Computational Modeling of High-Level Cognition and Brain Function

Marcel Adam Just,* **Patricia A. Carpenter, and Sashank Varma**

Center for Cognitive Brain Imaging, Carnegie Mellon University, Pittsburgh, Pennsylvania

◆ ══════════════════════════ ◆

Abstract: This article describes a computational modeling architecture, 4CAPS, which is consistent with key properties of cortical function and makes good contact with functional neuroimaging results. Like earlier cognitive models such as SOAR, ACT-R, 3CAPS, and EPIC, the proposed cognitive model is implemented in a computer simulation that predicts observable variables such as human response times and error patterns. In addition, the proposed 4CAPS model accounts for the functional decomposition of the cognitive system and predicts fMRI activation levels and their localization within specific cortical regions, by incorporating key properties of cortical function into the design of the modeling system. *Hum. Brain Mapping 8:128–136, 1999.* © 1999 Wiley-Liss, Inc.

Key words: brain function; fMRI; PET; 4CAPS

◆ ══════════════════════════ ◆

INTRODUCTION

Functional neuroimaging has resulted in an explosion of new findings about molar brain function, stimulated by new technologies (functional magnetic resonance imaging—fMRI), new analytic methods, and a boldness to examine a wide variety of different types of thinking. However, there has not been a concomitantly rapid development of integrative cognitive theories. Such theoretical frameworks cannot only integrate many disparate findings, but also generate new predictions in the process of testing and refining particular assumptions of the theory and task models. We report one of the first attempts to apply some of the types of computational models that account for high-level cognition to brain activation patterns. In this report, we enumerate some of the neuroimaging results that are central to such a framework and describe a computational architecture, 4CAPS, that is being developed from this perspective.

Several key properties of the cortical system stand out and guide the design of 4CAPS, which is a production-system architecture with several connectionist features. Before describing these properties, we note that production systems are the simulation medium of choice for high-level cognition. Some of the prominent ones, including SOAR [Newell, 1990], ACT-R [Anderson, 1993], EPIC [Meyer and Kieras, 1997], and 3CAPS [Just and Carpenter, 1992], account for the error patterns and processing times in problem solving, reasoning, decision making, memory and learning, language comprehension, and visual thinking. 3CAPS models, the predecessor of 4CAPS, account for errors and response times in analogical problem solving [Carpenter et al., 1990], and normal and aphasic sentence comprehension [Haarmann et al., 1997; Just and Carpenter, 1992]. The grain size of the analysis in

*Correspondence to: Dr. Marcel Just, Center for Cognitive Brain Imaging, Psychology Department, Carnegie Mellon University, Pittsburgh, PA 15213. email *just+@cmu.edu*

this modeling is particularly suited to the nature of fMRI data acquired in cognitive tasks.

CORTICAL PROPERTIES AND MODEL PROPERTIES

In this section, we describe some of the signature properties of cortical function and relate them to the design of 4CAPS.

Thinking is work. Like all other biological processes, the computational work underlying thinking must be accompanied by some resource utilization. fMRI appears to measure a facet of this utilization, albeit indirectly, by reflecting the accumulation of oxygenated hemoglobin in areas with neural activity, with a time lag of ~2 sec. Resource utilization within a neural system subsumes several dimensions, including neurotransmitter function and various metabolic support systems, as well as the connectivity and structural integrity of the system [see Parks et al., 1988]. In this sense, a brain region can be considered a resource pool.

The computational activity in 4CAPS is also resource consuming, in the sense that all of the information processing and maintenance functions operate by consuming an entity called "activation." The use of the word "activation" here is not the same as the "brain activation" measured by fMRI or positron emission tomography (PET). Rather, the use of the term "activation" in cognitive science for this purpose dates back to the 1960s [Collins and Quillian, 1969] when it specified the availability of a concept. The activation in a 4CAPS component is a limited resource that is drawn on by computational activities, namely, by propagating activation to representational elements in the course of information processing. One of the main innovations of 4CAPS is that the resource or capacity utilization in a given unit of time in each component 4CAPS system (described below) is intended to correspond to the amount of brain activation observed with a neuroimaging measure in the corresponding area during the corresponding time interval.

Furthermore, the size of the resource supply for a component is assumed to vary among normal individuals and assumed to vary pathologically in the case of some special populations (such as stroke patients), and such differences are believed to modulate cognitive performance. This analysis does not differentiate between greater resource availability being due to practice, learning, or physiological factors, such as bioenergetic resources.

