The CAPS Family of Cognitive Architectures

Sashank Varma

The Oxford Handbook of Cognitive Science
Edited by Susan E. F. Chipman

Print Publication Date: Oct 2017 Subject: Psychology, Cognitive Psychology
Online Publication Date: Nov 2014 DOI: 10.1093/oxfordhb/9780199842193.013.002

Abstract and Keywords

Over the past 30 years, the CAPS family of architectures has illuminated the constraints that shape human information processing. These architectures have supported models of complex forms of cognition ranging from problem solving to language comprehension to spatial reasoning to human–computer interaction to dual-tasking. They have offered pioneering explanations of individual differences in the normal range and group differences in clinical populations such as people with autism. They have bridged the divide between the mind and brain, providing unified accounts of the behavioral data of cognitive science and the brain imaging data of cognitive neuroscience. This chapter traces the development of the CAPS family of architectures, identifying the key historical antecedents, highlighting the computational and empirical forces that drove each new version, and describing the operating principles of the current architecture and the dynamic patterns of information processing displayed by its models. It also delineates directions for future research.

Keywords: Cognitive architecture, working memory, resource constraints, cognitive neuroarchitecture, cortical center, graded specialization, dynamic spillover, contralateral takeover, underadditivity, underconnectivity

Introduction

The Collaborative Activation-Based Production System (CAPS) family of cognitive architectures has developed continuously over the past 30 years.[1] That it has done so during a period that has seen incredible changes in cognitive science, from the rise of cognitive neuroscience to the addressing of applied problems in human–computer interaction, is a testament to the power of its core assumptions. In addition to providing a unified theory of cognition, as any cognitive architecture does, the CAPS family has made four important contributions to cognitive science. First, these architectures have demonstrated the utility of *constraints* on information processing in explaining the shaping of human cognition. Second, they have explained individual differences in human cognition, offering mechanisms whose parametric variation explains differences between individuals and among groups. Third, they have bridged the historically and philosophically important barrier between mind and brain, offering unified accounts of the behavioral data of cognitive science and the brain imaging data of cognitive neuroscience. Fourth, they have supported models of the most important but least tractable forms of cognition, including language understanding, problem solving, and spatial reasoning.

This chapter has four goals. The first is to sketch the evolution of the CAPS family over the past 30 years, focusing on the scientific imperatives that drove each new version and the associated models. The second goal is to describe the current version, 4CAPS, which was the first cognitive architecture capable of explaining behavioral and brain imaging data collected from both normal adults and neurological patients, including those with brain (p. 50) lesions and those with autism. The third goal is to describe how the operating principles of 4CAPS work together to provide insightful accounts of the empirical data. The fourth goal is to discuss limitations of 4CAPS and how they motivate some of the questions being addressed in ongoing research.

History

The CAPS family of architectures proposes that cognitive information processing is fundamentally shaped by constraints. This section unpacks this proposal. It first sketches the historical development of the architecture concept within cognitive science and the antecedent and contemporary systems that embody constraint satisfaction. It then describes how the notion of constraints has been developed in the CAPS family, from *informational* constraints in CAPS to *resource* constraints in 3CAPS. This sets the stage for the description of the *cortical* constraints of 4CAPS.

Cognitive Architecture

A central tenet of cognitive science is that cognition is a form of information processing, and therefore its theories typically take the form of computational models. Although computational models can be expressed in conventional programming languages, most are expressed in computational formalisms that are informed by what is known about the representations, processes, and control structure of the mind. Perhaps the earliest such example is the Information Processing Language, dating to the late 1950s (Newell, Shaw, & Simon, 1958; Newell & Tonge, 1960). This language differed from its contemporaries (e.g., Fortran) in including computational mechanisms that supported faculties of the mind such as associative retrieval from memory and dynamic expansion of problem spaces (Simon, 1998).

In the early 1970s, Newell collaborated with Gordon Bell, lead designer of several iterations of the seminal PDP and VAX minicomputers, on what would become a leading textbook of computer engineering (Bell & Newell, 1971). The book argued for the explicit definition of *computer architectures*, or sets of instructions that define the interface between physical hardware and virtual programs. Newell (1973*a*) imported this notion into cognitive science, arguing for the definition of *cognitive architectures*, or sets of computational mechanisms defining the interface between the biological substrate of the brain and the mental representations and processes of the mind. He offered production systems as a candidate architecture (Newell, 1973*b*).

Newell's research group pursued architectures during the 1970s within an artificial intelligence (AI) context. The result was the OPS family of production system architectures, which emphasized computational efficiency over psychological plausibility. However, the notion of architecture took root within cognitive science more generally, and the late 1970s and the 1980s saw the development of multiple candidate systems (Anderson, 1976; 1983; Holland, Holyoak, Nisbett, & Thagard, 1986; Rumelhart & McClelland, 1986). It was in this milieu that the CAPS family of architectures was born.

Constraint Satisfaction

Architectures make claims about the *style* of cognitive information processing (Varma, 2011). The CAPS family of architectures construes cognition as constraint satisfaction. From this perspective, a task is performed not by executing a sequence of operations to compute a solution, but rather by applying a set of constraints to the space of feasible solutions to identify a satisfactory solution.

The first cognitive science system to cast cognition as constraint satisfaction was the Pandemonium model of visual perception (Selfridge, 1959). This model was organized hierarchically. At the lowest level, a number of "demon" recognized basic features of an image, each "shouting" when its particular feature was present. These shouts were "heard" by demons in the next level, which recognized pairs of features and shouted when these were present. As shouts propagated upward through the hierarchy, higher order patterns were recognized, with one emerging at the highest level as the best interpretation of the scene.

Pandemonium proved hugely influential. Newell remarked that its decentralized model of information processing "turned my life" (McCorduck, 1979, p. 134). Constraint satisfaction is at the heart of Quillian's (1968) semantic network model of semantic memory, in which activation spreads from nodes corresponding to retrieval cues and converges on the most central related node, corresponding to the memory to be retrieved. Waltz (1975) applied this technique to the problem of understanding Blocks World scenes, showing that, in many cases, local constraint satisfaction is sufficient for unambiguous recognition.

The next milestone in the development of the constraint satisfaction processing style was (p. 51) the HEARSAY system for speech understanding (Erman, Hayes-Roth, Lesser, & Reddy, 1980). The acoustic signal is noisy and underspecified. Speech understanding therefore requires applying constraints from all levels of language—phonetic, lexical, syntactic, referential, and so on—to tame the problem of ambiguity and the combinatorial explosion it causes. Language is more complex and structured than visual recognition and semantic memory, and the network architectures of Selfridge, Quillian, and Waltz are insufficient. HEARSAY therefore proposed a modular architecture in which different "knowledge sources" (essentially productions) shared hypotheses on a "blackboard" (essentially working memory). Hypotheses interacted with and constrained each other, allowing one interpretation of the utterance to emerge as most likely.

The sense of constraint satisfaction embodied in HEARSAY inspired the next generation of architectures, including classifier systems (Holland et al., 1986), the Rumelhart and McClelland (1982; McClelland & Rumelhart, 1981) word recognition model, the construction-integration model of language understanding (Kintsch, 1988), and, farther afield, the connection machine of Waltz and Pollack (1985) and the analogy programs of Hofstadter's (1995) group. The first CAPS architecture also belonged to this class.

CAPS (1981–1992)

The CAPS architecture was developed by Just, Carpenter, and Thibadeau to express the READER model of language understanding (Just & Carpenter, 1987; Thibadeau, Just, & Carpenter, 1982). Like HEARSAY, READER computed multiple levels of language, from orthographic processing of word forms to referential understanding of discourse, and employed informational constraints to tame ambiguity. CAPS was developed to express this processing style. It employs a hybrid mixture of symbolic and activation-based mechanisms.

Symbolically, CAPS is a production system interpreter, derived from OPS4 (Forgy, 1979). In production system interpreters, cognitive representations are encoded as working memory elements (wmes), cognitive processes are encoded as production rules, and the control structure is typically serial. In more detail, a wme is a declarative representation, like a logical proposition or connectionist feature vector. It has a type and a set of attributes, each of which can take a symbolic or numerical value. Productions are condition–action pairs. The condition aspect specifies a pattern of wmes. This pattern includes unary tests on the attribute values of individual wmes and n-ary tests across the attribute values of multiple wmes; these n-ary tests are implemented via *variable bindings*. The action aspect specifies actions to be performed on working memory, typically adding a new wme or removing an existing wme. The unit of time is the recognize–act cycle. During the recognize phase, the condition sides of all productions are matched against all wmes, generating a set of instantiated productions. There are typically multiple instantiations. The control structure is the scheme by which a subset of instantiations are selected and fired (i.e., their action sides executed), changing the contents of working memory for the next cycle.

