation-based representations, graded processing, and parallel control. Broadly speaking, there

is a family resemblance between 4CAPS models and localist connectionist networks, with declarative elements functioning as units, productions as weighted links, and activation propagating in parallel. Specifically, each declarative element has an associated activation level that is a continuous variable. (The term *activation* is used in this paragraph to refer to the current availability of a representation, as in Collins & Quillian, 1969. The word also refers to *brain activation* as measured by fMRI or PET. The relation between the two meanings is described below.) The activation of a declarative element is a vector, not a scalar. There is a vector component corresponding to each center, with the value indicating the current activation contribution of that center to that element. (By convention, "the" activation of a declarative element refers to the sum of its component values—i.e., the activations contributed by the centers that are collaboratively processing it.) The productions in 4CAPS have some connectionist properties. Declarative elements must not only satisfy the symbolic constraints of condition sides; their activation levels must also exceed the specified thresholds. The primary action that a production executes is also graded—directing activation from one declarative element to another, modulo some continuous weight. Productions are also graded in time: They do not complete their processing all at once, but rather direct activation reiteratively over an interval of time.[2] Another connectionist property of 4CAPS is its parallel control scheme: At each point in time, called a *cycle*, all productions whose condition aspects are satisfied are

fired in parallel. This is true both within individual centers and across all centers. In summary, each 4CAPS center combines the strengths of a production system (symbolic variable-binding and processing of structured representations) with the graded, activation-based, and parallel operation of a connectionist network.

0. Thinking is the product of the concurrent activity of multiple centers that collaborate in a large-scale cortical network. The first operating principle is that the cognitive system, which appears unitary at the psychological level, is composed at the cortical level of multiple information-processing centers. Centers in 4CAPS are intended to correspond approximately to the cortical epicenters that compose large-scale distributed cortical networks, as proposed by Mesulam (2000). (The types of computations we focus on are performed in the cortex and are manifested as cortical activation measured by fMRI. It is very likely that subcortical structures also play a role in the forms of cognition we address, and it is also possible that some subcortical centers are governed by the same operating principles as cortical centers. However, this is outside the scope of the research and theory described here.) Each 4CAPS center maps to a cortical area that has demonstrated a cluster of activation over a range of cognitive tasks. For example, one center might correspond to Broca's area. 4CAPS does not stipulate a privileged spatial scale at which to decompose functionally coherent brain areas into centers. On the basis of clusters of activation observed in cognitive tasks, one could say that there are approximately 10–20 possible cortical centers of activity per hemisphere, or 20–40 centers overall. But some of these clusters could be further decomposed into functionally and spatially distinct microcenters, and some clusters of activation that adjoin each other and respond to similar variables could be combined into a single larger center. For example, Broca's area is sometimes treated as a single area, and at other times it is usefully partitioned into pars opercularis and pars triangularis. The preference for one treatment or the other is a matter of decomposing or aggregating to a useful cognitive scale. We assume decomposition along these lines into M centers.

Each center can perform a set of cognitive functions. An example of a cognitive function is the parsing (recognition) of a noun phrase. A cognitive function is implemented by a set of declarative elements and productions. Returning to the previous example, the declarative elements would represent phrase markers according to a particular linguistic theory (e.g., X-bar theory), and the productions would parse these representations according to a particular algorithm (e.g., bottom up). Whenever possible, we describe 4CAPS models at the abstract level of cognitive functions and bypass the details of the declarative elements and productions that implement them. This is in keeping with the focus of this article on the operating principles that shape large-scale networks.[3]

For the most part, we attribute functions to centers in a way that is consistent with the consensus view (e.g., that Broca's area is specialized for language). However, we add the notion that what unifies the cognitive functions

performed by a center is not their task domain, but their processing style. *Processing style* refers to the qualities of representations and of the operations that transform them. For example, whereas Broca's area (particularly pars opercularis) is typically construed as being involved in syntactic and articulatory processing, we construe it as constructing and selectively accessing structured (i.e., sequential and hierarchical) representations. This more general characterization of its processing style and cognitive functions is more consistent with the range of tasks for which it activates—for example, for the apprehension of musical rhythm as well as of language (Koelsch et al., 2002).

Centers collaborate by processing (i.e., contributing to the activations of) common declarative elements. In practice, two patterns of collaboration occur frequently. In the first pattern, centers act as peers (co-routines), reciprocally generating partial products that are exchanged back and forth and articulated into task outputs. We will see this in the model of sentence comprehension described below, in which centers corresponding to Wernicke's and Broca's areas collaborate intensely at all representational levels. In the second pattern, centers interact hierarchically, with one center implicitly supervising others. We will see this below in the model of spatial problem solving, in which a center corresponding to right DLPFC in some cases establishes goals for other centers (corresponding to more posterior brain areas) to pursue. In both instances, the locus of collaboration is the same: shared declarative elements whose activation vectors are contributed to by multiple centers.

1. Each center can perform multiple cognitive functions, and conversely, some cognitive functions can be performed by more than one center. The multiple cognitive functions that can be performed by each center have in common a similar information-processing style. For example, a center corresponding to the intraparietal sulcus (IPS) might be specialized for functions that are spatial and transformational in nature—such as rotation, translation, and scaling—and more generally for performing geometric computations. And although a center may have the capability of performing multiple cognitive functions, it is typically differentially specialized for the multiple functions that it can perform. In this context, *specialization* refers to efficiency of resource usage. More precisely, the specialization of center i for cognitive function j is denoted S_{ij}, where $S_{ij} \in [1, \infty)$. Perfect specialization is indicated by a value of 1.0, such that performing one unit of the cognitive function requires 1.0 units of the center's activation resources. Larger values indicate lesser specializations; for instance, a specialization of 2.0 means that performing one unit of a cognitive function requires 2.0 units of a center's activation resources. In the limit, a specialization of ∞ represents an inability to perform a cognitive function (because each center has a finite resource supply, as described below). If there are N cognitive functions, the total resource consumption of center i at a particular point in time is

$$\sum_{j=1}^{N} \left(A_{ij} \times S_{ij} \right),$$

where A_{ij} represents the amount of cognitive function j performed by center i.

Conversely, some cognitive functions can be performed by more than one center, since we claim that there is some genuine redundancy or overlap in the system. The canonical example is that some pairs of contralateral homologous areas appear capable of performing the same cognitive functions, such that if one is damaged the other can be recruited in its place. (We will return to this below in a model of sentence comprehension.) Different centers that can perform the same function typically differ in their *relative specializations* for the function. In other words, there is a rank order of the efficiency with which the function is performed *across* centers implicit in the S_{ij}. The canonical assignment of the function is determined by these relative specializations, so that, resources permitting, the function will be assigned to the center most specialized for, and thus most efficient at, performing it (i.e., S_{ij} is minimal). (However, centers have finite resource supplies that introduce dynamic, second-order influences on assignment, as described below.) The more specialized a center is for a given cognitive function, the fewer resources it requires to perform the function, and thus the lower its capacity utilization. This is consistent with the frequent finding that people with higher skill levels tend to exhibit lower activation levels (Büchel, Coull, & Friston, 1999; Just, Carpenter, & Miyake, 2003; Reichle, Carpenter, & Just, 2000).

This proposal is not modular or localist because there does not exist an isomorphism between centers and cognitive functions. Nor is it a form of equipotentiality, because not every center can perform every cognitive function. Rather, each center can typically perform multiple cognitive functions, and some cognitive functions can be performed by more than one center; the choice of which center performs a particular cognitive function is, to a first-order approximation, a matter of relative specialization.

2. Each center has a limited capacity of computational resources, constraining its activity. Computation is fueled by activation. Declarative elements must possess above-threshold activation to match the condition aspects of productions, and fired productions direct activation to declarative elements. Each center possesses a finite supply of activation, reflecting the fact that all biological systems are subject to a limit on resource availability. Specifically, the resource capacity of center i is denoted C_i, and the following constraint on resource consumption is enforced at all times:

$$\sum_{j=1}^{N} \left(A_{ij} \times S_{ij} \right) \leq C_i. \tag{1}$$

Resource limitations have an impact when a difficult task is being performed and resource demands exceed the resource supplies available at the well-specialized centers. Conceptually speaking, two adaptations are possible. The first is that excess cognitive functions can spill over from centers experiencing resource shortfalls to centers with overlapping functional specializations that have resources available. The dynamic spillover of cognitive functions is a unique feature of 4CAPS and is described in detail below. The second adaptation to resource short-

falls, first implemented in 3CAPS (Just & Carpenter, 1992), is to deallocate activation from declarative elements associated with completed cognitive functions and to reallocate this activation to cognitive functions experiencing shortfalls. For example, if a resource shortfall is experienced while processing a sentence of a text, activation can be deallocated from representations of the surface structures of prior sentences. One consequence is the gradual forgetting-by-displacement of previously activated declarative elements by more recently activated declarative elements. Both adaptations—spillover of excess workload and deallocation/reallocation—are consequences of the same resource allocation algorithm, described below.

3. The assignment of functions to centers changes dynamically, adapting to resource availability and to the functional demands of the task. The topology of a large-scale network is not fixed, but instead changes dynamically during task performance. From the perspective of an individual center, two factors determine the degree to which it participates in the network. The first factor is the match between the cognitive functions to be performed and the center's specializations for those functions: Is the center relatively well specialized for some of the pending functions? The second factor is resource availability: Does the center possess spare resources for fueling execution of the pending functions for which it is well specialized? 4CAPS incorporates an algorithm for weighing these factors, assigning pending cognitive functions to centers on the basis of a joint consideration of functional specializations and resource availability. The assignment guarantees that the large-scale network performs as much of each function as possible while respecting all resource constraints. The degree to which a center participates in the network is the indirect result of this assignment.

4CAPS maps the assignment of cognitive functions to centers onto a *linear programming* problem and solves it using the *simplex* algorithm, a standard technique in operations research (Cormen, Leiserson, Rivest, & Stein, 2001; Dantzig & Thapa, 1997). The canonical linear programming problem is to assign a set of manufacturing tasks (e.g., the production of widgets of various types) to a set of manufacturing sites (e.g., factories or machines). Manufacturing sites vary in the relative efficiencies with which they produce the various types of widgets and in their overall production capacities. The objective is to allocate production resources with maximal economic efficiency—that is, in a way that minimizes costs and thus maximizes profit earned from widget production. The widgets to be manufactured correspond in 4CAPS to the cognitive functions to be performed, and the manufacturing sites correspond to centers. The objective is to assign cognitive functions to centers in a way that maximizes cognitive throughput while minimizing resource consumption.

More formally, at each point during task performance, some mixture of the N cognitive functions must be performed. We denote the activation requested for cognitive function j as R_j. The goal is to assign the N cognitive functions to the M centers. Recall that A_{ij} denotes the amount of cognitive function j performed by center i; these are

the values to be determined in solving the assignment problem. The assignment is constrained in two ways. M constraints, one for each center, take the form of Equation 1 above. Each states that a center can supply no more than its capacity. Another N constraints, one for each cognitive function j, take the following form:

$$\sum_{i=1}^{M} A_{ij} \leq R_{j}. \qquad (2)$$

Each constraint states that the amount of cognitive function j performed across all centers must be as close as possible to (but no more than) the requested amount R_j. Solving the assignment problem requires finding values for A_{ij} that satisfy the constraints from Equations 1 and 2. (The assignment problem actually contains a third constraint as well, described below.) Because many assignments satisfy the constraints (e.g., $A_{ij} = 0$ for all i and j), we require a measure of the goodness of an assignment to be able to choose the "best" one. In linear programming, this measure is called the *objective function*, and it takes the form of a linear combination of the A_{ij} values to be maximized,

$$\sum_{i=1}^{M} \sum_{j=1}^{N} \left(W_{ij} \times A_{ij} \right). \qquad (3)$$

The only remaining problem is how to choose W_{ij}. Recall that all other things being equal, we prefer to assign the cognitive function j to the center i with the highest relative specialization—that is, whose S_{ij} is minimal. (This center corresponds to the canonical cortical area for performing the function.) This preference can be encoded in the objective function by defining the weights as $W_{ij} := 1/S_{ij}$.

To review, at each point in time, 4CAPS faces the problem of assigning pending cognitive functions to centers. It recasts this as a linear programming problem and solves it using the simplex algorithm. The result is an assignment of cognitive functions to centers (i.e., values for A_{ij}) that (1) respects the capacity constraints of each center; (2) performs as much of each cognitive function as possible; and (3) all other things being equal, assigns cognitive functions to those centers most specialized for them. These constraints determine the initial topology of the large-scale network recruited to perform a task and, more interestingly, its dynamic evolution over time in response to changing resource availability and functional demands. Although these dynamics are complex, four cases occur again and again in the models described in subsequent sections, and therefore warrant special consideration.

A. When a task of low difficulty is being performed, the cognitive functions to be executed are unlikely to exhaust the resources of the centers most specialized for them. These centers (and just these centers) will be recruited to form the canonical (or default) large-scale network for the task. In this case, the assignment of cognitive functions to centers is driven solely by first-order considerations—namely, the relative specializations of different centers. For example, assume two centers with equal resource supplies ($C_1 = C_2 = 6$) in which the first center is more specialized for a lone cognitive function than is the second center ($S_{11} = 1$ and $S_{21} = 2$). Furthermore, assume that the resource demand is relatively light (e.g., $R_1 = 3$). This defines the following linear programming problem:

maximize: $A_{11} + \frac{1}{2}A_{21}$ (objective function)

subject to: $A_{11} \leq 6$ (first center)
         $2A_{21} \leq 6$ (second center)
         $A_{11} + A_{21} \leq 3$ (cognitive function)
         $A_{11}, A_{21} \geq 0$ (default linear programming constraints),

depicted in Figure 5A. The constraints are shown as dashed lines. They demarcate the boundaries of the shaded *feasibility region*—the set of assignments (A_{11}, A_{21}) that satisfy the constraints. The simplex algorithm effectively positions the objective function line, shown as a solid line, so that it touches a vertex of the feasibility region. This vertex represents values of A_{11} and A_{21} that maximize the objective function. When the resource demands of the cognitive function are relatively light, the objective function line touches (i.e., is maximized) at a vertex that assigns the cognitive function entirely to the well-specialized first center (i.e., $A_{11} = 3$ and $A_{21} = 0$). A huge proportion of fMRI studies of cognition in normal participants correspond to this case.

B. As task difficulty increases from low to moderate, the maintenance and processing of representations consumes an increasing amount of resources. Over this range, 4CAPS predicts increasing activation in well-specialized centers. Returning to the example, as the resource demand of the cognitive function (R_1) increases from 4 to 5 to 6 units of activation, the shape of the feasibility region changes, and consequently so does the vertex at which the objective function is maximized. This is shown in Figure 5B. However, the well-specialized first center continues to possess a resource supply sufficient to satisfy the increasing resource demands, and continues to be assigned all of the cognitive function (i.e., $A_{21} = 0$). This case arises in "graded" fMRI studies that quantitatively vary the cognitive workload or task difficulty.

C. Next, consider the case in which a task is difficult. For example, when the resource demand of the cognitive function (R_1) is 7 units of activation, the feasibility region changes shape yet again, as shown in Figure 5C. The vertex where the objective function is now maximized represents an assignment of 6 units of the cognitive function to the well-specialized first center and 1 unit to the less-specialized second center. In other words, resource demands now exceed the resource supply of the first center, saturating it and causing the spillover of excess resource demands to the second center. The second center is recruited into the large-scale network, which changes its topology to a noncanonical form. When resource demands decrease following some peak, resources will become available in well-specialized centers, and previously overflowing cognitive functions will return to them. This corresponds to a transition from the feasibility region depicted in Figure 5C back to those
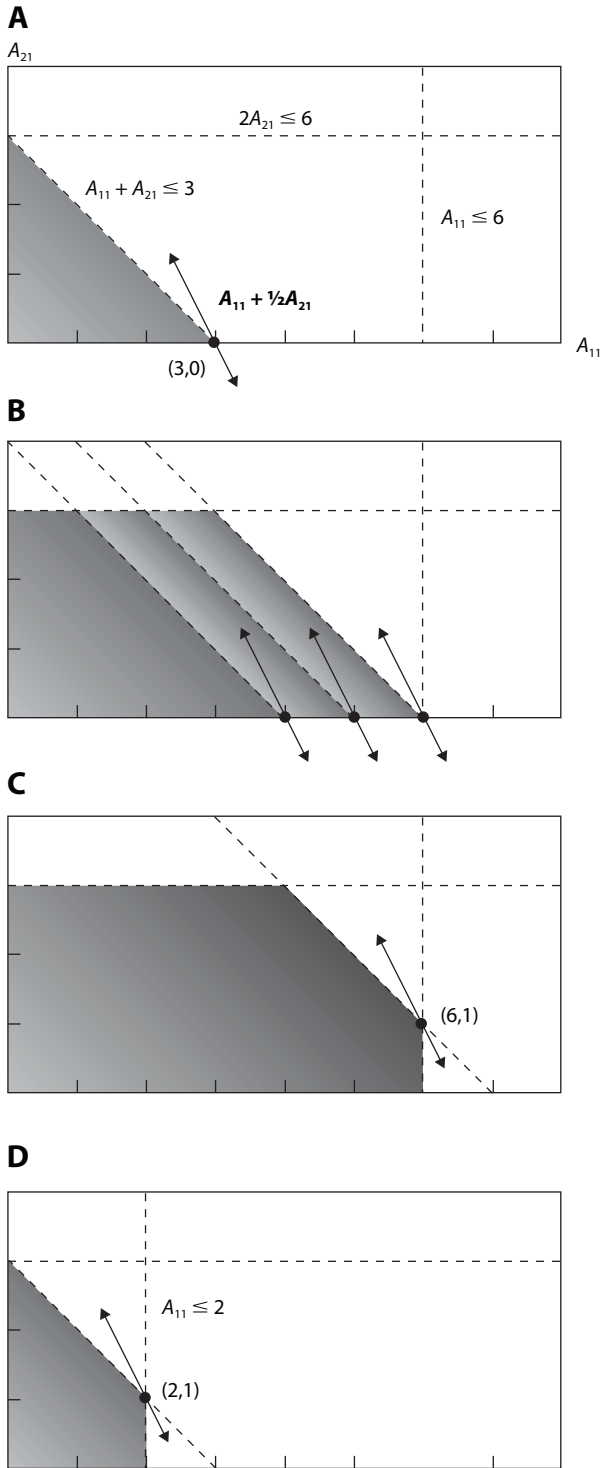
A

B

C

D

Figure 5. Graphical interpretation of the linear programming assignment problem. (A) When resource demands are low, the cognitive function is entirely assigned to the well-specialized first center. (B) As resource demands increase, the first center is increasingly recruited. (C) When resource demands exceed the resource supply of the first center, processing spills over to the less-specialized second center, which has now been recruited into the large-scale network. (D) When the resource supply of the first center is drastically reduced following damage to the corresponding brain area, spillover occurs earlier.

depicted in 5B. This case corresponds to some fMRI studies of performance in normal participants doing extremely difficult tasks that they can still perform accurately, and to globally compromised systems, such as Alzheimer's patients who can still perform a simpler task accurately.

D. The fourth case concerns compromise of a single center. Focal lesions to cortical areas are simulated by drastically reducing the resources available in the corresponding model center. This increases the likelihood of spillover. For example, if the resource supply of the first center is reduced from 6 units of activation to 2 units, then even light resource demands ($R_1 = 3$) will require recruitment of the less-specialized second center to help perform the cognitive function. This is shown in Figure 5D. This stands in contrast to the case depicted in Figure 5A, where the same light resource demands on the intact model are entirely satisfied by the first center, and therefore do not require recruitment of the second center.

Two types of intercenter interaction have been described above, and the distinction between them merits comment. The first is the straightforward collaboration referred to by Principle 0 and described earlier, in which the collaborating centers interact either as peers or hierarchically. The locus of collaboration in this case is the shared declarative elements whose activation vectors are contributed to by the multiple centers. In this form of interaction, the two collaborating centers are both drawn into the processing because the *different* cognitive functions that they are capable of performing are both needed. The second form of center interaction is the spillover of some of the processing from a more-specialized to a less-specialized center when the more-specialized center is running out of resources (an instance of dynamic change in network topology mandated by resource limitations, referred to by Principle 3). In this form of intercenter interaction, the second center is drawn into the processing because it is capable of performing types of cognitive functions *similar* to those of the first center. What both types of interaction have in common is that the representations being operated on by two centers are shared between the centers.

4. The communications infrastructure that supports collaborative processing is also subject to resource constraints, construed here as bandwidth limitations. In addition to limitations on the resource consumption of individual centers, there are also limitations on the joint activity of groups of centers. We term the former *intracenter* constraints and the latter *intercenter* constraints. An intercenter constraint makes resource availability in one center subject to resource consumption in other centers. Below, we develop a formally unified account of intracenter and intercenter resource constraints, although we acknowledge that the underlying biological mechanisms are likely to be quite different. We also offer an interpretation of intercenter constraints as bandwidth limitations on the communications infrastructure over which centers/areas collaborate.