Thinking is a team sport. Almost every cognitive task involves the activation of a network of brain regions (say, 4–10 per hemisphere), rather than a single area. This network phenomenon also can be observed in the non-focal nature of cognitive deficits in patients who have focal lesions [Mesulam, 1990]. An emerging consensual view is that cognitive tasks are subserved by large-scale cortical networks that consist of spatially separate computational components, each with its own set of relative specializations that collaborate extensively to accomplish cognitive functions. For example, visual sentence comprehension is subserved by a large-scale network that includes left inferior frontal gyrus (Broca's area), left posterior superior temporal gyrus (Wernicke's area), angular gyrus, extrastriate and primary visual cortex, and in some circumstances, left middle frontal gyrus (left DLPFC) and the right hemisphere homologues of Broca's and Wernicke's areas. No single area "does" sentence comprehension. Furthermore, the collaboration among areas is hypothesized to be highly interactive (as functional connectivity analyses indicate), making the resulting cognitive computations an emergent property of several collaborating team members.

One of the main innovations in 4CAPS is that it is composed of a number of collaborating component computational systems, each intended to correspond to the functions of a cortical area (such as Broca's area). Each component system is itself a production system (as described below), and the multiple production systems operate concurrently with each other. Furthermore, the systems are internally parallel; all satisfied production rules can fire (act) simultaneously. The highest level large-scale networks, composed of several component production systems, are intended to correspond to high-level cognitive systems (such as a language network). One goal for a computational model is to generate and test specific hypotheses about the nature of the collaboration among the components of the network or team.

Thinking is self-organizing. The neural underpinnings of cognitive performance are dynamically configured and allocated as the computational demands change either in magnitude or in quality. One facet of the dynamics is that the intensity and volume (hence the precise location) of brain activation in a given cortical area increase with the computational load, within some dynamic range. For example, there is more activation when a reader comprehends a structurally complex sentence than a structurally simpler one [Just et al., 1996]. The increasing activation shows up within neural regions, such as Wernicke's or Broca's areas,

and it also shows up in increasing involvement of other regions, such as their right hemisphere homologues. (The co-modulation of the activation of multiple cortical regions in response to the manipulation of some variable, such as the structural complexity of a sentence, provides one source of evidence that these regions are members of a large-scale network).

A second mechanism of change is the entry of additional components into the large-scale cortical network underlying task performance. For example, the activation of the dorsolateral prefrontal cortex (often associated with functions such as goal monitoring) in sentence comprehension may depend on whether a task requires a substantial amount of reasoning and goal management. The increase of activation of this component is incremental, reflecting the nature and amount of the task demand. Both types of dynamic recruitment may contribute to the brain plasticity that underlies normal development and adaptations to brain damage.

A 4CAPS model dynamically recruits the appropriate team of components to perform a task and dynamically modifies the recruitment if the task changes in the size or the quality of its computational demand. What lies at the heart of this self-organization is that at various levels, the processes are evoked automatically whenever their enabling conditions arise. The within-component recruitment consists of evoking more processing and, hence, generating more resource consumption. The recruitment of additional components consists of the initiation of processing (and resource consumption) in a previously noncollaborating component. The 4CAPS model of sentence comprehension described below automatically recruits the participation of a component corresponding to left DLPFC if the comprehension demands imposed by a sentence outstrip the resources of the Wernicke and Broca components.

Higher cortical areas have multiple specializations. Each area of association cortex may participate in the execution of a repertoire of related processes rather than just a single process. Functional neuroimaging meta-analysis routinely attribute more than a single function to a cortical region [e.g., Gabrieli et al., 1998; Grafman, 1995]. For example, DLPFC is credited with working memory-related functions, large-scale inhibitory functions, and executive functions. The multifunctionality might be only apparent because a region operates at a level sufficiently abstract that it subsumes differences among what are currently called different processes. For example, both planning and inhibition, which are different from one perspective, can be subsumed under the rubric of ''process coordination.''

This multifunctionality is not just an artifact of aggregating neuroimaging over a large region of interest. We routinely find individual voxels ($3.125 \times 3.125 \times 5$ mm at 3.0 T) that are involved in more than one type of processing, such as lexical and syntactic processing. 4CAPS instantiates the multifunction hypothesis by having each component contain productions that execute a variety of related processes with a common style or domain of processing.