CAPS also incorporates activation-based mechanisms similar to those utilized by the semantic networks of the 1970s and the localist connectionist networks that were emerging in the early 1980s. Each wme possesses an activation level indicating its relevance for future processing. The condition aspects of productions contain thresholds, and for a wme to match, it must both pass the unary and n-ary symbolic tests and possess an activation greater than threshold. With respect to the action aspects, the primary action of CAPS productions is to direct activation from one wme to another wme, multiplied by a weight. Positive weights produce excitatory activation and negative weights inhibitory activation.

Finally, the control structure of CAPS is fully parallel, with all instantiations fired on each cycle.[2]

The hybrid nature of CAPS supports graded representations (in that some representations are more active than others), graded processing (in that activation levels are excited or inhibited incrementally over time), and a concurrent control structure. These are the necessary characteristics for implementing a soft constraint satisfaction style of information processing. READER understands language not in a bottom-up fashion, but interactively. Orthographic, lexical, syntactic, thematic, and referential representations are constructed concurrently. Ambiguity within each level (e.g., given a polysemous word form, what is the correct lexical sense?) is reduced by informational constraints between levels (e.g., prefer the lexical entry that is more consistent with the current syntactic structure). In this way, informational constraints incrementally adjust the activation levels of language representations at multiple levels until a coherent understanding emerges.

Another important CAPS model—or rather pair of models—addressed problem solving on (p. 52) the Ravens progressive matrices test (hereafter “the Ravens”). This effort contributed to the evolution of the CAPS family of architectures, and so it is important to understand its achievements. Cattell’s (1963) theory of intelligence postulates two subfactors. *Crystallized intelligence* consists of all the knowledge and skills a person has learned. *Fluid intelligence* is the ability to solve abstract and novel (i.e., knowledge-independent) problems. The most prominent measure of fluid intelligence is the Ravens. In this test, people view a sequence of visual matrices, induce the hypothetical rules that govern their structure, and apply them to infer their missing components. The psychometric approach to intelligence and tests like the Ravens belong to differential psychology, which focuses on individual differences between people. By contrast, cognitive psychology belongs to experimental psychology and typically focuses on average performance, regarding individual differences as noise to be minimized. A few prominent exceptions aside, these two approaches have historically had little to say to each other (Cronbach, 1957).

To bridge this gap, Carpenter, Just, and Shell (1990) conducted an eye-tracking experiment of Ravens problem solving, formulated theories of average and superior performance, and instantiated these theories in two CAPS models. FAIRRAVEN solved Ravens problems by proposing multiple rule hypotheses in parallel, consistent with the constraint satisfaction processing style of CAPS. This approach worked well for easier problems but not for harder problems requiring the coordination of multiple rule hypotheses. Carpenter et al. (1990) proposed that the key individual difference in Ravens problem solving is the ability to exert executive control. BETTERRAVEN augmented FAIRRAVEN with a handful of productions that serialized problem solving through the articulation of goals in working memory. The proposal that individual differences in cognitive information processing can be understood as individual differences in executive control and working memory was an important bridge between differential and experimental psychology. It was also an important force in the development of the next member of the CAPS family of architectures.

3CAPS (1992–1999)

The CAPS architecture supports models that enforce informational constraints between representations. To this, the 3CAPS (Capacity-Constrained CAPS) architecture adds support for modeling resource constraints on cognitive information processing.[3]

Development of 3CAPS was spurred by experiments by Just and Carpenter in the late 1980s on individual differences in sentence comprehension (Carpenter & Just, 1989; King & Just, 1991; MacDonald, Just, & Carpenter, 1992). Why do some people have difficulty understanding complex syntactic constructions? Why do some people have difficulty understanding sentences that are temporarily ambiguous? As with Ravens, this research brought the question of individual differences to the forefront of cognitive science. The findings showed that individual differences in working memory capacity predicted reading time profiles on structurally complex sentences, peaking when readers had to maintain and process multiple unattached noun phrases, and predicted reading time profiles on temporarily ambiguous sentences, peaking when readers had to maintain and process multiple interpretations.

Under the direction of Just and Carpenter, Varma added a capacity constraint to the working memory of CAPS, producing a transitional architecture that was internally designated CAPS89 (Varma, 1990). This mechanism constrained the total activation available for storage and processing of representations. More precisely, at the end of each cycle, the storage demand was computed by summing the activations of all wmes. The processing demand was computed by summing all activations being directed by firing the action sides of all instantiations. The total demand was the sum of the storage and processing demands. If the total demand exceeded the capacity, then all storage and processing demands were scaled back proportionately so that the total allocated activation was equal to the capacity. That is, the activation of each wme in working memory was scaled back, as was each new direction of activation.

When performing a relatively easy task, the total demand does not exceed the capacity, and cognitive information processing in 3CAPS proceeds as in CAPS. However, when performing a relatively difficult task, the total demand can exceed the capacity, especially at the most difficult points. When this happens, the activation levels of wmes are scaled back. If this occurs over multiple cycles, then wme activations will fall below the threshold to participate in processing. This is a kind of "forgetting by displacement." Another consequence of the total demand exceeding the capacity is that the activations directed by firing instantiations are also scaled back. This increases the time required to activate new wmes above (p. 53) threshold, slowing processing. These dynamics have implications for individual differences in working memory capacity, which are simulated by varying the activation capacity of the system. Lower capacities result in critical wmes being displaced from working memory, leading to errors, and also result in slower processing.

A number of important models were written in 3CAPS. First among these is CC READER (Just & Carpenter, 1992; Just & Varma, 2002). Contrary to what its name suggests, it is not a reimplementation of the original READER model. Rather, it is a capacity-constrained model of sentence comprehension, performing lexical, syntactic, and thematic processing. CC READER can comprehend a range of sentence types. It takes a parallel approach to ambiguity, constructing all possible interpretations until there is sufficient information to disambiguate processing and identify a single interpretation. In this, it is consistent with the constraint satisfaction processing style of the CAPS family.

CC READER updated a classic proposal in psycholinguistics: that the gap between linguistic competence (i.e., knowledge of language) and linguistic performance (i.e., how that knowledge is applied during comprehension and production) is explained by limitations of cognitive information processing, including the number of "slots" in short-term memory for storing intermediate phrase markers during parsing. CC READER replaced the older construct of short-term memory, which emphasizes only storage, with the newer construct of working memory, which postulates a joint constraint on storage and processing (Baddeley & Hitch, 1974; Daneman & Carpenter, 1980). 3CAPS was the first formal account of working memory, and CC READER the first formal psycholinguistic account of how working memory limitations degrade linguistic competence to linguistic performance.

3CAPS enabled cognitive scientists to move beyond Navon's (1984) influential critique of resource constraints as a "theoretical soup stone"; that is, a vacuous theoretical construct supplying no explanatory power. It rendered this critique moot by formalizing resource constraints within a broader architectural account of cognitive information processing. This mechanism was subsequently taken up by ACT-R (Anderson, Reder, & Lebiere, 1996), where it was implemented more narrowly as a constraint on the activation emanating from goal wmes. To take another example, Byrne (1998) developed the SPAN architecture that posited that the limiting mechanism in working memory was not resources, but processing speed (e.g., Salthouse, 1996).

There were several commentaries on 3CAPS and CC READER offering smaller scale, alternative accounts of a few of the phenomena that these models explained. Waters and Caplan (1996) argued that there were in fact two separate working memory limitations on sentence comprehension, one on "obligatory" verbal processing and the other on "controlled" verbal processing. Leaving aside the details of their proposal, 3CAPS was in fact neutral on the question of whether working memory was a unitary or fractionated construct. In fact, it provided a facility for defining multiple working memories and specifying the resources of each independently. This capability was included in 3CAPS to account for findings on the existence of separable verbal and spatial components of working memory and intelligence. MacDonald and Christiansen (2002) argued that individual differences in sentence comprehension were driven not by resource differences but by experience differences and produced a toy connectionist model of one of the findings regarding structurally complex sentences that CC READER explained. Their approach was never developed into a full-fledged account of the full range of data for which CC READER accounted, nor were its flaws ever addressed (Caplan & Waters, 2002; Just & Varma, 2002).

A number of important 3CAPS models were developed during the 1990s. The 3CI model, an implementation of the construction-integration model of discourse comprehension (Kintsch, 1988) within 3CAPS, provided a better account of the empirical data when the short-term memory of CI was replaced with the working memory of 3CAPS (Goldman & Varma, 1995; Goldman, Varma, & Côté, 1996). A 3CAPS model of Tower of Hanoi (TOH) problem solving accounted for individual differences in performance as a function of individual differences in working memory capacity for maintaining goals (Just, Carpenter, & Hemphill, 1996). 3CAPS was also the basis for several models of human–computer interaction (Byrne & Bovair, 1997; Huguenard, Lerch, Junker, Patz, & Kass, 1997).