Intercenter constraints are motivated by a number of empirical findings. Of particular importance are studies of dual-tasking that employ complex component tasks that

are implemented by two largely separate large-scale networks. A common finding is that the activation observed when performing two such tasks concurrently is less than the sum of the activations observed when performing them in isolation (see, e.g., Just, Carpenter, Keller, et al., 2001, depicted in Figure 3). Below, we describe a 4CAPS model that can account for this finding only if it is augmented with intercenter resource constraints, demonstrating their necessity. Here, the focus is on how intercenter resource constraints are incorporated into the assignment problem.

The number of intercenter resource constraints that operate in the brain and the membership of each (i.e., the centers they conjoin) are open questions. For simplicity of exposition, we assume a single intercenter constraint on the resources available to all centers. (This assumption is refined in a subsequent section with respect to the dual-task data.) This cortexwide intercenter constraint is expressed as

$$\sum_{i=1}^{M} \sum_{j=1}^{N} \left(A_{ij} \times S_{ij} \right) \leq C_{\text{CORTEX}}. \tag{4}$$

It states that the joint resource consumption of all M centers as they execute all N cognitive functions is bound by C_{CORTEX}. (More generally, if there are P groups of centers, each governed by an intercenter resource constraint, then P such constraints must be formulated.) The complete assignment problem, then, is defined by Equations 1 (intracenter resource constraints), 4 (intercenter resource constraints), 2 (functional resource demands), and 3 (the objective function). 4CAPS solves this problem at each point in time (i.e., during each cycle). At one level of analysis, the result is an allocation of intracenter and intercenter resources for the execution of pending cognitive functions. At another, coarser level of analysis, the result defines the topology of the large-scale network on a moment-by-moment basis.

5. The activation of a cortical area, as measured by imaging techniques such as fMRI and PET, varies as a function of its cognitive workload. The characterization of resource consumption developed above permits a more formal definition of cognitive workload in terms of capacity utilization, which is the entity that brain activation indexes. The *capacity utilization* of a center is the proportion of its resource supply being consumed at a point in time:

$$CU_i = \frac{\sum_{j=1}^{N} \left(A_{ij} \times S_{ij} \right)}{C_i}. \tag{5}$$

The critical measurement assumption of 4CAPS is that the capacity utilization of a center indexes neural computation in the corresponding cortical area. This assumption will be tested below by considering for a number of models whether the capacity utilizations of their centers account for the activations (as measured by fMRI or PET) in the corresponding cortical areas.

Capacity utilization as defined by Equation 5 is an instantaneous measure. It can be averaged over intervals of time spanning multiple trials to account for the results of

neuroimaging experiments that employ block designs, in which multiple trials of the same type are performed in succession.

Capacity utilization can also be used to account for moment-by-moment fluctuations in brain activation during the performance of a single trial, as measured by experiments that employ event-related designs. This requires explicitly accounting for the temporal relation between neural computation, which is effectively instantaneous at the time scale at which 4CAPS operates, and the associated hemodynamic response measured by fMRI, which is delayed and distributed in time. The hemodynamic response is accounted for in a straightforward manner. First, a capacity utilization time series $CU_i(t)$ is acquired for each center i by sampling at the same frequency that images are acquired in the event-related study being modeled (e.g., one every 1.5 sec). Second, the capacity utilization time series is convolved with a hemodynamic response function $h(t)$ to generate a predicted activation time series $fMRI_i(t)$.

$$fMRI_i(t) = \sum_{x=1}^{t} CU_i(x) h(t - x). \tag{6}$$

Other researchers have established that the hemodynamic response function is well approximated by a gamma function with a fixed delay δ (Aguirre, Zarahn, & D'Esposito, 1998; Boynton, Engel, Glover, & Heeger, 1996).

$$h(t) = \frac{\left[(t-\delta)/\tau \right]^{n-1} e^{-(t-\delta)/\tau}}{\tau(n-1)!} \quad \text{when } t \geq \delta,$$

$$\text{or } 0 \text{ otherwise.} \tag{7}$$

We adopt this approximation, as well as published parameter estimates ($\delta = 2.5$, $\tau = 1.25$, $n = 3$). The predicted activation time series of a center can be compared with the activation time series observed in the corresponding cortical area.

Modeling strategy. The operating principles embodied in the 4CAPS cognitive neuroarchitecture are evaluated below in models of sentence comprehension, Tower of London problem solving, mental rotation, and complex dual-tasking. Space limitations make it impossible to describe each model in sufficient detail to permit independent reconstruction. Readers interested in this level of description are referred to the 4CAPS Web site for documentation and source code for all models. It is also impossible to completely evaluate each model here. Readers interested in assessments against both behavioral and brain imaging data that have been collected from both normal and special populations; in formal sensitivity analyses that identify the mechanisms that bear the explanatory weights; and in comparisons against competing models of the same domain are referred to the articles devoted to the individual models. Although we pay some attention to the details of our models, their correspondence to the data, and their relation to the competing accounts, the focus here is on how the models exemplify the operational principles embodied in 4CAPS, and in particular on their dynamic response to changing resource availability and functional demands.

The models are primarily evaluated by comparing the capacity utilizations of their centers with the activations

of the corresponding brain areas. A number of measures of brain activation are considered. The first measure, the number of voxels activated above a fixed threshold in various conditions, is volumetric. The second way in which brain activation is quantified is by intensity, or more precisely, the percent change in signal intensity in an experimental condition relative to a baseline fixation condition, usually for the voxels that activate above some fixed threshold in some specified region of interest. The third measure, the sum of the activation changes in the activated voxels, combines volumetric and intensity information. No one measure is privileged; the models are evaluated against whichever measures are reported by the studies being addressed. The use of multiple measures can be justified for two reasons. First, because the measures are at least monotonically related to one another, their differences can be glossed over without excessive loss of accuracy. Second, and more importantly, our goal is not to develop neurobiological microtheories of the measures (linking neural computation to different transformations of the BOLD signal), so that 4CAPS models could make *absolute* predictions. Rather, we aim to account for the *relative* activations of brain areas across a number of conditions that vary on task variables such as difficulty. For this reason, we are content to compare capacity utilizations with different measures of brain activation using correlations. This hides the differences between the measures in silent intercept terms. Modeling these differences is left as a topic for future research. (For an example of a modeling framework that supports absolute predictions of fMRI measures, see Husain, Tagamets, Fromm, Braun, & Horwitz [2004].)

The major free parameters of 4CAPS models—and the only ones estimated in the data fits reported below—are intracenter resource capacities and, when applicable, intercenter resource capacities. The parameter estimates used to fit each model to each data set and the data sets themselves are available at the 4CAPS Web site (www.ccbi.cmu.edu/4CAPS).

III. A 4CAPS MODEL OF SENTENCE COMPREHENSION

This section illustrates the operating principles embodied in 4CAPS in the context of a model of sentence comprehension. The model's ability to perform sentence comprehension and account for behavioral performance is inherited from successful CAPS and 3CAPS models of this domain (Just & Carpenter, 1987, 1992; Thibadeau et al., 1982) and other contemporary psycholinguistic accounts. The proposals concerning the neural distribution of cognitive function are relatively novel. The model is also evaluated by comparing the capacity utilizations of its centers with the activations of the corresponding cortical areas. The focus is on how the topology of the language network changes dynamically with resource availability and with the functional demands of sentence comprehension. The details of the model's implementation and its relation to a broader range of data (and competing accounts

of these data) are provided in Varma and Just (2007) and the 4CAPS Web site.

Sketch of a Theory of Sentence Comprehension

Sentence comprehension is defined as a mapping from a sequence of word percepts, the inputs, to a thematic representation of meaning (indicating who did what to whom), the outputs. (The perceptual processes that yield word percepts and the discourse processes that integrate thematic representations across sentences are outside the scope of the sentence comprehension model.) This mapping is actually a composition of mappings linked by intermediate representational levels: lexical, syntactic, and thematic. The representations are of a conventional sort. Lexical representations are sets of feature values that encode aspects of word meaning such as grammatical class (used to construct syntactic representations) and animacy (used to construct thematic representations). Other types of semantic features of words (such as perceptual, motor, or functional properties of the physical objects to which concrete nouns refer) are not yet represented in the model. Syntactic representations conform to the X-bar formalism (Chomsky, 1970) adopted by many psycholinguistic models. Thematic representations follow the familiar case–role scheme first introduced by Fillmore (1968). The composite mappings and intermediate representational levels are governed by a number of theoretical tenets shared with many contemporary behavioral models: Processing is (1) parallel, in that representations within and across levels are constructed simultaneously; (2) immediate, in that strategic choices are not postponed indefinitely; and (3) resource constrained, in that the number of representations that can be simultaneously maintained and processed is limited (Gibson, 1998; Jurafsky, 1996; Just & Carpenter, 1992; Kintsch, 1998; Lewis, 1993; Smolensky, 1999; Thibadeau et al., 1982; Vosse & Kempen, 2000).

There is a tendency for theories of sentence comprehension to adopt the functional carving of language offered by linguistic decompositions and to map linguistic functions (such as syntactic analysis) to cortical areas in a one-to-one manner. By contrast, the operating principles embodied in 4CAPS suggest that (1) each cortical area performs functions that typify its processing style and that can apply to multiple linguistic functions, and (2) a linguistic function that appears unitary at the behavioral level can be implemented by a network of collaborating brain areas. We therefore emphasize the multiple overlapping and complementary functional specializations of the core components of the language network: Wernicke's area, Broca's area, and their right-hemisphere homologues. (Of course, other brain areas also contribute to sentence comprehension, but these are beyond the scope of the model.) We also emphasize the patterns by which these areas collaborate.

Specifically, we propose that one form of collaborative processing occurs between Wernicke's and Broca's areas, which combine to *retrieve* and *structure* language representations at all levels. Specifically, we hypothesize that Wernicke's area is specialized for the associative retrieval/

design function. Existing representations serve as cues for the retrieval of language-based associations—that is, of prior knowledge, whether declarative or procedural. Wernicke's area collects these associations along with perceptual inputs and existing representations into designs (or blueprints) for new language representations. We hypothesize that Broca's area is specialized for the structure-building function. It uses the association-based designs activated by Wernicke's area to manufacture new structured representations (or constituent structures). These representations are notable because they possess internal structure—that is, sequentially or hierarchically embedded constituents organized by named relations. More generally, Broca's area is postulated to construct structured representations not only for the syntactic analysis of language, but also for other levels of language and for other types of tasks requiring such representations, such as the processing of musical rhythm (Koelsch et al., 2002).

The associative retrieval/design and structure-building functions attributed to Wernicke's and Broca's areas, respectively, are applied at various levels: to word percepts, to produce lexical designs and then lexical feature constituents; to lexical features, to produce syntactic designs and then syntactic constituents; to subordinate syntactic constituents, to produce syntactic designs and then superordinate syntactic constituents; and to superordinate syntactic constituents, to produce thematic designs and then thematic constituents. In short, the complementary functional specializations of Wernicke's and Broca's areas require them to collaborate at all levels of sentence comprehension. The attribution of an associative retrieval/design function to Wernicke's area and a structuring function to Broca's area represents a break from the dominant linguistic carving of language. However, these attributions are consistent with older linguistic and neuropsychological traditions (Jakobson, 1971; Luria, 1981). They also resonate with the common finding that manipulating a single linguistic variable (such as syntactic difficulty or lexical difficulty) modulates activation in both areas (see, e.g., Keller et al., 2001)—a pattern recognized in recent reviews of the neuroimaging literature on language processing (e.g., Bookheimer, 2002; Friederici, 2002; Kaan & Swaab, 2002).

Another form of collaborative processing is more dynamic in nature, occurring when a cortical area most specialized for a required cognitive function lacks the resources to perform it. When this happens, a cortical area with a secondary specialization for the cognitive function and having adequate resources will be recruited into the large-scale network to perform the excess processing. This form of collaborative processing occurs frequently in the language network. Specifically, we hypothesize that the right-hemisphere homologues of Wernicke's and Broca's areas have secondary specializations for the retrieval/design and structuring functions, respectively. When the resources of a well-specialized left-hemisphere area are exhausted, processing spills over to its right homologue, which is recruited into the language network. This can happen because a sentence is especially difficult, in which case the right homologue enters the language network on a temporary basis (see, e.g., Just, Carpenter, Keller, et al., 1996). It can also happen when a stroke-induced focal lesion chronically depletes the resources available in a left-hemisphere area, in which case the contralateral area enters the language network on a more permanent basis (see, e.g., Thulborn et al., 1999). Both forms of dynamic recruitment are consequences of the assignment problem that 4CAPS solves at each point in time, as the sentence comprehension model demonstrates below.

The Sentence Comprehension Model

The sentence comprehension model consists of four centers, corresponding to Broca's and Wernicke's areas and their right-hemisphere homologues. The centers, their primary functional specializations, and their canonical and noncanonical patterns of collaboration are shown in Figure 6. Collaboration is extensive, a consequence of (1) the two-step procedure by which new representations are first designed and then constructed and (2) the localization of these steps to different centers, so that no one center bears complete responsibility for any level of language processing. Moreover, as the right side of the figure shows, collaboration is dynamic: When the functional demands exceed the resources of the well-specialized left-hemisphere centers, processing spills over to the less-specialized right-hemisphere centers.

The Associative center. The Associative center corresponds to Wernicke's area. It retrieves prior knowledge associations cued by perceptual inputs and existing language representations. It combines these elements into designs for new language representations. Associative does not itself construct these new representations, because it lacks the required constituent-structuring capabilities. Rather, it performs the first step of the process by collecting together related information.

Associative performs the retrieval/design function for all levels of language. At the lexical level, it takes a word percept as input and retrieves associated information such as its grammatical class and animacy. The center combines these elements into designs for new lexical feature representations. Associative also combines existing lexical feature representations into designs for new word-level phrase representations, combines existing phrase representations into designs for new superordinate phrase representations, and combines existing superordinate phrase representations into designs for new clause representations. For example, one production looks for the pattern of a determiner D adjacent to a noun N. When existing representations match this pattern, the production fires, creating a design for a new phrase XP that combines them. We denote this design $\{XP, D, N\}$ to indicate its lack of internal (i.e., constituent) structure.

Associative also participates in the design of thematic representations. Thematic processing is expectation driven. When a noun phrase representation appears, Associative designs an expectation that the verb that heads the clause will assign a thematic role to the phrase. Similarly, when a verb representation appears, Associative designs an expectation that the verb will assign a thematic role
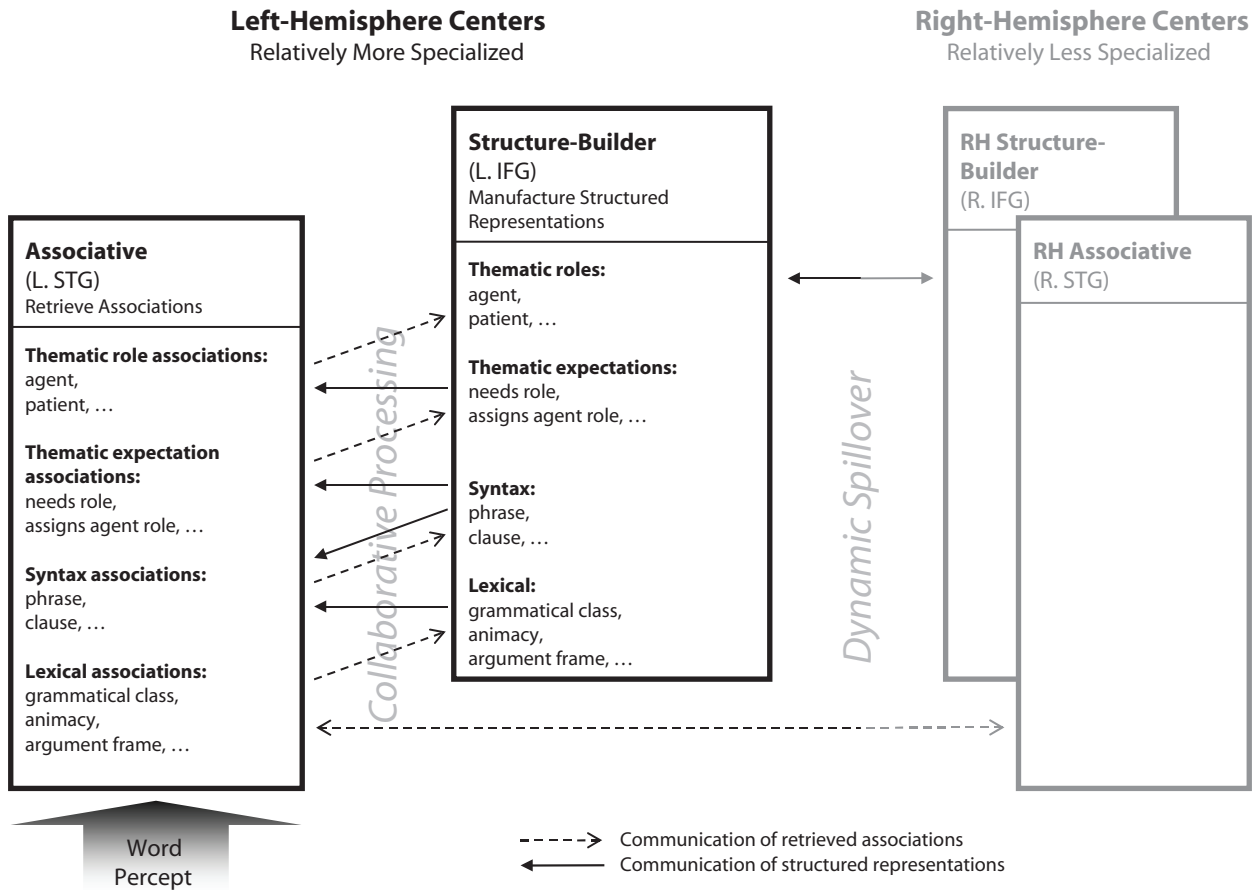
Left-Hemisphere Centers
Relatively More Specialized

Right-Hemisphere Centers
Relatively Less Specialized

Structure-Builder
(L. IFG)
Manufacture Structured
Representations

Thematic roles:
agent,
patient, …

Thematic expectations:
needs role,
assigns agent role, …

Syntax:
phrase,
clause, …

Lexical:
grammatical class,
animacy,
argument frame, …

RH Structure-Builder
(R. IFG)

RH Associative
(R. STG)

Dynamic Spillover

Associative
(L. STG)
Retrieve Associations

Thematic role associations:
agent,
patient, …

Thematic expectation associations:
needs role,
assigns agent role, …

Syntax associations:
phrase,
clause, …

Lexical associations:
grammatical class,
animacy,
argument frame, …

Collaborative Processing

Word
Percept

- - - -> Communication of retrieved associations
⟵ Communication of structured representations

Figure 6. Centers of the sentence comprehension model, their primary language specializations, and their canonical and dynamic patterns of collaboration.

for each element in the verb's argument frame. Expectations are resolved in accord with the *theta role criterion* of government–binding theory (Chomsky, 1981). Specifically, Associative looks for matching receiving and assigning expectations, and when it finds a pair, combines them into a design for a new thematic role representation.

The Structure-Builder center. The Structure-Builder center corresponds to Broca's area. In the proposed division of labor, Associative performs the first step of designing new representations, and Structure-Builder the second step of manufacturing these designs into constituent-structured representations. Specifically, Structure-Builder takes the perceptual inputs, existing representations, and prior knowledge associates that are loosely collected together in designs and activates new representations that organize these constituents according to named relations; it does this for all levels of language. To continue the example above, {*XP, D, N*} is only a design; it leaves unspecified the type of the new phrase representation and the manner in which the embedded determiner and noun representations are organized. Structure-Builder takes this design, constructs a new representation of the "noun phrase" type, assigns the determiner to the "specifier" position, and assigns the noun to the "head" position, producing <*NP*

specifier:D head:N>. This new representation is available for the next iteration of design; for example, it may be combined with existing representations and retrieved prior knowledge associates into a design for a new clause representation. In this way, Associative and Structure-Builder collaborate to articulate all of the representational levels required to comprehend sentences.

Right-hemisphere centers. The model also contains RH Associative and RH Structure-Builder centers corresponding to the right-hemisphere homologues of Wernicke's and Broca's areas, respectively. The right-hemisphere centers perform cognitive functions similar to those of their left-hemisphere counterparts, but at lower levels of specialization (efficiency). For example, although Structure-Builder is the most specialized center for the construction of new thematic representations, RH Structure-Builder can also perform this function, though less efficiently; that is, it requires more resources to construct the same representation. Of course, the right-hemisphere centers are more than just weak sisters; they have their own functional specializations that are gradually coming to be understood, such as prosodic processing and discourse comprehension (Beeman, 1998; Long & Baynes, 2002; Mason & Just, 2004; Meyer, Alter, Fried-

erici, Lohmann, & von Cramon, 2002). These language-related functions for which the RH homologues are most specialized are outside the scope of the present model.