A given specialization may occur in more than one cortical area, although at different levels of efficiency. More than one area may be able to process the same information and produce the same or a similar result, although the precise qualitative nature of the processing, and hence its efficiency (resource use), would generally differ. This overlap is presumed to arise as a consequence of the way that the specializations arose developmentally. Elman and colleagues [1996] suggest that initial differences at infancy in cell types across cortical areas can result in some areas gradually becoming more proficient (and hence specialized) in particular types of processing, namely, those types of processing for which the given cell types are most suited. For example, cells in more than one brain area may initially attempt to process speech sounds, but the area containing cells that are particularly sensitive to the fine timing distinctions that differentiate phonemes may eventually become specialized for speech processing. Another area also may have initially attempted to process speech and eventually lost the competition for the specialization, but may nevertheless have retained a residual capability to process speech, albeit less efficiently. Thus there might be some overlap in function between areas, such as between the left and right homologues of the language network.

Such overlap in function between areas raises the question of how the processing is dynamically allocated to a given area when a given computation to be performed comes along. Different regions may attempt to execute the same process in parallel with each other, albeit with different efficiencies and processing heuristics. Even if two components initiate the computation for a given part of a task, the more specialized component would normally perform faster and more accurately and provide the result to the rest of the collaborating large-scale cortical network earlier. A less specialized system for a given computation might play a more important role if the more specialized system were less available due to structural damage (lesioned), or if its resources were already consumed by performing another ongoing computation. For example, the *right* superior temporal gyrus (STG) may
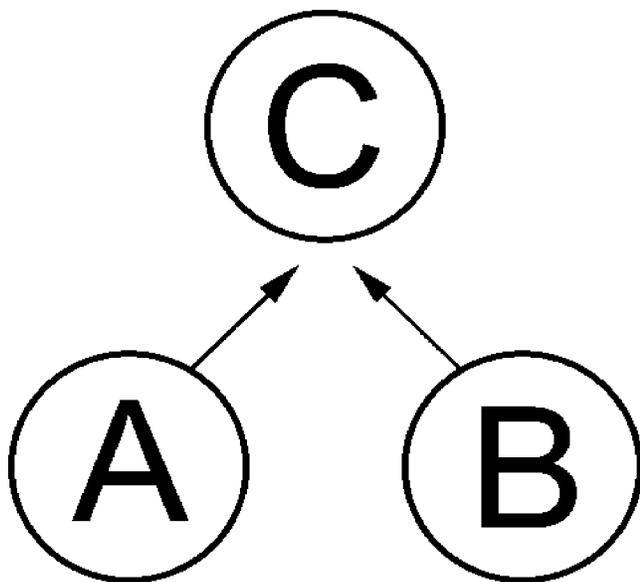
Figure 1.

The figure, with Table I, illustrates how productions can function as links that incrementally propagate activation among the nodes of a network. It shows how information from nodes A and B is integrated to activate C, a function accomplished by the two productions in Table I. If the condition and action of a production are constants, the production functions like a link in a connectionist network. If the condition and action are variables, the link formed by the production is dynamically formed in the course of processing.

play a larger role in language comprehension if left STG were damaged (say, by stroke), or if the computational load were unusually high; both phenomena have been observed with fMRI [Just et al., 1996; Thulborn et al., 1999). Thus in the current 4CAPS version, the assignment of which component performs a given process that may entail some overlapping assignments is determined dynamically on the basis of relative specialization and resource availability. The resulting cognitive processing is the emergent product of relatively specialized computational components engaged in a closely knit collaboration (the interactive team sport).

Representations and processes are graded. This assumption speaks to the continuous nature of the elements and processes of thinking in keeping with the nature of its physiological substrate. As noted above, many contemporary cognitive theories assume that representational elements have associated with them activation levels that vary continuously, corresponding to their level of availability or degree to which they are in play. A concept can be in one of a variety of activation states,

from low (say, the state of one's mother's maiden name before the topic was mentioned) to high (say, the concept of activation itself during the reading of this sentence). Similarly, the nature of the information processing is assumed to be graded. A process is not executed all at once. Rather, it gradually performs its function over several cycles of activity. Neural network models are prominent examples of graded representation and processing, in which the processing occurs over many cycles of activation propagation through a network [Rumelhart et al., 1986]. The representational elements in 4CAPS similarly have activation levels. Also, the processing in 4CAPS is graded in the sense that activation is propagated to elements by the productions over several cycles. As noted above, the activation in 4CAPS plays the additional role of the limited resource entity. These properties underlie the mechanisms' ability to account for the distribution of processing times and error probabilities.

