

## Reassessing Working Memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996)

Maryellen C. MacDonald  
University of Southern California

Morten H. Christiansen  
Southern Illinois University

M. A. Just and P. A. Carpenter's (1992) capacity theory of comprehension posits a linguistic working memory functionally separated from the representation of linguistic knowledge. G. S. Waters and D. Caplan's (1996) critique of this approach retained the notion of a separate working memory. In this article, the authors present an alternative account motivated by a connectionist approach to language comprehension. In their view, *processing capacity* emerges from network architecture and experience and is not a primitive that can vary independently. Individual differences in comprehension do not stem from variations in a separate working memory capacity; instead they emerge from an interaction of biological factors and language experience. This alternative is argued to provide a superior account of comprehension results previously attributed to a separate working memory capacity.

The concept of a working memory resource or capacity for temporary storage and manipulation of information has played an important role in many theories of cognition, particularly theories of language processing (e.g., Baddeley, 1986; Engle, Cantor, & Carullo, 1992; Just & Carpenter, 1992). The particular approach advocated by Just and Carpenter (1992) is one in which linguistic working memory capacity directly constrains the operation of language comprehension processes, and that variation in the capacity of linguistic working memory within the normal population is a primary source of individual differences in language comprehension. Just and Carpenter further suggest that reductions in working memory capacity in aging can explain reduced language comprehension and production abilities among normal elderly adults and that aphasic patients' language comprehension deficits following brain injury may be explained by a deficit in working memory capacity rather than by a loss of linguistic knowledge (Miyake, Carpenter, & Just, 1994).

Just and Carpenter's work has been extremely successful in emphasizing the importance of individual differences in language research, but their approach is not without controversy. Criticisms of their claims have taken several forms. Waters and Caplan (1996) suggested that there are at least two different working

memory capacities that subservise language use, and they have sharply criticized the data that Just and Carpenter interpreted in support of a single working memory capacity (see also Caplan & Waters, 1999b; cf. Just, Carpenter, & Keller, 1996). Similarly, several aphasia researchers have suggested that a reduction of working memory capacity is not an adequate description of these patients' deficits (Caplan & Waters, 1995; R. C. Martin, 1995; but cf. Miyake et al., 1995). Cognitive aging researchers have argued that reduced working memory capacity does not capture the declines associated with normal aging (Stoltzfus, Hasher, & Zacks, 1996) and that other factors, such as increased interference in memory (Hasher & Zacks, 1988) or slowed perceptual and cognitive processes (Salthouse, 1996) are the real sources of cognitive decline. Other researchers, such as Ericsson and Kintsch (1995), have rejected the emphasis on working memory size in accounting for individual differences, as this approach neglects the role of practice and skill. Developmental psychologists have made similar points about child development, that improved performance in cognitive tasks may result from more opportunities to practice and develop expertise rather than from growth of working memory capacity (Munakata, McClelland, Johnson, & Siegler, 1997; Roth, 1984). Finally, Navon (1984) argued that capacity- or resource-based approaches are inherently flawed in that they are typically vague, they have often been evaluated only in light of positive evidence, and this positive evidence is generally consistent with nonresource explanations. On Navon's view, the notion of processing resources is a theoretical "soup stone" that contributes no explanatory power to cognitive theorizing.

In this article, we join a number of these critics in doubting the usefulness of the traditional concept of working memory. We evaluate both Just and Carpenter's original claims for a single linguistic working memory and claims in the Waters and Caplan (1996) critique of Just and Carpenter and their reply (Just et al., 1996). A significant portion of the exchange in the latter two articles addresses the strength of the data that Just and Carpenter presented, but that is not our focus here. Like Waters and Caplan,

---

Maryellen C. MacDonald, Department of Psychology, University of Southern California; Morten H. Christiansen, Department of Psychology, Southern Illinois University.

Morten A. Christiansen is now at Department of Psychology, Cornell University.

This research was supported by National Science Foundation Grant SBR-95-11270 and National Institute on Aging Grant RO1 AG-11773 to the University of Southern California. We thank Mark Seidenberg, David Plaut, and Walter Kintsch for helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Maryellen C. MacDonald, who is now at Department of Psychology, University of Wisconsin—Madison, 1202 West Johnson Street, Madison, Wisconsin 53706-1696. E-mail: mcmacdonald@facstaff.wisc.edu

we agree that Just and Carpenter sometimes overinterpreted marginal data, but like Just and Carpenter and Just et al. (1996), we agree that the bulk of the evidence suggests that there are real though sometimes subtle individual differences in linguistic processing abilities within the normal population. We also do not doubt that there are individual differences in the normal population on working memory tasks, and we even do not doubt that language processing performance and performance on linguistic working memory tasks often correlate with one another. Our differences with Just and Carpenter and Waters and Caplan instead rest in our interpretation of these data: We claim that the distinction commonly drawn between *language-processing* tasks and *linguistic working memory* tasks is an artificial one, and that all of these tasks are simply different measures of language processing skill. We further argue that differences in skill within the normal population can be attributed to (a) variations in exposure to language—some individuals engage in language comprehension, particularly reading, more than others—and (b) biological differences that affect processing accuracy, such as differences in the precision of phonological representations that are developed for spoken and written input.

Our claims are grounded in a connectionist approach to language comprehension that treats traditional distinctions between linguistic knowledge, linguistic processing, and processing capacity in a very different way. Just and Carpenter's and Waters and Caplan's claims, however, are based on more symbolic processing architectures. In all cases, the choice of processing architecture has a direct effect on claims for the role of working memory in language comprehension. We first illustrate this relationship by sketching the Just and Carpenter and Waters and Caplan positions on language and working memory. We then detail our own approach and discuss how it permits a reinterpretation of a broad range of individual differences in language comprehension that have previously been thought to emerge from variations in the size of a linguistic working memory capacity. We argue that our framework offers a more coherent account of these data than is available within more traditional approaches.

#### The Just and Carpenter (1992) and Waters and Caplan (1996) Approaches to Working Memory and Language Comprehension

Both Just and Carpenter and Waters and Caplan adopt a view of working memory as both a storage space and a processing arena for executing computations (Baddeley & Hitch, 1974). Just and Carpenter suggested that their own view of working memory is roughly analogous to the linguistic portion of Baddeley's (1986) central executive, and they did not posit any modality-specific rehearsal buffers such as Baddeley's phonological loop. Whereas Just and Carpenter embraced a single working memory for language comprehension, Waters and Caplan argued for two separate working memories: one dedicated to "obligatory," unconscious psycholinguistic processes involved in on-line comprehension and another dedicated to controlled, verbally mediated tasks. These different positions about the number of working memory components appear to be deeply rooted within each group's views of language comprehension.

#### *Just and Carpenter's Approach*

Just and Carpenter favor an approach to language comprehension in which information from different levels of representation is allowed to interact during processing (Just & Carpenter, 1987). They link their approach to a computational model called CC READER, developed within the 3CAPS production system framework (Haarmann, Just, & Carpenter, 1997; Just & Thibadeau, 1984; Thibadeau, Just, & Carpenter, 1982).<sup>1</sup> In this model, declarative knowledge, such as the lexicon, and procedural knowledge in the form of a set of condition-action rules (*productions*) are stored in long-term memory. A separate working memory space is used to process and store current input and partial products of ongoing computations. In line with other early production system approaches to cognitive modeling (e.g., Anderson, Kline, & Lewis, 1977; Holland, 1975; Newell, 1980), the Just and Carpenter model permits several productions to fire in parallel within a given processing cycle and allows for gradual instantiations of information in working memory. Information in working memory is assigned a numerical activation (or confidence) level, and information without sufficient activation decays from working memory. During the processing of a sentence, the model uses the items in working memory to build explicit syntactic and semantic representations for the purpose of interpretation. This information may be used by several different kinds of productions (lexical, syntactic, etc.), permitting interaction across levels of processing when productions from different levels are executed in parallel. Thus, the advocacy of a single working memory for verbal comprehension stems from Just and Carpenter's belief in the interactive nature of the underlying processes.

