1 Models of Working Memory

An Introduction

PRITI SHAH AND AKIRA MIYAKE

Working memory plays an essential role in complex cognition. Everyday cognitive tasks – such as reading a newspaper article, calculating the appropriate amount to tip in a restaurant, mentally rearranging furniture in one's living room to create space for a new sofa, and comparing and contrasting various attributes of different apartments to decide which to rent – often involve multiple steps with intermediate results that need to be kept in mind temporarily to accomplish the task at hand successfully. "Working memory" is the theoretical construct that has come to be used in cognitive psychology to refer to the system or mechanism underlying the maintenance of task-relevant information during the performance of a cognitive task (Baddeley & Hitch, 1974; Daneman & Carpenter, 1980). As reflected by the fact that it has been labeled "the hub of cognition" (Haberlandt, 1997, p. 212) and proclaimed as "perhaps the most significant achievement of human mental evolution" (Goldman-Rakic, 1992, p. 111), it is a central construct in cognitive psychology and, more recently, cognitive neuroscience.

Despite the familiarity of the term, however, it is not easy to figure out what working memory really is. To begin with, the term *working memory* is used in quite different senses by different communities of researchers. In the behavioral neuroscience and animal behavior fields, for example, the term is associated with the radial arm maze paradigm. In this paradigm, a hungry animal (usually a rat) is placed in a multipronged maze and searches for food located at the end of each arm. If the animal has a good "working memory" and can remember which arms it has already visited, it should not return to those arms because the food there is already gone. Thus, in this context, working memory has a specific operational definition different from that generally used by cognitive psychologists: "the ability of an animal to keep track of its location in space by remembering where it has been" (Olton, 1977, p. 82; see Gagliardo, Mazzotto, & Divac, 1997, for a recent study of "working memory" in this sense).

The confusion remains even within the discipline of cognitive psychology. First of all, there is not always a clear-cut distinction between working mem-

We thank Andrew Conway, Peter Delaney, Arthur Graesser, Reid Hastie, Janice Keenan, Daniel Kimberg, Peter Polson, Satoru Saito, and Sashank Varma for reading an earlier version of this chapter and giving us useful comments and suggestions.

ory and the still prevalent concept of "short-term memory" or STM (Brainerd & Kingma, 1985; Engle, Tuholski, Laughlin, & Conway, in press; Klapp, Marshburn, & Lester, 1983; see also a collection of articles in the March, 1993, issue of *Memory & Cognition* on STM). Textbooks, in particular, often contradict one another and are sometimes even internally inconsistent in their discussion of the distinction between STM and working memory. Adding to the confusion is that a number of different metaphors are used to refer to working memory and to highlight different characteristics of the concept, including the "box" or "place" metaphor, the "workspace" or "blackboard" metaphor, the "mental energy" or "resources" metaphor, and the "juggling" metaphor.

To make things even worse, the working memory literature is filled with seemingly contradictory claims. For example, some articles emphasize the unitary nature of working memory (e.g., Engle, Cantor, & Carullo, 1992), whereas others focus on its non-unitary nature and argue for a more domain-specific view of working memory (e.g., Daneman & Tardif, 1987). Some articles put forth a theory in which individual differences in working memory capacity are conceptualized in terms of variation in the total amount of mental resources available (e.g., Just & Carpenter, 1992), whereas others claim that long-term knowledge and skills provide a better account of individual differences in working memory (e.g., Ericsson & Kintsch, 1995). The common practice of capitalizing on differences in viewpoints is understandable in terms of the sociology of science, but it is not always clear from these articles whether such different conceptualizations are fundamentally incompatible or merely reflect differences in emphasis.