4CAPS ARCHITECTURE

4CAPS uses a production system flow of control like that of its predecessor system, 3CAPS [Just and Carpenter, 1992]. All procedural knowledge is contained in an unordered set of *if-then* production rules. The "if" part specifies an enabling condition for the rule to fire. The "then" part specifies the actions to be taken when the production fires. Processes occur incrementally by raising (or sometimes lowering) the activation of the associated action elements. Figure I and Table I illustrate how productions can function like links in a connectionist network. A specific example, in Table II, illustrates how productions accomplish a small part of sentence parsing. One sentence-parsing production has as a condition the encoding of a definite article (*the*) and as its action, increasing the activation associated

TABLE I. Two productions that constitute the network in Figure I

Production 1:	
Conditions	If the action level of A reaches a threshold and C isn't activated to threshold ⟶
Action	Increment the activation level of C by an amount that is proportional to A's activation.
Production 2:	
Conditions	If the action level of B reaches a threshold and C isn't activated to threshold ⟶
Action	Increment the activation level of C by an amount that is proportional to B's activation.

TABLE II. Simplified production system for noun phrase parsing

Production 1:
  Condition   If a determiner (e.g., "the") is encoded $\longrightarrow$
  Action   Start a representation of a noun phrase
Production 2:
  Condition   If an adjective follows a determiner $\longrightarrow$
  Action   Assume it will modify the head noun of the noun phrase (e.g., "the happy . . .)
Production 3:
  Condition   If a noun eventually follows a determiner $\longrightarrow$
  Action   Represent it as the head of the noun phrase (e.g., "the happy farmer")

with the representation of a noun phrase. As the processing unfolds, productions act like dynamic links associating conditions and actions. Multiple productions may collaborate, because on each cycle the productions whose enabling conditions are satisfied all fire, which may change the state of the system for the next cycle. There is no central control of the processing flow because the production rule firings are self-scheduling. 4CAPS has several important connectionist properties: parallel firings of more than one satisfied rule, "graded" representational elements with activation levels, and re-iterative activation propagation to elements by the productions over several cycles (graded processing). The resource utilization in a given unit of time in each component 4CAPS system, measured in terms of the units of activation consumed by the processing and storage relative to the total capacity, is intended to correspond to the amount of brain activation observed with a neuroimaging measure in the corresponding component during the corresponding interval.

Production systems have Turing machine power, but their ability to simulate high level cognition efficiently (such as the processing of grammatical rules) derives from the binding of variables, which provides the ability to generalize over class instances and easily to use complex rules formed by any Boolean combination of conditions.

INITIAL 4CAPS MODEL OF SENTENCE PROCESSING

The first testbed for the 4CAPS system was a small-scale model of sentence comprehension, intended to simulate both behavioral and fMRI results. The model generates a measure of resource utilization

in its component modules that corresponds to amounts of activation (as modulated by sentence structure) in the relevant brain areas as reported in Just et al. [1996]. In that study, participants read sentences of three types that are superficially similar but differ in structural complexity and, hence, in the amount of computational demand that they impose. The sentences below use the same words to exemplify the three types; in the actual study, each sentence involved different words:

1. Active conjoined *The reporter attacked the senator and admitted the error.*
2. Subject relative clause *The reporter that attacked the senator admitted the error.*
3. Object relative clause *The reporter that the senator attacked admitted the error.*

Type 1 sentences contain active clauses that are simply conjoined. Type 2 sentences, of intermediate complexity, contain a relative clause that interrupts a main clause and requires additional maintenance. In the most complex sentences, Type 3, the main clause is interrupted, and the first noun plays different roles in the two clauses (as the subject of the main clause and the object of the relative clause). Type 3 sentences produce longer processing times, higher error rates, than the less complex Type 2 [Just and Carpenter, 1992; King and Just, 1991]. The model accounts for the word-by-word processing times and error rates on these types of sentences. The fMRI findings showed that, in addition, the amount of cortical activation in Wernicke's and Broca's areas (left posterior STG and left IFG) increases with sentence complexity [Just et al., 1996]. The activation image for one slice of a single participant is shown in Figure 2 and the group results are graphed in Figure 3. These fMRI-measured activation across-sentence conditions in multiple cortical areas can be accounted for by the 4CAPS sentence-comprehension model in addition to the response times and error rates.