Just and Carpenter (1992) suggest that CC READER's activation and parallel computation properties are "connectionist" (see, e.g., p. 124). The use of this term is somewhat misleading, as the original exposition of the 3CAPS framework (Thibadeau et al., 1982) clearly indicated that the inspiration for incorporating these properties into the model did not stem from connectionist work but rather from other production models of cognitive processing (e.g., Anderson et al., 1977; Newell, 1980). Thus, the CC READER model is firmly rooted within the tradition of production systems in that its computations are inherently symbolic and maintain the distinction between working memory resources and knowledge of language. Indeed, when simulating individual differences in the CC READER model, Just and Carpenter chose different values of the maximum activation parameter to simulate readers with good and poor reading span performance (Daneman & Carpenter, 1980), while holding linguistic knowledge and processing operations constant. Carpenter, Miyake, and Just (1994) argued that this two-way division between resources, on the one hand, and linguistic knowledge and processing, on the other, represent an important theoretical claim, one that distinguished Just and Carpenter's proposal from unimplemented resource theories:

The capacity theory we have proposed provides a mechanistic account for a wide variety of phenomena, by specifying a precise computational mechanism within which resources (activation), storage, and

<sup>1</sup> Various versions of the 3CAPS-based model have been developed, with some implementational differences among them. Here we sketch the key computational elements common to all versions.

processing are all well defined. By virtue of this specification, the theory can explain a variety of phenomena without the vacuity against which Navon cautioned. (p. 1110)

The separation between working memory and knowledge can perhaps be seen most clearly in Just and Carpenter and colleagues' treatment of language impairments (Haarmann et al., 1997; Miyake et al., 1994; Miyake, Carpenter, & Just, 1995). If working memory and knowledge of language are separate, it logically follows that one of these may be impaired independent of the other in brain injury or disease:

Stated from the perspective of the CC READER model, the theory assumes that, while the lexicon and the production rules in the system are still intact, the maximum amount of activation available for the storage and processing of linguistic information is far more severely limited in the aphasic system than in the normal system. (Carpenter et al., 1994, p. 1093)

This reduced-resources account of aphasic sentence comprehension has recently been instantiated in a 3CAPS computational model (Haarmann et al., 1997).

Just and Carpenter's approach to the role of experience in individual differences is less specified. Although they acknowledged that experience might play some role in individual differences, they provided no details concerning how experience could influence processing. Without an explicit learning theory, it is not clear how one could readily incorporate a comprehensive role for experience into the 3CAPS framework. Moreover, although Just and Carpenter noted that it is theoretically possible to recast resource accounts in terms of processing efficiency or knowledge differences, they explicitly rejected this move. They suggested that the individual differences they discuss "are better explained in terms of total capacity than process efficiency" (Just & Carpenter, 1992, p. 145). Consistent with other working memory researchers (e.g., Daneman, 1988), they view the effects of linguistic experience as largely restricted to vocabulary size, with limited effects on on-line comprehension tasks: "knowledge-based explanations are less useful in accounting for the on-line processing profile of comprehension. Moreover, the potential role of vocabulary and knowledge is not so salient in sentence processing tasks, in which the structure and vocabulary are familiar and restricted" (Carpenter et al., 1994, p. 1110). Our account directly challenges this claim, in that we argue that experience has a reach well beyond vocabulary size, forming critical expertise in syntactic structures (Chang, Dell, Bock, & Griffin, 1999) and in probabilistic constraints (Pearlmutter & MacDonald, 1995) that govern on-line language comprehension.

#### *Waters and Caplan's Approach*

Waters and Caplan view language processing in a different way than Just and Carpenter. Their approach to comprehension is more closely tied to generative linguistic theory (e.g., Chomsky, 1965), particularly to claims about the modular nature of language processes and the linguistic knowledge they use (Fodor, 1983; Frazier, 1987). Apart from a module for syntactic processing, Waters and Caplan (1996) envisaged components for other functions, including "acoustic-phonetic conversion, lexical access, assignment of intonational contours, determination of sentential semantic values such as thematic roles, and determination of discourse-level se-

mantic values such as topic and coherent coreference" (p. 770). Information from these different components is brought together and processed in a common working memory to build explicit syntactic and semantic representations for the purpose of interpretation. Thus, the representation of linguistic knowledge is highly modularized and separated from processing in working memory. Waters and Caplan's approach is tightly linked to work in neuropsychology, where similar assumptions of multiple specialized language components are common in the literature and are often invoked in accounts of patients who seem to have "modular" deficits (e.g., Martin, Shelton, & Yaffee, 1994).

In contrast to Just and Carpenter, Waters and Caplan argued that the reading span task does not assess the working memory used for language comprehension. Rather, they suggested it taps separate working memory resource for conscious controlled processing. This view of what the reading span task is measuring forms the heart of the controversy between Just and Carpenter and Waters and Caplan over the individual-differences data. Waters and Caplan argued that the lack of certain statistical interactions between reading span, syntactic complexity, and other factors such as external processing load points to two separate working memories.

Although Waters and Caplan criticized the CC READER model for having too many free parameters, they provided no computational account of their own dual working memory approach. This makes their position particularly vulnerable to Navon's (1984) concerns about the vacuity of resource theories. The fact that Waters and Caplan proposed two different and underspecified processing resources only compounds this problem (Christiansen & MacDonald, 1999). Moreover, Just et al. (1996) correctly noted that delineating the two working memories according to whether processing is conscious or not is very problematic, as language processing seems to move along a continuum between conscious and unconscious processing. A related problem concerns how experience affects Waters and Caplan's conscious-unconscious dichotomy, as the authors appear to have had no clear position regarding the role of experience in comprehension.

As with Just and Carpenter, Waters and Caplan's distinction between knowledge and working memory can be seen in their approach to language deficits. For Waters and Caplan, aphasic deficits are in some cases the result of deficits in specific linguistic components (Caplan & Hildebrandt, 1988; Caplan & Waters, 1995) and in other cases due to reductions in psycholinguistic working memory (Caplan & Waters, 1999a). Thus although Just and Carpenter and Waters and Caplan strongly disagreed on how working memory systems are demarcated and, consequently, on the nature of impairments in aphasic patients (Caplan & Waters, 1995; Miyake et al., 1994, 1995), they shared a commitment to the separation between linguistic knowledge and working memory. We directly challenge this claim and argue that the two are inseparable. In our view, neither *knowledge* nor *capacity* are primitives that can vary independently in theory or computational models; rather they emerge from the interaction of network architecture and experience. Like Just and Carpenter and Waters and Caplan, our approach to working memory is firmly rooted in our approach to language processing.

#### Connectionist Approaches to Language Processing

Connectionist language comprehension looks very different from Just and Carpenter's (1992) and Waters and Caplan's (1996)

symbolic approaches. Within the connectionist framework, the processing of input is achieved not through the action of rules or productions operating on declarative knowledge in a computational workspace but rather through the passing of activation through a multilayer network. In this framework, the network's capacity to process information varies as a function of the input (e.g., whether the material is complex or simple), the properties of the network (how activation is passed through weights, etc.), and the interaction of these properties—how much the network has experienced similar input before.<sup>2</sup>

With the entire network contributing to processing, and with processing capacity tied to the efficiency of passing activation across large numbers of units, where is working memory? To the extent that it is useful to talk about working memory within these systems, it is the network itself; it is not some separate entity that can vary independently of the architecture and experience that governs the network's processing efficiency. This approach bears some similarity to one in which individual differences emerge from experience rather than from variations in a separate capacity (Ericsson & Kintsch, 1995), but the two approaches are not identical. Ericsson and Kintsch, like Just and Carpenter and Waters and Caplan, postulated a working memory for temporary storage and processing, separated from the representation of long-term knowledge. On our account, the long-term knowledge of language is not functionally separated from the locus of processing. We will show that this integration has important implications for individual differences.