Collaborative processing and dynamic network reconfiguration. The Associative and Structure-Builder centers collaborate closely to construct lexical, syntactic, and thematic representations, as described above and depicted by the ascending set of arrows on the left side of Figure 6. The canonical pattern of collaboration results from their complementary retrieval/design and structuring functions. This pattern is expected during the comprehension of relatively simple sentences by relatively intact brains and is depicted schematically in Figures 5A and 5B. If sentences become more complex, and therefore resource demands increase, or if centers are damaged, and therefore resource supplies are reduced, then the pattern of collaboration can change, as is depicted schematically in Figures 5C and 5D. In particular, if a center is well specialized for a cognitive function but lacks the resources to perform it, some or all of the function will spill over to a less-specialized center that can perform the function and has adequate resources. The dynamic entry of new centers into the language network, which causes the network to assume a noncanonical topology, is depicted by the arrows on the right side of Figure 6. Dynamic entry is typically on a just-in-time and as-needed basis. When resources become available in well-specialized centers later in the sentence, the spilled-over processing returns to those centers and the less-specialized centers exit the network.

We now illustrate the dynamic entry of centers with the processing of a center-embedded object-relative sentence: *The senator that the reporter attacked admitted the error.* Because the relative clause *that the reporter attacked* is embedded in the center of the main clause, the emerging representation of the main clause must be buffered while the interrupting relative clause is processed. This translates into large resource demands toward the end of the relative clause, causing the spillover of excess processing from well-specialized left-hemisphere centers lacking spare resources to less-specialized right-hemisphere centers possessing spare resources. To illustrate this, the capacity utilizations of the Structure-Builder and RH Structure-Builder centers during the processing of this sentence are plotted at the top of Figure 7. (For convenience of exposition, the words within noun phrases are grouped together and the average capacity utilization during their processing is plotted. Also, the concurrent and interleaved processing of the Associative and RH Associative centers is not depicted.) Recall that thematic roles, the output of sentence comprehension, are computed via constraint satisfaction over pairs of thematic expectations, each keyed to a particular clause. The periods during which thematic expectations and clauses are active, and thus imposing resource demands, are depicted by the horizontal arrows at the bottom of Figure 7.

During the processing of the first five words of the sentence, representations for the main clause, the relative clause, and their attendant thematic expectations are de-
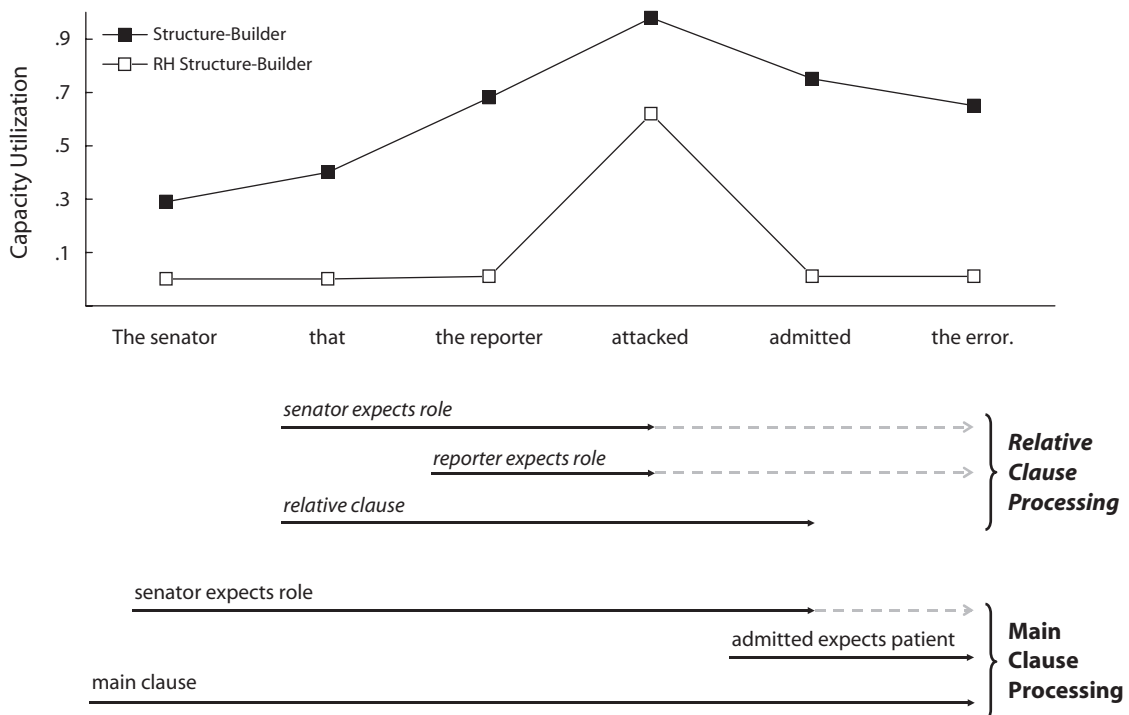

Figure 7. Capacity utilizations of the Structure-Builder and RH Structure-Builder centers of the sentence comprehension model during processing of an object-relative sentence. Below the graph, the thicker lines indicate the processing of the clauses and the thematic expectations, as well as the duration during which these representations draw on the resources of the centers. (The adjoining dashed lines indicate the resulting thematic role representations.)

signed, constructed, and maintained. The resource supply of the well-specialized Structure-Builder is sufficient to meet the mounting resource demands, as indicated by its steadily rising capacity utilization. Because no processing spills over to the less-specialized RH Structure-Builder, its capacity utilization is 0. In other words, during the initial portion of the sentence, the centers of the language network collaborate according to the canonical pattern.

The next two words, *attacked* and *admitted*, mark the end of the relative clause and resumption of the main clause. The design, construction, and maintenance of thematic expectations triggered by these verbs impose additional resource demands, and for the first time, the resource supply of Structure-Builder is insufficient, as indicated by its capacity utilization saturating at 1. For the first time, excess processing spills over to RH Structure-Builder, which has secondary specializations for the construction of new representations (and the maintenance of existing representations). This center enters the language network for the first time, as indicated by its nonzero capacity utilization. As a result, the topology of the network changes to a noncanonical pattern.

The entry of RH Structure-Builder is on a just-in-time and as-needed basis. Pairs of receiving and assigning thematic expectations are matched, and new thematic role representations are designed and constructed, as indicated by the dashed arrows at the bottom of Figure 7. The satisfied thematic expectations are then suppressed, easing resource demands. In addition, *admitted* marks the resumption of the main clause, and the representation of the relative clause can be suppressed, further easing resource demands. At this point, Structure-Builder once again possesses sufficient resources for the construction and maintenance of all structured representations. There is no additional processing to spill over to RH Structure-Builder, which exits the language network, as indicated by its capacity utilization returning to 0. Centers collaborate according to the canonical pattern for the remainder of the sentence, because resource demands never again exceed the resource supply of Structure-Builder. The dynamic entry and subsequent exit of centers into the language network are emergent outcomes of the assignment problem that 4CAPS solves at each point in time.

Empirical Evaluation of the Sentence Comprehension Model

The sentence comprehension model can be evaluated against behavioral and brain imaging data collected from normal young adults and from patients with lesions. In the comparisons below, the description of the empirical studies is necessarily brief. The main goal here is to illustrate how the operating principles embodied in 4CAPS account for the dynamic nature of the language network as a function of resource availability and the changing function demands of sentence comprehension.

Behavioral data. The model makes behavioral predictions at multiple levels of aggregation: overall sentence reading times, phrase-by-phrase reading times, and word-by-word reading times. At the broadest level, it predicts that the time to read a sentence is a function of the resource demands it imposes. For example, a sentence with a center embedding requires additional resources to maintain the initial part of the main clause when it is interrupted during processing of the embedded clause. The human reading times we collected for 11 types of sentences of increasing structural complexity (broadly defined), as well as the corresponding model processing times, are shown in Figure 8. The correlation between these data points is .97 ($p < .01$), establishing the model's general ability to account for what makes a sentence take longer to comprehend. A detailed evaluation of the model's correspondence to the behavioral data, including phrase and word reading times on center-embedded and reduced-relative sentences, is provided by Varma and Just (2007). Generally speaking, the 4CAPS model performs as well in this regard as a previous 3CAPS model of sentence comprehension (Just & Carpenter, 1992).

Brain activation data. The capacity utilizations of model centers can be compared with the activations of corresponding cortical areas during the processing of various sentence types. Below, we evaluate the model against data collected in studies that employ both block and event-related designs and that test both intact normal and brain-damaged populations.

Structural complexity in a block design. Just, Carpenter, Keller, et al. (1996) used a block design to assess the effect of structural complexity on the activation levels of the core brain areas of the language network. Three types of sentences were presented to normal young adults:

Conjoined active:	*The senator attacked the reporter and admitted the error.*
Subject-relative:	*The senator that attacked the reporter admitted the error.*
Object-relative:	*The senator that the reporter attacked admitted the error.*

These sentences are nearly identical in their lexical content but differ in their syntactic structures. As described above, comprehending a center-embedded object-relative sentence requires multiple clauses and thematic expectations to be maintained and processed, taxing the resources of the Structure-Builder and Associative centers. At the points of highest resource demand, processing spills over from the well-specialized left-hemisphere centers of the model to the less-specialized right-hemisphere homologues, which dynamically enter the network. The (clausal and thematic expectation) resource demands are progressively lower for subject-relative and conjoined active sentences. As a result, the resource shortfalls in the left-hemisphere centers are progressively less severe for the latter sentence types, and the degree to which the right-hemisphere centers are activated is correspondingly attenuated.

The capacity utilizations of the model's centers and the fMRI-measured activations of the corresponding brain areas exhibit the same profile—namely, an increase with
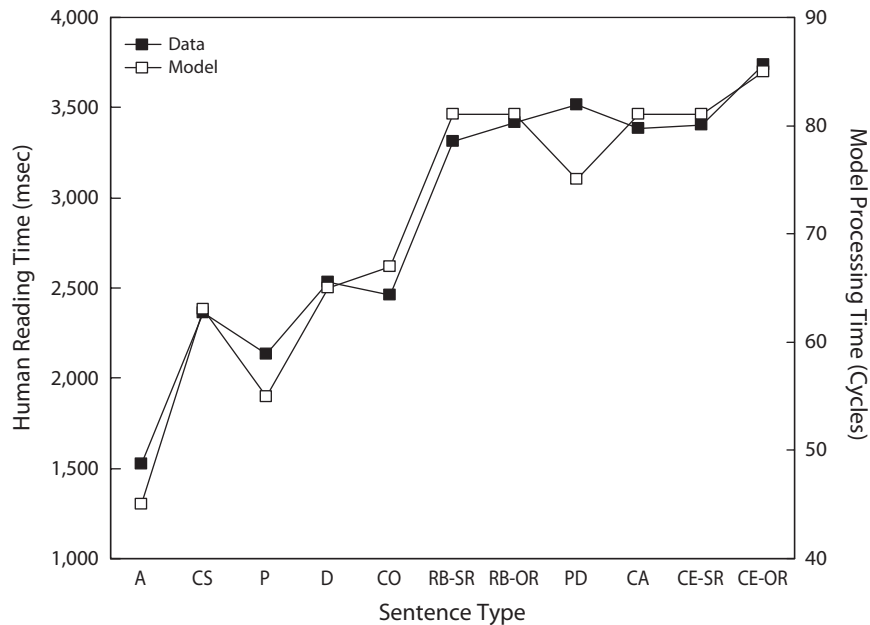
Figure 8. Sentence reading times of human subjects and processing times of the sentence comprehension model for 11 sentences of increasing difficulty. The sentence types are A, active; CS, cleft subject; P, passive; D, dative; CO, cleft object; RB-SR, right-branching subject-relative; RB-OR, right-branching object-relative; PD, passive dative; CA, conjoined active; CE-SR, center-embedded subject-relative; and CE-OR, center-embedded object-relative.

sentence complexity. The activations observed in Wernicke's area and its right-hemisphere homologue and the capacity utilizations of the corresponding model centers, Associative and RH Associative, are plotted in Figure 9. The activations observed in Broca's area and its right-hemisphere homologue and the capacity utilizations of the corresponding model centers, Structure-Builder and RH Structure-Builder, are plotted in Figure 10. The model reproduces the two major empirical findings. First, the capacity utilizations of the left-hemisphere centers, which are well-specialized for language functions, increase with sentence complexity. This is exactly the pattern depicted schematically in Figure 5B above. Second, there is increasing recruitment of the less-specialized right-hemisphere centers with increasing sentence complexity, due to excess processing spilling over to these centers, as shown in Figure 5C above. The close correspondence between the brain activations and the center capacity utilizations is reflected in the .98 ($p < .01$) correlation between the 12 points of comparison in Figures 9 and 10. The model provides a good account of dynamic changes to the topology of the language network with increasing resource/functional demands.

Syntactic ambiguity in an event-related design. Mason, Just, Keller, and Carpenter (2003) used an event-related paradigm to measure the time course of brain activation every 1.5 sec during the comprehension of classic reduced-relative garden path sentences, such as *The experienced soldiers warned about the dangers conducted the midnight raid*. The ambiguity arises with *warned*, which can be either the past participle or the past tense form of the verb *warn*. Under the past participle interpretation,

the verb begins a reduced relative clause. This is the unpreferred (less frequently occurring) interpretation. The second clause then imposes additional resource demands. This contrasts with the preferred interpretation, that *warned* is the past tense main verb of the sentence, which is more frequent and, because of its being a single clause, less resource demanding. This study presented four types of sentences to normal young adults that varied orthogonally in whether they were ambiguous or unambiguous and whether they were ultimately resolved in favor of the unpreferred (two-clause) interpretation or the preferred (one-clause) interpretation.

Amb–unpref: *The experienced soldiers warned about the dangers conducted the midnight raid.*

Unamb–unpref: *The experienced soldiers who were told about the dangers conducted the midnight raid.*

Amb–pref: *The experienced soldiers warned about the dangers before the midnight raid.*

Unamb–pref: *The experienced soldiers spoke about the dangers before the midnight raid.*

Each sentence was followed by a true–false comprehension probe.

Like many contemporary models of sentence comprehension, the 4CAPS model processes ambiguities in

A   Wernicke's Area (L. STG) and Right Homologue

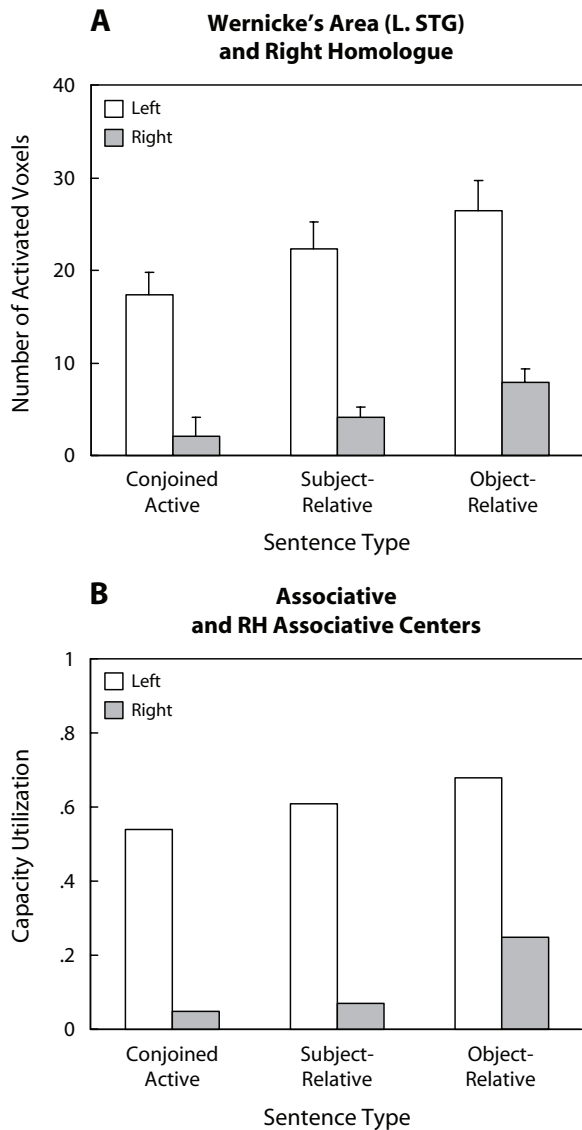## B   Associative and RH Associative Centers

Figure 9. Comparison of (A) fMRI activation of Wernicke's area and its right-hemisphere homologue (from Just, Carpenter, Keller, et al., 1996) and (B) capacity utilization of the corresponding Associative and RH Associative centers of the sentence comprehension model, on sentences of increasing difficulty.

parallel. When the ambiguous verb *warned* is read, the model generates both the past participle and past tense lexical senses. Both interpretations are elaborated over subsequent words until disambiguating information is encountered, at which point the incorrect interpretation is suppressed, freeing resources, and the correct interpretation is retained. The additional resource demands of designing, constructing, and maintaining multiple interpretations manifest themselves as higher capacity utilizations for the ambiguous sentences than for the corresponding unambiguous sentences, especially in the ambiguous region. (The prediction can also be made by a serial model that sometimes has to do extra work [e.g., reanalysis] in the ambiguous case—namely, reprocessing the ambigu-

ous segment under the alternative interpretation. Note also that a parallel model with limited resources devolves to a serial model.) This ambiguity effect is more pronounced for the unpreferred sentences because of the additional resource demands associated with the second clause.

Recall that the capacity utilization time series of a model center can be transformed into a predicted activation time series that can be compared with the observed activation time series in the corresponding cortical area. This is a multistep process. First, a conversion rate between "model time" and real (i.e., human processing) time is estimated. A conversion rate of 20 cycles of the model time per 1.5 sec of real time was estimated by calibrating the number of cycles the model requires to process the four sentence types with the human reading times obtained in a self-paced reading study using the same stimuli (MacDonald, Just, & Carpenter, 1992). Second, a capacity utilization time series for each model center is constructed by sampling at the same frequency that Mason et al. (2003) used. The capacity utilization time series for the Associative center when processing the unambiguous–preferred sentence is shown in Figure 11A. Third, the capacity utilization time series must be convolved with the hemodynamic response function (Equation 7) to produce a sequence of individual hemodynamic responses. (Recall that previous research has established that the shape of this function is well approximated by a time-delayed gamma function with delay $\delta = 2.5$ sec and gamma parameters $\tau = 1.25$ sec and $n = 3$; Aguirre et al., 1998; Boynton et al., 1996.) This is shown for the Associative center in Figure 11C. Each hemodynamic response corresponds to one of the instantaneous capacity utilization values shown in Figure 11A, delayed and smoothed in time. Finally, the individual hemodynamic responses must be summed according to Equation 6 in order to produce a predicted fMRI time series. This is shown for the Associative center in Figure 11E. (The two peaks of the curve correspond to the processing of the sentence and the processing of the comprehension probe, respectively.)

The fMRI time series observed in Wernicke's area for each of the four sentence types used by Mason et al. (2003) are shown in Figure 12A. There is a main effect of ambiguity, with ambiguous sentences producing higher activation levels than unambiguous sentences, and a main effect of preference, with sentences resolved in favor of the unpreferred interpretation producing higher levels of activation than sentences resolved in favor of the preferred interpretation. The predicted fMRI time series in Associative, the corresponding model center, for each of the four sentence types are shown in Figure 12B. Qualitatively speaking, the model produces a similar ordering of capacity utilizations across the four sentence types as well as a similar activation profile over time. Quantitatively speaking, the correlation between the observed and predicted fMRI time series data is .86 ($p < .01$). (Although not shown, the correlation between Broca's area and Structure-Builder is comparably good: $r = .91$, $p < .01$.) The model displays greater capacity utilizations on ambiguous than on unambiguous sentences because of the extra resource demands of dealing with (designing,

A   Broca's Area (L. IFG) and Right Homologue

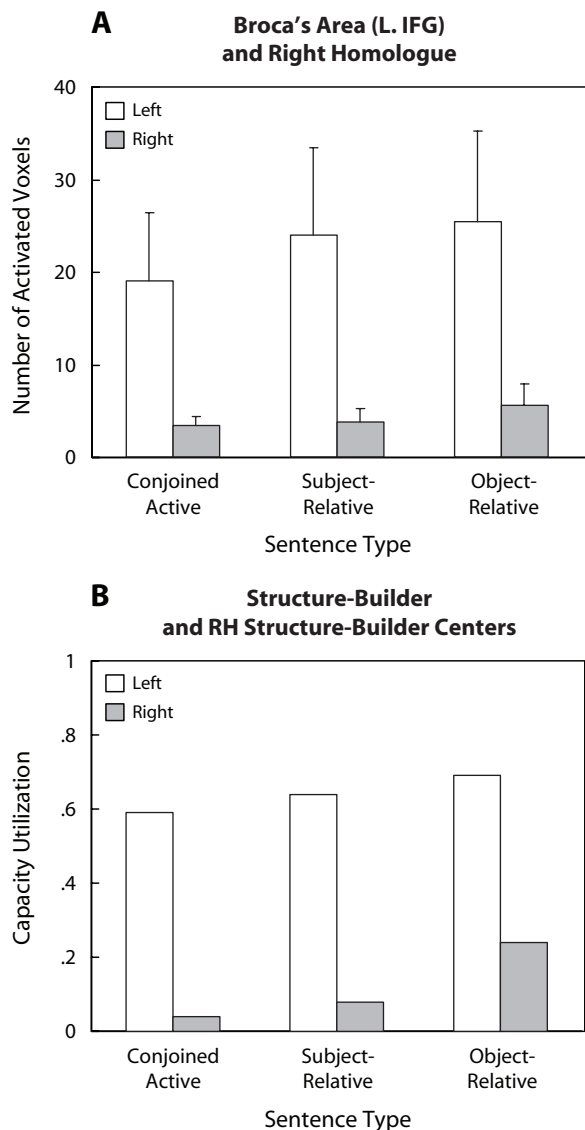## B   Structure-Builder and RH Structure-Builder Centers

Figure 10. Comparison of (A) fMRI activation of Broca's area and its right-hemisphere homologue (from Just, Carpenter, Keller, et al., 1996) and (B) capacity utilization of the corresponding Structure-Builder and RH Structure-Builder centers of the sentence comprehension model, on sentences of increasing difficulty.

constructing, and maintaining) two interpretations. It displays greater capacity utilizations on unpreferred than on preferred sentences because of the extra resource demands of processing the second clause. These simulations illustrate the ability of a 4CAPS model to account for the time course of activation as measured by event-related fMRI.