The original implementation of 3CAPS, called CAPS89, was programmed in Maclisp, a dying dialect of the Lisp programming language that made it difficult to distribute the architecture to other researchers. For this reason, in 1992, 3CAPS was reimplemented in Common Lisp, the ANSI standard version of the Lisp language.[4] This implementation (p. 54) adopted the syntax of OPS5 (Forgy, 1982), allowing researchers to learn the symbolic aspect of the architecture through tutorials on OPS5 (Brownston, Farrell, Kant, Martin, 1985; Cooper & Wogrin, 1988). Learning the connectionist aspect required reading one of several short, unpublished manuals (Thibadeau, 1982; Varma, 1990; 1992). In 1995, Henk Haarmann wrote a comprehensive tutorial manual for 3CAPS. The value of this tutorial, and of Haarmann's community-building efforts more generally, was demonstrated by the success of 3CAPS workshops held in 1995 in Pittsburgh and 1996 in Boulder. Each attracted 20–30 attendees from across the country and the world. These workshops were led by Haarmann, and the guest speakers included Just, Carpenter, Varma, and Michael Byrne.

As a harbinger of what was to come, by the late 1990s, researchers began applying 3CAPS to model neuropsychological data, specifically the cognitive impairments that follow brain lesions. Connectionist models simulated lesions by deleting a random set of connections between units. By contrast, 3CAPS models simulated lesions by drastically reducing the available resources. This had the effect of introducing errors (because the activation levels of wmes were scaled back and ultimately displaced from working memory) and slowing processing (because productions required additional cycles to activate wmes). Using this approach, Haarmann, Just, and Carpenter (1997) modeled the canonical impairment observed in agrammatic aphasia: declining comprehension accuracy with increasing syntactic complexity (Caplan, Baker, & Dehaut, 1985). Goel, Pullara, and Grafman (2001) similarly reduced working memory resources to model the impaired TOH problem solving of patients with frontal lesions, specifically their larger deviations from optimality and longer solution times compared to normal controls and patients with non-frontal lesions.

The 4CAPS Cognitive Neuroarchitecture (1999–Present)

Can we measure and model the *intensity* of thought? 4CAPS has its origins in Just and Carpenter's investigations of the mid-1990s into what it means to think hard. Their initial attempts followed-up on an old proposal of Kahneman's (1973) that pupil diameter is an online measure of processing intensity. They initially conceptualized intensity as the ratio of one's resource demand to one's resource supply (Just & Carpenter, 1993). This has two interesting properties. First, the more difficult a task, and therefore the greater the resource demands, the greater the intensity of thought. The second property concerned individual differences: the smaller one's resource capacity, the harder he or she has to think when performing a task.

The rise of functional magnetic resonance imaging (fMRI) shortly thereafter provided a more direct way to measure processing intensity and reoriented Just and Carpenter's research agenda toward cognitive neuroscience. Recall that 3CAPS had previously been applied to neuropsychological data. By drastically reducing the resource supply, 3CAPS models were able to account for the impaired sentence comprehension that followed lesions to Broca's area and the impaired TOH problem solving that followed lesions to the frontal lobe. However, this approach suffers from two problems. First, it does not extend to modeling brain activation as measured by fMRI and positron emission tomography (PET). Second, it makes the localist assumption that cognitive constructs (i.e., working memory resources) map to brain areas (e.g., Broca's area) in a one-to-one manner. However, this assumption is almost certainly incorrect. As earlier lesion studies (Luria, 1966; Mesulam, 1990) and contemporary imaging studies (Just, Carpenter, Keller, Eddy, & Thulborn, 1996) reveal, cognition is the product of a network of interacting brain areas, the membership and topology of which change dynamically as a function of task demands. The next member of the CAPS family of architectures, 4CAPS, addresses these problems.

At the time Just and Carpenter were beginning to measure the intensity of thought using fMRI, Varma had been independently investigating the implementation of multiple working memories in a new system termed the object-oriented production system interpreter (OOPSI). OOPSI permitted the definition of multiple production system models, each with its own resource supply, which interacted to perform cognitive tasks. To evaluate the compatibility of these research projects, Varma spent May 1997 in Pittsburgh in discussions with Just and Carpenter. The result was a set of operating principles defining a new style of cortical information processing. An initial attempt to imbue OOPSI with these principles was successful, resulting in the first version of the 4CAPS (Cortical 3CAPS) architecture. Over the next 2 years, they co-articulated 4CAPS and a model of sentence comprehension, refining the operating principles into a minimal set that was sufficient for accounting for the behavioral and brain imaging data for one complex task.

(p. 55) 4CAPS is a cognitive *neuro*architecture. It accounts for behavioral and brain imaging data collected from a range of individuals—normal adults, patients with brain lesions, people with autism, and so on. Like its predecessors, it is a hybrid symbolic-connectionist architecture that highlights the role of informational and resource constraints on cognitive information processing. Its new claims concern cortical information processing. Cognition is understood as the emergent product of multiple collaborating brain areas, and information processing is profoundly shaped by resource constraints. In more detail, the cortex consists of a set of interconnected areas, each with its own functional specializations. There are resource limitations on *computation* within individual brain areas and on *communication* between brain areas. Cortical information processing can be understood as optimally distributing responsibility for performing requisite functions across brain areas such that computational and communicative resource demands are minimized while cognitive throughput is maximized. This proposal is articulated in later sections; the reader interested in more details is referred to Just and Varma (2007).

Centers and Cognitive Functions

A 4CAPS model consists of a set of *centers*, each corresponding to a cortical area (i.e., a gyrus or sulcus). Each center is a hybrid symbolic-connectionist computational system—essentially an encapsulated 3CAPS production system. With respect to the symbolic side, representations are encoded as wmes and processes as productions. The activation dynamics—the activation levels of wmes and the constraint on total activation—work similarly to CAPS and 3CAPS, although the actual algorithms are different, as described later. Denote the number of centers by M.

A 4CAPS model performs a cognitive task such as sentence comprehension or problem solving. Each task can be decomposed into a set of *cognitive functions* (or, more simply, “functions”). A function is a convenient abstraction for the set of wmes and productions that together implement a cognitively interesting operation such as parsing the syntax of a sentence or maintaining a goal stack. Denote the number of functions by N.

This raises the question of how functions are mapped to centers. Several mappings are possible. Modularity proposes a one-to-one mapping: each function is implemented by exactly one center, and each center implements exactly one function. This extreme form of localism is assumed, for example, by the Wernicke-Geschwind model of language understanding (Geschwind, 1970) and by the ACT-R architecture (Anderson, 2007). Equipotentiality, by contrast, proposes an N-to-M mapping, with every function implemented by every center (and vice versa). This assumption is exemplified by Lashley’s (1950) notion of mass action and by connectionist architectures (Rumelhart & McClelland, 1986). 4CAPS proposes a third mapping, one intermediate between modularity and equipotentiality: centers are multiply and gradedly specialized for functions. Each center can perform multiple functions, some more efficiently than others. Conversely, each function can be performed by multiple centers, although some are more efficient than others. More precisely, the *specialization* of center i for function j is denoted $S_{ij} \in [1, \infty)$. This is the amount of resources required to perform one unit of the function. Perfect specialization is indicated by the value 1, lesser specializations are indicated by larger values, and a complete inability of center i to perform function j is indicated by the value ∞ (because resources are limited, as we will see below). If the units of function j performed by center i are denoted A_{ij}, then the resource demand of function j on center i is $A_{ij} \times S_{ij}$, and the total resource demand on center i is:

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

We specify how the A_{ij} are computed below.

Resource Constraints on Computation and Communication

4CAPS inherits from 3CAPS a focus on resource constraints. Because it is a neuroarchitecture rather than an architecture, these constraints are enforced at the cortical level rather than the cognitive level. There are two sets of resource constraints. There are resource constraints on computation within centers. Each center possess a finite supply of resources—of activation, as in 3CAPS—for fueling storage and processing. This reflects a fundamental biological limitation on the energetic resources (e.g., glucose) available in each brain area—an upper bound on what the vascular system can supply. Denote the resource capacity of center i as C_i. The following computational constraint is enforced at all times:

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

(1)

This mandates that the resource demand on a center cannot exceed its resource supply.

(p. 56) There are also resource constraints on communication between centers. Performing a task recruits a network of centers with relevant functional specializations. These centers collaborate through the sharing of wme representations. This sharing across a network is implemented by the white-matter tracts that connect brain areas. There are bandwidth limitations on this communication that reflect the integrity of the underlying white-matter tracts. Rather than modeling the communication network on a point-to-point basis, 4CAPS posits constraints on the *joint* resource consumption of multiple collaborating centers. For ease of exposition, we assume a joint resource constraint on the resource consumption of all centers:

$$\sum_{i=1}^{N} \sum_{j=1}^{M} (A_{ij} \times S_{ij}) \leq C_{CORTEX}$$

(2)

This ensures that resource consumption across all centers cannot exceed the entire cortical supply, denoted C_{CORTEX}. The number of such intercenter constraints, and the centers constrained by each, are empirical matters (Just & Varma, 2007).