The claim that capacity and knowledge are inseparable does not mean that connectionist networks cannot vary in capacity (or knowledge). A number of researchers have investigated capacity variation and the nature of individual differences within connectionist architectures. These differences are instantiated in networks in various ways, including variations in the number of hidden units that the network has available to learn and process information (Harm & Seidenberg, 1999; Patterson, Seidenberg, & McClelland, 1989), variation in the amount of training the network receives (Munakata et al., 1997), variation in the efficiency with which the network is able to pass information among units (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997), and variation in the amount of "noise" in the input signal presented to the network (St. John & Gernsbacher, 1998). Thus, connectionist architectures are fully compatible with the notion that individuals and networks may vary in processing capacity, but in all cases, these manipulations affect the behavior of the whole network, both its processing and its representation. Again, the impact of these architectural choices can be seen clearly in conceptualizations of brain damage (e.g., N. Martin & Saffran, 1997; Patterson, Graham, & Hodges, 1994). Because any architectural change affects an entire connectionist network, connectionist models do not treat aphasia or other deficits as being a deficit in either working memory capacity or linguistic knowledge—they are inseparable in these networks, and any architectural change will have effects on both the processing capacity of the network and the nature of the representations embodied in the network.

### An Alternative Approach to Individual Differences

Our connectionist-based account makes two claims about the sources of the intertwined individual differences in linguistic

knowledge and processing capacity. First, a substantial amount of individual differences in language-processing ability within the normal population is due to variation in experience with language. This claim represents a major difference between our approach and much other research in working memory, as working memory size, not experience, is typically thought to be the major determinant of individual differences in the field (but cf. Ericsson & Kintsch, 1995). Second, we argue that the biological differences that do exist are not in the capacity of a separate working memory. For example, individuals vary in the nature of the phonological representations that are developed during language processing. We argue that innate differences in phonological representations affect processes well beyond the representation of speech sounds, and that they have important consequences in development for the accrual of linguistic experiences. Extending this account to the normal adult lifespan, we also suggest that processing speed (Salt-house, 1994, 1996) is an important architectural factor in language comprehension in elderly adults. In the remainder of the article, we use these experiential and biological-architectural factors to show how individual differences in language-processing ability could emerge in a connectionist framework.

Our position completely recasts the debate between Just and Carpenter (1992) and Waters and Caplan (1996). Just and Carpenter (and Just et al., 1996) offered evidence for the role of a separate linguistic working memory in seven different domains, whereas Waters and Caplan argued that the working memory capacity assessed in this research is not central to core language comprehension processes and that the data argue instead for multiple working memory capacities. As we elaborate our own account, we review each of these seven domains and show how none of them demands any number of separate working memory capacities but is instead best accounted for by our own approach. We begin by examining the notion of *working memory task* that underlies so much of Just and Carpenter's and Waters and Caplan's accounts.

### What Do Working Memory Tasks Measure?

The connectionist framework we adopt has implications for interpretation of performance on working memory tasks. Just and Carpenter's central measure of linguistic working memory is Daneman and Carpenter's (1980) reading span task, in which participants read sentences aloud and remember sentence-final words for later recall. Both Just and Carpenter and Waters and Caplan agree that the reading span task provides a measure of some kind of working memory capacity, but Just and Carpenter see this capacity as central to language comprehension whereas Waters and Caplan do not. By contrast, we do not view the reading span task as measuring any fixed working memory resource. In our

<sup>2</sup> It is important to distinguish the connectionist framework from more general constraint-satisfaction approaches to language processing (e.g., Kintsch, 1998; MacDonald et al., 1994; Tanenhaus & Trueswell, 1995). It is possible to model constraint-satisfaction processes using a number of different architectures, including both symbolic and connectionist approaches (MacDonald et al., 1994). Thus, general notions of constraint satisfaction are not inherently incompatible with a separate working memory system, but once constraint satisfaction processes are instantiated in a connectionist architecture, the network incorporates both representation and processing, without a separate working memory.

view, the reading span task is simply a measure of language processing skills, like measures of lexical decision latency, reading times, and so forth (Ericsson & Kintsch, 1995, and R. C. Martin, 1995, make related points). We do not claim that there is a unitary construct called *working memory capacity* measured by working memory tasks any more than we claim that lexical decision tasks measure *lexical decision capacity* separate from language comprehension abilities. Reading span, lexical decision, and reading are all just language-processing tasks, with slightly different task demands, and experiment participants marshal their comprehension abilities in different ways to meet those demands. Some tasks may be more appropriate than others for investigating particular research questions, but our claim is that no task has a more privileged status in a theory than any other. Just and Carpenter and Waters and Caplan, however, view reading span and other working memory tasks as special measures of a particular computational capacity, and they debate its role in comprehension. Just and Carpenter cite correlations between reading span and a comprehension measure like reading time as evidence that the computational capacity measured by reading span is used in comprehension. Waters and Caplan question the strength of these relationships and argue instead that the capacity measured by the reading span task is used in conscious “postinterpretive” processes, not in on-line comprehension. We, however, eschew all causal attributions here. We would not want to conclude from a correlation between lexical decision times and some other comprehension measure that a separate lexical decision capacity is used in language comprehension, and we also do not want to take correlations between working memory tasks and comprehension tasks to mean that fluctuations in some separate working memory capacity are causing changes in comprehension abilities. For us, all of these tasks are comprehension tasks, and correlations emerge because the task demands of the ones traditionally called working memory tasks and the ones traditionally called comprehension tasks are such that performance is affected by similar factors, such as the participants’ language-processing experiences. We next review the role of experience, followed by attention to the role of phonological representations and other biological factors in language-processing tasks, including working memory tasks.

### *Reappraising the Role of Experience*

Within the connectionist language-processing framework, individual differences in language-processing skill emerge from the interaction of experience and biological factors. We turn first to linguistic experience and argue that it has far-reaching and complicated effects that go well beyond variations in vocabulary size, the traditional focus of experience effects within working memory literature (Carpenter et al., 1994; Daneman, 1988). An important part of this argument concerns how connectionist models process input, particularly how they handle the joint effects of the frequency of a particular linguistic pattern and the “regularity” of the pattern, that is, its similarity to other patterns in the language.

### *Experience in Connectionist Models*

The complex effects of experience on comprehension have been well studied in the domain of visual word recognition. In these models, knowledge of the mappings between spelling and pronunciation are inherently tied to the processing that the network

performs (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989). During word recognition, activation from a layer of orthographic representation units passes to a layer of phonological units through sets of intermediate “hidden” units. The amount of activation that is passed to the next level is dependent on which orthographic units have been activated by the input and the weights on the connections between the units. These weights are set by prior learning experience, with greater experience yielding more reliable mappings between input and output. As a result, high-frequency words are recognized more rapidly and accurately than are low-frequency words. Moreover, the effect of experience (i.e., frequency) is greater for irregular words with exceptional orthography–phonology correspondences such as *have* and *deaf* (cf. *cave*, *save*, and *leaf*, *sheaf*) than for regular words that follow the dominant letter–sound correspondences in the language. This Frequency  $\times$  Regularity interaction arises because performance on a regular word is helped not only by prior experience with this word but also by prior experience with words that have similar spelling–sound correspondences. By contrast, pronunciation success for irregular words, which have unusual correspondences, is much more strongly determined by experience with the particular word.