Recovery from brain damage. Several PET studies have found right-hemisphere activation for language processing following recovery from aphasia caused by left-hemisphere stroke (Karbe et al., 1998; Weiller et al., 1995). This phenomenon is generally attributed to contralateral takeover of function after damage to left-hemisphere language areas (Finger, Buckner, & Buckingham, 2003;

Heiss, Thiel, Kessler, & Herholz, 2003). In a recent fMRI study of 2 stroke patients, one with damage to Wernicke's area and the other with damage to Broca's area, the right-hemisphere homologue of the damaged area showed substantial activation during sentence comprehension (Thulborn et al., 1999). We focus on 1 patient (Case 1) who exhibited a dense expressive aphasia immediately after a left middle cerebral artery stroke, which produced a large focal lesion in Broca's area. (The other patient is more difficult to evaluate because he suffered from epilepsy in addition to the stroke.) However, 6 months after the stroke, language function had substantially recovered, and the patient performed with moderately high accuracy in an fMRI sentence comprehension study requiring the silent reading of simple sentences (mean length 5.5 words) and the answering of true–false comprehension probes. The main finding was that Broca's area, which activates in normal readers, exhibited no activation in the patient, reflecting the catastrophic effect of the lesion, whereas the contralateral area was clearly activated. This pattern presumably occurs because the cognitive functions formerly performed by Broca's area have spilled over to its right homologue on a permanent basis. By contrast, the undamaged Wernicke's area and its right-hemisphere homologue showed the typical pattern found in normal right-handed participants, with substantially more activation in the left- than in the right-hemisphere area. These data are shown in Figure 13A.

The patient's Broca's lesion was modeled by depriving the corresponding model center, Structure-Builder, of all of its activation resources. Recall that when a center well specialized for a cognitive function lacks the resources to perform that function, the center with the next highest relative specialization for that function that possesses adequate resources will be recruited into the network to perform the function, although less efficiently. This was depicted schematically in Figure 5D. The capacity utilizations of the model's four centers are shown in Figure 13B. Because the lesioned Structure-Builder center lacks the resources to participate in processing, its capacity utilization is 0 and the function of constructing new structured representations spills over entirely to RH Structure-Builder, as indicated by the high capacity utilization in that center. (In this sense, 4CAPS models are like sports teams: When one member is injured, a team can continue to function by substituting a less effective player with qualitatively similar skills.) By contrast, the unlesioned Associative and RH Associative centers function as in normal comprehension, with the Associative center performing the bulk of the retrieval/design function for which it is specialized, and RH Associative assisting only when the resource demands of this function exceed the resource supply of Associative. The correlation between the model and the data is .99 ($p < .01$). In summary, the sentence comprehension model shows promise for accounting for how large-scale networks adapt to damage to their canonical components.

Summary of the Sentence Comprehension Model

The sentence comprehension model accounts for overall sentence reading times and patterns of brain activation (as
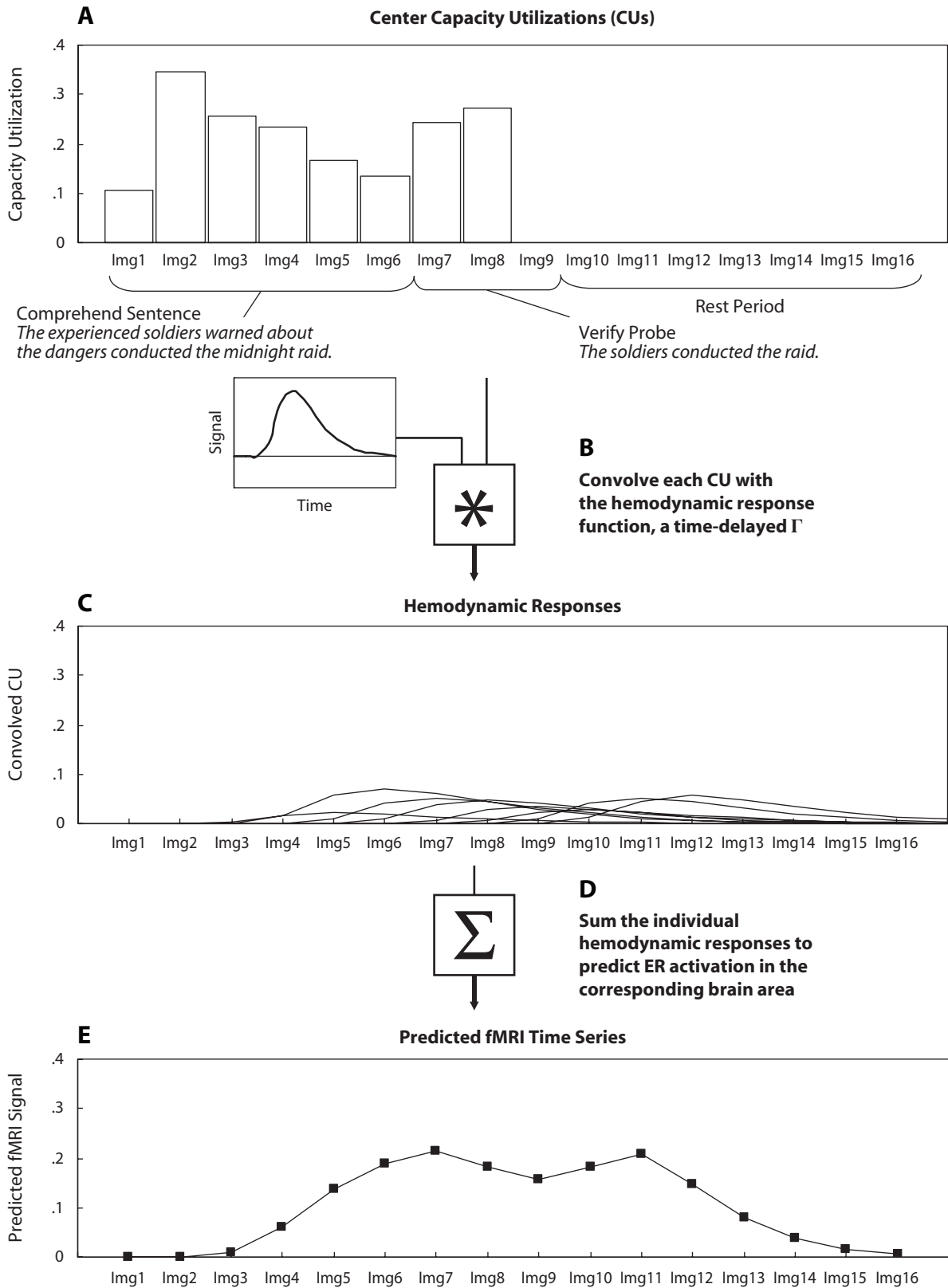
A

Center Capacity Utilizations (CUs)

Capacity Utilization

.4

.3

.2

.1

0

Img1 Img2 Img3 Img4 Img5 Img6 Img7 Img8 Img9 Img10 Img11 Img12 Img13 Img14 Img15 Img16

Comprehend Sentence
*The experienced soldiers warned about
the dangers conducted the midnight raid.*

Rest Period

Verify Probe
The soldiers conducted the raid.

Signal

Time

B

**Convolve each CU with
the hemodynamic response
function, a time-delayed Γ**

✱

C

Hemodynamic Responses

Convolved CU

.4

.3

.2

.1

0

Img1 Img2 Img3 Img4 Img5 Img6 Img7 Img8 Img9 Img10 Img11 Img12 Img13 Img14 Img15 Img16

D

Σ

**Sum the individual
hemodynamic responses to
predict ER activation in the
corresponding brain area**

E

Predicted fMRI Time Series

Predicted fMRI Signal

.4

.3

.2

.1

0

Img1 Img2 Img3 Img4 Img5 Img6 Img7 Img8 Img9 Img10 Img11 Img12 Img13 Img14 Img15 Img16

Figure 11. (A) Capacity utilization of the Associative center of the sentence comprehension model, sampled every 1,500 msec; (B) hemodynamic response function (a time-delayed gamma); (C) convolution of each capacity utilization sample with the hemodynamic response function; (D) summation of the convolved capacity utilizations; (E) predicted fMRI time series, which can be compared with the fMRI time series observed in Wernicke's area.
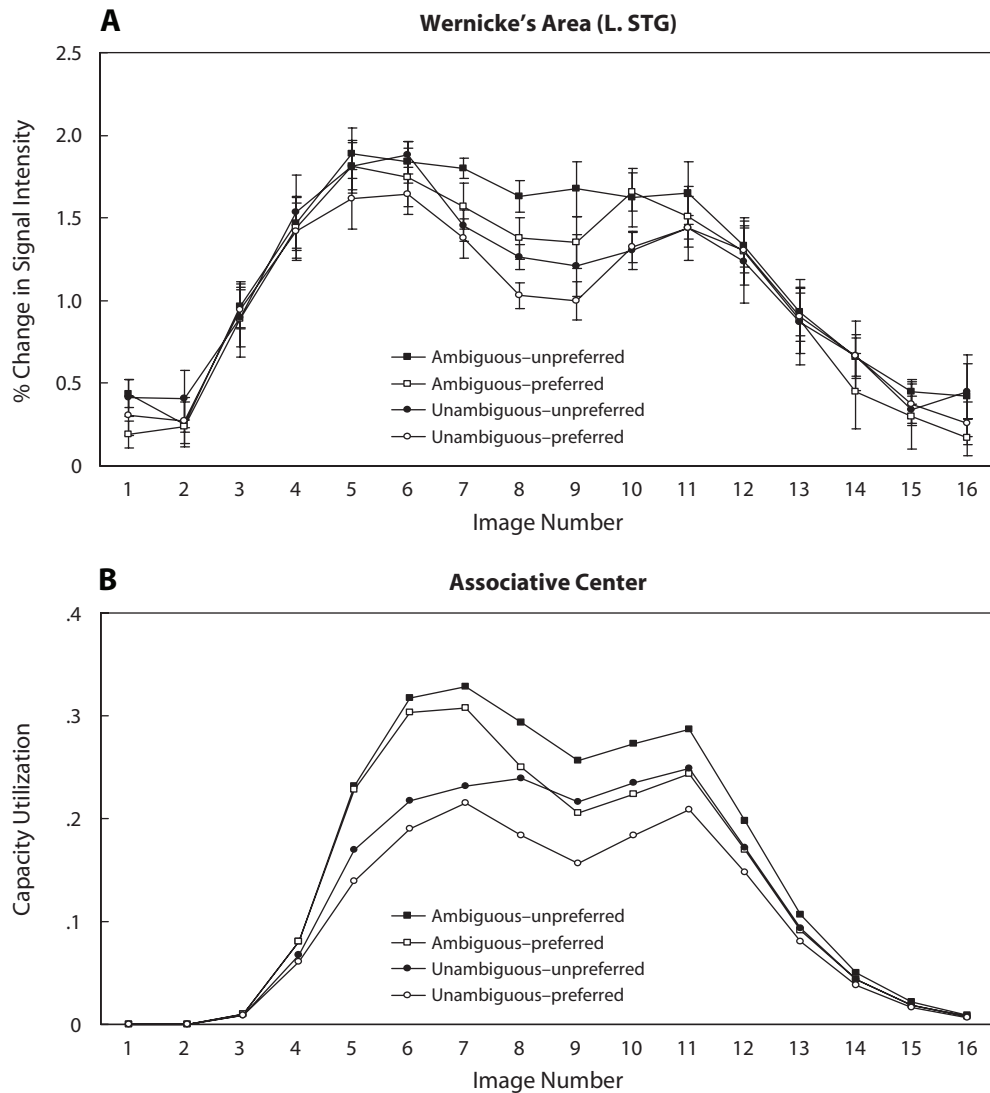
Figure 12. (A) fMRI activation time series observed in Wernicke's area (from Mason et al., 2003) during comprehension of ambiguous and unambiguous sentences that are resolved in favor of either the preferred or the unpreferred interpretation. (B) The predicted fMRI time series of the corresponding Associative center of the sentence comprehension model.

measured by fMRI), specifically for mean activation levels as measured in block-design studies and moment-to-moment fluctuations as measured in event-related studies. It also accounts for the pattern of activation after stroke-induced lesioning of a left-hemisphere language area—specifically, the contralateral takeover of cognitive function. The model's empirical success follows largely from the operating principles embodied in the underlying 4CAPS architecture: The language network adapts to changes in resource availability within centers and to changes in the functional demands of sentence comprehension. In particular, the same resource allocation algorithm explains both the relatively transient recruitment of right-hemisphere areas when comprehending structurally complex sentences (Just, Carpenter, Keller, et al., 1996)

and the relatively permanent recruitment of these areas following damage to their left-hemisphere homologues (Thulborn et al., 1999).

IV. 4CAPS MODELS OF SPATIAL PROBLEM SOLVING

This section describes a general model of spatial problem solving. The general model synthesizes two existing theories: a computational theory of human problem solving and a neuropsychological theory of executive function. The general model is instantiated in two domains, Tower of London (TOL) problem solving and mental rotation of Shepard–Metzler (1971; henceforth "SM") figures. In addition to extending the scope of 4CAPS beyond sentence

A   fMRI Data (Broca/L. IFG Lesion)

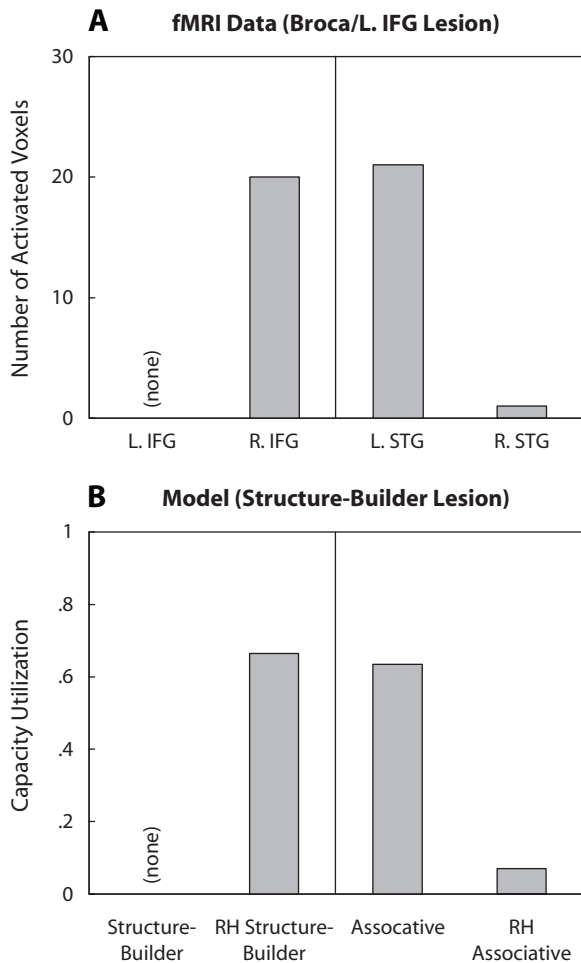**B   Model (Structure-Builder Lesion)**

Figure 13. (A) The number of voxels activated during sentence comprehension in four areas of the brain of a stroke patient (from Thulborn et al., 1999). The patient's stroke produced a lesion in Broca's area (L. IFG)—hence, the complete absence of activation there. (B) The capacity utilizations of the corresponding centers of the sentence comprehension model. These simulations were produced by "lesioning" the Structure-Builder center (i.e., depleting its resources)—hence, the complete lack of capacity utilization there.

comprehension, the TOL and mental rotation models illustrate how the operating principles account for another dynamic property of large-scale cortical networks: shifting strategies with increasing task difficulty.

The general model inherits its problem-solving mechanisms from Soar (Newell, 1990), a computational account of human problem solving whose sufficiency has been demonstrated by successful models of a range of tasks (Rosenbloom, Laird, & Newell, 1993). These mechanisms include a hierarchy of *goals* that guide proposal of *operators*, selection between operators based on *preferences*, and application of the selected operator to the current *state* to produce a new current state. The general model also inherits elements from Shallice's (1982) theory of executive function, in which the decision of what to do next (i.e., which operator to select and apply) results from the interplay of two streams of cognitive control. The *routine*

stream is parallel, bottom up, and perceptually driven. It is driven by low-level schemas that are triggered by new perceptual inputs. By contrast, the *nonroutine* stream is serial, top down, and goal driven. It is driven by high-level schemas that comprise the supervisory attention system (SAS). Resolving the two streams of control is the job of the contention scheduler. Shallice and Burgess (1996) attribute the routine and nonroutine streams of control in a general way to posterior and prefrontal areas, respectively.

Neither Soar nor Shallice's (1982) theory alone provides an adequate basis for the 4CAPS general model of spatial problem solving. Although Soar offers a comprehensive account of problem solving, it is unclear how to map its computational mechanisms to brain function. Conversely, although Shallice has localized the SAS and contention scheduler to brain areas in a general way, it is unclear whether these constructs are powerful enough to account for complex problem solving (as opposed to simple forms of executive function). We have therefore synthesized them into a single model that inherits their respective strengths, offsets (some of) their respective weaknesses, and situates them within the resource-constrained computational environment of 4CAPS.

The 4CAPS general model of spatial problem solving uses the computational mechanisms of Soar to implement the routine and nonroutine control streams of Shallice's (1982) theory. We refer to the routine stream as the *perceptual mode* and the nonroutine stream as the *strategic mode*, for reasons that will become clear below. The perceptual mode is driven by visuospatial analysis, which results in the proposal of perceptual operators that increase the configurational similarity between the current problem state and the desired end state. The perceptual mode alone can solve simple problems—that is, those whose starting states can be transformed into their ending states by pure hill-climbing (using the terminology of artificial intelligence). Its limited competency is consistent with the success of patients with frontal lobe lesions on simple tests of executive function (Owen, Downes, Sahakian, Polkey, & Robbins, 1990). However, if an operator is selected but cannot be applied because some of its preconditions are not satisfied, the perceptual mode *impasses* and the strategic mode assumes some control. The strategic mode proposes goals to establish the unsatisfied preconditions and strategic operators that accomplish these goals. If the preconditions of a strategic operator are themselves unsatisfied, then goals to establish them are recursively proposed, producing the goal hierarchies characteristic of complex problem solving. As the strategic mode progressively takes actions to satisfy outstanding preconditions, the perceptual mode gradually regains control, and the original perceptual operator is applied.

Four centers compose the 4CAPS model of spatial problem solving, as shown in Figure 14. The RH Spatial and LH Spatial centers correspond to right- and left-hemisphere superior parietal cortex, respectively, including superior parietal lobule (SPL) and, particularly, the intraparietal sulcus area (IPS). We propose that, broadly speaking, RH Spatial is functionally specialized for directing visuospatial attention and LH Spatial for the con-

struction, transformation, and suppression of visuospatial representations (Carpenter et al., 1999; Chelazzi & Corbetta, 2000; Mesulam, 1990). RH Spatial is the seat of the perceptual mode. It compares the current and ending states and proposes perceptual operators that increase the configurational similarity between them (without regard for whether the preconditions of these operators are satisfied). LH Spatial serves as the visuospatial workspace of the model. The starting and ending states become available to it from visual/perceptual processing that is outside the scope of the model (presumably performed by occipital cortex). LH Spatial applies the selected operator to the current state, if possible, producing a new current state. This center also maintains a record of the intermediate states generated during problem solving. Implicit in this record is information required to perform certain spatial problem-solving tasks, such as administrations of the TOL task in which participants are asked to determine the minimum number of moves required to solve a problem. The proposed functional distinction between the RH Spatial and LH Spatial centers, and thus between the corresponding superior parietal areas, is based on evidence from a variety of neuroimaging and neuropsychological studies; for more background on this model, see Newman et al. (2003).