Collaborative Processing

Performing a task requires executing a set of functions. The set is partially ordered by informational dependencies between the functions (i.e., when an output of one is an input of another). This recruits a network of centers with relevant functional specializations. Two factors determine the participation level of a particular center at a particular point in time. The first is whether the center is specialized for the functions pending execution: the better its specializations, the greater the likelihood of participation. The second is whether the center possesses free resources to fuel the execution of pending functions: the more free resources, the greater the likelihood of participation. As the functional demands of task performance change over time, the cortical network configures and reconfigures itself, shifting computation efficiently between centers.

These dynamics can be formalized as follows. At each point during task performance, there exists an agenda of functions to be performed, determined by hybrid symbolic-connectionist information processing within individual centers. The *assignment problem* is how to allocate the resources of centers to perform the functions in a way that minimizes overall resource consumption while maximizing cognitive throughput. This amounts to computing the A_{ij}, which denote the amount of function j performed by (i.e., assigned to) center i. This is done through constraint satisfaction.

More precisely, there are three sets of constraints on the A_{ij}. First, there are M intracenter constraints, one for center i, stipulating that the computational demand on the center's resources does not exceed its resource supply C_i; these are shown in Equation (1). Second, there are intercenter constraints that stipulate that the communication demand does not exceed bandwidth limitations; these are shown in Equation (2). Third, there are N constraints, one for each function j, stipulating that as much of the resource demand for the function as possible, denoted R_j, is satisfied by the resource supplies of the various centers:

$$\sum_{i=1}^{M} A_{ij} \leq R_j$$

(3)

Many assignments satisfy constraints (1), (2), and (3), such as A_{ij}= 0 for all i and j. To select among them, we define a measure of the goodness of an assignment. This *objective function* is a linear combination of the A_{ij}:

$$\sum_{i=1}^{M} \sum_{j=1}^{N} (W_{ij} \times A_{ij})$$

(4)

Setting the weights as $W_{ij}= 1/S_{ij}$ ensures that the objective function will be maximized when functions are assigned to the centers that are most specialized for them.

The constraints (1), (2), and (3) and the objective function (4) define a linear programming (LP) problem. The proposal is that the brain effectively solves such an LP problem at each point in time, assigning functions to brain areas. 4CAPS solves the assignment LP problem using the simplex algorithm (Dantzig & Thapa, 1997). The resulting assignment has two important properties. First, it respects resource constraints on cortical computation and communication. Second, it minimizes overall resource consumption by assigning cognitive functions to the centers most specialized for their performance, all other things being equal. As a result, it maximizes the throughput of cortical information processing. We return to the question of whether resource allocation in 4CAPS is too optimal when considering limitations and directions for future research.

Capacity Utilization

One goal of the 4CAPS neuroarchitecture is to formalize the intensity of thought, another to account (p. 57) for fMRI data. Both are achieved by defining the *capacity utilization* of center i, denoted CU_i, as follows.

$$CU_i = \frac{\sum_{j=1}^{N} (A_{ij} \times S_{ij})}{C_i}$$

The CU of a center—the proportion of its resource supply currently being used to fuel execution of cognitive functions—is an index of how hard it is working. The key linking assumption is that the CU of a center predicts activation in the corresponding brain area as measured by fMRI and related methodologies. This is a conceptually coherent hypothesis because the fMRI signal reflects how hard a brain area is working, specifically the vascular response to the expenditure of bioenergetic resources (Logothetis, 2003).

4CAPS models can account for the results of fMRI studies that employ block designs. In these studies, participants process sets of similar stimuli. The average activation in a brain area while processing a set of similar stimuli is predicted by the average CU in the corresponding model center when processing the same set.

4CAPS models also account for the results of event-related fMRI studies, where multiple volumes are acquired during the processing of each stimulus, generating an activation time series for each brain area. Note that fMRI does not measure neural computation directly, but rather the vascular system's response to neural computation. The hemodynamic response is sluggish—delayed and then distributed in time. It can be approximated by a γ function with a fixed delay δ (Aguirre, Zarahn, & D'Esposito, 1998; Boynton, Engel, Glover, & Heeger, 1996).

$$h(t) = \frac{((t-\delta)/\tau)^{n-1} e^{-(t-\delta)/\tau}}{\tau(n-1)!} \text{ when } t \geq \delta, 0 \text{ otherwise}$$

We adopt this approximation of the hemodynamic response function and published parameter estimates ($\delta = 2.5$, $\tau = 1.25$, $n = 3$). Event-related data are predicted by first sampling the *CU* of a center at the same frequency with which brain volumes are acquired (e.g., one image every second) to generate a $CU_i(t)$ time series. This is a moment-to-moment prediction of neural computation. This is then convolved with the hemodynamic response function $h(t)$ to generate a predicted activation time series:

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

The predicted activation time series of a center predicts the observed activation time series in the corresponding brain area.

4CAPS models account for other neuroscience measures. A lesion to a brain area can be simulated by drastically reducing the resource supply of the corresponding center. The increasing error rates and slower performance that follow can be compared to the behavioral impairments exhibited by patients. In addition, 4CAPS will dynamically reassign the execution of functions formerly performed by the lesioned center to other centers with lesser specializations but possessing sufficient resources, a prediction that can be evaluated against fMRI data from patients. To take another example, the negative impacts of reduced white-matter tract integrity in autism and dyslexia can be simulated by reducing the communication bandwidth between centers. We will see examples of these applications to neuropsychological data later.

4CAPS Models: The Dynamics of Cortical Information Processing

4CAPS models have been constructed for a range of cognitive domains, demonstrating the generality of the neuroarchitecture. Like their CAPS and 3CAPS predecessors, they account for behavioral data—response times and error rates—collected from typical adults and patients with brain lesions. Their novel contribution is in accounting for brain activation and functional connectivity data collected from these and other populations.

Although 4CAPS models necessarily make proposals about the localization of cognitive functions to brain areas, these proposals are not the focus of this section. Rather, it is to illustrate how the operating principles embodied in 4CAPS explain the dynamics of cortical information processing. We articulate four dynamic patterns here and illustrate them with model simulations. The patterns are not *assumptions* engineered into 4CAPS, but rather *consequences* of its operating principles.

Parametric Response to Increasing Computational Demands

The first dynamic pattern exhibited by 4CAPS models is that with increasing task difficulty comes increased *CU*s in the centers most specialized for the cognitive functions to be performed. This is a simple consequence of the fact that more difficult tasks make greater resource demands and the definition of *CU* as the ratio of resource demands to resource supply. (p. 58)

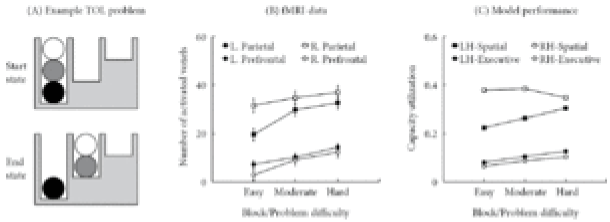

Fig. 3.1 Parametric response to increasing computational demands. (A) Example Tower of London (TOL) problem. The task is to transform the start state into the end state by repeatedly moving one ball from the top of one pocket to the top of another. (B) As the number of moves required to solve a problem increases, so does activation in bilateral prefrontal cortex and left parietal cortex (Newman et al., 2003). (C) The capacity utilizations of the corresponding centers of the 4CAPS model display the same dynamic pattern. The correlation between human and model performance across the 12 points of comparison—four brain areas by three difficulty levels—is 0.96 ($p < .01$).

Parametric response is seen in the 4CAPS model of Tower of London (TOL) problem solving (Newman, Carpenter, Varma, & Just, 2003). An example TOL problem is shown in Figure 3.1A. The task is to transform the start state into the end state by moving one ball at a time from the top of one pocket to the top of another pocket. Difficult problems cannot be solved from *perceptual* problem solving alone (i.e., by directly moving balls to their locations in the end state). They additionally require *strategic* problem solving (i.e., formulating plans for clearing blocking balls so that deeply buried balls can be moved). Patients with frontal lesions can solve simple TOL problems but not difficult ones, suggesting that the frontal lobes are neural correlates of strategic problem solving (Shallice, 1982). More recently, fMRI studies have investigated the neural correlates of TOL problem solving in typical adults. Newman et al. (2003) had participants solve blocks of problems of increasing difficulty. Activation was measured in two regions of interest (ROIs) for perceptual problem solving, left and right intraparietal sulcus (IPS) and superior parietal lobule (SPL), and two ROIs for strategic problem solving, left and right dorsolateral prefrontal cortex (DLPFC). Increasing activation with increasing block difficulty was found in all regions except right parietal cortex (see Figure 3.1B).