A crucial component of the Frequency  $\times$  Regularity interaction for our account of individual differences is that the precise nature of the interaction can vary with overall exposure to the language. For example, Seidenberg (1985) examined the effects of reading skill on reading regular and irregular words and found that for good readers, irregular words were no harder than regulars except in the very low frequency range, but poor readers had longer latencies for irregular words than for regular words except in the very high frequency range. Seidenberg interpreted these results to indicate that good readers, who read a great deal, had a larger band of effectively high-frequency irregular words for which context-sensitive pronunciations could be computed rapidly, compared with the poor readers, who read less and were less practiced computing the irregular pronunciations. In other words, variation in the amount of exposure to language has two important consequences: (a) it has differential effects on the processing of easy and more difficult material, and (b) it can create individual differences that appear qualitative in nature, such as the different Frequency  $\times$  Regularity patterns that Seidenberg found in good and poor readers. This leads us to the crux of our claim about many of the individual differences in language processing in the normal population: Whereas Just and Carpenter and others have attributed these differences to variations in working memory capacity, we claim that the differences are instead different patterns of Frequency  $\times$  Regularity interactions, owing largely to differential exposure to language—that higher skilled comprehenders, who score well on the reading span test, are ones who read more or read more complex material than lower skilled comprehenders. We begin to elaborate this argument through four examples of syntactic and discourse processing that Just and Carpenter cited as evidence for the existence of a separate working memory system for language.

### *Interpreting Complex Unambiguous Sentences*

For a variety of reasons, sentences with object-relative clauses, such as *The reporter that the senator attacked admitted the error*,

are more difficult to comprehend than sentences with subject relatives, such as *The reporter that attacked the senator admitted the error* (Bever, 1970; Gibson & Ko, 1998; Holmes & O'Reagan, 1981; King & Just, 1991). Just and Carpenter (1992, pp. 128–129) discussed a subset of the results from King and Just (1991): High-span participants, those who scored well on Daneman and Carpenter's (1980) reading span task, had a different pattern of reading times on subject and object relatives compared with the low-span participants. Because we attempt to simulate these data below, it is important to understand the King and Just results in detail. Just and Carpenter (1992) illustrate the relevant King and Just data in their Figure 9 (p. 140). The important reading time effects were localized on the main verb of the subject- and object-relative sentences (the word *admitted* in Just & Carpenter's Figure 9). Just and Carpenter offered three pieces of evidence for the importance of working memory capacity in the interpretation of these structures: (a) King and Just found shorter reading times for high-span participants than low-span participants on the main verb, averaging over relative clause type; (b) there were shorter reading times on the main verb for subject than for object relatives, collapsing over reading span; and (c) there was an interaction between reading span and relative clause type, such that there was little or no effect of reading span on the easier subject relatives, but low-span participants had longer main verb reading times than did high-span participants on the more difficult object relatives. King and Just and Just and Carpenter interpreted these results to indicate that all readers had sufficient working memory capacity to comprehend the easier subject relatives, but only the high-span readers had enough working memory capacity to interpret the more difficult object relatives. Just and Carpenter simulated these reading patterns with CC READER, varying the amount of permissible activation available for computations, holding all other factors constant. The simulation data are also shown in Just and Carpenter's Figure 9.

Waters and Caplan's (1996) main criticism of these data (pp. 764–765) is a potentially serious one, in that they argued that King and Just's (1991) effects may be spurious. Waters and Caplan noted that King and Just failed to report several critical statistical tests, and thus it is not clear that high- and low-span participants really performed differently on these two sentence types. It is true that the three effects on the main verb in Just and Carpenter's (1992) Figure 9 (main effect of reading span, main effect of relative clause type, and Span  $\times$  Clause Type interaction) are not actually supported by statistics; instead, King and Just reported statistics for data collapsed across two different reading conditions: "normal" self-paced reading (the data in Just & Carpenter's Figure 9) and another condition in which participants were reading while maintaining a one-word memory load. It is likely that collapsing over the two reading procedures strengthened some effects, and thus Waters and Caplan are correct that there is uncertainty about which of the three effects are actually reliable for the exact Figure 9 data. Clearly it would be preferable to have all relevant statistics, but it is our feeling that the analyses that are available do show individual differences in interpreting subject- and object-relative clauses, especially given the fact that the general pattern—little effect of reading span in the easy condition, larger effect in the harder condition—is consistent with data that Just and Carpenter present for other constructions (see also Just et al., 1996, but cf. Caplan & Waters, 1999b).

Thus, we agree with Just and Carpenter on the general form of the reading patterns, but we differ from both Just and Carpenter and Waters and Caplan in the interpretation of these data. We do not take the Span  $\times$  Clause Type interaction that Just and Carpenter report to be evidence for the role of a separate working memory capacity in relative clause interpretation but rather as an example of a Frequency  $\times$  Regularity interaction in sentence processing, similar to the word recognition examples described above (see Juliano & Tanenhaus, 1994; Pearlmutter & MacDonald, 1995, for other extensions of the Frequency  $\times$  Regularity interaction to sentences). Subject relatives are relatively regular in their word order because this structure has the same word order as simple active one-clause sentences, which are very frequent in English. Comprehension processes for interpreting subject relatives, such as interpreting a postverbal noun phrase as the object of the verb, are therefore aided by a comprehender's experience with simple sentences, for which many of same processes are used. Object relatives, however, have a more irregular word order (e.g., the direct object in the relative clause precedes the verb), and thus experience with simple sentences is less relevant. Successful interpretation of object relatives therefore depends strongly on direct experience with object relatives. The amount of exposure to the language, particularly to sophisticated reading material in which relative clauses are likely to occur, should therefore have a larger effect on comprehension of the "irregular" object relatives than the "regular" subject relatives. We suggest that high-span participants, who are more successful with object relatives than are low-span participants, tend to be individuals who read more and so have more experience with relative clauses. The claim is not that high-span participants simply read more object relatives than do low-span participants; rather our claim is that high-span participants read more of all kinds of relatives and that this extra experience has more effect on processing object relatives than subject relatives.

Several sources of evidence support this experience-based account. First, Roth (1984) provided children with extra experience and feedback on processing relative clauses, including subject and object relatives like those studied by King and Just (1991). She found that trained but not untrained groups improved in comprehension rates over pretest levels. Roth argued that the training with relative clauses led to improved processing efficiency and suggested that developmental changes in relative clause processing should be traced to linguistic experience, not growth of working memory. Second, Christiansen (1992) and Weckerly and Elman (1992) have suggested that differences in difficulty across the two relative clause types could emerge from the sequential nature of linguistic input and the properties of the language-processing mechanism itself, without any extrinsic limitations on a working memory system. Weckerly and Elman (1992) presented simulations in which a simple recurrent network (SRN; Elman, 1990) was trained on simple one-clause sentences and sentences with one or more subject or object relatives. Focusing on comparisons between sentences involving two levels of recursion, they found better performance on right-branching subject-relative constructions than on center-embedded object relatives. Christiansen and Chater (1999) replicated these results with sentences having only one level of recursion.

We expect that a similar approach can be extended to account for the individual differences that King and Just observed. A

straightforward test of the experience hypothesis within the connectionist framework is to assess performance of a network after different amounts of training (Munakata et al., 1997). In the simulations reported below, we show how SRNs early in training exhibit a pattern of processing similar to that of the low-span participants in the King and Just (1991) study, whereas the networks' performance resembles that of high-span participants after additional training.

### *Simulating Individual Differences in Language Comprehension Through Experience*

Our simulations involved SRNs whose architecture is essentially that of a standard feedforward connectionist network, but with the addition of an extra layer of "context" units. The networks used in our simulations, shown in Figure 1, have 31 input and output units as well as 60 hidden and context units. At a particular time step,  $t$ , an input pattern is activated on the input layer. Activation is then propagated forward through the hidden layer to the output layer. During training, the pattern on the output layer is then compared with the desired output, and network weights are adjusted in proportion to the size of this discrepancy using the back-propagation learning algorithm (Rumelhart, Hinton, & Williams, 1986). At the next time step,  $t + 1$ , the activation of the hidden unit layer at time  $t$  is copied back to the context layer and paired with the current input. This means that the current state of the hidden units can influence the processing of subsequent inputs, giving the network the ability to deal with integrated sequences of input presented successively.