A variety of models and theories proposed earlier reflect such diverse – and one might say, disparate – perspectives on the nature, structure, and functions of working memory (e.g., Anderson, Reder, & Lebiere, 1996; Baddeley, 1986; Barnard, 1985; Cowan, 1988; Ericsson & Kintsch, 1995; Just & Carpenter, 1992; Schneider & Detweiler, 1987). Attempts to figure out what characteristics working memory has and how it is organized by carefully reading these theoretical articles sometimes leave one even more confused than before. We ourselves experienced this frustration prior to editing this volume and would imagine that our frustration might be somewhat analogous to what Eysenck (1986) once felt about various psychometric theories of intelligence: "Discussions concerning the theory, nature, and measurement of intelligence historically have resulted more in disagreement than in agreement, more in smoke than in illumination" (p. 1). Many people might agree that this quote would continue to make sense if the phrase *working memory* were substituted for the word *intelligence*. Indeed, this suspicion has been confirmed by one embarrassing question repeatedly raised by different colleagues and students of ours, all aware of different conceptions of working memory: "What is working memory, anyway?"

We believe that the time has come to take a step toward clarifying this confusing state of affairs in the field. In this volume, we tackle this challenge by systematically comparing existing influential models and theories of working memory. As a casual skimming of the subsequent chapters (Chapters 2 to 11) makes clear, the models included in this volume represent a wide range of theoretical perspectives that, on the surface, look quite different from one another. Our primary goal is to closely examine how these different models characterize working memory and elucidate some commonalities among them – commonalities that may help us better define and understand working memory.

The specific approach we decided to adopt for this purpose is to ask each theorist to address the same set of important (and often controversial) theoretical questions that have been guiding working memory research. This "common-question" approach to theory comparison has rarely been used in cognitive psychology. To the best of our knowledge, the only book that has explicitly used this approach is a volume edited by Baumgartner and Payr (1995), entitled *Speaking Minds: Interviews with Twenty Eminent Cognitive Scientists*. In that volume, Baumgartner and Payr interviewed leading cognitive scientists and asked them the same set of theoretical questions such as, "Do you think the Turing Test is a useful test (or idea)?" We found their approach quite effective in elucidating the commonalities and differences of various researchers' opinions because the shared questions provide a useful common ground against which different theorists' ideas can be compared and contrasted. Because existing models of working memory differ radically in their scope and focus, we thought that, without such shared questions, it might be difficult to compare seemingly disparate models and identify their commonalities.

WHY THEORY COMPARISON? In our view, systematically comparing and contrasting different models of working memory in terms of the common set of designated questions has several important merits that are worth pointing out in addition to the ones mentioned above. First, systematic issue-by-issue comparisons help clarify common misconceptions or misinterpretations of different models of working memory. Theoretical articles often provide detailed specifications of some aspects of a model, but give cursory or no treatment to other aspects of the model. Although there is nothing inherently wrong with this common practice, it invites a lot of guessing on the part of readers about various issues, sometimes leading to confusion and even wrong interpretations.[1] Asking different theorists to address all major aspects of working memory may reduce the confusion and misinterpretations.

[1] For example, one common misinterpretation prevalent in the literature is that Just and Carpenter's (1992) model assumes a unitary, domain-general notion of working memory. The model described in the 1992 paper included only one "resource pool," but it only reflected the fact that the model was restricted to the domain of language comprehension. Just and Carpenter themselves had a more domain-specific view of working memory, assuming at least a distinction between language and visuospatial working memory. Another common misinterpretation concerns the processing capabilities of the visuospatial sketchpad system in Baddeley's (1986) model. Although Baddeley himself has consistently argued that it can actively manipulate mental images, some researchers have portrayed Baddeley's sketchpad system as a pure storage buffer without any processing capability.