The initial comprehension model consists of three collaborating components, that are intended to correspond to the functioning of Broca's area (left inferior frontal gyrus), Wernicke's area (left superior temporal gyrus), and left DLPFC (dorsolateral prefrontal cortex). The processes of visual and perceptual encoding are outside of the scope of this model, which takes as its input perceptual representations that activate word meanings and processes them to construct a full syntactic and semantic representation of the sentence. The model processes the successive words of a sentence, one at a time, attempting to interpret each word as fully as possible in the context of the preceding
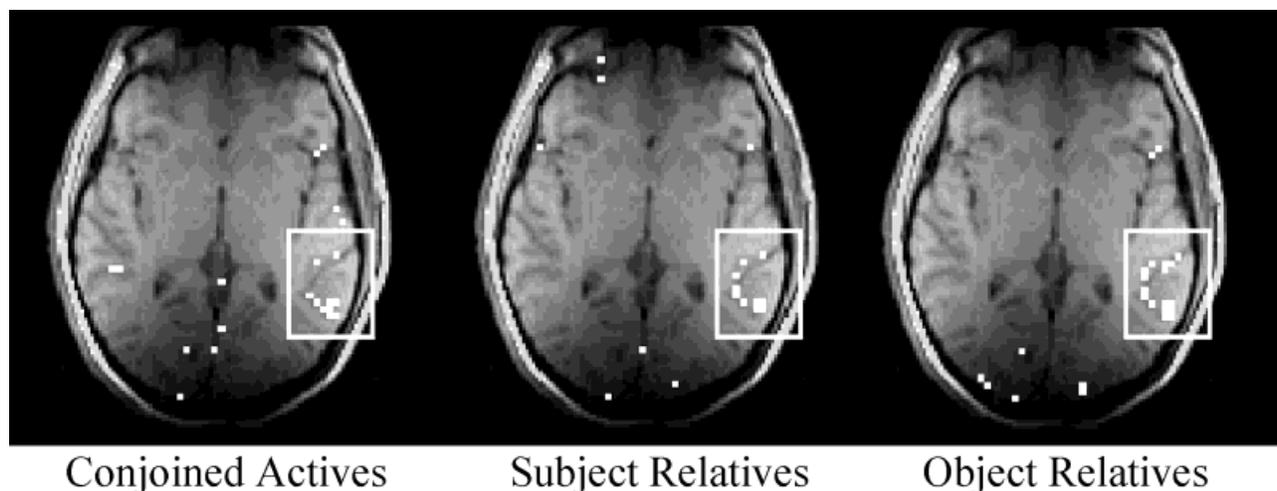
Figure 2.
Thresholded fMRI activation images for only the most activated slice through Wernicke's area (indicated by the box) from one participant. The number of activated voxels (shown in white) generally increases with sentence complexity. (Adapted from Brain activation modulated by sentence comprehension, Just et al., 1996, Science, 274, Fig. 2, p. 115. © 1996 American Association for the Advancement of Science, reprinted with permission.)

words while incrementally constructing a representation of the sentence, just as human readers do, as their eye fixations indicate [Just and Carpenter, 1980]. Each word is processed at a lexical, syntactic, and thematic level, retrieving and interpreting stored and newly computed information [see Haarmann et al., 1997, for a description of a single-component version of the model]. For example, on encountering the first word of each sentence (the word *the*), the model recognizes it as a determiner, hypothesizes that a noun phrase is being initiated, and constructs a representation of a noun phrase. On encountering the second word (*reporter*), the model recognizes it as a noun, with the potential of being the head of the noun phrase. This particular noun also has the potential of being the agent or recipient of an action, information that will come into play at the times when the matching between nouns and verbs occurs. When any processing occurs (i.e., when a representation is constructed or its activation level is modified), some of the limited activation resource is consumed, constituting the model's resource consumption or capacity utilization. In the case of the three types of sentences processed in this study, their lexical computational demands are almost identical, but they differ primarily in their syntactic computations. In particular, the subject and object relative sentences require that the processing of the first clause be interrupted and that the intermediate products be maintained while the embedded clause is processed,

adding extra work to the pairing of nouns with their verbs. The object relative sentences further entail the first noun playing a different thematic role in the main and the embedded clause (agent and patient, respectively), again resulting in additional computational work. The workload is distributed among the three collaborating components.