*Method.* In the simulations, we trained 10 SRNs to predict the next word in a sentence, using sentences generated from a probabilistic context-free grammar with a 30-word vocabulary. Each network had a different set of initial starting weights, and each was trained on a different variation of the training corpus, corresponding roughly to the fact that different individuals have somewhat different initial conditions prior to learning and are exposed to different variations of the primary linguistic input. The grammar that generated the training corpora covered a small fragment of English and included subject noun-verb agreement, present and past tense verbs that differed with respect to their argument structure (transitive, intransitive, and optionally transitive), and subject- and object-relative clauses (allowing multiple embeddings with complex agreement structures). The grammar and its associated probabilities can be found in the Appendix along with sample sentences and other simulation details. Crucially, subject- and object-relative constructions occurred with equal prob-

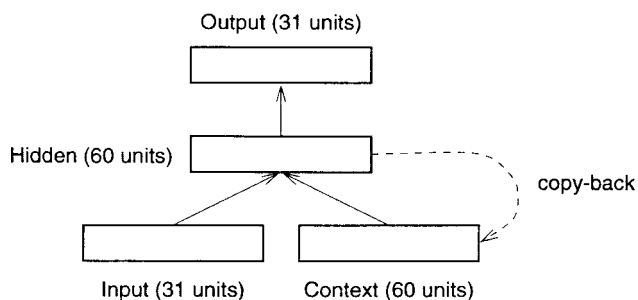
ability; each represented about 2.5% of the sentences in the training set. During training, sentences containing relative clauses were randomly interleaved with simple transitive and intransitive sentences. Each of the 10 training corpora contained 10,000 sentences with a mean length of 4.5 words, ranging from 3 to 27 words per sentence. Each pass through a corpus corresponded to one training epoch. To simulate comprehenders with different degrees of exposure to linguistic input, we evaluated the networks after one, two, and three epochs of training.

Following successful training, an SRN will output a probability distribution of possible next words given the previous context. For example, after having processed the sequence *The lawyer* and following training on sentences from the grammar in the Appendix, the SRNs should activate the units corresponding to the grammatical continuations, in this case past tense verbs, singular present tense verbs, and the relative pronoun *that*. To assess how well the network has learned the grammar, one can group the units into lexical classes and determine their theoretically appropriate level of activation given the probabilities associated with the grammar. In the present example, the singular verb class and the past tense verb class each should reach an activation level of .475, and the relative pronoun *that* should receive an activation of .05 according to the probabilities with which words in these classes were presented during training. At each point in a sentence, the network's ability to process the input can be measured in terms of how much the network output deviates from the theoretical probability distribution. We used grammatical prediction error (GPE; Christiansen & Chater, 1999) as a conservative measure of how well the network predicted next words (see the Appendix for details). GPE scores range between 0 and 1 and can be mapped qualitatively onto reading times, with low GPE values reflecting a prediction for short reading times and high values indicating long predicted reading times.

*Results and discussion.* The 10 trained networks generally performed well on the various constructions allowed by the grammar. For example, on 10 simple one-clause test sentences, the average GPE values for one, two, and three epochs of training were .19, .16, and .12, respectively. Here we focus on the results concerning the processing of singly embedded subject- and object-relative clauses, which we assessed by examining the networks' performance on 10 sentences of each type that had not been previously presented to the networks during training.

The critical question in analyzing our networks is how the pattern of data obtained from these experience-based simulations compares with the reading time patterns of high- and low-span participants obtained by King and Just (1991) and with Just and Carpenter's CC-READER simulations of the King and Just data, all shown in Just and Carpenter's Figure 9 (1992, p. 140). Our Figure 2 shows the average GPE values from the 10 networks for the two relative clause constructions and is designed to be compared with Just and Carpenter's Figure 9. It is difficult to make precise comparisons between our results and Just and Carpenter's data, first because some key statistics for their human data are not available, as discussed above, and second because Just and Carpenter presented no statistics for the CC-READER simulations, relying only on visual comparisons between the human and simulation graphs in their Figure 9 to make their points. Given the scarcity of comparable statistics, we must also rely on visual comparisons but have included error bars (standard errors) to indicate the reliability of our effects in Figure 2.

Three lines are shown in Figure 2, indicating the networks' performance following one, two, and three epochs of training. On the assumption that network performance early in training (one epoch) roughly corresponds to that of low-span participants, and that network performance later on (three epochs) roughly corre-



*Figure 1.* The basic architecture of the simple recurrent network used in the subject- and object-relative simulation. Arrows with solid lines denote trainable weights; the arrow with the dashed line denotes the copy-back connections.

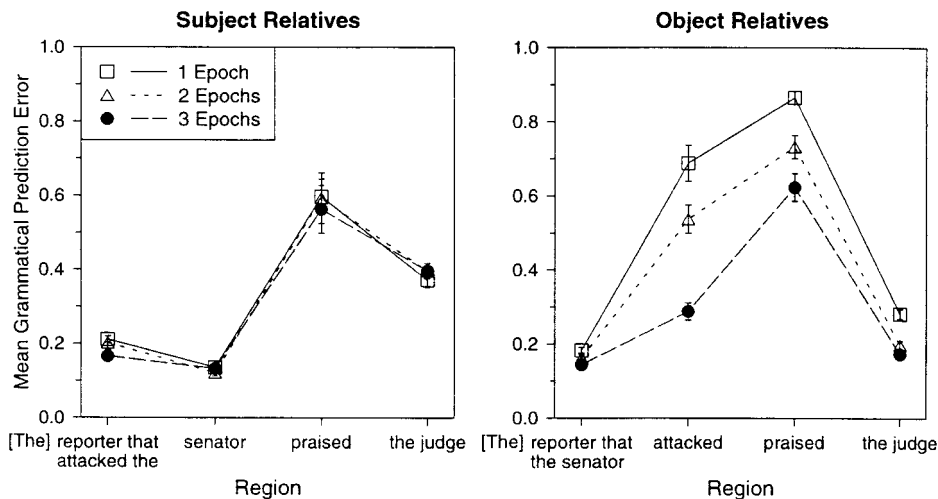


Figure 2. Network performance on sentence involving singly embedded subject- and object-relative clauses. Grammatical prediction error scores were averaged over 10 novel sentences of each kind and grouped into four regions to facilitate comparisons with the human data (see Just & Carpenter, 1992, Figure 9, p. 140). Error bars represent standard error.

sponds to that of high-span participants, the pattern of performance resembles that found in the human data discussed by Just and Carpenter, including at the main verb region where King and Just reported their three critical effects.<sup>3</sup> First, averaging across sentence type, the networks with more training had lower error rates than less-trained networks at the critical main verb region, analogous to the main effect of reading span that King and Just (1991) reported. Second, averaging across training epoch, the SRNs had higher error rates with object than subject relatives at the main verb, yielding a main effect of relative clause type. Third, there was little effect of training for the “regular” subject relatives and a large effect of training for the “irregular” object relatives at the main verb. This result captures the Span  $\times$  Clause Type interaction discussed by Just and Carpenter and stands in contrast to their own simulation results. Just and Carpenter’s CC-READER simulations, which are shown in the top panels of their Figure 9, appear to have yielded main effects of span and sentence type but not the crucial interaction in that the effect of span appears no larger on object than subject relatives at the main verb or at any other region.

Our results highlight the potentially complex role of experience in individual differences. Both types of relative clauses were encountered equally frequently in the training corpora, but the superior performance on the subject relatives stems from the networks’ abilities to generalize to rare structures as a function of experience with similar, more common simple sentences. The extent and nature of this Frequency  $\times$  Regularity interaction changed as a function of the overall experience of the network in that additional experience helped performance with object relatives more than with subject relatives.

At this point it is important to consider an alternative view, that our model is simply a connectionist implementation of a capacity-based account of individual differences. We are not claiming, of course, that there is no such thing as capacity; clearly any network (and any human) can be described as having a particular capacity to process information, and individual networks and people can vary in their capacities. What sets our account apart from Just and

Carpenter’s and Waters and Caplan’s is that we have two claims about what capacity is and is not. First, capacity is not some primitive, independent property of networks or humans in our account but is instead strictly emergent from other architectural and experiential factors. Second, capacity is not independent of knowledge, so that one cannot manipulate factors underlying the capacity of a network (e.g., hidden unit layer size, activation function, weight decay, connectivity pattern, training) without also affecting the knowledge embedded in that network. These two claims are not merely terminological changes but rather make our account qualitatively different from the working memory accounts advocated by Just and Carpenter and Waters and Caplan, for which one or more capacities can vary independently of knowledge, experience, and other factors. In their view, capacity is something that enables a certain level of processing ability or skill, whereas for us, capacity is a synonym for that skill. Just and Carpenter’s and Waters and Caplan’s intermediate step is superfluous in our account.