The RH Executive and LH Executive centers correspond to right- and left-hemisphere DLPFC, respectively, including medial prefrontal cortex. RH Executive serves a role analogous to the SAS of Shallice's (1982) theory. It is specialized for plan formulation, which it accomplishes via hierarchical planning (to again use the terminology of artificial intelligence), decomposing difficult problems into conjunctions of simpler problems through the proposal of goals. It also proposes strategic operators that achieve these goals. As the seat of the strategic mode, it is responsible for the creation, maintenance, and suppression of goals and strategic operators. Plan formulation is distinguished from plan execution, for which LH Executive is specialized. It selects between the operators proposed by other centers, including the perceptual operators of RH Spatial and the strategic operators of LH Executive. Selection is via Soar-like productions that assert heuristic preferences about the relative goodness of pairs

of operators (e.g., an operator that achieves a goal is preferred to one that does not). This center also consolidates over preferences to select the most preferred operator. In adjudicating between perceptual and strategic operators, and therefore between the perceptual and strategic control modes, LH Executive serves the same functional role as the contention scheduler of Shallice's theory. The differential specializations attributed to RH Executive and LH Executive, and thus to right and left DLPFC, are based on a number of empirical findings. For example, plan formulation in the Tower of Hanoi (TOH) puzzle is more impaired in patients with right prefrontal damage than in those with left prefrontal damage (Morris, Miotto, Feigenbaum, Bullock, & Polkey, 1997a). By contrast, when multiple solution sequences exist and the conflict between them must be resolved, plan execution is more adversely affected in patients with left prefrontal lesions than in those with right prefrontal lesions (Morris, Miotto, Feigenbaum, Bullock, & Polkey, 1997b). Additional evidence for the proposed functional distinction lies in their differential patterns of functional connectivity with parietal areas (Newman et al., 2003). (Source code for the general model of spatial problem solving is available at the 4CAPS Web site. Varma [2006] provides a detailed prose description of the productions.)

The key property here of the general model is the smooth and dynamic transfer of control between the perceptual and strategic modes. The two modes are not mutually exclusive; the model does not remain in either one or the other, and it does not shift abruptly between them. Even when goals are present in RH Executive and that center proposes strategic operators to achieve them, RH Spatial continues to propose perceptual operators. Selecting between strategic and perceptual operators—resolving the conflict between the two control modes—is a function that LH Executive performs dynamically at each point in time as a consequence of the operating principles embodied in 4CAPS.

The TOL Model

The TOL task seems particularly appropriate for evaluating cognitive neuroarchitectures because it has been investigated using both behavioral and brain imaging meth-
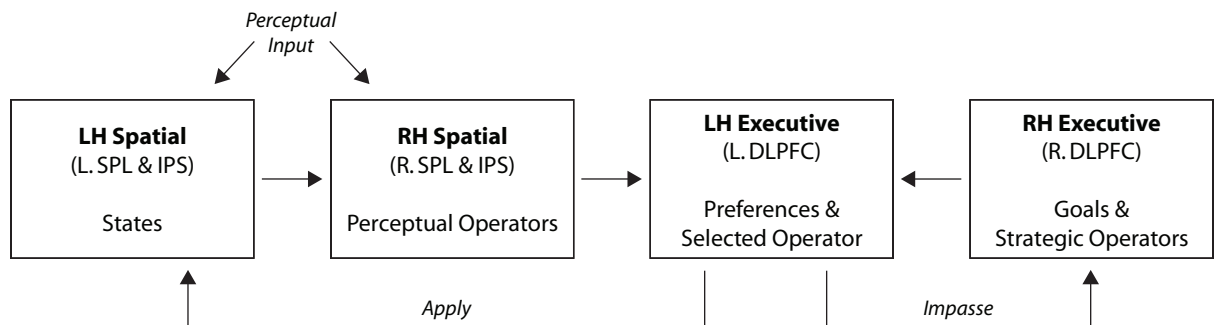

Figure 14. 4CAPS model of spatial problem solving.

ods with both normal and patient populations. A sample TOL problem is shown at the bottom of Figure 4C. It is defined by a starting configuration (on the left) and an ending configuration (on the right). Each configuration consists of three pockets of varying depths accommodating between one and three balls and a distribution of three balls across the pockets, with no pocket containing more balls than its depth permits. Solving a TOL problem (optimally) requires transforming the starting configuration into the ending configuration via a (minimum-length) sequence of legal moves. A legal move transfers a ball from the topmost position of one pocket to the topmost position of another pocket with available space. A variant of the TOH task, TOL is more useful for evaluating patients with frontal lesions because its instructions are simpler and it permits finer gradations in problem difficulty. Algorithms for solving TOL problems have not been formally specified and empirically documented, as they have for TOH problems (see, e.g., Simon, 1975); there exist only a handful of models, each employing a different ad hoc strategy (Dehaene & Changeux, 1997; Polk, Simen, Lewis, & Freedman, 2002). For the TOL model, we have adapted a strategy used in the TOH task (the so-called *sophisticated perceptual strategy*) that includes both strategic and perceptual processing, so that the smooth shift between the two control modes can be illustrated. A description of this strategy, the productions that realize it, and their distribution over the four model centers are provided in Appendix A of Newman et al. (2003). (The source code for the TOL model, as well as a related TOH model, is available from the 4CAPS Web site.)

To illustrate the strategy and its execution by the TOL model, a problem-solving trace is shown in Figure 15. In this trace, the perceptual mode falters because of an unsatisfied precondition, and there is a partial shift to the strategic mode to establish this precondition (and a partial shift back after the precondition is established). During the first cycle, RH Spatial compares the current configuration (which is also the starting configuration) with the ending configuration and proposes two perceptual moves that potentially increase the similarity between them: Move A would move the white ball to the middle pocket, and Move B, the gray ball to the middle pocket. During the second cycle, LH Executive asserts a preference for B over A on the basis of the TOL heuristic that it is better to move balls into "deeper" positions in the ending configuration than into "shallower" positions. During the third cycle, LH Executive selects B on the basis of this preference. However, B cannot be performed, because one of its preconditions—that there be no ball above the one to be moved—is not satisfied. An impasse occurs, and on the fourth cycle, RH Executive activates a goal to resolve the impasse by unblocking the gray ball. This marks a partial shift to the strategic mode, with its greater executive supervision.

During the fifth cycle, RH Executive proposes a strategic Move C to resolve the impasse. C makes progress toward a goal, but does not increase the perceptual similarity between the current and ending configurations. Also on the fifth cycle (because the model has not completely shifted to the strategic mode), RH Spatial concurrently proposes the same two perceptual moves it proposed during the first cycle. During the sixth cycle, LH Executive asserts preferences concerning the three moves. The strategic Move C emerges as most preferred, and on the seventh cycle it is selected. In this way, LH Executive adjudicates between the perceptual and strategic controls modes on a moment-by-moment basis. Because all of C's preconditions are satisfied, LH Spatial performs it on the eighth cycle, transforming the current configuration into a new current configuration. Notice that the gray ball is now unblocked (i.e., the goal has been achieved). Therefore, this goal will be suppressed on the next cycle (not shown) and the influence of the strategic mode will slowly decline, an important dynamic property of 4CAPS. Perceptual-mode processing is sufficient to solve the problem in two additional moves (also not shown).

Empirical evaluation. The TOL model was evaluated against the behavioral and brain imaging data that Newman et al. (2003) collected from normal young adults who solved TOL problems with (optimal) solutions of between one and six moves. Problems were presented in blocks of increasing difficulty: "easy" blocks consisted of five 1- and 2-move problems and one 3-move problem; "moderate" blocks consisted of one 1-move problem and three 3- and 4-move problems; and "difficult" blocks consisted of one 1-move problem and two 5- and 6-move problems. Both the human problem solvers and the TOL model displayed increasing solution times with increasing problem difficulty, as shown in Figure 16A, with a correlation of .96 ($p < .01$) between them.[4] It is perhaps not surprising that the more moves a problem requires, the more cycles of processing the model requires. What is surprising is that the model produces this linear relation while smoothly shifting from predominantly perceptual-mode processing on simpler problems to a mixture of perceptual- and strategic-mode processing on more difficult problems. Moreover, this shift is dynamic. It is not mandated in a top-down, all-or-nothing fashion by an executive controller. Rather, it emerges on a moment-by-moment basis as LH Executive selects between the perceptual moves proposed by RH Spatial and the strategic moves (if any) proposed by RH Executive.

The capacity utilizations of the four model centers can be compared with the activations in the corresponding cortical areas—left and right DLPFC and superior parietal cortex—for easy, moderate, and difficult blocks. The brain activations are shown in Figure 16B and the capacity utilizations in Figure 16C. Both right and left DLPFC show a reliable increase in activation with increasing block difficulty. The corresponding model centers, RH and LH Executive, show a similar increase in capacity utilization. Although it is tempting to dismiss the increasing capacity utilization as an overall difficulty effect, the model offers a more nuanced explanation. The increase in RH Executive is due to the increasing need for plans to establish unsatisfied preconditions, which requires the resource-consuming creation and maintenance of more and more goals and strategic operators. The increase in LH Executive is due to the increasing difficulty of controlling plan execution, because selection is between not

Problem Perception	Cycle 1 Propose Moves	Cycle 2 Assert Preferences	Cycle 3 Select Move	Cycle 4 *Activate New Goal*	Cycle 5 Propose Moves	Cycle 6 Assert Preferences	Cycle 7 Select Move	Cycle 8 Perform Move
	RH Spat	**LH Exec**	**LH Exec**	**RH Exec**	**RH Spat, RH Exec**	**LH Exec**	**LH Exec**	**LH Spat**
Starting/Current Puzzle Configuration → Ending Puzzle Configuration	Perceptual Move A / Perceptual Move B	Perceptual Move B > Perceptual Move A	Selected Move: B	*Unblock Ball ○ Goal*	Perceptual Move A / Perceptual Move B / *Unblock Ball Mv C*	Perceptual Move B > Perceptual Move A / *Unblock Ball Mv C* > Perceptual Move A / *Unblock Ball Mv C* > Perceptual Move B	Selected Move: C	Current Puzzle Configuration
	Propose two perceptual moves, A and B, on the basis of surface differences between the current and ending configurations.	Assert preferences between perceptual moves A and B.	On the basis of the preference, select the most preferred move (B).	Cannot perform the selected move, so activate a goal to establish unsatisfied precondition.	Propose the same two perceptual moves as before plus a strategic move C based on the goal.	Assert preferences between all pairs of perceptual and strategic moves.	On the basis of the preferences, select the most preferred move—the strategic move C.	Perform the selected move C, producing a new current configuration.

Perceptual Mode   **Perceptual and Strategic Modes**

Figure 15. A trace of the Tower of London model, illustrating the smooth transfer of control between the perceptual and strategic modes.

just perceptual moves, but both perceptual and strategic moves, which requires the resource-consuming creation of more and more preferences.

Next, consider the superior parietal data, which are important because they do not show an overall difficulty effect, and therefore represent a more difficult test of the TOL model. There is a reliable increase in activation with problem difficulty in left superior parietal cortex, but no (reliable) effect in its right-hemisphere homologue, as shown in Figure 16B. Qualitatively speaking, the model reproduces these patterns: The capacity utilization of LH Spatial increases with problem difficulty, whereas that of RH Spatial is relatively flat, as shown in Figure 16C. More importantly, the model offers a principled explanation for the differential activation patterns of the two superior parietal areas. LH Spatial is the visuospatial workspace of the model, and as such must create and maintain all intermediate puzzle configurations as a record of the number of moves required to solve the problem—the required re-

sponse in the Newman et al. (2003) administration of the TOL task. As the number of moves increases, so does the number of intermediate puzzle configurations, and therefore the resource demands on this center. By contrast, the resource demands on RH Spatial are relatively flat (or even decline slightly) with problem difficulty. As the seat of the perceptual control mode, this center is always proposing perceptual moves that increase the similarity of the current and ending puzzle configurations. Because the number of such moves is relatively constant with increasing problem difficulty (or even declines slightly as the visual similarity between start and goal becomes less salient), the capacity utilization of this center is relatively flat, which is consistent with the statistically flat activation profile displayed by the corresponding right superior parietal area. The qualitative match between the data and model for the 12 points of comparison in Figures 16B and 16C is consistent with the .96 ($p < .01$) correlation between them.

A Behavioral Performance
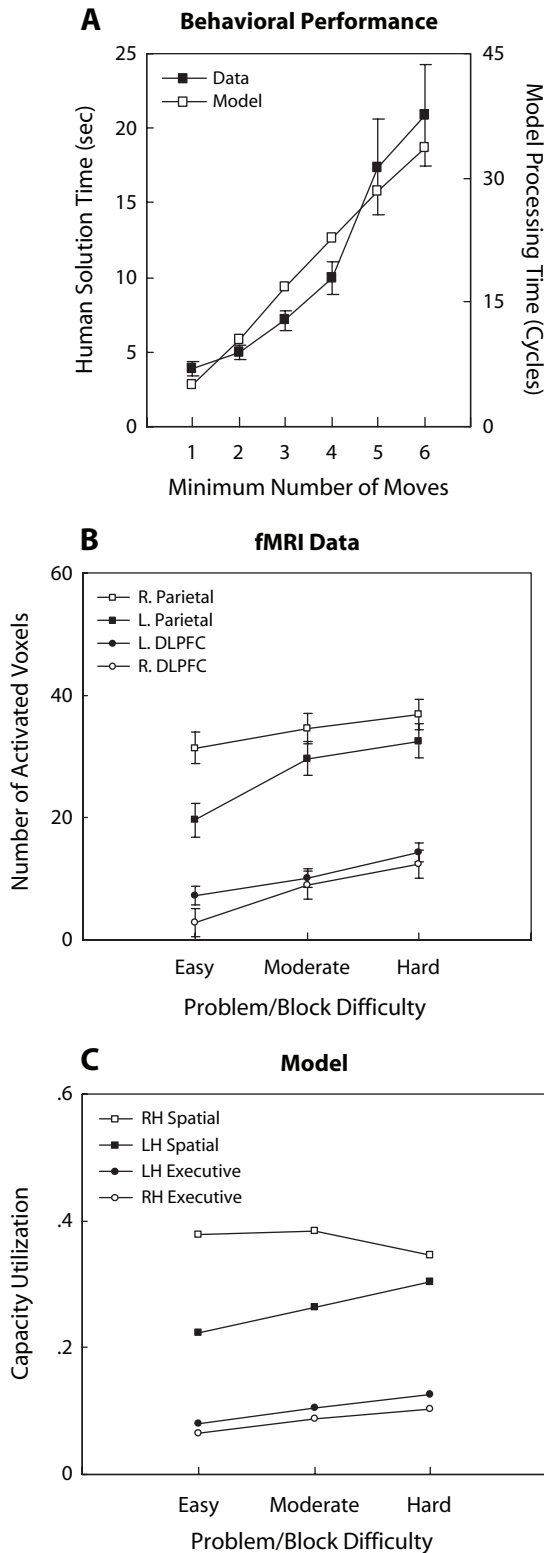

B fMRI Data

C Model

Figure 16. (A) Human solution times (from Newman et al., 2003) and Tower of London (TOL) model processing times on blocks of problems of increasing difficulty. (B) fMRI activations of left and right superior parietal cortex (including the intraparietal sulcus) and DLPFC (from Newman et al., 2003). (C) Capacity utilization of the corresponding LH and RH Spatial and Executive centers of the TOL model.

In sum, the 4CAPS model of TOL problem solving accounts well for the behavioral and brain imaging data of one study (Newman et al., 2003). Its important contribution here is illustrating a property of large-scale cortical networks not seen before: the smooth shift between perceptual and strategic control in response to task difficulty. (This property is also displayed by a 4CAPS model of the closely related TOH task that accounts for relevant behavioral and brain imaging data collected from normal young adults and patients with frontal lesions [Varma, 2006].)

The Mental Rotation Model

A second 4CAPS spatial model performs the Shepard–Metzler (1971) mental rotation task and accounts for the fMRI data in this task (an increase in parietal activation with rotation angle; Carpenter et al., 1999). The participants' task is to decide whether two figures are the same or different in their structure, except for a rotation, as shown at the bottom of Figure 4B. Like the TOL model (and the sentence comprehension model before it), the mental rotation model displays increasing capacity utilization with increasing task difficulty. The model is notable for the selectivity of this increase, confined to just the spatial centers, a prediction that can be evaluated against neuroimaging data.

The mental rotation problem consists of two SM figures, each composed of four segments joined end-to-end at 90° angles. The participant must decide whether the figures are identical (except for a rotation) or mirror images. The classic behavioral finding, that response time is a linearly increasing function of angular disparity, is commonly interpreted as indicating that the figures must be rotated into alignment so that they can be directly compared (Shepard & Metzler, 1971). The 4CAPS mental rotation model utilizes a strategy that Just and Carpenter (1985) formulated on the basis of a detailed analysis of eye movement behavior during rotation of SM figures (and instantiated in a CAPS model). The eye movements identified four successive processing stages:

- *initial search* for a pair of potentially corresponding segments of the figures;

- *transformation* of one segment by stepwise mental rotation until it is aligned with the other segment (the sequence of rotation steps comprises a rotation trajectory, which is stored);

- *subsequent search* for a second pair of potentially corresponding segments; and

- *confirmation* that when one segment of the second pair is rotated through the trajectory constructed previously, it is aligned with the other segment.

This strategy produces the classic linear relation between angular disparity and response time, because of the stepwise construction of a rotation trajectory during transformation and the stepwise application of this trajectory during confirmation.

The mental rotation model implements the four-stage strategy within the general model of spatial problem solving. It is composed of the same centers as the TOL model, which correspond to the same cortical areas and possess the same functional specializations. The mental rotation model differs from the TOL model in relying much less on the strategic-mode processing of the executive centers and more on the perceptual-mode processing of the spatial centers.

The mental rotation model solves an SM problem as follows. Initially, the SM figures that constitute the problem are encoded in LH Spatial by visual/perceptual processing outside the scope of the model. The representational scheme is that of Marr and Nishihara (1978). During initial search, RH Spatial examines the figures and proposes pairs of segments, one from each figure, that might be matching segments. During transformation, RH Spatial proposes rotation steps that reduce the difference in the orientations of the segments. Each step is of a fixed size, a parameter of the model that was fixed a priori at 10° (but within some range the parameter has little effect on the match to the fMRI data). Rotation steps are proposed and executed until the pair of corresponding segments is brought into orientation alignment. LH Spatial maintains a record of the rotation steps taken; these constitute the rotation trajectory. During the next, subsequent search, phase, RH Spatial examines the figures and proposes a second pair of segments, one from each figure, that might be corresponding segments. Finally, during confirmation, RH Spatial proposes rotation steps for one segment of the second pair using the rotation trajectory constructed during the transformation stage. RH Spatial's proposal of possible rotation steps and LH Spatial's execution and maintenance of the record of rotation steps consume resources proportional to the angular disparity.

The executive centers play minimal roles during mental rotation. At each point in time, RH Executive maintains a goal specifying the current stage of the strategy (i.e., initial search, transformation, subsequent search, and confirmation). LH Executive selects between potentially matching pairs of segments during initial search and subsequent search and between proposed rotation steps during transformation and confirmation. Resource demands in both centers are relatively low and relatively constant (i.e., independent of angular disparity). Of course, other brain areas, not included in the model, also contribute to mental rotation, such as inferior temporal areas that probably encode the detailed structure of the figures without showing a monotone relation between activation level and angular disparity (Carpenter et al., 1999), as well as additional executive areas (such as anterior cingulate).

Empirical evaluation. Normal young adults solved SM problems with angular disparities (for *same* trials) ranging from 0° to 120° (Carpenter et al., 1999). The mental rotation model provides a good account of the behavioral measures: The human response times and model processing times (which were obtained without fitting any parameters) are shown in Figure 17A. The close qualitative match is confirmed by a correlation of .99 ($p < .01$).

The fMRI-measured activation in left and right superior parietal cortex (including the IPS) increased approxi-

mately linearly with angular disparity, whereas activation in left and right DLPFC was minimal and relatively unaffected by angular disparity, as shown in Figure 17B. (These DLPFC activation data are from the original data analysis, and were not reported in Carpenter et al., 1999.) The mental rotation model matched this pattern. Capacity utilization in the spatial centers increased linearly with angular disparity, but was relatively flat in the executive centers, as shown in Figure 17C. The qualitative match between the data and the model for the 16 points of comparison in Figures 17B and 17C is supported by the .99 ($p < .01$) correlation between them. The fit of the model was obtained by estimating four free parameters, the resource capacities of the four centers. (These parameters are not responsible for the predicted differential response of the spatial and executive centers with increasing angular disparity; nor do they determine the slopes of the lines plotted in Figure 17C, only their intercepts. Nor is the rotation rate parameter, fixed a priori at 10°, responsible for the good fit; it can only affect the degree of a sloped line, not whether the line is sloped or flat.) The rotation model's fit is based on the differential pattern of increasing superior parietal activation but flat prefrontal activation found with increasing angular disparity. Thus, the 4CAPS model of mental rotation provides an account of a smooth and selective increase in resource consumption in just the spatial centers (corresponding to superior parietal cortex) with increasing angular disparity.

Summary of the Spatial Problem-Solving Models

The distinguishing characteristic of the models is their ability to smoothly shift from one mode of processing to another as the functional and resource demands of task performance require. The TOL model exhibits a smooth shift from perceptual-mode processing to strategic-mode processing as task difficulty increases. The signature of this smooth shift is increasing capacity utilization in the executive and LH Spatial centers, and thus the corresponding bilateral prefrontal and left parietal areas, with increasing task difficulty. Notably, there is no need for a central executive controller to mandate the shift. The mental rotation model displays a smooth increase in the resource demands of the spatial centers, and thus the corresponding superior parietal areas, with increasing angular disparity. Both forms of dynamic adaptation to changing demands follow from the general 4CAPS operating principles.