Newman et al. (2003) constructed a 4CAPS model of TOL problem solving. They first defined a general model of frontoparietal interaction. Two Executive centers are important for strategic problem solving. RH-Executive corresponds to right DLPFC and is specialized for proposing goals (i.e., formulating plans) when perceptual problem solving falters and for proposing strategic operators that achieve these goals. LH-Executive corresponds to left DLPFC and is specialized for applying heuristics to choose among proposed operators (e.g., prefer those that achieve goals over those that do not). Two Spatial centers are important for perceptual problem solving. Corresponding to right IPS/SPL is RH-Spatial, which is specialized for proposing perceptual operators that increase the similarity between the current state and the ending state. Corresponding to left IPS/SPL is LH-Spatial, which is specialized for applying the operator selected by LH-Executive to the current state to generate the next state. Newman et al. (2003) instantiated the frontoparietal model for the domain of TOL problem solving by instantiating states as puzzle configurations, operators as ball movements, and so on. The TOL model solved the same problems used in the fMRI study. The *CU*s of its four centers are shown in Figure 3.1C. All but RH-Spatial show a parametric response with block difficulty, mirroring the activations observed in the corresponding brain areas.

The parametric relation between task difficulty and brain activation can also be observed on a moment-to-moment basis. Mason, Just, Keller, and Carpenter (2003) conducted an event-related fMRI experiment in which participants read sentences that varied in difficulty on two dimensions: whether they contained one or two clauses and whether they were unambiguous or ambiguous. Immediately after reading each sentence, they answered a question to test their comprehension. Prior research has shown that two-clause sentences are more difficult than one-clause sentences and that ambiguous sentences (p. 59) are more difficult than unambiguous sentence (MacDonald et al., 1992). These were behavioral experiments, and difficulty was reflected in slower reading times. The novel prediction of the fMRI experiment was that more difficult sentences would elicit greater activation in two key components of the language network, Wernicke's area (left posterior superior temporal gyrus; pSTG) and Broca's area (left inferior frontal gyrus; IFG). Figure 3.2A shows the observed activation time series for left pSTG, with images acquired every 1.5 seconds. The prediction was corroborated, with the highest activations observed for the notoriously difficult two-clause, ambiguous ("reduced relative") sentence. (The predictions also held for left IFG, not shown.)

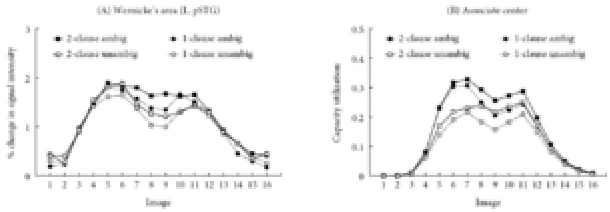

Fig. 3.2 Parametric response to increasing computational demands as revealed by event-related functional magnetic resonance imaging (fMRI). (A) As sentence difficulty (i.e., number of clauses and presence of ambiguity) increases, so does activation in Wernicke's area (left posterior superior temporal gyrus [L. pSTG]) on a moment-to-moment basis (Mason et al., 2003). (B) Capacity utilization in the corresponding Associate center of the 4CAPS model shows the same dynamic pattern when sampled at the same rate (one image every 1.5 seconds). The correlation between human and model performance across the 64 points of comparison—four sentence types by 16 images—is 0.86 ($p < .01$).

Just and Varma (2007; Varma & Just, in preparation) developed a 4CAPS model capable of accounting for these and other findings on sentence comprehension. It consists of centers corresponding to left pSTG, left IFG, and their right-hemisphere homologs. The functional specializations of these centers were assigned based on prior neuropsychological and neuroimaging studies. The Associate center, which corresponds to left pSTG, is specialized for combining existing language representations (e.g., phrases) into designs for new, superordinate language representations (e.g., clauses). The Structure center, which corresponds to left IFG, is specialized for manufacturing structured language representations from these associative designs. The analysis of language processing into associative design and structured manufacturing functions and the assignment of these functions to left pSTG and left IFG, respectively, is a novel proposal of the 4CAPS model. Another novel proposal is that the right-hemisphere homologs of these areas, mapping to the RH-Associate and RH-Structure centers, are also specialized for the associative design and structured manufacturing functions, respectively, but to a lesser degree.

The model was run on the same sentences used in the Mason et al. (2003) study. The *CU* time series of each center was computed for each sentence type, and each was convolved with the hemodynamic response function to yield a predicted fMRI time series. Those for the Associate center are shown in Figure 3.2B. The model provides a good account of the data. Critically, the predicted fMRI time series shows the same relative ordering as the observed fMRI time series, with increasing sentence difficulty (i.e., multiple clauses, presence of ambiguity) producing higher average and peak activations.

Dynamic Spillover

The second dynamic pattern exhibited by 4CAPS models is a sequelae of the first. As task difficulty increases, so do the resource demands on centers that are well specialized for the cognitive functions to be performed—this is the first pattern. When the task is sufficiently difficult, the resource demands will exceed the resource supplies of well-specialized centers, and processing will spill over to centers that are less specialized for the functions to be performed but which possess spare resources. Spillover is dynamic and on an as-needed basis: when the most difficult aspect of a task has been completed and resource demands have eased to the point where the resource supplies of well-specialized centers are again sufficient, then processing will migrate back to them. (p. 60)

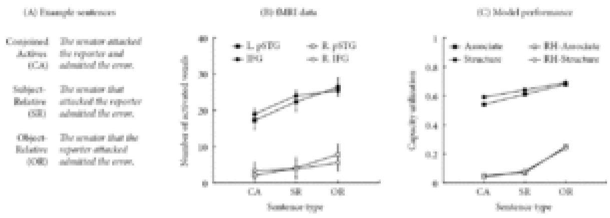

Fig. 3.3 Dynamic spillover. (A) Sentences of increasing structural complexity. (B) As structural complexity increases, so does the activation in classic left-hemisphere language areas—Wernicke's area (left posterior superior temporal gyrus [L. pSTG]) and Broca's area (left inferior frontal gyrus [L. IFG])—precipitating dynamic spillover of activation to their right-hemisphere homologs (Just et al., 1996). (C) The capacity utilizations of the corresponding centers of the 4CAPS model exhibit the same dynamic pattern. The correlation between human and model performance across the 12 points of comparison—four brain areas by three sentence types—is 0.98 ($p < .01$).

Just et al. (1996) documented dynamic spillover in an fMRI study in which adults read sentences of increasing structural complexity:

- conjoined actives: [*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*].

The conjoined actives sentence concatenates two clauses. It makes relatively light resource demands because processing of the first clause finishes before processing of the second clause begins. By contrast, the subject-relative sentence embeds one clause inside the other. This makes heavier resource demands because processing of the first clause is interrupted, and its partial products (i.e., its subject *the senator*) must be maintained while the second clause is processed so that they are available when processing of the first clause resumes. The object-relative sentence is more complex still because, in addition, the partial products of the second clause (i.e., the subject *the reporter* and object *the* *senator*) must be maintained for longer until the verb *attacked* is finally encountered and its grammatical and thematic relations can be computed.

Just et al. (1996) found increasing activation in left pSTG and left IFG with increasing structural complexity (see Figure 3.3B). This is the first dynamic pattern. Critically, they also found increasing activation in right pSTG and right IFG. This is the second dynamic pattern—the spillover of excess processing from well-specialized left-hemisphere language areas with insufficient resources to their less-specialized right-hemisphere homologs possessing spare resources. Just, Carpenter, and Varma (1999) simulated these data using the 4CAPS model of sentence comprehension (see Figure 3.3C). They found increasing *CU*s in the Associate and Structure centers with increasing structural complexity, consistent with the first dynamic pattern. Critically, they also found increasing *CU*s in the RH-Associate and RH-Structure centers with increasing structural complexity. This reflects increasing spillover of excess resource demands for processing embedded clauses and maintaining subjects and objects deeper into relative clauses.

Contralateral Takeover

The third pattern, *contralateral takeover*, can be understood as a variant of the second pattern, dynamic spillover, operating over a longer time scale. A phenomenon observed in patients with stroke-induced lesions is that the cognitive functions previously performed by a damaged area shift permanently to the homologous area in the contralateral hemisphere (Finger, Buckner, & Buckingham, 2003). The damage can be viewed as a drastic reduction in the available resource supply. Without adequate resources, the area will not be able to perform the functions for which it is specialized, and other areas specialized for the same functions, albeit less so, will be recruited into the large-scale network on a more-or-less permanent basis. Contralateral takeover follows from the resource allocation algorithms embedded in 4CAPS. Specifically, when the resources of centers well specialized for the execution of cognitive functions are drastically reduced, (p. 61) processing spills over to less-specialized centers with available resources, increasing their *CU*s.