Second, systematic issue-by-issue comparisons can also crystallize which seemingly conflicting theoretical claims are indeed mutually incompatible (rather than merely complementary) and, hence, must be resolved in future research. The identification of mutually incompatible claims not only sharpens the focus of research, but also provides an important basis for rigorous tests of competing ideas by way of "competitive argumentation" (VanLehn, Brown, & Greeno, 1984) – pitting competing models (or alternative versions of a single model) *directly* against one another to analyze and clarify theoretical issues often left implicit. VanLehn et al. clearly articulate the importance of this approach:

> To show that some constraint is crucial is to show that it is *necessary* in order for the theory to meet some criteria of adequacy. To show that it is *sufficient* is not enough. . . . [W]hen there are two theories, one claiming that principle X is sufficient and another claiming that a different, incompatible principle Y, is sufficient, sufficiency itself is no longer persuasive. One must somehow show that X is better than Y. Indeed, this sort of competitive argumentation is the only realistic alternative to necessity arguments (VanLehn et al., 1984, p. 240).

Last but not least, unification is always an aim of science, as Newell (1990) pointed out in his book *Unified Theories of Cognition*. Identifying which seemingly conflicting theoretical claims are in fact complementary could help unify or synthesize different models, possibly leading to a unified theory of working memory. As the broad range of the eight designated questions we discuss in the next section indicates (see Table 1.1), the key theoretical issues in current working memory research interface many (if not all) aspects of cognitive psychology. To borrow Haberlandt's (1997) expression, the study of working memory is essentially "a microcosm of the field of cognition" (p. 213). Thus, systematic comparisons of different models of working memory may even contribute strongly to the development of unified theories of cognition.

The Eight Designated Theoretical Questions for This Volume

The eight designated theoretical questions that provide a basis for the theoretical comparisons offered in this volume are listed in Table 1.1. They touch on all major theoretical issues of central importance to working memory research, including those that are currently highly controversial (particularly, Questions 3 and 4). We motivate the eight designated questions below, one by one, by discussing their importance in working memory research and providing a brief historical review. In addition, we offer some guidelines that might help readers in their endeavor of comparing and evaluating the models presented in this volume.

Table 1.1. The Eight Designated Questions for This Volume

(1) Basic Mechanisms and Representations in Working Memory
How is information encoded into and maintained in working memory? What is the retrieval mechanism? Also, how is information represented in working memory? Is the representation format for different types of information (e.g., verbal or visuospatial information) the same or different?

(2) The Control and Regulation of Working Memory
How is the information in working memory controlled and regulated? What determines which information is stored and which is ignored? Is the control and regulation of working memory handled by a central control structure (e.g., the central executive)? If so, what are the functions of the control structure? If your model does not postulate a central control structure, how do the control and regulation of information emerge?

(3) The Unitary Versus Non-Unitary Nature of Working Memory
Is working memory a unitary construct, or does it consist of multiple separable subsystems? If the latter is the case, then what are the subsystems of working memory and how do they interact with one another? What evidence or theoretical considerations justify your view?

(4) The Nature of Working Memory Limitations
What are the mechanisms that constrain the capacity of working memory (e.g., a limited supply of activation, processing speed, decay, inhibition, interference, skills)? If your model postulates multiple subsystems within working memory, does the same set of constraining mechanisms apply to each subsystem? What evidence or theoretical considerations have motivated the postulation of those capacity-constraining mechanisms?

(5) The Role of Working Memory in Complex Cognitive Activities
How is working memory implicated in the performance of complex cognitive tasks, such as language comprehension, spatial thinking, mental arithmetic, and reasoning and problem solving? What complex cognitive phenomena have you examined from the perspective of your model, and, according to your analysis, what role(s) does working memory play in these tasks? How does your model account for the performance limitations associated with these tasks?

(6) The Relationship of Working Memory to Long-Term Memory and Knowledge
What is the relationship between working memory and declarative long-term memory? Are they structurally separate entities? Or is working memory simply an activated portion of long-term memory? How do they interact with each other? How does working memory also relate to procedural skills? How might working memory limitations or functions be influenced by learning and practice?

continued

Table 1.1, continued

(7) The Relationship of Working Memory to Attention and Consciousness
What is the relationship between working memory and attention? Do these terms refer to the same construct? Or are they somehow separate from each other (either partially or completely)? If so, what differentiates them, and how do they interact with each other? Also, how does working memory relate to consciousness or awareness?