V. DUAL-TASK PERFORMANCE AND INTERCENTER CONSTRAINTS

This section focuses on the 4CAPS operating principle concerning the resource constraints that affect the system as a whole rather than individual centers (or brain areas). We refer to such constraints as *intercenter* resource constraints. The claim that cognition is the emergent product of a large-scale cortical network naturally implies informational interdependencies among the network's nodes. Intercenter constraints place a limit on these interdependencies in a way that is separate from the limitations on
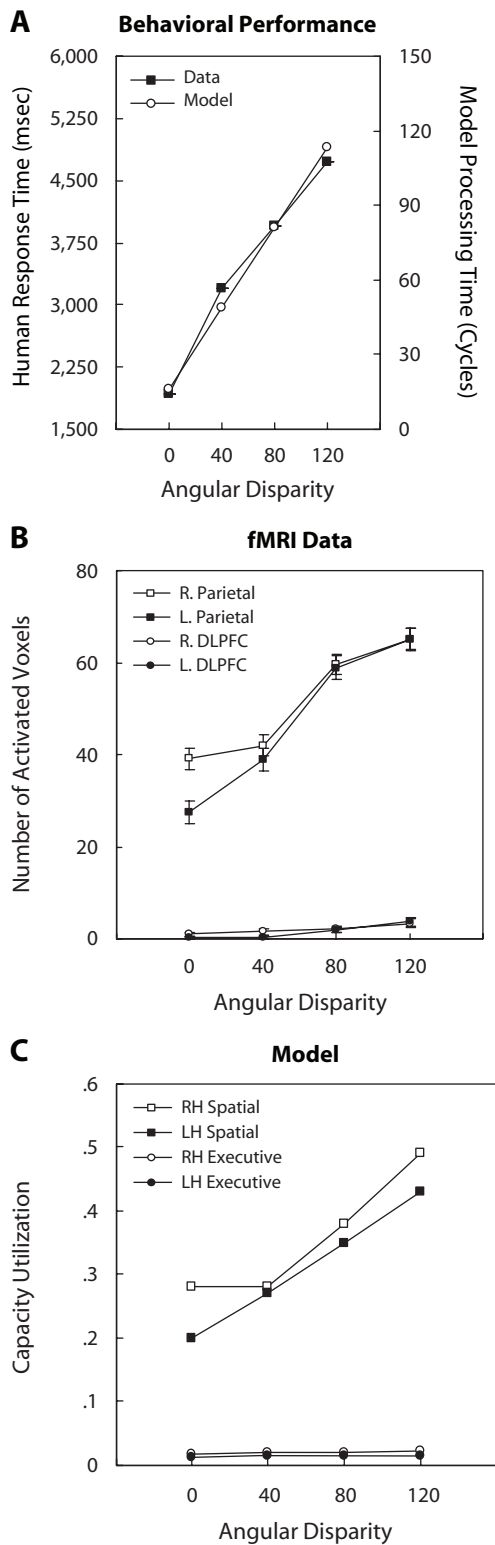
Figure 17. (A) Human response time (from Carpenter et al., 1999) and mental rotation model processing times for solving Shepard–Metzler (1971) problems of increasing angular disparity. (B) fMRI activation of left and right superior parietal cortex (including the intraparietal sulcus) and DLPFC (from Carpenter et al., 1999). (C) Capacity utilization of the corresponding LH and RH Spatial and Executive centers of the mental rotation model.

the nodes themselves. Specifically, 4CAPS proposes intercenter constraints, interpreted as bandwidth limitations, that govern intercenter communication. (Alternate interpretations of intercenter constraints will be considered at the end of this section.) Under an intercenter resource constraint, multiple centers draw their resources from a shared and limited supply. As a result, the resources available to any one center can depend on the resource demands of other centers. In particular, if the joint resource demands of centers governed by the same intercenter constraint exceed the available supply, the demands of individual centers will not be fully satisfied. This section evaluates the role of intercenter constraints in the context of a 4CAPS model of complex dual-tasking.

Specifically, we consider studies of dual-tasking in which two complex tasks are performed concurrently, and in which the two component tasks are underpinned by largely independent cortical networks. Because each cortical network requires communication bandwidth to link its constituent centers, concurrent use of both networks imposes heavy demands on the shared communications infrastructure. We can therefore evaluate the necessity of intercenter constraints by determining whether 4CAPS models that include such constraints fit the dual-tasking data better than those that do not. (Additional data sources for evaluating intercenter constraints will be considered in the General Discussion.)

There are many cognitive neuroimaging studies of dual-tasking, and not surprisingly, they find different effects. By *effect*, we mean the relation between the pattern of activation observed during dual-tasking versus that observed during single-tasking. We focus here on perhaps the most surprising of these effects, *underadditivity*, which occurs when the activation when performing two tasks concurrently is *less* than the sum of the activation when performing each task in isolation (see, e.g., Just, Carpenter, Keller, et al., 2001; Loose, Kaufmann, Auer, & Lange, 2003; Newman, Keller, & Just, 2007; Szameitat, Schubert, Müller, & von Cramon, 2002). As described below, underadditive activations during dual-tasking can be explained by intercenter resource constraints. (Other effects that have been observed in studies of dual-tasking, and the conditions under which 4CAPS can produce these effects, will be briefly considered at the end of this section.)

Two 4CAPS models of dual-tasking are described below. The *base* model represents the hypothesis that intercenter resource constraints are not necessary, whereas the *augmented* model represents the simplest possible alternative hypothesis: that all centers draw their resources from a single shared supply, interpreted as a bandwidth limitation on communication within the cortex. Only the augmented model will prove capable of accounting for the underadditive effects found in some studies of complex dual-tasking.

The dual-task fMRI data that were modeled come from a study that used two component tasks that draw on independent networks (Just, Carpenter, Keller, et al., 2001). The language task was verification of auditorily presented sentences, such as *Botany is a biological science and it*

A   fMRI Data

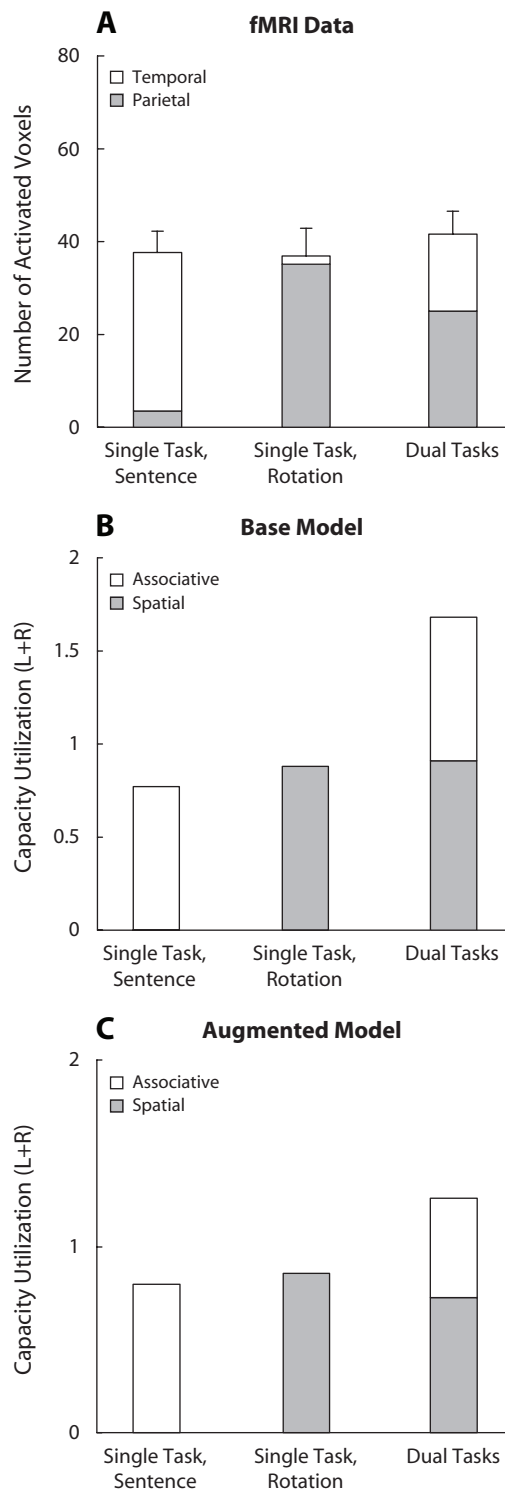## B   Base Model

C   Augmented Model

Figure 18. (A) Underadditive activation in the dual-task condition (from Just, Carpenter, Keller, et al., 2001). (B) The base dual-task model, with no intercenter resource constraints, shows additive capacity utilization. (C) The base model augmented with a single Cortex intercenter constraint shows underadditive capacity utilization.

deals with the life, structure, and growth of plants. The spatial task was the SM mental rotation task described

above. Participants performed the language and spatial tasks separately in single-task conditions and concurrently in a dual-task condition. It is important to note that each component task requires many seconds of processing for each item, and therefore dual-tasking requires concurrent execution of two threads of thought; this contrasts with dual-tasking studies that employ simple component tasks, and that can be performed by rapidly switching between quickly processed items. The activations of bilateral temporal and bilateral parietal areas during single- and dual-tasking (shown earlier in Figure 3) are reproduced in Figure 18A. (Because no interesting laterality effects were observed, the activations were summed across hemispheres. Also, because the activations of frontal areas were relatively low and constant, they are not shown.) In the language task performed in isolation, posterior STG and posterior middle temporal gyrus (MTG) showed bilateral activation, consistent with other fMRI studies of auditory language comprehension (Michael, Keller, Carpenter, & Just, 2001; Schlosser, Aoyagi, Fulbright, Gore, & McCarthy, 1998), whereas the parietal areas were not activated. The opposite pattern was observed during the spatial single task (there was activation in bilateral parietal areas but not in posterior STG/MTG), consistent with the attribution of visuospatial processing to parietal areas, as described above (Carpenter et al., 1999; Cohen et al., 1996). In the dual-tasking condition, in which both tasks were performed concurrently, activation was observed in both temporal and parietal areas, but it was 25% to 40% less (respectively) than the sum of the activations observed in the single-task conditions. The new phenomenon to account for is the underadditivity.

Dual-Task Models

The base model is simply the conjunction of the sentence comprehension (with auditory input) and mental rotation models described above. It consists of the eight centers that together comprise these models: Associative and RH Associative correspond to left posterior STG/MTG (i.e., Wernicke's area) and its right-hemisphere homologue, respectively; Structure-Builder and RH Structure-Builder correspond to left inferior frontal gyrus (IFG; i.e., Broca's area) and its right-hemisphere homologue, respectively; LH Executive and RH Executive correspond to left and right DLPFC, respectively; and LH Spatial and RH Spatial to left and right superior parietal areas (SPL/IPS), respectively. No additional functional specializations were specified for these centers, and no other centers (corresponding to additional brain areas) were added. The distinguishing characteristic of the base model is the lack of intercenter resource constraints.

The augmented model extends the base model with a single intercenter constraint on the joint resource consumption of all eight centers. This intercenter constraint is interpreted as a bandwidth limitation on intercenter communication, and therefore on collaborative processing, within the entire cortex. The proposal is that (1) a shared communications infrastructure exists, (2) centers belonging to the two independent networks must nevertheless communicate using the shared infrastructure, and (3) the

limited bandwidth of this infrastructure impairs collaborative processing *within* the networks.

Empirical Evaluation

The eight intracenter resource supplies (one for each center) and, in the case of the augmented model, the single intercenter resource supply were chosen to maximize the R^2 between the capacity utilizations of the model centers and the activations of the corresponding brain areas across the three conditions of the Just, Carpenter, Keller, et al. (2001) study. The capacity utilizations of contralateral homologue centers (e.g., Associative and RH Associative) were summed across hemispheres, corresponding to the aggregation of the fMRI data. In the base model, which has no intercenter constraint, the capacity utilizations of the associative and spatial centers (corresponding to temporal and parietal areas, respectively) are additive from the two single-task conditions to the dual-task condition, as shown in Figure 18B. Thus, the base model fails to capture the critical underadditivity of the data, as a comparison of Figures 18A and 18B reveals (although the R^2 between the model and data is still 79%, $p < .02$). Next, consider the augmented model, with its cortexwide intercenter constraint. The capacity utilizations of its associative and spatial centers across the three conditions are shown in Figure 18C. Unlike the base model, the augmented model displays underadditive capacity utilizations in the dual-task condition. It also fits the fMRI data quite well, both in the absolute sense of variance accounted for ($R^2 = 95\%$, $p < .01$). These simulations suggest that intercenter resource constraints are necessary for accounting for the underadditive activations observed in some studies of complex dual-tasking.

Sensitivity Analysis

The superior fit of the augmented model relative to the base model raises the question of which topological property of the intercenter constraints is critical for producing the observed underadditivity; in other words, which subset of centers must draw on the same intercenter resource supply (or supplies)? A sensitivity analysis was performed in order to answer this question. Four alternative augmented models were constructed, each possessing different intercenter constraints, and therefore making a different claim about the limited-bandwidth communications channels that support collaborative processing.

1. *Domain-based intercenter constraints*. Intercenter constraints might apply to sets of cortical areas specialized for common task domains, such as the linguistic domain. Therefore, a *linguistic* intercenter constraint was imposed on centers corresponding to areas that perform language processing (bilateral posterior STG/MTG and IFG), and a *spatial* intercenter constraint was imposed on centers corresponding to areas that perform controlled visuospatial processing (bilateral DLPFC and SPL/IPS).

2. *Hemisphere-based intercenter constraints*. An intercenter constraint might apply to each cortical hemisphere. Therefore, a *left-hemisphere* intercenter constraint was defined for centers corresponding to left-hemisphere areas

(left IFG, posterior STG/MTG, DLPFC, and SPL/IPS), and a *right-hemisphere* intercenter constraint was defined for centers corresponding to right-hemisphere areas (right IFG, posterior STG/MTG, DLPFC, and SPL/IPS).

3. *Lobe-based intercenter constraints*. An intercenter constraint might apply to each lobe of the cortex. Therefore, a *frontal* intercenter constraint was defined for centers corresponding to frontal lobe areas (bilateral DLPFC and IFG), a *temporal* intercenter constraint was defined for centers corresponding to temporal lobe areas (bilateral posterior STG/MTG), and a *parietal* intercenter constraint was defined for centers corresponding to parietal lobe areas (bilateral SPL/IPS).

4. *Posterior–anterior intercenter constraints*. Separate intercenter constraints might apply to posterior and anterior areas. Therefore, a *posterior* intercenter constraint was defined for centers corresponding to parietal and temporal areas (bilateral posterior STG/MTG and SPL/IPS), and an *anterior* intercenter constraint was defined for centers corresponding to frontal areas (bilateral DLPFC and IFG).

Each alternative augmented model was fit to the Just, Carpenter, Keller, et al. (2001) data. Specifically, intracenter and intercenter resource capacities were chosen that maximized the R^2 between the capacity utilizations of the model centers and the activations of the corresponding cortical areas across the three conditions.

The first three alternative augmented models (domain, hemispheric, and lobe based) account for the fMRI data as well as, but no better than, the base model, which has no intercenter constraints, as is shown in Table 1. This suggests that the intercenter resource constraints of these alternative augmented models lack the critical topological property. By contrast, the posterior–anterior model fits the data as well as the original augmented model, which includes a single cortexwide intercenter constraint on the joint resource consumption of all eight centers; the R^2 for the posterior–anterior model is 96%. The topological property common to the posterior–anterior and cortex models but missing from the domain, hemispheric, lobe, and base models is a constraint on the joint resource consumption of the associative and spatial centers—that is, of bilateral posterior STG/MTG and IPS/SPL. Note that this is *not* a matter of the best-fitting models having more free parameters; if this were the case, the lobe-based model, which posits three intercenter constraints, would account for the most variance. Rather, the posterior–anterior and cortex models provide the best accounts of the underadditive activations observed in this study of complex dual-tasking because they force the temporal and parietal areas to use a shared resource (interpreted as the bandwidth of the communications link to the frontal areas).

Other Forms of Dual-Tasking

Various types of dual-tasking effects have been observed in other studies. We briefly consider how 4CAPS might be related to these other effects. One such effect is *additivity*—when activation during dual-tasking is the sum of the activations during single-tasking (see, e.g., Jaeggi

Table 1
Fit of the Base and Augmented Models to the
Just, Carpenter, Keller, et al. (2001) Data

Model	R^2	Number of Intercenter Constraints
Base	79%	0
Cortex intercenter	95%	1
Domain intercenters	79%	2
Hemispheric intercenters	79%	2
Lobe-based intercenters	79%	3
Posterior–anterior intercenters	96%	2

et al., 2003). Another effect is *overadditivity*—when activation during dual-tasking is greater than the sum of the activations during single-tasking (e.g., Herath, Klingberg, Young, Amunts, & Roland, 2001). Additivity and overadditivity have been primarily observed in studies of executive function; these studies focused on the control mechanisms that support dual-tasking—such as task set switching, selective attention, and working memory—which have been localized to prefrontal areas. The studies typically use relatively simple component tasks and examine the extent to which dual-tasking recruits prefrontal areas (e.g., D'Esposito et al., 1995). If the tasks are truly performed concurrently but are simple enough, the resource demands never exceed the shared resource supply; the dual task thus will not show underadditivity, and can be accounted for by a base model. If the dual tasks are brief enough to allow participants to use task switching, this enables the staggering of the demands on the shared resource over time, so that the limitations of the shared resource are not broached. If the tasks draw on the same underlying networks, overadditivity can occur, a phenomenon yet to be addressed by a 4CAPS model. As findings cumulate to indicate the range of effects of dual-tasking, the resource-sensitive framework of 4CAPS provides a range of analytic tools for modeling a dual draw on resources.

VI. COMPARISON WITH OTHER COGNITIVE NEUROARCHITECTURES

This section compares 4CAPS with other cognitive neuroarchitectures that aim for similarly broad coverage of cognitive psychology, neuropsychology, and neuroscience. The comparison reveals which operating principles are shared among different neuroarchitectures and which are unique to 4CAPS. The comparison also extends past the operating principles to more general neuroarchitectural issues.

Other Cognitive Neuroarchitectures

First, we briefly review three other cognitive neuroarchitectures proposed in the literature that support models of behavioral and brain imaging data. The first is the *large-scale neural model* developed by Horwitz and his colleagues (Horwitz, Friston, & Taylor, 2000; Horwitz & Tagamets, 1999). The basic computational element of this model is an excitatory–inhibitory neuronal pair, which is intended to correspond to a cortical hypercolumn. A 9×9 grid of such elements constitutes a population, and

one or more populations are used to model each brain area, which are defined at approximately the same spatial scale as 4CAPS centers (i.e., individual gyri and sulci). The individual neuronal elements are hand-configured to implement the functions attributed to them on the basis of electrophysiological considerations, particularly from single-unit recording studies of monkeys (see, e.g., Funahashi, Bruce, & Goldman-Rakic, 1990). Neuroanatomical considerations guide how neuronal elements are connected within each area and across different areas. To simulate the results of a neuroimaging study, the model is given the same stimulus or stimulus sequence that participants receive. The activities of the neuronal elements are aggregated over populations and convolved with a hemodynamic response function to produce a predicted fMRI time series for each brain area, which is compared with the observed data. A large-scale neural model of the delayed-match-to-sample task has simulated the observed activation patterns in V1/V2, V4, inferior temporal cortex, and prefrontal cortex (Tagamets & Horwitz, 1998) and has also been used to account for the results of a transcortical magnetic stimulation study (Husain et al., 2002).

Another cognitive neuroarchitecture with explanatory goals similar to those of 4CAPS is the *synthetic PET* framework developed by Arbib and colleagues (see, e.g., Arbib, Bischoff, Fagg, & Grafton, 1994). Synthetic PET models are cast at the same spatial scale as 4CAPS models, with components that correspond to brain areas such as PFC and IPS. Each brain area is modeled as an array of neural elements, and each element is modeled as a leaky integrator. The functional specializations attributed to neural elements are derived from single-unit recording studies. Critical to synthetic PET is the notion of a *schema*, which is a collection of knowledge and skills that together implement a complex perceptual, cognitive, or motor function (Arbib, 2003). Each schema encompasses a number of disparate representations and processes, and its implementation is typically distributed over multiple brain areas. For example, the hand-shape recognition schema is distributed over a network of temporal areas that includes superior temporal sulcus, middle temporal gyrus, and medial superior temporal cortex. Conversely, each brain area contributes to multiple schemas. The distribution of a schema over multiple brain areas lends cortical information processing its collaborative flavor. The activation in a simulated brain area is computed by summing the activations of its individual neural elements.