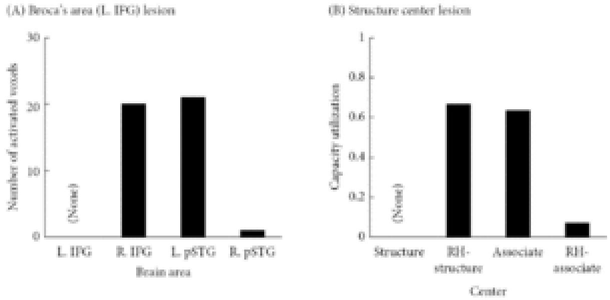

Fig. 3.4 Contralateral takeover. (A) Activation in Broca's area (left inferior frontal gyrus [L. IFG]), Wernicke's area (left posterior superior temporal gyrus [L. pSTG]), and their right-hemisphere homologs during comprehension of single-clause sentences in a patient with a L. IFG lesion (Thulborn et al., 1999). (B) Capacity utilizations of the corresponding centers of the 4CAPS model following "lesioning" of the Structure center by zeroing its resource supply, exhibiting the same dynamic pattern. The correlation between human and model performance across the four points of comparison is 0.99 ($p < .01$).

Thulborn, Carpenter, and Just (1999) documented contralateral takeover in a patient who suffered a stroke that damaged his left IFG. He initially experienced a dense expressive aphasia, but recovered much of his language function over the following 6 months. He was then imaged while reading simple five- and six-word sentences. The activation observed in left IFG, left pSTG, and their right-hemisphere homologs is shown in Figure 3.4A. He displayed the typical pattern of left-lateralized activation in pSTG. Strikingly, the pattern reversed in IFG. This was interpreted as evidence that the processing normally performed by the left IFG had been taken over by the undamaged right IFG.

Just and Varma (2007) simulated these data using the 4CAPS model of sentence comprehension described earlier. They zeroed the resource supply of the Structure center corresponding to the damaged left IFG and ran the model on the same sentences the patient read. The *CU*s of the model's four centers are shown in Figure 3.4B. The model provided a good match to the data, displaying left-lateralized activation in pSTG and right-lateralized activation in IFG. This was a direct consequence of the 4CAPS resource allocation algorithms. The Associate center corresponding to the undamaged left pSTG possessed sufficient resources to fuel most of the associative design of new language representations. By contrast, the Structure center lacked the resources to manufacture these designs into new representations, and thus this cognitive function shifted to the contralateral RH-Structure center, which was less specialized for this function but possessed sufficient resources.

It should be noted that, like their 3CAPS predecessors, 4CAPS models can account for the behavioral deficits caused by brain lesions. For example, Goel and Grafman (1995) had intact controls and patients with (left and/or right) frontal lesions solve nine TOH problems of increasing difficulty. The probability that each problem was solved within 2 minutes for each group was computed (see Figure 3.5A). Not surprisingly, solution probabilities were lower on more difficult problems. Critically, the frontal patients had lower solution probabilities than the intact controls. To address these data, Varma (2006) instantiated the frontoparietal model described earlier for the domain of TOH problem solving. States were instantiated as TOH puzzle configurations and operators as disk movements. Goals and operator selection heuristics were specialized to implement the sophisticated perceptual strategy typically used by human problem solvers (Simon, 1975). Patient performance was simulated by drastically reducing the resource supply of Executive centers corresponding to the damaged frontal areas. The solution probabilities for the control and lesioned models are shown in Figure 3.5B. Reducing the resource supply of RH-Executive (corresponding to a right frontal lesion) diminished planning ability, and reducing (p. 62) the resource supply of LH-Executive (corresponding to a left frontal lesion) diminished the efficacy of operator selection. These negative impacts on model performance match the impairments observed in the human data.

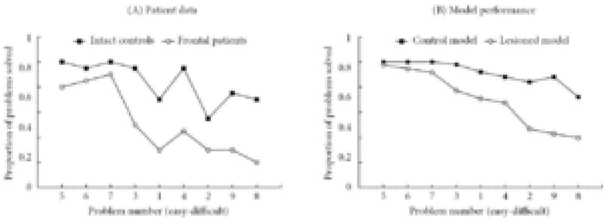

Fig. 3.5 CAPS models account for behavioral impairments caused by brain lesions. (A) Proportion of Tower of Hanoi (TOH) problems solved in 2 minutes by intact controls and patients with frontal lesions (Goel & Grafman, 1995). (B) Performance of the control 4CAPS model and one "lesioned" by reducing the resource supplies of the corresponding Executive centers. The correlation between human and model performance across the 18 points of comparison—two groups/models by nine problems—is 0.90 ($p < .01$).

Dynamic Response to Increasing Communication Demands

The dynamic patterns described thus far are the product of resource constraints on computation within individual centers. The final dynamic pattern arises because of resource constraints on communication between centers. Specifically, intercenter resource constraints implement bandwidth limitations on collaborative processing.

Intercenter resource constraints are not normally a factor when performing an individual task, such as sentence comprehension, because this recruits a single cortical network. However, performing dual tasks—for example, pairing sentence comprehension with mental rotation—recruits two cortical networks simultaneously, increasing the communication demands between centers to the point at which bandwidth limitations become visible. For example, Just, Carpenter, Keller, Emery, Zajac, and Thulborn (2001) had participants perform sentence comprehension and mental rotation tasks individually as single tasks and concurrently as dual tasks. The sentence comprehension task was to verify auditorily presented sentences as true or false. The mental rotation task was to decide whether two three-dimensional block figures were identical to or mirror images of each other. A large literature indicates that people perform this task by mentally rotating the figures, attempting to bring them into alignment (Shepard & Metzler, 1971). The ROIs were two components of the sentence comprehension network, left and right pSTG, and two components of the spatial reasoning network, left and right SPL. As expected, during single-task performance, there was greater pSTG activation during sentence comprehension and greater SPL activation during mental rotation (Figure 3.6A). The surprising finding was *underadditivity* during dual-task performance, with the activation in pSTG less than that during single-task sentence comprehension and the activation in SPL less than that during single-task mental rotation task. This is in spite of the fact that behavioral performance remained high.

Just and Varma (2007) developed a 4CAPS model of these results. They first developed a mental rotation model by instantiating the frontoparietal model described earlier in this domain. The states, operators, goals, and operator selection heuristics were instantiated according to the mental rotation theory of Just and Carpenter (1985). In particular, RH-Spatial proposes incremental rotations (i.e., operators) along the x, y, and z axes at each step along the rotation path. LH-Spatial generates intermediate figure representations (i.e., states) for each step along the rotation path. The greater the angular disparity, the more incremental rotations RH-Spatial must propose and the more intermediate figure representations LH-Spatial must construct and maintain. The model provides a good account of the behavioral data and the brain imaging data on mental rotation, exhibiting longer rotation times and greater Spatial center *CU*s with greater angular disparity (Carpenter, Just, Keller, Eddy, & Thulborn, 1999; Shepard & Metzler, 1971). (p. 63)

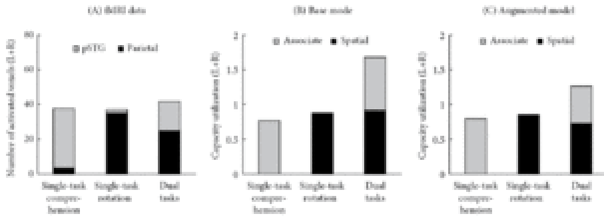

Fig. 3.6 Dynamic response to increasing communication demands. (A) Activation in Wernicke's area (left posterior superior temporal gyrus [L. pSTG]) and its right-hemisphere homolog, and in bilateral parietal cortex, when auditory sentence comprehension and mental rotation are performed individually, and underadditive activation when they are performed simultaneously (Just et al., 2001). Capacity utilizations of the corresponding model centers in (B) the base model and (C) the augmented model. Only the augmented model, which limits the communication bandwidth between centers by enforcing a cortex-wide resource constraint, displays the observed underadditivity. The correlation between human and augmented model performance across the six points of comparison—two brain areas by three conditions—is 0.97 ($p < .01$).