(8) The Biological Implementation of Working Memory
How does your model relate to various neuroscience findings on working memory (e.g., studies of brain-damaged patients, neuroimaging data, electrophysiological measures, animal studies)? How might your view of working memory be implemented in the brain?

Question 1: Basic Mechanisms and Representations in Working Memory

How is information encoded into and maintained in working memory? What is the retrieval mechanism? Also, how is information represented in working memory? Is the representation format for different types of information (e.g., verbal or visuospatial information) the same or different?

The traditional view of human memory (e.g., Atkinson & Shiffrin, 1968; Waugh & Norman, 1965) offers an elegant account of the basic mechanisms (encoding, maintenance, and retrieval) and representations in working memory or, rather, STM. According to this view, there are a number of structurally separate components or stores through which information is transferred. A subset of the information in the sensory registers is chosen for later processing via *selective attention* and is transferred into a short-term store (STS) (encoding). The information in the STS is considered fragile and decays quickly, so *rehearsal* is necessary to keep it within the STS (maintenance) and to transfer it to a more durable long-term store (LTS). The information in the STS is assumed to be accessible relatively quickly and effortlessly (retrieval), but there may be a slight slowdown of retrieval speed as a function of the number of items within the STS (Sternberg, 1966). Once lost from the STS, information cannot be retrieved unless it is encoded in the LTS. Retrieval from the LTS, however, is generally considered a slower and more effortful process than that from the STS.

As for the representation issue, the traditional view emphasizes speech-based codes (i.e., acoustic, phonological, or verbal) as the predominant memory code in STM, as reflected in the fact that most of the STM experiments in the 1960s and 1970s were done using verbal materials, despite the fact that Atkinson and Shiffrin (1968) themselves explicitly acknowledged the possibility of other STM codes (e.g., visual, spatial). The emphasis on speech-based codes in STM is contrasted with meaning-based (semantic) codes considered dominant in LTM.

This traditional view is simple and intuitively makes sense, but the story is too simplistic. It could be argued that the overall framework of the "modal" model is defensible (Healy & McNamara, 1996; Pashler & Carrier, 1996), but the basic mechanisms and representations of the model need to be modified, qualified, or elaborated further, given the recent empirical and theoretical advances in the field, particularly those associated with working memory research. The first question, thus, asked each contributor to outline his or her current view of the encoding, maintenance, and retrieval mechanisms as well as the nature of the representational codes in working memory.

Although all 10 theory chapters in this volume (Chapters 2 to 11) address the basic mechanism and representation issue, the types of answers they provide and the manners in which they answer this question vary considerably. Indeed, of all the eight questions, answers to the first question seem to be the most difficult to discern, perhaps because it is such a basic issue that, with the exception of Cowan (Chapter 3), the question is answered in an implicit, highly distributed fashion. Even though some chapters have a section dedicated to it, the arguments relevant to the basic mechanism and representation issue tend to be made in many different places throughout the chapter. Moreover, the chapters describing computational models tend to answer this question by outlining the overall architecture and the basic assumptions of their respective models. Although most of their descriptions have important implications for the issue of basic mechanisms and representations, those descriptions are not always directly cast in terms of the concepts we used in formulating this first question (i.e., encoding, maintenance, retrieval, and representational format). Thus, readers interested in comparing and contrasting the answers to Question 1 should keep these provisos in mind when they go through the chapters in this volume.

Question 2: The Control and Regulation of Working Memory

How is the information in working memory controlled and regulated? What determines which information is stored and which is ignored? Is the control and regulation of working memory handled by a central control structure (e.g., the central executive)? If so, what are the functions of the control structure? If your model does not postulate a central control structure, how do the control and regulation of information emerge?