#### *Multiple Versus Single Interpretations of Syntactic Ambiguities*

Just and Carpenter (1992, pp. 130–132; see also the Waters & Caplan, 1996, reply on pp. 765–766) presented evidence for a working memory capacity based on MacDonald, Just, & Carpen-

<sup>3</sup> The one region where the SRN does not correspond well to King and Just’s human data is at the last word of the subordinate clause in the subject-relative sentences, where the SRN showed less processing difficulty than King and Just’s participants. This discrepancy may be due to variations in the materials, particularly to the length of the subject-relative clause, which varied in King and Just’s items but was more uniform in our materials. Gibson and Ko (1998) also used uniformly short subject relatives and found little processing difficulty at this position in self-paced reading studies, and our simulations correspond quite well to their reading data in this and other regions.



ter's (1992) study of syntactic ambiguity resolution. MacDonald et al. investigated interpretation of main verb (MV)–reduced relative (RR) ambiguities in which a verb such as *warned* either may be the main verb of the sentence, as in *The experienced soldiers warned about the dangers before the midnight raid*, or may introduce a reduced relative clause modifying the preceding noun phrase, as in *The experienced soldiers warned about the dangers conducted the midnight raid*. When the ambiguous sentence was resolved with the simple MV interpretation, high-span participants had longer reading times in the disambiguating region compared with the equivalent region of unambiguous control sentences, whereas low-span participants showed no difference in reading time as a function of ambiguity. MacDonald et al. and Just and Carpenter interpreted this counterintuitive result as indicating that high-span participants were constructing both MV and RR syntactic interpretations of the ambiguity and holding them in parallel in working memory, yielding longer reading times compared with an unambiguous sentence for which only a single syntactic interpretation could be constructed and held in memory. They argued that low-span readers could not hold two syntactic interpretations in working memory and quickly settled on the MV interpretation, with the result that their reading times did not differ from those for the unambiguous sentences. Waters and Caplan reported failures to replicate these effects, but Pearlmuter and MacDonald (1995) replicated the important results with ambiguities resolved with the simple MV interpretation (they did not test the alternative RR interpretation). Thus, we believe it is possible to design experiments in which high-span participants show greater sensitivity to stimulus manipulations than do low-span participants, but we disagree with both Waters and Caplan and Just and Carpenter about the proper account of these data.

Just and Carpenter's account of the MacDonald et al. (1992) data is notable for its clear relationship to working memory: High- and low-span participants appear to have a qualitative difference in their approach to ambiguity resolution as a direct result of the number of syntactic structures they can hold in their working memory. This approach is at variance with other results in the literature, however. First, there is good evidence against the claim that comprehenders build explicit syntactic structures and hold them in memory in the way that Just and Carpenter suggest (see MacDonald, Pearlmuter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995, for reviews). Second, there exists an alternative, constraint-based interpretation of the MacDonald et al. (1992) data, based on variations in the amount of exposure to language input and subsequent variations in sensitivity to probabilistic constraints that guide ambiguity resolution. Pearlmuter and MacDonald (1995) assessed how probabilistic constraints (such as whether it is more plausible for experienced soldiers to warn someone or be warned) affected the viability of alternative syntactic interpretations. They traced the high-span participants' reading patterns to their sensitivity to very subtle probabilistic constraints in the stimulus sentences, involving conditional probabilities over several critical words in an ambiguous region. Low-span participants' reading patterns, however, reflected only very simple constraints. Pearlmuter and MacDonald suggested that the degree of sensitivity to complex combinatorial constraints could result from variations in exposure to language input: More experienced readers have a greater range of complex constraints that can be computed efficiently, compared with less experienced

readers. Trueswell, Sekerina, Hill, and Logrip (1999) found exactly this pattern with children, in that both 5- and 8-year olds used strong local constraints in interpreting ambiguous spoken sentences, but only the older children used more complex discourse constraints that were available in the input. Thus, contrary to Just and Carpenter's claims, the original MacDonald et al. (1992) data do not support an account in which comprehenders hold syntactic structures in working memory. Instead, individual differences in ambiguity resolution performance are readily interpreted as differences in processing skill, as a function of experience.

### *Use of Context in Ambiguity Resolution*

The same interpretation can be given to studies of individual differences in the ability to use contextual information to guide ambiguity resolution. Just and Carpenter (1992, pp. 125–128) investigated high- and low-span readers' use of contextual information in resolving MV-RR ambiguities with the difficult reduced relative interpretation. They argued that only high-span participants had sufficient memory capacity to interpret the contextual information concerning the plausibility of the subject noun's being the agent or the patient of the adjacent verb. Thus, for the sequence *the soldiers warned . . . , the soldiers* serves as the agent in the simple MV interpretation (the soldiers are warning someone), and this noun phrase is the patient in the difficult RR interpretation, such that someone is warning the soldiers. Waters and Caplan (1996) argued that the results instead "could be consistent with the view that high span participants can combine different types of information more efficiently in sentence comprehension than low span participants" (p. 764). This hypothesis is generally compatible with our own view, but Waters and Caplan offer no explanation for why high-span participants are more efficient or why the efficiency explanation is preferable to the one Just and Carpenter advocate.

Our account makes this skill-through-experience position explicit. Increased experience allows rapid computation of combinatorial probabilistic information that can be more constraining than simple frequency information, whereas less experienced comprehenders (i.e., low-span participants) rely more heavily on simple frequency information and do not compute complex constraints rapidly. It is not that low-span comprehenders are completely ignorant of complex constraints; indeed, Pearlmuter and MacDonald (1995) showed that both high- and low-span comprehenders demonstrated good knowledge of complex constraints when they had unlimited time to think about the material. This result is reminiscent of Seidenberg's (1985) investigation of the Frequency  $\times$  Regularity interaction with good and poor readers: All participants could read the words accurately, but reading latencies reflected the differential effects of reading experience on regular and irregular words. Our claim is that contextual constraints such as the plausibility of a noun serving as the agent of a verb are like the constraints needed to pronounce irregular words, in that the plausibility computation depends on the frequency of a conjunction (the noun and verb in the case of plausibility, and the conjunction of different letters in reading). The amount of experience with combinatorial constraints, at both word and sentence levels, affects the rapidity of their computation.

### *Interpreting Pronouns*

Our approach to Just and Carpenter's (1992) data concerning interpretation of pronominal reference (pp. 133–134) again rests on differences in processing skill as a function of experience. Just and Carpenter cited Daneman and Carpenter's (1980) results showing that high-span participants could determine the referent of pronouns further back in a discourse than could low-span participants, as judged by accuracy on questions about a paragraph that participants had recently read. Just and Carpenter interpreted these data to indicate that only the high-span participants had sufficient memory capacity to maintain the pronoun's antecedent in memory, whereas low-span participants could not find distant antecedents for pronouns because the antecedent had been lost from working memory. Waters and Caplan (1996, pp. 767–768) claimed that this work did not challenge their multiple-resources account, and they suggested that pronominal reference in these cases was affected by postinterpretive reasoning processes. The notion of forgetting information as working memory demands increase (whether in a single or multiple working memory systems) is an attractive metaphor for understanding differences in readers' abilities to interpret pronouns, but these data also do not demand a capacity account. As with other examples of ambiguity resolution, resolution of pronominal reference is a constraint satisfaction process in which several syntactic-, semantic-, and discourse-level constraints can guide computation of the correct pronoun–antecedent pairing (e.g., Givón, 1976; Gordon, Grosz, & Gilliom, 1993). The pronoun in Daneman and Carpenter's distant-antecedent paragraphs referred to an entity that was not the current discourse topic at the point at which the pronoun appeared in the paragraph. Pronouns typically refer to prominent, recent antecedents, whereas distant antecedents are generally referenced by repeating the name or noun phrase that was originally used to introduce the entity into the discourse. Using a pronoun to refer to a distant entity is extremely rare and awkward (Ehrlich & Rayner, 1983; Sanford & Garrod, 1981); such a usage can be viewed as the referential equivalent of resolving an ambiguity with a low-frequency, complex interpretation. As we have seen, high-span comprehenders are more likely to know the subtle constraints that guide interpretation of rare usages in the language, and this appears to be no less true at the discourse level than at the syntactic level. For example, Long, Oppy, and Seely (1997) suggested that varying amounts of experience with texts underlie differences between skilled and less skilled readers' abilities to compute complex discourse-level information and draw inferences. The pronominal reference data are readily interpreted as a function of processing skill and do not demand constraints on one or more external working memory capacities.