Broca's area simulation. The model of Broca's area is based on a hypothesized set of relative specializations of that area, namely, implicit speech and participation in the articulatory loop [see Awh et al., 1996, for a summary] and silent word-generation [Hinke et al., 1993]. Broca's area is also associated with syntactic processing in lesion studies [Caramazza and Berndt, 1988] and in neuroimaging studies [Just et al., 1996; Stromswold et al., 1996]. To account for these disparate functions, Mesulam [1990] characterizes Broca's area as being at the syntactic-articulatory pole of a region. Our proposal provides a unifying view of these two roles as well as others. In the current proposal, Broca's area is credited with internal structure generation. This component of the model is responsible for generating (either constructing or re-activating) the types of representations that underlie language processing and other types of serial-order-based propositional processing. In this scheme, Broca's area is not the architect of the language-related representations, but rather the contractor that executes the construction or reactivation of
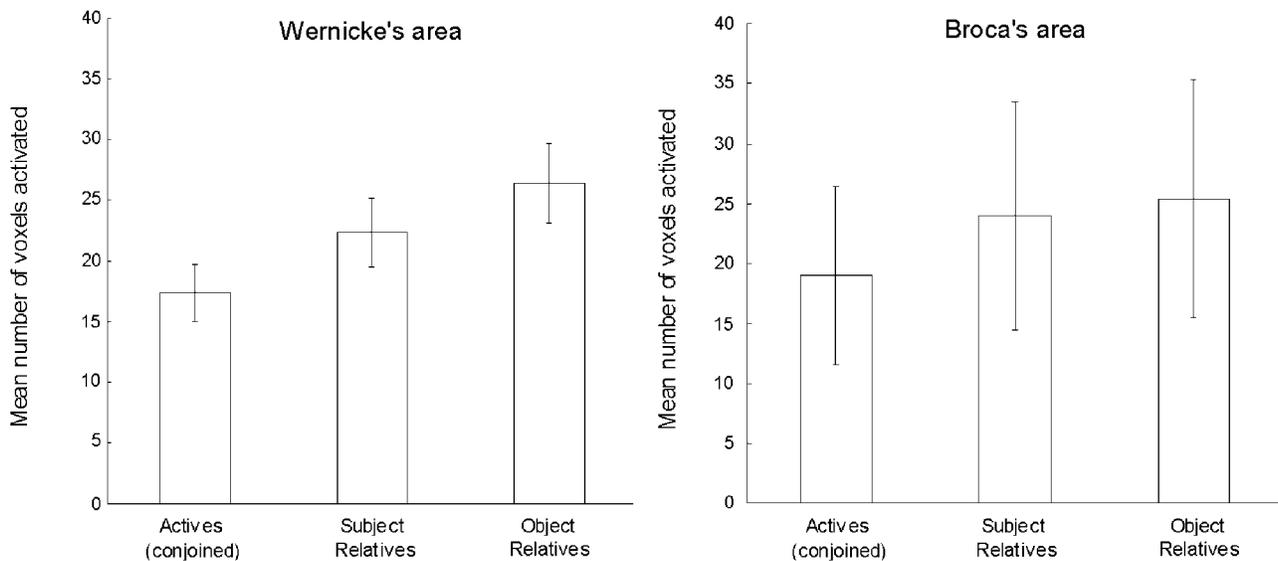
Figure 3.

The average number of activated voxels across participants indicates that the processing of more complex sentences leads to an increase in the volume of neural tissue that is highly activated in more than one area. The left panel indicates the average number of activated voxels in the Wernicke's area (left posterior STG) and standard errors of the means over 15 participants. The right-hand panel indicates the average number of activated voxels in Broca's area (left inferior frontal gyrus) and standard errors of the means over only five participants. (Adapted from Brain activation modulated by sentence comprehension, Just et al., 1996, Science, 274, Fig. 1, p. 115. © 1996 American Association for the Advancement of Science, reprinted with permission.)

a representation based on a design or plan that is initially proposed by another system component (possibly Wernicke's area). None of Broca's presumed secondary specializations are implemented in this initial model.

Although we have not yet modeled the right hemisphere homologue of Broca's area, we credit it some related capabilities, but at lower levels of proficiency. This area becomes activated if a comprehension task becomes extremely difficult [Just et al., 1996], or if Broca's area becomes lesioned [Frakowiak et al., 1997; Thulborn et al., 1999], suggesting that it is relatively less specialized for some of the same functions as Broca's area. At the same time, Broca's homologue may have functions for which it is relatively more specialized than Broca's area, perhaps including the processing of prosody. This hypothesized type of distribution of relative specializations among cortical areas is intermediate between complete specialization and complete equipotentiality.

Wernicke's area simulation. Wernicke's relative specialization is operationalized as language interpretation and elaboration. Its input is information that has come through the perceptual system, or that has been internally generated by itself working in combination with the Broca component. The interpretive function con-sists of retrieving additional associated information that is relevant to the input information, filling in slots in a knowledge structure pertaining to the input structure, or elaborating the design of a knowledge structure. The interpretive function may be thought of as using an exemplar or content-addressable memory to perform syntactic, lexical, thematic, and referential computations. For example, the syntactic function of the Wernicke component is to retrieve or hypothesize an augmentation of the syntactic elements of an existing representation, which can be either a single word or a complex syntactic structure. In addition, the Wernicke component also performs the design (but not the construction) of propositional representations that embody the synthesis of the retrieval. Finally, the Wernicke component helps maintain representations in sound-based form, but does not construct those representations.