The third cognitive neuroarchitecture we consider is *ACT-R*, which has been developed by Anderson and his colleagues (Anderson et al., 2004). They approach cortical information processing from a direction opposite the one used in the large-scale neural model and synthetic PET. Rather than build from a biological foundation of anatomical, electrophysiological, and imaging studies of brain function, Anderson et al. retrofitted an existing cognitive architecture to new results in cognitive neuroimaging. ACT-R is a production system interpreter consisting of a set of modules. Each module maps to a different brain area and is defined at approximately the same spatial scale as 4CAPS centers. Each module implements one general

function of a production systems interpreter, which translates to multiple domain-specific functions. There is a procedural module that selects which production rule to apply at each moment in time, a goal module that performs goal management functions, a declarative memory module that stores episodic and semantic information, and a number of sensory and motor modules. The internal state of each module is summarized by the contents of its buffer. Because a buffer can hold just one chunk, and because modules communicate by inspecting each other's buffers, ACT-R can be viewed as imposing a rudimentary limitation on intermodule communication. ACT-R modules do not collaborate as true peers, but instead are controlled by the procedural module, which inspects the buffers of the other modules, matches the highest-utility production rule against their contents, and applies this rule, potentially changing these contents. The activity of a module, indexed by the degree to which its representations are being processed, is used to predict activation in the corresponding brain area, both in an average sense and (when convolved with a hemodynamic response function) over finer increments of time.

Shared and Distinguishing Operating Principles

Each cognitive neuroarchitecture can be evaluated on whether it does or does not possess each of the operating principles embodied in 4CAPS. This is shown in Table 2. Surprisingly, a number of operating principles are shared by all four neuroarchitectures, but the rest are unique to 4CAPS. First, let us consider the shared operating principles. All four neuroarchitectures claim that cognition is the product of multiple collaborating brain areas. All four also attribute multiple cognitive functions (or schemas in the case of synthetic PET) to each brain area, in contrast to a localist or modular scheme that is limited to a one-to-one mapping. All four include constraints on interarea communication, although the nature of these constraints ranges from the limited pairwise connections of the large-scale neural model and synthetic PET, to the limited buffers of ACT-R modules, to the intercenter resource constraints of 4CAPS. Finally, all four use measures of the activity of model components to predict activation in the corresponding brain areas. These are the consensus operating principles of cortical information processing.

Next, consider the three operating principles that are unique to 4CAPS, and from which the empirical successes

of the sentence comprehension, TOL, mental rotation, and dual-tasking models derive. The first is that similar cognitive functions are performed by multiple brain areas. Note that this is not a claim about distributed implementation; all neuroarchitectures agree that what are unitary cognitive functions at the psychological level are realized at the cortical level by networks of collaborating brain areas. Rather, it is a claim about the redundancy of the neural implementation of cognitive functions.

The second operating principle unique to 4CAPS is that the computational power of individual cortical areas is limited. Cortical areas, like all biological systems, are subject to resource constraints, and these constraints help explain the patterns of activation observed in intact and lesioned brains as tasks become more difficult. 4CAPS extends the notion of intracenter resource constraints to that of intercenter resource constraints on the joint processing of multiple brain areas, which are interpreted here as bandwidth limitations on communication across a shared infrastructure. This is a more molar view of communication than the point-to-point channels of the large-scale neural model and synthetic PET and the shared buffers of ACT-R—and also the only one capable of accounting for the underadditive activations observed under certain dual-tasking conditions (e.g., when performing two complex tasks that rely on independent cortical networks; Just, Carpenter, Keller, et al., 2001).

The third operating principle that is unique to 4CAPS concerns the dynamic nature of collaborative processing. The activations of cortical areas fluctuate during the performance of a difficult task, which can take many seconds (or longer). 4CAPS explains these fluctuations as the result of changing resource availability and the changing functional demands of task performance. There are four general cases.

1. Increasing task difficulty can result in increasing resource consumption in centers well-specialized for pending cognitive functions. This form of dynamic response was illustrated in Figure 5B above. An example was seen in the mental rotation model, in which the capacity utilizations of the spatial centers, which are well specialized for visuospatial processing, increase linearly with angular disparity, paralleling the activations observed in the corresponding parietal areas.

2. Increasing task difficulty can require a qualitatively different type of processing. This form of dynamic re-

Table 2
Operating Principles Instantiated by Competing Cognitive Neuroarchitectures

Operating Principle	4CAPS	Large-Scale Neural Model	Synthetic PET	ACT-R
0. Multiple collaborating centers	Yes	Yes	Yes	Yes
1a. Multiple functions per center	Yes	Yes	Yes	Yes
1b. Multiple centers per function	Yes	No	No	No
2. Resource-constrained computation	Yes	No	No	No
3. Dynamic assignment	Yes	No	No	No
4. Bandwidth-limited communication	Yes	Yes	Yes	Yes
5. Capacity utilization	Yes	Yes	Yes	Yes

sponse, also illustrated in Figure 5B above, was exhibited by the TOL model, in which difficult problems require strategic problem solving, and hence increasing recruitment of the executive centers specialized for plan articulation and execution. This corresponded to the increasing activations observed in the corresponding cortical areas, right and left DLPFC.

3. If the difficulty of a task is increased to the point at which the resources of centers well-specialized for pending cognitive functions are exhausted, centers with secondary specializations for these functions (and possessing adequate resources) will be recruited into the large-scale cortical network. This form of dynamic response was illustrated in Figure 5C above. An example occurred in the sentence comprehension model, in which increasing syntactic complexity produced resource shortfalls in the well-specialized centers corresponding to Wernicke's and Broca's areas, causing their less-specialized right-hemisphere homologues to be recruited into the language network.

4. The occurrence of a lesion that drastically reduces the resources of centers well-specialized for pending cognitive functions will prompt recruitment of centers with secondary specializations for those functions (and possessing adequate resources). This form of dynamic entry, illustrated in Figure 5D above, was exhibited by the sentence comprehension model: Lesioning the Structure-Builder center to simulate the effect of a lesion to the corresponding cortical area (Broca's) resulted in the early recruitment of the RH Structure-Builder center into the large-scale network, matching the recruitment of the right-hemisphere homologue of Broca's area by a stroke patient.

The 4CAPS allocation algorithm, which assigns cognitive functions to centers at each point in time, provides a unified account of all four dynamic changes to the topologies of large-scale cortical networks.

Other Dimensions of Comparison

The various neuroarchitectures can be compared on the basis of other properties than the 4CAPS operating principles, as shown in Table 3. This comparison reveals two different clusterings of the neuroarchitectures. The first groups 4CAPS and ACT-R together, in opposition to the large-scale neural model and synthetic PET. 4CAPS and ACT-R share a number of points in common: Both have their origins in cognitive science; focus on the functional decomposition of cognition; employ hybrid computational mechanisms that combine symbolic mechanisms, such as variable binding, and connectionist mechanisms, such as activation-based representations; and target high-

level forms of cognition, such as language processing and complex problem solving. In contrast, the large-scale neural model and synthetic PET have been primarily shaped by the results of anatomical and physiological studies of primate brains; employ neuron-like computational elements (e.g., leaky integrators) and incorporate biological constraints such as the relative proportion of excitatory versus inhibitory connections between brain areas; and target low-level forms of cognition that are by no means unique to humans, such as grasping. The different emphases of the two clusters result in complementary strengths and weaknesses. Although the computational sufficiency of 4CAPS and ACT-R in accounting for complex forms of cognition is relatively well demonstrated, it is an open question whether their hybrid computational mechanisms will allow for plausible neural implementations. Conversely, although the computational mechanisms of the large-scale neural model and synthetic PET are consistent with what is known about neural information processing, whether these mechanisms can scale to explain language processing, problem solving, mathematical reasoning, and other forms of complex cognition is also an open question.

A different clustering of the neuroarchitectures emerges when the natures of brain areas and the collaborative patterns they enter into are considered. 4CAPS, the large-scale neural model, and synthetic PET all share the view that each brain area is a full-fledged processor, a "distributed system in which storage and data processing are intertwined" (Arbib, 2003, p. 33). This view is uncontroversial in neuroscience. By contrast, ACT-R is composed of a number of modules, each implementing a restricted kind of functionality: The declarative memory module provides only storage, the procedural module provides only processing, and the goal module provides only control. This very different conception of the computational power of individual brain areas is the result of backward-mapping a cognitive architecture that was originally developed with reference to the behavioral data of cognitive psychology. As a result, ACT-R posits a more questionable mapping of cognitive functions to brain areas. A related difference between the other neuroarchitectures and ACT-R concerns the locus of control within large-scale cortical networks. According to 4CAPS, the large-scale neural model, and synthetic PET, brain areas can collaborate in a peer-to-peer fashion, and control is therefore shared or heterarchical (see, e.g., Mesulam, 1990). In particular, 4CAPS posits that the matching and firing of productions (evoking relevant processes when their enabling

Table 3
Dimensions of Variation Between Competing Cognitive Neuroarchitectures

Operating Principle	4CAPS	Large-Scale Neural Model	Synthetic PET	ACT-R
Empirical basis	Cognitive/ functional	Biological/ structural	Biological/ structural	Cognitive/ functional
Computational mechanisms	Hybrid	Neural	Neural	Hybrid
Level of phenomena	High	Low	Low	High
Centers full-fledged processors?	Yes	Yes	Yes	No
Control	Heterarchical	Heterarchical	Heterarchical	Hierarchical

conditions arise) is a fundamental capability of cortical systems, distributed throughout the cortex. By contrast, in ACT-R, the procedural module (localized to a subcortical area, the basal ganglia) coordinates the interaction between the other modules, first accessing their buffers, then matching the production with the highest utility against their chunks, and finally firing this production. This hierarchical control scheme is another outcome of ACT-R's origins in symbolic cognitive science.

Summary

In comparison with competing cognitive neuroarchitectures, 4CAPS makes three unique claims: that similar functions are implemented by different brain areas, that computation with individual brain areas is resource constrained, and that the pattern of collaborative processing is a dynamic function of changing resource availability and changing functional demands. These operating principles are critical for the ability of 4CAPS models to account for the neuroimaging data on the complex cognition of normal adults and patients with focal lesions. However, these differences should not overshadow the many commonalities between 4CAPS and the large-scale neural model, synthetic PET, and ACT-R, which together may indicate a scientific consensus about neuroarchitecture.

GENERAL DISCUSSION

We conclude with a discussion of several insights on related research topics that follow from the new theory and are not readily visible from other perspectives.

Fluid Intelligence

It has been tempting in the past to attribute individual differences in fluid intelligence to supervisory processes associated with the prefrontal cortex (see, e.g., Duncan et al., 2000). However, 4CAPS suggests that fluid intelligence is due not only to the capabilities of individual areas (in the prefrontal cortex or elsewhere), but also to the dynamic flexibility with which these areas assemble themselves into collaborative networks in response to a novel cognitive challenge. In the 4CAPS perspective, the fluidity of intelligence lies not just in flexible supervisory software, but also in an adaptive organization of the underlying neural hardware. From this perspective, it is not surprising that the volumes of white and gray matter in the brain are correlated with g to similar degrees (Posthuma et al., 2002). Intelligence lies not only in the nodes, but also in the connections of the large-scale networks.

Intercenter Resource Constraints

The interpretation of intercenter resource constraints adopted here—as bandwidth limitations on intercenter communications—can be evaluated further. The measurement of functional connectivity can be related to various measures of structural connectivity, such as white matter morphometry and diffusion tensor imaging (DTI) of anatomical connectivity, that together document the communications infrastructure of the thinking human brain.[5] Such data can be used to further test the bandwidth limitations interpretation adopted here, particularly in certain patient populations. For example, Just, Cherkassky, Keller, and Minshew (2004) initially proposed an *underconnectivity* account of autism based on findings of lower functional connectivity in that population than in matched controls. More recent studies have demonstrated a correlation across participants with autism between their frontal–parietal functional connectivity and properties of their anatomical connectivity (i.e., the size of the relevant portion of the corpus callosum; Just, Cherkassky, Keller, Kana, & Minshew, 2007; Kana, Keller, Cherkassky, Minshew, & Just, 2006). These findings point to intercenter connectivity, both functional and anatomical, as a key part of the infrastructure of cognition. For example, if people with autism have reduced intercenter communication, they should be particularly impaired in dual-tasking versus single-tasking—an inference that has already been empirically demonstrated (García-Villamisar & Della Sala, 2002). This general approach—understanding "disconnection syndromes" as reductions in intercenter communications resources—naturally complements the understanding of the cognitive impairments following focal lesions as reductions in intracenter computational resources, as we illustrated earlier in the sentence comprehension model.

More complex modeling of intercenter constraints could take into account the hierarchical organization of the cortex. For example, the frontal lobe is composed of IFG and other areas; IFG is composed of BA 45 (pars triangularis) and other areas; and so on. If resources are shared within some such hierarchy of centers such that subordinate centers draw their resources from superordinate centers, this can then be modeled by nested intercenter resource constraints.

Dynamic Assignments to Centers

Recall that 4CAPS assigns responsibility for cognitive functions to centers in a way that minimizes overall resource consumption while satisfying as many resource demands as possible. 4CAPS currently casts the assignment problem as a linear programming problem and solves it "centrally" using the simplex algorithm. Solving the assignment problem in a centralized manner, however, seems neurally implausible. Therefore, an avenue for future research is to replace the present scheme with a more "local" scheme in which centers communicate with other centers in the same "neighborhood," taking on and passing off pending cognitive functions as their available resources wax and wane. A number of candidate schemes have been developed by operations and computer science researchers interested in the parallel solution of linear programming problems (Alon & Megiddo, 1994; Lustig & Rothberg, 1996; Maros & Mitra, 2000) and by artificial intelligence researchers interested in cooperative computation by resource-constrained agents (Wellman, 1993; Wooldridge & Dunne, 2006). A more distributed scheme for responsibility assignment would be more neurally plausible.

Plasticity

A leading developmental account of the ontogenetic development of brain area specializations fits well with the perspective offered here. It has been proposed that initial differences (in infancy) of cell types across brain areas may predispose some areas toward certain processing styles, and therefore toward certain kinds of cognitive functions (see Elman et al., 1996, among others). For example, an area containing neurons that are particularly suitable for making fine distinctions in the timing of stimuli may eventually become highly specialized for speech processing, because of the need to differentiate phonemes that differ in their fine timing characteristics. Other areas may also have attempted to process speech, and although they eventually lose in the competition to become the most specialized, some of these competing areas may retain a residual capability (i.e., a secondary specialization) to process speech, albeit less efficiently (perhaps requiring, e.g., the participation of more neurons and/or the consumption of more resources). This may be the developmental path that leads to redundancy of function across brain areas, but with different relative specializations, such as between the left and right homologous areas in the language network.

If the brain area most specialized for a cognitive function is damaged, especially early in its development, the mechanism for shifting that function to a less-specialized area is already in place. For example, if the left hemisphere becomes damaged early in development, homologous regions in the right hemisphere activate during the performance of language tasks (Booth et al., 1999; Müller et al., 1998). In 4CAPS, this shift is a natural consequence of the algorithm by which cognitive functions are assigned to centers on the basis of resource availability and relative specializations, as was demonstrated by the sentence comprehension model's account of the Thulborn et al. (1999) stroke lesion data. Another form of plasticity entails the development of a new strategy to perform a task without the participation of a damaged area (Frackowiak, Friston, Frith, Dolan, & Mazziotta, 1997; Grafman & Litvan, 1999). Whether this form of plasticity can be incorporated into the assignment algorithm 4CAPS currently uses is a question for future research.

Learning

Although plasticity is a general form of learning, we can also consider the term in its more narrow senses. One such sense is the strengthening and tuning of associations with experience. Although it is common to think of individual associations as residing in particular brain areas, 4CAPS suggests another form of associative learning—namely, the optimization of intercenter coordination. For example, in one fMRI study (Büchel et al., 1999), participants learned an association between pictures of objects and their location on the display, two attributes that are associated with two different cortical areas (inferior temporal and posterior parietal areas, respectively). The functional connectivity between these areas increased with learning, and the larger changes in functional connectivity over days of practice occurred at the same time as the larger changes in the behavioral measure of learning. The implication is that repeated coactivation of centers leads to more efficient forms of collaboration between them. An avenue for future research is therefore to formulate learning algorithms that operate at the intercenter level.

Another sense of learning is the induction of new strategies. 4CAPS suggests a new mechanism for strategy innovation—namely, the dynamic formation of novel and noncanonical large-scale networks. This can occur when there is an unusual shortfall of available resources, whether because of lesion or concurrent load. It can also occur when there is an unusual surplus of available resources, whether because of development or the reduction of a typical load. In both cases, a novel noncanonical network can be assembled to perform the task at hand—one composed of different brain areas, which therefore implements a new cognitive algorithm. Modeling the induction of new strategies as a result of resource shortfalls and surpluses is another avenue for future research.

Conclusion

We conclude with the observation that, despite dealing with neuroimaging data, this article contains no images depicting brain activation. Instead, there are many graphs depicting the *functional relation* between some facet of brain activation (such as its intensity level) and some hypothesized property of the underlying cognitive processing. The operating principles embodied in 4CAPS enable progress from visual depictions of activation phenomena (which are undeniably a stunning accomplishment of science and technology) to formal accounts of the mechanisms underlying the phenomena. This approach represents a first step in understanding how resource constraints and system dynamics shape cortical and cognitive information processing. Although the primary responsibility of a scientific theory is to the data, an important additional purpose here is to provide a conceptual framework for bringing together the neuroscientific and psychological levels of explanation. In this regard, 4CAPS provides a first-order description of the neuroarchitecture of cognitive information processing and attempts to illuminate the interface between the brain and the mind.

REFERENCES

AGUIRRE, G. K., ZARAHN, E., & D'ESPOSITO, M. (1998). The variability of human, BOLD hemodynamic responses. *NeuroImage*, **8**, 360-369.

ALON, N., & MEGIDDO, N. (1994). Parallel linear programming in fixed dimension almost surely in constant time. *Journal of the ACM*, **41**, 422-434.

ANDERSON, J. R., BOTHELL, D., BYRNE, M. D., DOUGLASS, S., LEBIERE, C., & QIN, Y. (2004). An integrated theory of the mind. *Psychological Review*, **111**, 1036-1060.

ARBIB, M. A. (2003). Towards a neurally-inspired computer architecture. *Natural Computing*, **2**, 1-46.

ARBIB, M. A., BILLARD, A., IACOBONI, M., & OZTOP, E. (2000). Synthetic brain imaging: Grasping, mirror neurons and imitation. *Neural Networks*, **13**, 975-997.

ARBIB, M. A., BISCHOFF, A., FAGG, A. H., & GRAFTON, S. T. (1994). Synthetic PET: Analyzing large-scale properties of neural networks. *Human Brain Mapping*, **2**, 225-233.

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.

BEEMAN, M. (1998). Coarse semantic coding and discourse comprehension. In M. Beeman & C. Chiarello (Eds.), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience* (pp. 255-284). Mahwah, NJ: Erlbaum.

BOOKHEIMER, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, **25**, 151-188.

BOOTH, J. R., MACWHINNEY, B., THULBORN, K. R., SACCO, K., VOYVODIC, J., & FELDMAN, H. M. (1999). Functional organization of activation patterns in children: Whole brain fMRI imaging during three different cognitive tasks. *Progress in Neuropsychopharmacology & Biological Psychiatry*, **23**, 669-682.

BOYNTON, G. M., ENGEL, S. A., GLOVER, G. H., & HEEGER, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, **16**, 4207-4221.

BRAVER, T., COHEN, J. D., JONIDES, J., SMITH, E. E., & NOLL, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*, **5**, 49-62.

BÜCHEL, C., COULL, J. T., & FRISTON, K. J. (1999). The predictive value of changes in effective connectivity for human learning. *Science*, **283**, 1538-1541.

BYRNE, M. D., & BOVAIR, S. (1997). A working memory model of a common procedural error. *Cognitive Science*, **21**, 31-61.

CABEZA, R., & NYBERG, L. (1997). Imaging cognition: An empirical review of PET studies with normal subjects. *Journal of Cognitive Neuroscience*, **9**, 1-26.

CABEZA, R., & NYBERG, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, **12**, 1-47.

CALLICOTT, J. H., MATTAY, V. S., BERTOLINO, A., FINN, K., COPPOLA, R., FRANK, J. A., ET AL. (1999). Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cerebral Cortex*, **9**, 20-26.

CARPENTER, P. A., JUST, M. A., KELLER, T., EDDY, W. F., & THULBORN, K. R. (1999). Graded functional activation in the visuospatial system with the amount of task demand. *Journal of Cognitive Neuroscience*, **11**, 9-24.

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**, 404-431.

CHELAZZI, L., & CORBETTA, M. (2000). Cortical mechanisms of visuospatial attention in the primate brain. In M. S. Gazzaniga (Ed. in Chief), *The new cognitive neurosciences* (2nd ed., pp. 667-686). Cambridge, MA: MIT Press.

CHOMSKY, N. (1970). Remarks on nominalization. In R. A. Jacobs & P. S. Rosenbaum (Eds.), *Readings in English transformation grammar* (pp. 184-221). Waltham, MA: Ginn.

CHOMSKY, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

COHEN, M. S., KOSSLYN, S. M., BREITER, H. C., DIGIROLAMO, G. J., THOMPSON, W. L., ANDERSON, A. K., ET AL. (1996). Changes in cortical activity during mental rotation: A mapping study using functional MRI. *Brain*, **119**, 89-100.

COLLINS, A. M., & QUILLIAN, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior*, **8**, 240-248.

CORMEN, T. H., LEISERSON, C. E., RIVEST, R. L., & STEIN, C. (2001). *Introduction to algorithms* (3rd ed.). Cambridge, MA: MIT Press.

DANTZIG, G. B., & THAPA, M. N. (1997). *Linear programming: 1. Introduction*. New York: Springer.