### *Summary*

Our account of these four sentence- and discourse-processing domains views individual differences as emerging from variations in experience. Whereas comprehenders with limited experience can rapidly compute simple constraints to resolve ambiguities, only more experienced comprehenders have encountered complex combinatorial constraints often enough to compute them efficiently. This skill-through-experience account of comprehension is superior to both Just and Carpenter's and Waters and Caplan's capacity accounts. Our approach offers a unified treatment of the

Frequency  $\times$  Regularity interaction in word recognition, sentence processing (the first three examples above), and discourse processing (the fourth example) in that in all cases, increasing exposure to the language yields greater sensitivity to subtle combinatorial constraints. By contrast, Just and Carpenter's claim that only high-span participants have the capacity to remember antecedents, build syntactic structure, or compute complex constraints does not extend well to the word recognition case. It is difficult to argue that poor readers' special difficulty with exception words is due to overloaded memory capacity, because these effects are typically demonstrated in tasks in which participants are simply reading isolated words aloud.<sup>4</sup> Waters and Caplan similarly have no position on word recognition, and their interpretation of the four domains offers no unified approach to individual differences, in contrast to our own account.

### *Combined Effects of Biological and Experiential Factors*

Having now presented the substantial influence of experiential factors in creating individual differences, we now examine architectural constraints and their interaction with linguistic experience. We argue that many of the individual differences in the normal young adult population that cannot be accounted for by experience alone can be accounted for by individual differences in the nature of phonological representations and by the interaction of this architectural factor and experiential factors. Our claim is that some comprehenders have a more precise representation of phonological information than others and that this representational precision, not some external working memory capacity, underlies a number of individual differences in comprehension performance. We are not claiming that phonological precision is the only relevant biological difference among individuals; aging leads to another biological change, reductions in processing speed, and clearly some individuals experience disease, stroke, and so forth, which have profound effects on neural computations.<sup>5</sup> We pursue these arguments through a reevaluation of three sets of data that Just and Carpenter have offered in support of their account.

### *Extrinsic Memory Load*

In language comprehension experiments with an extrinsic memory load, participants read several sentences while retaining a

<sup>4</sup> Just and Carpenter do incorporate basic word frequency effects in their model in that higher frequency words are activated more rapidly than are lower frequency words in CC-READER. However, this difference has nothing to do with working memory (both high- and low-span simulations have exactly the same frequency parameters), highlighting the extent to which innate working memory capacities and linguistic experience (word frequency) are independent components in Just and Carpenter's approach.

<sup>5</sup> One extreme example is aphasia, in which stroke victims display a variety of language-processing problems. Interestingly, Miyake et al. (1994) point out that language experience prior to the onset of aphasia is likely to affect the nature of individual performance deficits, but they invoke a severely constrained version of the Just and Carpenter capacity model in their explanation of these individual differences, which does not provide any means for readily incorporating variations in experience. In contrast, our approach provides a more suitable framework for studying potential interactions between prior experience and poststroke performance deficits (see Plaut, 1997, for an instantiation of this idea).

series of words or digits for later recall. Just and Carpenter (1992) reviewed data from King and Just (1991) and other extrinsic load studies. They hypothesized that the extent to which an extrinsic load interferes with comprehension is a function of three factors: the size of the load, the participant's reading span, and the syntactic complexity of the stimulus sentences (pp. 132–133). Waters and Caplan (1996), however, argued that the load task and normal language comprehension are largely unrelated in terms of their demands on working memory systems, and they suggested that Just and Carpenter's data were weak, in particular that they did not yield the crucial three-way interactions that should have underlain Just and Carpenter's claims (p. 767). To the extent that there are effects of extrinsic load on comprehension (and there clearly are some), the data are important because they suggest that a prototypical memory task—remembering a series of words through a short distraction interval—can tap properties of language comprehension. Of course the mere fact that performance on the reading span task and on the extrinsic load task correlate with one another is not particularly interesting, because the two tasks are essentially identical—both require the participant to simultaneously comprehend language while retaining a load of words or digits for later recall. The only difference between the two tasks is in the choice of dependent measure: In studies on the effect of extrinsic load on language comprehension, the dependent measure is comprehension accuracy, whereas it is the number of words recalled in the reading span task. The correlation itself therefore has little interest, but it is important to understand how the extrinsic load and comprehension processes interfere with one another and whether this interference argues for any separate working memory capacities.

Our account of the interference effects of load relies on the fact that maintaining a set of unrelated words requires substantial activation of phonological representations. Phonological representations are a key to understanding the nature of the reading span task and the extrinsic load studies because (a) phonological and articulatory representations must be activated in order to utter the words for the load task; (b) phonological activation is an important component of written and spoken sentence comprehension, particularly for certain difficult sentence structures; (c) the extent to which phonological representations are important during comprehension of difficult syntactic structures is likely to vary inversely with experience, such that phonological information is more crucial for less experienced comprehenders; and (d) there appear to be notable individual differences in the “precision” of phonological representations computed during language comprehension, and these differences are thought to owe both to reading experience and to biological factors. We discuss the evidence for each of these claims below, first addressing Just and Carpenter's extrinsic load data and then examining implications of these claims for Daneman and Carpenter's (1980) reading and listening span tasks.

*Phonological representations in extrinsic load and in sentence processing.* As many classic studies of short-term memory attest, phonological activation is crucial to rehearsing a set of unrelated words (e.g., Conrad, 1964; Conrad & Hull, 1964). We suggest that this activation is part of the articulatory planning processes that are a normal part of speech production. Phonological codes are also crucial for language comprehension, even with written presentations. Several different roles have been suggested for phonological codes during comprehension, including recognizing printed words

(Van Orden, 1987) integrating information across saccades during reading (Pollatsek, Lesch, Morris, & Rayner, 1992), accessing or reinterpreting word meaning in sentence contexts (Folk & Morris, 1995; Keir & Duffy, 1999), sentence processing (Montgomery, 1995), and integrating information from different parts of a text during reading (Folk & Morris, 1995).<sup>6</sup> Thus, we directly challenge Waters and Caplan's claim that load and sentence comprehension are unrelated; on our account they are intertwined by the crucial involvement of phonological representations.

A key part of our argument is the fact that activated phonological codes have been shown to exert strong interference on one another (Bock, 1987; Dell & O'Seaghdha, 1992), probably owing to the nature of speech production processes: To avoid excessive speech errors, articulatory planning processes must strongly activate only one phonological-articulatory unit at a time and inhibit activation of recently uttered and upcoming units (Dell, Burger, & Svec, 1997). This inhibition can affect comprehension as well: Repeated occurrences of a word-initial phoneme in a sentence (as in tongue twisters), slow sentence comprehension compared with sentences without such repetition (McCutchen, Dibble, & Blount, 1994). It is therefore not surprising that in comprehension during extrinsic load, the phonological activation required for the load task interferes with phonological activation in language comprehension, and vice versa. The effect of load size that Just and Carpenter observed directly follows from this account, as larger extrinsic loads create more phonological activation to compete with the activation needed for sentence processing.