In our theoretical framework, the Wernicke and Broca components operate as equals working in close collaboration. This proposal is based on the molar findings that the two areas are co-activated and co-modulated during sentence processing [Just et al., 1996]. Additional evidence for the collaboration is that there are many voxels in these two areas that have activation that is highly correlated across time, such that these voxels form a functionally connected cluster.
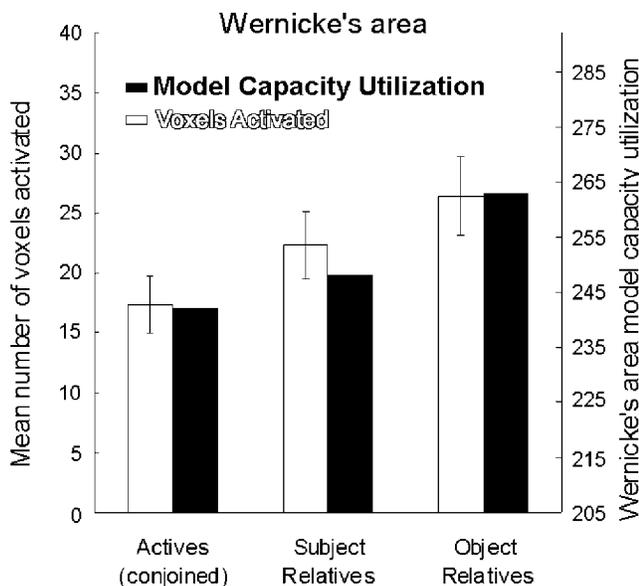
Figure 4.
Comparison of the fMRI-measured brain activation in Wernicke's area (left posterior STG) and the capacity utilization in the Wernicke component of the 4CAPS processing model during the processing of three sentences of increasing complexity.

The division of labor between the two components may be similar with respect to various levels of language processing, including syntactic, lexical, and thematic processing. When each ensuing word of a sentence is encoded, it is processed reciprocally by the two components. The activated representations in the Wernicke component are passed to the Broca component for construction into larger or higher order structures. Reciprocally, the structures built by the Broca component are passed to the Wernicke component for further elaboration. Cyclical processing continues until quiescence occurs (defined as small rate of change across cycles). Then, the next word of the input sentence is processed.

The primary predictions of the computational model pertain to the activation in Wernicke's area and Broca's area for each of the three sentence types. In the current model, the resource utilization of both the Broca and Wernicke components increases monotonically across the three sentence types, constituting a match to the corresponding fMRI measures of activation, as shown in Figure 4 for the Wernicke component (the capacity utilization is modulated very similarly in the Broca component). In addition, the two components together account for the word-by-word processing time for several different types of sentences (11 types in total) and the distribution of error probabilities across the

sentence types [see Haarmann et al., 1997, for error modeling specifics].

Modeling prefrontal cortex. A third component that occasionally participates in sentence comprehension, the left dorsolateral prefrontal cortex (DLPFC), illustrates the self-organizing nature of the processing. DLPFC's functions fit the rubric of "executive processes," such as goal management and planning, the inhibition of inappropriate high level goals and their associated actions, and latching of various memory buffers. In the model, the Wernicke and Broca components can handle common, automatic language tasks, such as processing single words and simple syntactic structures using their inherent capabilities, with little or no participation of DLPFC. However, as the resources of the language components are consumed by an increasingly complex language task, the DLPFC component is increasingly recruited. The recruitment occurs by means of an automatic detection that the Wernicke and Broca components are approaching the limits of their resource pools and beginning to lose (forget) their subgoals. This detection recruits DLPFC participation with its greater executive capabilities, such that DLPFC plays two roles in sentence comprehension: goal management and memory buffer latching. In the sentence comprehension study, DLPFC may be recruited as the model proceeds into an increasingly center-embedded part of a sentence (requiring the storage of loose ends from the interrupted sentence constituents). The storage of the preceding sentence information and goal management information may deplete most of the Wernicke and Broca resources for some individuals. The remainder of the sentence is then processed with DLPFC performing the goal management (keeping track of which noun phrases are waiting to encounter their matching verb phrases) and ensuring the storage and later matching-up of those incomplete clause segments. Additional functions to be implemented for this component may include the instantiation of schemas that integrate over large units of time in the course of comprehension [cf. Grafman, 1995].