From the beginning of modern working memory (or STM) research, the issue of control and regulation has been considered of central importance. Indeed, the notion of "control processes" was already present in Atkinson and Shiffrin's (1968) modal model of human memory. The control processes in that model, however, were limited to those involved in pure memorization, such as rehearsal, coding, and search strategies. In contrast to the traditional, storage-oriented notion of STM, working memory is considered a more processing-oriented construct and is sometimes conceptualized as the "workspace" or "blackboard" of the mind in which the active processing and temporary storage of task-relevant information dynamically take place. Such a view of working memory necessitates a more sophisticated account of control mechanisms that go beyond simple memorization strategies. In addition, there is an increasingly popular view of working memory as consisting of multiple subsystems. This view requires a satisfactory explanation of how these different subsystems are regulated so that working memory as a whole functions smoothly. The second designated question, therefore, asked the contributors to specify the mechanisms of control and regulation.

The classic answer to this "control and regulation" question is to postulate a central control structure like the central executive, as Baddeley and Hitch (1974) did in their influential multicomponent model. As often pointed out by critics, however, this approach has the danger of implicating a mysterious little "homunculus" inside working memory. In addition, as Baddeley (1986) himself admitted, the central executive may have become almost synonymous with a theoretical "ragbag" for all functions not attributable to the peripheral slave systems (i.e., the phonological loop and the visuospatial sketchpad).

Donald (1991) vividly described these problems associated with the notion of the central executive: "The 'central executive' is a hypothetical entity that sits atop the mountain of working memory and attention like some gigantic Buddha, an inscrutable, immaterial, omnipresent homunculus, at whose busy desk the buck stops every time memory and attention theorists run out of alternatives" (p. 327). The challenge, therefore, is to more precisely specify the mechanisms underlying the control and regulation of information in working memory without postulating an explicit homunculus-like entity.

The chapters in this volume provide interesting answers to this formidable challenge. As we discuss in more detail in the concluding chapter (Miyake & Shah, Chapter 13), the general approaches to the control and regulation issue represented in this volume range from specifying the subcomponents or subfunctions of a "central executive," through relying on regulatory mechanisms inherent in the underlying computational architecture, to conceptualizing control and regulation as an emergent property (i.e., a natural consequence of dynamic interactions among different subsystems). Readers might be interested in speculating on how well these models as a whole manage to address the "homunculus" or "ragbag" problem, a focus of the commentary provided in Chapter 12 by Kintsch, Healy, Hegarty, Pennington, and Salthouse.

Question 3: The Unitary Versus Non-Unitary Nature of Working Memory

Is working memory a unitary construct, or does it consist of multiple separable subsystems? If the latter is the case, then what are the subsystems of working *memory and how do they interact with one another? What evidence or theoretical considerations justify your view?*

The issue of whether working memory is unitary or non-unitary has been a source of controversy in the working memory literature. Some researchers have emphasized the unitary nature of working memory (e.g., Anderson et al., 1996; Engle et al., 1992; Kyllonen & Christal, 1990), whereas others have emphasized its non-unitary nature (e.g., Daneman & Tardif, 1987; Martin, 1993; Monsell, 1984; Shah & Miyake, 1996). Within the non-unitary camp, different researchers fractionate working memory in different ways, and there has been little consensus as to the number of subsystems and the nature of each subsystem. Some researchers, for example, are relatively conservative, proposing only a few domain-specific subsystems such as those for verbal and visuospatial storage or processing (e.g., Baddeley, 1986), whereas others postulate other subsystems or types of codes or representations, such as auditory, motor, lexical, semantic, syntactic, and so on (e.g., Barnard, 1985; Martin, 1993; Schneider & Detweiler, 1987). Some accounts go even further, postulating separable subsystems at a much finer level of analysis. One study of aphasic language comprehension, for example, argues for the independence of processing resources for computing a verb's thematic representations and those for computing the syntactic trace-antecedent relations (Shapiro, Gordon, Hack, & Killackey, 1993). The third designated question, thus, asked the contributors to discuss their current thoughts on the unitary or non-unitary nature of working memory.