They next constructed a family of dual-task models. The *base* model simply conjoins the sentence comprehension and mental rotation models. There were multiple *augmented* models, each enforcing different intercenter constraints. We consider only the simplest such model here, which enforced a single bandwidth limitation on all cortical communication, and refer readers to Just and Varma (2007) for information on the other variants. The question was whether the bandwidth limitation of the augmented model was necessary and sufficient for explaining underadditivity during dual tasking. The summed *CU*s of the left and right Associate centers and the left and right Spatial centers during single-tasking and dual-tasking are shown in Figure 3.6B for the base model and Figure 3.6C for the augmented model. The augmented model displays the underadditivity observed in the human data, demonstrating the sufficiency of intercenter resource constraints, and the base model does not, suggesting their necessity.

Conclusion

The CAPS family of architectures has made a number of important contributions to cognitive science. CAPS demonstrated the importance of informational constraints on complex cognition. 3CAPS added resource constraints to explain limitations on and individual differences in cognitive information processing. The newest member, 4CAPS is revealing how cortical resource constraints shape the dynamics of brain function. These architecture have supported models of language understanding, problem solving, intelligence, spatial reasoning, human–computer interaction, dual-tasking, and other domains. These models explain behavioral and brain imaging data collected from typical adults and neurological patients and have driven empirical and theoretical research within cognitive science and cognitive neuroscience.

Future Directions

The CAPS family has goals that differ from those of other architectural families, and it must be evaluated on its own terms. For example, the CAPS family has never specified the computational mechanisms of declarative long-term memory (LTM), unlike the ACT family (Anderson, 2007). And although several 3CAPS models of human–computer interaction have been developed (Byrne & Bovair, 1997; Huguenard et al., 1997), the CAPS family has never offered a comprehensive account of perceptual and motor processing, unlike EPIC (Meyer & Kieras, 1997). These omissions appear to be limitations, and the development of declarative LTM and perceptual and motor centers could be directions for future research. However, our priority in the future development of 4CAPS will be driven more by the goal of addressing new findings on the neural bases of complex cognition.

Connectivity

No cognitive task, not even the simplest perceptual judgment, is performed by one and only one brain area. Every task recruits a network of brain areas that collaborate, exchanging and co-articulating representations. Communication is key, and just as resource constraints shape computation (p. 64) within brain areas, they also shape communication between brain areas. 4CAPS currently implements bandwidth limitations on intercenter communication as resource constraints on the joint computation of multiple centers. This is a useful start, enabling the modeling of underadditivity during dual-tasking, as we saw earlier. But it is only a start.

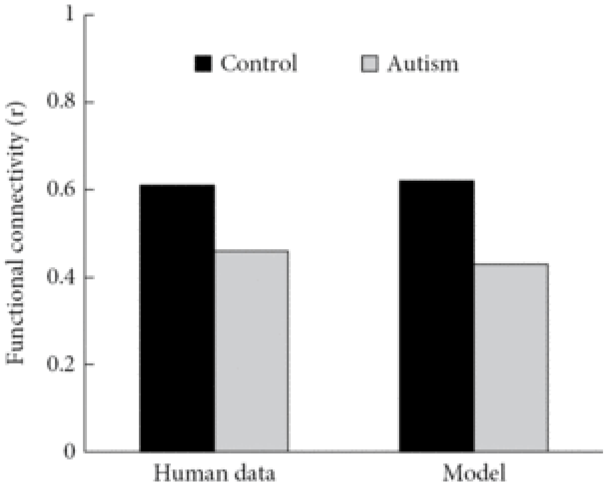

Fig. 3.7 Connectivity and cortical communication. Functional connectivity between bilateral dorsolateral prefrontal cortex (DLPFC) and bilateral parietal cortex during Tower of London (TOL) problem solving is higher in typical controls than in people with autism (Just et al., 2007). The 4CAPS model also displays reduced functional connectivity when the communication bandwidth between frontal (Executive) centers and posterior (Spatial) centers is reduced (and the autonomy of posterior centers increased).

An emerging goal of cognitive neuroscience is to map the communication network of the brain—the *human connectome* (Sporns, Tononi, & Kötter, 2005). Early research identified the *functional* connectivity of the brain by looking for correlated activation in different areas, which was interpreted as evidence of collaborative processing (e.g., Diwadkar, Carpenter, & Just, 2000). More recent research is mapping *structural* connectivity through diffusion tensor imaging (DTI) of the white-matter tracts by which brain areas communicate. An important goal for future research is to replace the current indirect implementation of intercenter communication with one more closely aligned with the DTI data.

Just, Keller, Malave, Kana, and Varma (2012) took a first step toward this goal in a model of TOL problem solving and autism. The *underconnectivity* theory of autism proposes that cognitive impairments in this population are the product of reduced structural connectivity (i.e., less coherent white-matter tracts) between frontal and posterior brain areas (Just, Cherkassky, Keller, & Minshew, 2004). Consistent with this theory, when solving TOL problems, the functional connectivity between DLPFC and parietal cortex is lower in people with autism than in neurotypical controls (see Figure 3.7). Just et al. (2012) implemented the underconnectivity theory in a 4CAPS model of TOL problem solving in autism. They added to the existing 4CAPS TOL model a bandwidth limitation on communication between frontal and posterior centers. Specifically, an intercenter constraint was defined on the joint resource consumption of the Executive and Spatial centers.[5] As we saw earlier, the four centers of the model collaborate extensively during problem solving through sharing of state, operator, goal, and operator selection representations. Enforcing the intercenter resource constraint disrupted collaborative processing. The result was reduced functional connectivity: lower average correlations between the predicted fMRI time series of each pair of Executive and Spatial centers. This closely matched the fMRI data (see Figure 3.7).

Although the Just et al. (2012) results are promising, a more general model of constraints on communication—of bandwidth limitations on point-to-point connections between centers corresponding to the white-matter tracts that connect brain areas—is required. It is possible that bandwidth limitations can be implemented within the current LP framework. The *transshipment problem* is an LP formalization of the problem of transporting goods over routes of varying carrying capacities in an optimal manner. A unified account of resource constraints on cortical information processing might be possible by formulating resource constraints on intracenter computation as a conventional LP problem, as 4CAPS currently does, and resource constraints on intercenter communication as a transshipment problem.

Optimality

4CAPS formalizes the allocation of cortical resources as an LP problem and computes an optimal solution using the simplex algorithm. This is currently done using centralized data structures and algorithms that have an unrealistically global view of cortical resources. One area for future research, then, is the development of new resource allocation algorithms that are more consistent with the distributed nature of cortical computation. These algorithms would specify how centers make local allocation decisions about when to accept extra processing from other centers and when to shift extra processing to other centers. These decisions (p. 65) would be more myopic, based exclusively on locally available information, and would sometimes be suboptimal at the global level. These algorithms might be informed by research on using distributed processing to solve large LP problems (Alon & Megiddo, 1994; Lustig & Rothberg, 1996; Maros & Mitra, 2000).

Envoi

Ranging beyond connectivity and optimality, a number of topics at the forefront of contemporary cognitive science and cognitive neuroscience represent exciting directions for future 4CAPS modeling efforts.

There are currently no 4CAPS models of learning. However, the operating principles of the neuroarchitecture appear to offer leverage on several important questions in neural plasticity:

- *Training effects*. A frequent outcome of training studies is reduced activation within relevant brain areas and increased functional connectivity between them (e.g., Büchel, Coull, & Friston, 1999). Can these training effects be modeled by adding learning algorithms to 4CAPS to tune the functional specializations of centers with experience?
- *Remediation effects*. Studies of the remediation of dyslexia and dyscalculia have revealed two neural correlates of behavioral improvements: increased activation in for-

merly underactivated areas (e.g., Eden et al., 2004) and increased integrity of white-matter tracts (Keller & Just, 2009). Can these remediation effects be captured by the same learning algorithms?

More speculatively, it is possible that the 4CAPS formalization of cortical computation and communication can be applied to understand the organization of cortex:

- *Cortical parcellation*. The centers of 4CAPS models correspond roughly to Brodmann areas—major sulci and gyri—that were historically defined based on architectonic considerations. Can this organization also be mathematically derived, as optimally balancing between the extremes of modularity/localism and equipotentiality/distributedness?
- *Cortical connectivity*. Cortical areas are connected by white-matter tracts into cortical networks. Are the topologies of these networks mathematically optimal, as when computer engineers lay out integrated circuits to minimize wire length and thus maximize throughput (Cherniak, Mokhtarzada, Rodriguez-Esteban, & Changizi, 2004)?
- *The resting state network*. The resting state network consists of those brain areas that have high activation levels during rest and low activation levels during task performance (e.g., Raichle, 2006). Does the resting state network fall out of 4CAPS models for free, as a default attractor state?

These are interesting question for future research. They demonstrate the continuing, unreasonable effectiveness of the CAPS family of cognitive architectures for anticipating and addressing central questions in cognitive science and cognitive neuroscience.

Acknowledgments

I thank Marcel Just and Susan Chipman for comments on a prior version of this chapter. I also thank Marcel Just for being the driving force behind the development of the CAPS family of architectures.