Conclusions

This model is a small first step in making some assumptions about cortical function explicit in terms of a computational model. Its most promising attribute for functional neuroimaging is its ability to relate cognitive computations to brain activation. Of course, many challenges remain for this approach, primarily increasing its scope and detail to account for most of

the cortical centers and many cognitive tasks. For example, some recent fMRI findings show a related type of collaboration between two cortical areas (parietal and inferior temporal) in the mental rotation of increasingly difficult items [Carpenter et al., 1999] that would be equally amenable to modeling within this framework. Another challenge is to model explicitly the functioning at somewhat lower levels of processing. Proceeding in this way, computational modeling could provide the theoretical framework for specifying the complex dynamic behavior of a set of understandable components in the brain, just as this approach has done in modeling complex dynamic systems in many other disciplines.

REFERENCES

Anderson JR. 1993. Rules of the Mind. Hillsdale, NJ: Erlbaum.

Awh E, Jonides J, Smith EE, Schumacher EH, Koeppe RA, Katz S. 1996. Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. Psychological Sci 7:25–31.

Caramazza A, Berndt RS. 1988. A multicomponent deficit view of agrammatic Broca's aphasia. In: Kean ML (ed): Agrammatism. New York: Academic Press, pp 27–63.

Carpenter PA, Just MA, Keller TA, Eddy W, Thulborn KR. 1999. Graded functional activation in the visuospatial system with the amount of task demand. J Cognitive Neurosci 11:9–24.

Carpenter PA, Just MA, Shell P. 1990. What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. Psychological Rev 97:404–431.

Collins AM, Quillian MR. 1969. Retrieval time from semantic memory. J Verbal Learning Verbal Behavior 8:240–248.

Elman JL, Bates EA, Johnson MH, Karmiloff-Smith A, Parisi D, Plunkett K. 1996. Rethinking innateness: a connectionist perspective on development. Cambridge: MIT Press.

Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Mazziotta JC, eds. 1997. Human Brain Function. San Diego: Academic Press.

Gabrieli JDE, Poldrack RA, Desmond JE. 1998. The role of leftprefrontal cortex in language and memory. Proc Natl Acad Sci USA 95:906–913.

Grafman J. 1995. Similarities and distinctions among current models of prefrontal cortical functions. Annals NY Acad Sci 769:337–368.

Haarmann HJ, Just MA, Carpenter PA. 1997. Aphasic sentence comprehension as a resource deficit: A computational approach. Brain and Language 59:76–120.

Hinke RM, Hu X, Stillman AE, Kim SG, Merkle H, Salmi R, Ugurbil K. 1993. Functional magnetic resonance imaging of Broca's area during internal speech. NeuroReport 4:675–678.

Just MA, Carpenter PA. 1992. A capacity theory of comprehension: Individual differences in working memory. Psychological Rev 99:122–149.

Just MA, Carpenter PA. 1980. A theory of reading: From eye fixations to comprehension. Psychological Rev 87:329–354.

Just MA, Carpenter PA, Keller TA, Eddy WF, Thulborn KR. 1996. Brain activation modulated by sentence comprehension. Science 274:114–116.

King J, Just MA. 1991. Individual differences in syntactic processing: The role of working memory. J Memory and Language 30:580–602.

Mesulam M-M. 1990. Large-scale neurocognitive networks and distributed processing for attention, language and memory. Annals Neurol 28:597–613.

Meyer DE, Kieras DE. 1997. Precis to a practical unified theory of cognition and action: some lessons from EPIC computational models of human multiple-task performance (TR-97/ONR-EPIC-08) Ann Arbor, University of Michigan, Department of Psychology, Electrical Engineering & Computer Science Department.

Newell A. 1990. Unified Theories of Cognition. Cambridge: Harvard University Press.

Parks RW, Lowenstein DA, Dodrill KL, Barker WW, Yoshii F, Chang JY, Emran A, Apicella A, Sheramata WA, Duara R. 1988. Cerebral metabolic effects of a verbal fluency test: a PET scan study. J Clinical and Experimental Neuropsychology 10:565–575.

Rumelhart DE, Hinton G, McClelland JL. 1986. A general framework for parallel distributed processing. In: Rumelhart DE, McClelland JL (eds): Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations. Cambridge: MIT Press, pp 318–362.

Stromswold K, Caplan D, Alpert N, Rauch S. 1996. Localization of syntactic comprehension by positron emission tomography. Brain and Language 52:452–473.

Thulborn KR, Carpenter PA, Just MA. 1999. Plasticity of language-related brain function during recovery from stroke. Stroke. 30:749–754.