This controversy has an interesting historical parallel in the domains of intelligence and attention.[2] In the case of intelligence, the Spearman–Thurston controversy is well known. Spearman (1904) argued that a single entity called *general intelligence* or *g* (conceptualized by Spearman as neurologically based "power" or "energy") underlies intellectual performances of various types, whereas Thurston (1938) argued that seven independent primary abilities can explain intellectual functioning well without postulating a general factor. Guilford (1967) went even further than Thurston, postulating 120 distinct ability factors in his "Structure of Intellect" model. Indeed, Eysenck's (1986) remark quoted earlier illustrates how radically different these psychometric theories looked from one another. Analogously, the "unitary versus non-unitary" debate also surrounded the resource (or capacity) theories of attention, some theorists

[2] The similarity of the unitary vs. non-unitary debate in intelligence and attention research is perhaps no coincidence. The factor analysis technique, used to develop psychometric theories of intelligence, and the analysis of performance-operating characteristics (POC), used to specify resource theories of attention, share some underlying commonalities (Heuer, 1985).

proposing a unitary view (e.g., Kahneman, 1973) and others a non-unitary view (e.g., Navon & Gopher, 1979; Wickens, 1984).[3]

At first glance, the answers to the question of the unitary versus non-unitary nature of working memory provided in this volume appear rather disparate. Some models strongly emphasize the unitary characteristics of working memory (e.g., Engle, Kane, & Tuholski, Chapter 4; Lovett, Reder, & Lebiere, Chapter 5), whereas others argue for a non-unitary position (e.g., Baddeley & Logie, Chapter 2; Barnard, Chapter 9; Schneider, Chapter 10). However, we would like to invite interested readers to evaluate if those seemingly different answers are fundamentally incompatible. The answer we offer in the final two chapters of this volume (Kintsch et al., Chapter 12; Miyake & Shah, Chapter 13) is "no." Whereas there are clearly some unresolved issues (see Chapter 13), we argue that a global consensus seems to be emerging and that some sort of synthesis may even be near.

Question 4: The Nature of Working Memory Limitations

What are the mechanisms that constrain the capacity of working memory (e.g., a limited supply of activation, processing speed, decay, inhibition, interference, skills)? If your model postulates multiple subsystems within working memory, does the same set of constraining mechanisms apply to each subsystem? What evidence or theoretical considerations have motivated the postulation of those capacity-constraining mechanisms?

The fourth designated question concerns the hallmark characteristic of working memory, identified and studied for over a century – the severe limitations in its capacity (e.g., Jacobs, 1887; James, 1890). In his classic book, William James (1890) stated that, unlike the virtually unlimited amount of knowledge that can be stored in a person's "secondary memory," only a small amount of information can be kept conscious at any one time in one's "primary memory." Moreover, the early scientific work reviewed by James suggests that there was much interest in the 1800s in just how much information can be temporarily maintained and for how long.

Despite the wide consensus on the existence of capacity limits in working memory, there has been little consensus, since the beginning, on the underlying mechanisms responsible for these limitations. For example, James (1890)

[3] Of course, this unitary vs. non-unitary debate still continues in both intelligence and attention research, but in a new direction. In intelligence research, the notion of *g* is still popular (Dennis & Tapsfield, 1996; Jensen, 1998), but modern versions of the non-unitary perspective that explicitly deny the necessity of postulating *g* also abound (e.g., Gardner, 1983/1993; see also Ceci, 1990/1996, for a detailed critique of *g*). In attention research, although the seemingly unitary characteristic of attention is noted by many researchers (e.g., Engle et al., Chapter 4; Lovett et al., Chapter 5), different ways of fractionating attention into different components or aspects have also been strongly advocated recently (e.g., Allport, 1993; Pashler, 1997; Posner & Raichle, 1994).