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 dimensions almost surely in constant time. *Journal of the ACM,* 41, 422–434.

Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford: Oxford University Press.

(p. 66) Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, 30, 221–256.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation, Vol. 8* (pp. 47–89). New York: Academic Press.

Bell, C. G., & Newell, A. (1971). *Computer structures: Readings and examples*. New York: McGraw-Hill.

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.

Brownston, L., Farrell, R., Kant, E., & Martin, N. (1985). *Programming expert systems in OPS5*. Reading, MA: Addison-Wesley.

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. (1998). Taking a computational approach to aging: The SPAN theory of working memory. *Psychology and Aging*, 13, 309–322.

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

Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition,* 21, 117–175.

Caplan, D., & Waters, G. S. (2002). Working memory and connectionist models of parsing: A response to MacDonald and Christiansen. *Psychological Review*, 109, 66–74.

Carpenter, P. A., & Just, M. A. (1989). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), *Information processing: The impact of Herbert A. Simon* (pp. 31–68). Hillsdale, NJ: Erlbaum.

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.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1–22.

Cherniak, C., Mokhtarzada, Z., Rodriguez-Esteban, R., & Changizi, K. (2004). Global optimization of cerebral cortex layout. *Proceedings of the National Academy of Sciences USA*, 101, 1081–1086.

Cooper, T., & Wogrin, N. (1988). *Rule-based programming in OPS5*. San Mateo, CA: Morgan-Kaufmann.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671–684.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.

Dantzig, G. B., & Thapa, M. N. (1997). *Linear programming 1: Introduction*. New York: Springer-Verlag.

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.

Eden, G. F., Jones, K. M., Cappell, K., Gareau, L., Wood, F. B., Zeffiro, T. A., et al. (2004). Neural changes following remediation in adult developmental dyslexia. *Neuron*, 44, 411–422.

Erman, L. D., Hayes-Roth, F., Lesser, V. R., & Reddy, D. R. (1980). The Hearsay-II speech understanding system: Integrating knowledge to resolve uncertainty. *ACM Computing Surveys*, 12, 213–253.

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 and Language*, 85, 385–395.

Forgy, C. L. (1979). *OPS4 user’s manual*. Paper 2136, Computer Science Department, Carnegie Mellon University.

Forgy, C. L. (1982). RETE: A fast algorithm for the many pattern/many objects pattern matching problem. *Artificial Intelligence*, 19, 17–37.

Geschwind, N. (1970). The organization of language and the brain. *Science*, 170, 940–944.

Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in “planning” functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia*, 33, 623–642.

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. Weaver, S. Mannes, & C. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 337–358). Hillsdale, NJ: Erlbaum.

Goldman, S. R., Varma, S., & Côté, N. (1996). Extending capacity-constrained construction integration: Toward "smarter" and flexible models of text comprehension. In B. K. Britton & A. C. Graesser (Eds.), *Models of text comprehension* (pp. 73–113). Hillsdale, NJ: Erlbaum.

Haarmann, H. J. (1995). *Introduction to programming in the 3CAPS modeling system*. Unpublished manual, Psychology Department, Carnegie Mellon University.

Haarmann, H. J., Just, M. A., & Carpenter, P. A. (1997). Aphasic sentence comprehension as a resource deficit: A computational approach. *Brain and Language,* 59, 76–120.

Hofstadter, D., & the Fluid Analogies Research Group. (1995). *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought*. New York: Basic Books.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

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.

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 and 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. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, 47, 310–339.

Just, M. A., Carpenter, P. A., & Hemphill, D. D. (1996). Constraints on processing capacity: Architectural or implementational? In (p. 67) D. Steier, & T. 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., & 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., & 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., Keller, T. A., Malave, V., Kana, R., & Varma, S. (2012). Autism as a neural systems disorder: A theory of fronto-posterior underconnectivity. *Neuroscience and Biobehavioral Reviews*, 36, 1292–1313.

Just, M. A., & Varma, S. (2002). A hybrid architecture for working memory. *Psychological Review*, 109, 54–64.

Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of cognition. *Cognitive, Affective, and Behavioral Neuroscience*, 7, 153–191.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Keller, T. A., & Just, M. A. (2009). Altering cortical connectivity: Remediation-induced changes in the white matter of poor readers. *Neuron*, 64, 624–631.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580–602.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163–182.

Lashley, K. S. (1950). In search of the engram. *Society of Experimental Biology Symposium*, 4, 454–482.

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

Luria, A. R. (1966). *Higher cortical functions in man*. London: Tavistock.

Lustig, I. J., & Rothberg, E. (1996). Gigaflops in linear programming. *Operations Research Letters,* 18, 157–165.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review*, 109, 35–54.

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.

Mason, R. A., Just, M. A., Keller, T. A., & Carpenter, P. A. (2003). Ambiguity in the brain: How syntactically ambiguous sentences are processed. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 29, 1319–1338.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375–407.

McCorduck, P. A. (1979). *Machines who think*. San Francisco: W. H. Freeman.

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

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, 104, 3–65.

Navon, D. (1984). Resources: A theoretical soup stone?*Psychological Review*, 91, 216–234.

Newell, A. (1973*a*). You can't play 20 questions with nature and win: Projective comments on the papers in this symposium. In W. G. Chase (Ed.), *Visual information processing* (283–308). New York: Academic Press.

Newell, A. (1973*b*). Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual information processing* (463–526). New York: Academic Press.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Chess-playing programs and the problem of complexity. *IBM Journal of Research and Development*, 2, 320–325.

Newell, A., & Tonge, F. M. (1960). An introduction to information processing language V. *Communications of the Association of Computing Machinery*, 3, 205–211.

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.

Quillian, R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 18–29). Cambridge, MA: MIT Press.

Raichle, M. E. (2006). The brain's dark energy. *Science*, 314, 1249–1250.

Rosenbloom, P. S. (1979). *The XAPS reference manual*. Unpublished manual, Computer Science Department, Carnegie Mellon University.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89, 60–94.

Rumelhart, D. E., & McClelland, J. L. (1986). PDP models and general issues in cognitive science. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed computing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 110–146). Cambridge, MA: MIT Press.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103, 403–428.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In D. V. Blake & A. M. Uttley (Eds.), *Proceedings of the symposium on mechanisation of thought processes* (pp. 511–529). London: H. M. Stationary Office.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B,* 298, 199–209.

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.

Simon, H. A. (1998). Allen Newell: 1927–1992. *IEEE Annals of the History of Computing*, 20, 63–76.

Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Computational Biology*, 1, e42.

(p. 68) Thibadeau, R. (1982). *CAPS short manual: Psylisp version*. Unpublished manual. Psychology Department, Carnegie Mellon University. Not currently cited.

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. (1990). *The CAPS89 manual*. Unpublished manual, Psychology Department, Carnegie Mellon University.

Varma, S. (1992). *The CLCAPS manual*. Unpublished manual, Psychology Department, Carnegie Mellon University.

Varma, S. (2006). *A computational model of Tower of Hanoi problem solving*. Unpublished doctoral dissertation, Psychology and Human Development Department, Vanderbilt University.

Varma, S. (2011). The design and evaluation of cognitive architectures. *Cognitive Science*, 35, 1329–1351.

Varma, S., & Just, M. A. (in preparation). The neuroarchitecture of sentence comprehension.

Waltz, D. L. (1975). Understanding line drawings of scenes with shadows. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 19–92). New York: McGraw-Hill.

Waltz, D. L., & Pollack, J. (1985). Massively parallel parsing: A strongly interactive model of natural language interpretation. *Cognitive Science*, 9, 57–84.

Waters, G., & Caplan, D. (1996). The capacity theory of sentence comprehension: A critique of Just and Carpenter (1992). *Psychological Review*, 103, 761–772.

Notes:

(1.) The development of CAPS family architectures and models has been generously funded by a number of sources, including (in roughly chronological order) the National Institute of Education, the National Institute of Mental Health, the Sloan Foundation, the Office of Naval Research, the Andrew W. Mellon Foundation, the McDonnell-Pew Program in Cognitive Neuroscience, the Air Force Office of Scientific Research, and the Multidisciplinary Research Program of the University Research Initiative.

(2.) The activation-based dynamics of CAPS derive in part from XAPS (Rosenbloom, 1979), a production system interpreter written by Rosenbloom after he left Newell’s group and spent 1 year at UCSD during the period when modern connectionism was emerging (Rosenbloom, personal communication).

(3.) For further details on the formalization of resource constraints in 3CAPS, see Just and Varma (2002).

(4.) This version was internally designated as CLCAPS for a time (Varma, 1992).

(5.) As an adaptation to this bandwidth limitation, they also increased the autonomy of the spatial centers.

Sashank Varma

Sashank Varma, Educational Psychology, University of Minnesota