DEHAENE, S., & CHANGEUX, J.-P. (1997). A hierarchical neuronal network for planning behavior. *Proceedings of the National Academy of Sciences*, **94**, 13293-13298.

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**, 279-281.

DIWADKAR, V. A., CARPENTER, P. A., & JUST, M. A. (2000). Collaborative activity between parietal and dorso-lateral prefrontal cortex in dynamic spatial working memory revealed by fMRI. *NeuroImage*, **12**, 85-99.

DUNCAN, J., & MILLER, E. K. (2002). Cognitive focus through adaptive neural coding in the primate prefrontal cortex. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 278-291). Oxford: Oxford University Press.

DUNCAN, J., SEITZ, R. J., KOLODNY, J., BOR, D., HERZOG, H., AHMED, A., ET AL. (2000). A neural basis for general intelligence. *Science*, **289**, 457-460.

ELMAN, J. A., BATES, E. A., JOHNSON, M. H., KARMILOFF-SMITH, A., PARISI, D., & PLUNKETT, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

FILLMORE, C. J. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory* (pp. 1-88). New York: Holt, Rinehart & Winston.

FINCHAM, J. M., CARTER, C. S., VAN VEEN, V., STENGER, V. A., & ANDERSON, J. R. (2002). Neural mechanisms of planning: A computational analysis using event-related fMRI. *Proceedings of the National Academy of Sciences*, **99**, 3346-3351.

FINGER, S., BUCKNER, R. L., & BUCKINGHAM, H. (2003). Does the right hemisphere take over after damage to Broca's area? The Barlow case of 1877 and its history. *Brain & Language*, **85**, 385-395.

FODOR, J. A., & PYLYSHYN, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, **28**, 3-71.

FRACKOWIAK, R. S. J., FRISTON, K. J., FRITH, C. D., DOLAN, R. J., & MAZZIOTTA, J. C. (1997). The cerebral basis of functional recovery. In R. S. J. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, & J. C. Mazziotta, *Human brain function* (pp. 275-299). San Diego: Academic Press.

FRIEDERICI, A. D. (2002). Towards a neural basis of auditory sentence comprehension. *Trends in Cognitive Sciences*, **6**, 78-84.

FRISTON, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, **2**, 56-78.

FUNAHASHI, S., BRUCE, C. J., & GOLDMAN-RAKIC, P. S. (1990). Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. *Journal of Neurophysiology*, **63**, 814-831.

GARCÍA-VILLAMISAR, D., & DELLA SALA, S. (2002). Dual-task performance in adults with autism. *Cognitive Neuropsychiatry*, **7**, 63-74.

GEORGOPOULOS, A. P., LURITO, J. T., PETRIDES, M., SCHWARTZ, A. B., & MASSEY, J. T. (1989). Mental rotation of the neuronal population vector. *Science*, **243**, 234-236.

GIBSON, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, **68**, 1-76.

GOEL, V., PULLARA, S. D., & GRAFMAN, J. (2001). A computational model of frontal lobe dysfunction: Working memory and the Tower of Hanoi task. *Cognitive Science*, **25**, 287-313.

GOLDMAN, S. R., & VARMA, S. (1995). CAPping the construction–integration model of discourse comprehension. In C. A. Weaver III, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 337-358). Hillsdale, NJ: Erlbaum.

GRAFMAN, J., & LITVAN, I. (1999). Evidence for four forms of neuroplasticity. In J. Grafman & Y. Christen (Eds.), *Neuronal plasticity: Building a bridge from the laboratory to the clinic* (pp. 131-139). Berlin: Springer.

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.

HAARMANN, H. J., JUST, M. A., & CARPENTER, P. A. (1997). Aphasic sentence comprehension as a resource deficit: A computational approach. *Brain & Language*, **59**, 76-120.

HAXBY, J. V., GRADY, C. L., UNGERLEIDER, L. G., & HORWITZ, B. (1991). Mapping the functional neuroanatomy of the intact human brain with brain work imaging. *Neuropsychologia*, **29**, 539-555.

HEISS, W. D., THIEL, A., KESSLER, J., & HERHOLZ, K. (2003). Distur-

bance and recovery of language function: Correlates in PET activation studies. *NeuroImage*, **20**, S42-S49.

HERATH, P., KLINGBERG, T., YOUNG, J., AMUNTS, K., & ROLAND, P. (2001). Neural correlates of dual task interference can be dissociated from those of divided attention: An fMRI study. *Cerebral Cortex*, **11**, 796-805.

HOMAE, F., YAHATA, N., & SAKAI, K. L. (2003). Selective enhancement of functional connectivity in the left prefrontal cortex during sentence processing. *NeuroImage*, **20**, 578-586.

HORWITZ, B., FRISTON, K. J., & TAYLOR, J. G. (2000). Neural modeling and functional brain imaging: An overview. *Neural Networks*, **13**, 829-846.

HORWITZ, B., RUMSEY, J. M., & DONOHUE, B. C. (1998). Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proceedings of the National Academy of Sciences*, **95**, 8939-8944.

HORWITZ, B., & TAGAMETS, M.-A. (1999). Predicting human functional maps with neural net modeling. *Human Brain Mapping*, **8**, 137-142.

HUGUENARD, B. R., LERCH, F. J., JUNKER, B. W., PATZ, R. J., & KASS, R. E. (1997). Working-memory failure in phone-based interaction. *ACM Transactions on Computer–Human Interaction*, **4**, 67-102.

HUSAIN, F. T., NANDIPATI, G., BRAUN, A. R., COHEN, L. G., TAGAMETS, M.-A., & HORWITZ, B. (2002). Simulating transcranial magnetic stimulation during PET with a large-scale neural network model of the prefrontal cortex and the visual system. *NeuroImage*, **15**, 58-73.

HUSAIN, F. T., TAGAMETS, M.-A., FROMM, S. J., BRAUN, A. R., & HORWITZ, B. (2004). Relating neuronal dynamics for auditory object processing to neuroimaging activity: A computational modeling and an fMRI study. *NeuroImage*, **21**, 1701-1720.

JAEGGI, S. M., SEEWER, R., NIRKKO, A. C., ECKSTEIN, D., SCHROTH, G., GRONER, R., & GUTBROD, K. (2003). Does excessive memory load attenuate activation in the prefrontal cortex? Load-dependent processing in single and dual tasks: Functional magnetic resonance imaging study. *NeuroImage*, **19**, 210-225.

JAKOBSON, R. (1971). Aphasia as a linguistic topic. In *Selected writings: Vol. 2. Word and language* (2nd ed., pp. 229-238). The Hague: Mouton.

JURAFSKY, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, **20**, 137-194.

JUST, M. A., & CARPENTER, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, **92**, 137-172.

JUST, M. A., & CARPENTER, P. A. (1987). *The psychology of reading and language comprehension*. Newton, MA: Allyn & Bacon.

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

JUST, M. A., CARPENTER, P. A., & HEMPHILL, D. D. (1996). Constraints on processing capacity: Architectural or implementational? In D. Steier & T. M. Mitchell (Eds.), *Mind matters: A tribute to Allen Newell* (pp. 141-178). Mahwah, NJ: Erlbaum.

JUST, M. A., CARPENTER, P. A., KELLER, T. A., EDDY, W. F., & THULBORN, K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, **274**, 114-116.

JUST, M. A., CARPENTER, P. A., KELLER, T. A., EMERY, L., ZAJAC, H., & THULBORN, K. R. (2001). Interdependence of nonoverlapping cortical systems in dual cognitive tasks. *NeuroImage*, **14**, 417-426.

JUST, M. A., CARPENTER, P. A., MAGUIRE, M., DIWADKAR, V., & MCMAINS, S. (2001). Mental rotation of objects retrieved from memory: A functional MRI study of spatial processing. *Journal of Experimental Psychology: General*, **130**, 493-504.

JUST, M. A., CARPENTER, P. A., & MIYAKE, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, **4**, 56-88.

JUST, M. A., CARPENTER, P. A., & VARMA, S. (1999). Computational modeling of high-level cognition and brain function. *Human Brain Mapping*, **8**, 128-136.

JUST, M. A., CHERKASSKY, V. L., KELLER, T. A., KANA, R. K., & MINSHEW, N. J. (2007). Functional and anatomical cortical underconnectivity in autism: Evidence from an fMRI study of an executive function task and corpus callosum morphometry. *Cerebral Cortex*, **17**, 951-961.

JUST, M. A., CHERKASSKY, V. L., KELLER, T. A., & MINSHEW, N. J. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity. *Brain*, **127**, 1811-1821.

JUST, M. A., & VARMA, S. (2002). A hybrid architecture for working memory: Reply to MacDonald and Christiansen (2002). *Psychological Review*, **109**, 55-65.

KAAN, E., & SWAAB, T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends in Cognitive Sciences*, **6**, 350-356.

KANA, R. K., KELLER, T. A., CHERKASSKY, V. L., MINSHEW, N. J., & JUST, M. A. (2006). Sentence comprehension in autism: Thinking in pictures with decreased functional connectivity. *Brain*, **129**, 2484-2493.

KARBE, H., THIEL, A., WEBER-LUXENBURGER, G., HERHOLZ, K., KESSLER, J., & HEISS, W.-D. (1998). Brain plasticity in poststroke aphasia: What is the contribution of the right hemisphere? *Brain & Language*, **64**, 215-230.

KELLER, T. A., CARPENTER, P. A., & JUST, M. A. (2001). The neural bases of sentence comprehension: A fMRI examination of syntactic and lexical processing. *Cerebral Cortex*, **11**, 223-237.

KINTSCH, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

KOELSCH, S., GUNTER, T. C., von CRAMON, D. Y., ZYSSET, S., LOHMANN, G., & FRIEDERICI, A. D. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage*, **17**, 956-966.

LEWIS, R. L. (1993). *An architecturally based theory of human sentence comprehension*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

LOGOTHETIS, N. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *Journal of Neuroscience*, **23**, 3963-3971.

LONG, D. L., & BAYNES, K. (2002). Discourse representation in the two cerebral hemispheres. *Journal of Cognitive Neuroscience*, **14**, 228-242.

LOOSE, R., KAUFMANN, C., AUER, D. P., & LANGE, K. W. (2003). Human prefrontal and sensory cortical activity during divided attention tasks. *Human Brain Mapping*, **18**, 249-259.

LURIA, A. R. (1981). *Language and cognition* (J. V. Wertsch, Ed.). New York: Wiley.

LUSTIG, I. J., & ROTHBERG, E. (1996). Gigaflops in linear programming. *Operations Research Letters*, **18**, 157-165.

MACDONALD, M. C., JUST, M. A., & CARPENTER, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, **24**, 56-98.

MAROS, I., & MITRA, G. (2000). Investigating the sparse simplex algorithm on a distributed memory multiprocessor. *Parallel Computing*, **26**, 151-170.

MARR, D., & NISHIHARA, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Series B*, **200**, 269-294.

MASON, R. A., & JUST, M. A. (2004). How the brain processes causal inferences in text: A theoretical account of generation and integration component processes utilizing both cerebral hemispheres. *Psychological Science*, **15**, 1-7.

MASON, R. A., JUST, M. A., KELLER, T. A., & CARPENTER, P. A. (2003). Ambiguity in the brain: What brain imaging reveals about the processing of syntactically ambiguous sentences. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1319-1338.

MCCLELLAND, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, **86**, 287-330.

MESULAM, M.-M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language and memory. *Annals of Neurology*, **28**, 597-613.

MESULAM, M.-M. (2000). Behavioral neuroanatomy: Large-scale networks, association cortex, frontal syndromes, the limbic system, and hemispheric specializations. In M.-M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (2nd ed., pp. 1-120). Oxford: Oxford University Press.

MEYER, M., ALTER, K., FRIEDERICI, A. D., LOHMANN, G., & von CRAMON, D. Y. (2002). fMRI reveals brain regions mediating slow pro-

sodic modulations in spoken sentences. *Human Brain Mapping*, **17**, 73-88.

MICHAEL, E. B., KELLER, T. A., CARPENTER, P. A., & JUST, M. A. (2001). fMRI investigation of sentence comprehension by eye and by ear: Modality fingerprints on cognitive processes. *Human Brain Mapping*, **13**, 239-252.

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.

MIYAKE, A., & SHAH, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.

MORRIS, R. G., MIOTTO, E. C., FEIGENBAUM, J. D., BULLOCK, P., & POLKEY, C. E. (1997a). The effect of goal–subgoal conflict on planning ability after frontal- and temporal-lobe lesions in humans. *Neuropsychologia*, **35**, 1147-1157.

MORRIS, R. G., MIOTTO, E. C., FEIGENBAUM, J. D., BULLOCK, P., & POLKEY, C. E. (1997b). Planning ability after frontal and temporal lobe lesions in humans: The effects of selection equivocation and working memory load. *Cognitive Neuropsychology*, **14**, 1007-1027.

MÜLLER, R. A., ROTHERMEL, R. D., BEHEN, M. E., MUZIK, O., MANGNER, T. J., CHAKRABORTY, P. K., & CHUGANI, H. T. (1998). Brain organization of language after early unilateral lesion: A PET study. *Brain & Language*, **62**, 422-451.

NEWELL, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

NEWMAN, S. D., CARPENTER, P. A., VARMA, S., & JUST, M. A. (2003). Frontal and parietal participation in problem solving in the Tower of London: fMRI and computational modeling of planning and high-level perception. *Neuropsychologia*, **41**, 1668-1682.

NEWMAN, S. D., JUST, M. A., & CARPENTER, P. A. (2002). Synchronization of the human cortical working memory network. *NeuroImage*, **15**, 810-822.

NEWMAN, S. D., KELLER, T. A., & JUST, M. A. (2007). Volitional control of attention and brain activation in dual task performance. *Human Brain Mapping*, **28**, 109-117.

NICOLELIS, M. A. L. (1997). Dynamic and distributed somatosensory representations as the substrate for cortical and subcortical plasticity. *Seminars in Neuroscience*, **9**, 24-33.

OWEN, A. M., DOWNES, J. J., SAHAKIAN, B. J., POLKEY, C. E., & ROBBINS, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, **28**, 1021-1034.

POLK, T. A., SIMEN, P., LEWIS, R. L., & FREEDMAN, E. (2002). A computational approach to control in complex cognition. *Cognitive Brain Research*, **15**, 71-83.

POSTHUMA, D., DE GEUS, E. J. C., BAARÉ, W. F. C., HULSHOFF POL, H. E., KAHN, R. S., & BOOMSMA, D. I. (2002). The association between brain volume and intelligence is of genetic origin. *Nature Neuroscience*, **5**, 83-84.

RAO, S. C., RAINER, G., & MILLER, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science*, **276**, 821-824.

REICHLE, E. D., CARPENTER, P. A., & JUST, M. A. (2000). The neural bases of strategy and skill in sentence–picture verification. *Cognitive Psychology*, **40**, 261-295.

RÖDER, B., STOCK, O., NEVILLE, H., BIEN, S., & RÖSLER, F. (2002). Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: A functional magnetic resonance imaging study. *NeuroImage*, **15**, 1003-1014.

ROSENBLOOM, P. S., LAIRD, J. E., & NEWELL, A. (1993). *The Soar papers: Research on integrated intelligence* (2 Vols.). Cambridge, MA: MIT Press.

RYPMA, B., PRABHAKARAN, V., DESMOND, J. E., GLOVER, G. H., & GABRIELI, J. D. E. (1999). Load-dependent roles of frontal brain regions in the maintenance of working memory. *NeuroImage*, **9**, 216-226.

SANES, J. N., & DONOGHUE, J. P. (1997). Dynamic motor cortical organization. *The Neuroscientist*, **3**, 158-165.

SCHLOSSER, M. J., AOYAGI, N., FULBRIGHT, R. K., GORE, J. C., & MCCARTHY, G. (1998). Functional MRI studies of auditory comprehension. *Human Brain Mapping*, **6**, 1-13.

SHALLICE, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London: Series B*, **298**, 199-209.

SHALLICE, T., & BURGESS, P. (1996). The domain of supervisory processes and temporal organization of behaviour. *Philosophical Transactions of the Royal Society of London: Series B*, **351**, 1405-1412.

SHEPARD, R. N., & METZLER, J. (1971). Mental rotation of three-dimensional objects. *Science*, **171**, 701-703.

SIMON, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, **7**, 268-288.

SMOLENSKY, P. (1999). Grammar-based connectionist approaches to language. *Cognitive Science*, **23**, 589-613.

SZAMEITAT, A. J., SCHUBERT, T., MÜLLER, K., & VON CRAMON, D. Y. (2002). Localization of executive functions in dual-task performance with fMRI. *Journal of Cognitive Neuroscience*, **14**, 1184-1199.

TAGAMETS, M.-A., & HORWITZ, B. (1998). Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. *Cerebral Cortex*, **8**, 310-320.

THIBADEAU, R., JUST, M. A., & CARPENTER, P. A. (1982). A model of the time course and content of reading. *Cognitive Science*, **6**, 157-203.

THULBORN, K. R., CARPENTER, P. A., & JUST, M. A. (1999). Plasticity of language-related brain function during recovery from stroke. *Stroke*, **30**, 749-754.

VARMA, S. (2006). *A computational model of Tower of Hanoi problem solving*. Doctoral dissertation, Vanderbilt University, Nashville, TN.

VARMA, S., & JUST, M. A. (2007). *The cognitive neuroarchitecture of sentence comprehension*. Manuscript in preparation.

VOSSE, T., & KEMPEN, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, **75**, 105-143.

WEILLER, C., ISENSEE, C., RIJNTJES, M., HUBER, W., MULLER, S., BIER, D., ET AL. (1995). Recovery from Wernicke's aphasia: A positron emission tomographic study. *Annals of Neurology*, **37**, 723-732.

WELLMAN, M. P. (1993). A market-oriented programming environment and its application to distributed multicommodity flow problems. *Journal of Artificial Intelligence Research*, **1**, 1-23.

WOOLDRIDGE, M., & DUNNE, P. E. (2006). On the computational complexity of coalitional resource games. *Artificial Intelligence*, **170**, 835-871.

NOTES

1. Throughout this section, we will use the term *L. STG* (left superior temporal gyrus) to denote the posterior temporal area of activation during sentence comprehension, which actually includes not only L. STG, but also the left middle temporal gyrus and the sulcus between them; we will also use the term in this sense interchangeably with *Wernicke's area*. We note that the latter term has not typically been defined elsewhere either precisely or consistently. We will also use the term *L. IFG* (left inferior frontal gyrus) interchangeably with *Broca's area*.

2. The iterativeness of a production can be viewed as a function of its weight, which specifies a multiplier (usually fractional) of the amount of activation it directs to its target declarative element per cycle. The larger the weight, the more activation directed per cycle, and thus the fewer cycles required to raise the activation level of the target element above threshold. In the interest of parsimony, all productions in all models were assigned the same weight, one large enough that productions required only a single cycle to raise the activations of their target elements above threshold. This has the advantage of yielding a clear interpretation of the simulation cycle counts reported below as the minimum number of processing cycles (a proxy for processing times) required to perform tasks. As production weights are reduced, the number of processing cycles to perform a task or process increases approximately linearly. There are a number of theoretically interesting consequences of increased iterativeness. One is cascaded processing (McClelland, 1979). Another is that spreading a process over several iterations provides an opportunity for other processes to interact with it. Documenting these consequences of iterative processing is beyond the scope of this article.

3. The details of how declarative elements and productions aggregate into cognitive functions are available in code and documentation on the 4CAPS Web site. Briefly, each declarative element is of a particular class (e.g., a phrase). Declarative elements of the same class and productions that process elements of that class (e.g., Determiner + Noun → NP) together constitute a cognitive function (e.g., *parsing*). The specialization of a center for a cognitive function refers to the efficiency with which declarative elements of the associated class are maintained and

processed by productions. The 4CAPS architecture knows about classes (and declarative elements and productions), not cognitive functions; the latter are an abstraction we adopt here to simplify the description of the architecture and its models.

4. Historically, the temporal mapping between human response times and the processing times of production system models (whether CAPS, ACT, Soar, or EPIC) is 25–100 msec/cycle, with a geometric mean of 50. The sentence comprehension model is in this range: The temporal mapping for Figure 8 is 60 msec/cycle. The mental rotation model described in the next section is also in this range, since the temporal mapping for Figure 17A is 30 msec/cycle. However, the TOL model is outside this range, with a temporal mapping for Figure 16A of 556 msec/cycle. This is because the TOL model is written at a coarser grain size than the other 4CAPS models. In principle, the logic of each TOL production could be decomposed into a sequence of finer-grained productions, and such a model would produce a temporal mapping consistent with other production system models. Such a decomposition should affect neither the model's fit to the human solution time data nor its fit to the brain activation data.

5. It is important to note that the assumption of extensive *functional* connectivity among centers does not require extensive *anatomical* connectivity. In particular, there is no need to postulate a number of point-to-point anatomical connections quadratically related to the number of brain areas. Just as a huge number of postal addresses are effectively connected by a relatively lean physical infrastructure, a large number of functional connections between brain areas can be efficiently implemented by a relatively small number of neural pathways arranged quasi-hierarchically (e.g., with the corpus callosum serving as a "trunk" line).