The interference from external load has a greater effect on the comprehension of some sentence types than others because some sentences appear to involve phonological representations to greater degrees than others do (Montgomery, 1995). Sentences that have impoverished discourse context (a common situation in psycholinguistic experiments), rare or awkward syntax, conflicting probabilistic constraints, or other challenges require greater involvement of phonological information during comprehension compared with simpler sentences. For example, object relatives, which are more challenging than subject relatives, are likely to rely more on phonological information than subject relatives, and thus the phonological activation necessary for the load task may interfere with the comprehension of object relatives more than subject

<sup>6</sup> The role of phonological representations in visual word recognition, auditory and written language comprehension, dyslexia, specific language impairment, and so on, form an enormous body of work spanning developmental, cognitive, and educational psychology and neuropsychology (see Baddeley, Gathercole, & Papagno, 1998). Although the majority of researchers in these fields agree that phonological representations form a crucial part of reading and higher level language comprehension, there are exceptions, particularly in patients who perform poorly on tests of phonological memory yet exhibit good sentence processing abilities (e.g., Martin et al., 1994; Waters et al., 1991). In these cases it is important to examine the relative difficulty of the phonological and sentence-processing tasks, as some sentence-processing tasks, such as grammaticality judgment, may be performed reasonably accurately without good comprehension. More generally, it is clear that different aspects of language comprehension involve phonological activation to varying degrees (e.g., Jared & Seidenberg, 1991), so it would not be surprising if some studies failed to find significant involvement of phonological codes in some combinations of participants, stimuli, and tasks.

relatives. Thus, the effect of sentence type that Just and Carpenter observed in load experiments can be traced to variations in the degree of involvement of phonological representations in different kinds of sentence structures.

The effect of reading span that Just and Carpenter observed can be explained in a similar fashion. The degree of interference between phonological representations used in sentence comprehension and in the extrinsic load task is expected to be larger for low-span participants than for high-span participants. Low-span participants comprehend object relatives more slowly and less accurately than do high-span participants (King & Just, 1991), owing in our view to their more limited language processing skill compared with high-span participants. As phonological representations appear to play a larger role in sentence comprehension with sentences that are syntactically difficult, the low-span participants' lower skill level entails that they will require substantial phonological activation to aid syntactic analysis for these sentences. These participants will thus suffer more from the competing phonological activation from the extrinsic load task than will high-span participants, whose greater skill should yield lower demands for phonological activation during comprehension of object relatives.

*Reading and listening span again.* Because the extrinsic load task and Daneman and Carpenter's (1980) reading span task are essentially identical save for the dependent variable, our claims about the load task have implications for reading span as well. We have suggested that the reading span task and its auditory equivalent, listening span, are basically measures of participants' abilities to do particular language-processing tasks. More specifically, these tasks measure the ability to comprehend sentences in the face of competing phonological activation from a series of words that are being prepared for articulation, and to maintain phonological activation for the words in the face of competing demands from sentence processing. This account predicts that measures of the accuracy of phonological representations, which are commonly assessed in auditory perception tasks that have little or no memory component, should be a good predictor of performance on the listening and reading span tasks. Just and Carpenter's working memory approach makes no such prediction, but several studies of reading acquisition have provided good evidence for the role of phonological representations in the listening span task (Gottardo, Stanovich, & Siegel, 1996; Leather & Henry, 1994). Moreover, this correlation can be traced specifically to the competing demands of sentence processing under extrinsic load, because measures of phonological skills do not correlate particularly well with simple word span tasks (Bowey, Cain, & Ryan, 1992; Leather & Henry, 1994; Mann & Liberman, 1984), and some patients who perform poorly on simple span tasks have reasonably good sentence comprehension (Martin et al., 1994; Waters, Caplan, & Hildebrandt, 1991).

The importance of phonological information in sentence processing and for performance on the listening span task (Gottardo et al., 1996; Leather & Henry, 1994) suggests that high- and low-span comprehenders differ in their representation of phonological information. The accuracy of phonological representations has been shown to be strongly dependent on reading experience (e.g., Bertelson, 1987), but there is clear evidence of a biological component from a number of different paradigms and laboratories. For example, Molfese and Molfese (1997) found reliable differences in

neonates' evoked potentials to speech stimuli and demonstrated that these differences predicted substantial variance in the children's verbal abilities 5 years later. Behavioral testing of phonological skills has similarly revealed measurable individual differences in prereading children who are matched for many experiential factors, and these differences correlated with children's auditory sentence comprehension abilities as well as their reading abilities (Mann, Cowin, & Schoenheimer, 1989). Abundant evidence also points to early abnormalities in phonological representations in children who are later diagnosed as dyslexic or as having other language disabilities such as specific language impairment (Shankweiler, Crain, Brady, & Macaruso, 1992; Shankweiler et al., 1995; Stanovich, 1992). If such innate individual differences exist and if these differences affect the acquisition of reading skills (Shankweiler et al., 1992; Stanovich, 1992) then the interaction between biological differences in phonological representations and the effects of experience is likely to be tremendously complicated. People with lower quality phonological representations may tend to find reading frustrating and thus be likely to have less print exposure than those who enjoy reading. Because reading is the major source of knowledge about sophisticated vocabulary and syntactic structure (Hayes, 1988), it is clear that reduced print exposure could rapidly compound biological individual differences. These interactions have not been adequately explored to date, in part because the working memory approaches favored by Just and Carpenter and Waters and Caplan have limited the interest in the role of experience in individual differences.

The complexity of the biological and experiential interactions make this domain a natural target for computational modeling, but current models have not incorporated the necessary architecture to fully explore this issue. We have argued that phonological representations have a crucial role in sentence processing, but none of the currently available sentence-processing models (e.g., Just & Carpenter, 1992; St. John & Gernsbacher, 1998; Tabor, Juliano, & Tanenhaus, 1997; Weckerly & Elman, 1992) have a phonological component. Similarly, models that use phonological representations seldom go beyond the individual word level (e.g., Dell & O'Seaghdha, 1992; Gupta & MacWhinney, 1997; McClelland & Elman, 1986). Moreover, none of these models explore the effects of experience. We took a first step by explicitly modeling experience in the simulations of relative clause processing described above, but we had no phonological component. Similarly, Harm and Seidenberg (1999) simulated the effects of reading experience in networks with differing amounts of phonological skill, but their network modeled only recognition of individual words. These first steps are encouraging, but such models have not yet attempted the nontrivial task of incorporating phonological representations in a sentence-processing model that simulates effects of both biology and experience.

#### *Functional Magnetic Resonance Imaging (fMRI) Data*

In their reply to Waters and Caplan, Just et al. (1996, pp. 774–775) reported fMRI data from participants performing comprehension and working memory tasks. In a read-only condition, participants read sentences silently and judged each as true or false, and in a read-and-maintain condition, participants also remembered the sentence-final words of two to four sentences, silently recalling them at the end of the set. Just et al. found that the

read-only condition activated both Broca's and Wernicke's areas significantly more than a control condition in which participants merely fixated on a visual cross. In the read-and-maintain condition, the activation in Broca's area was of similar magnitude, and Wernicke's area was significantly more activated than in the read-only condition. Just et al. interpreted this pattern as indicating that the same computational capacity is involved in both the reading span task and language comprehension and that these data provide evidence against Waters and Caplan's hypothesis that the reading span task invokes a separate memory system than the one involved in normal language comprehension. As we have never doubted that performance on the reading span task correlates with performance on language-processing tasks, our focus here is whether Just et al.'s data provide evidence for a computational capacity that is separate from linguistic representations.<sup>7</sup> The data do not provide this evidence in our view but rather reflect the activation of representations during language comprehension and, in the read-and-maintain condition, the additional phonological activation associated with the extrinsic load.<sup>8</sup> Broca's and Wernicke's areas have long been the hypothesized sites of linguistic representations; if we consider imaging data for these areas as reflecting instead the "potential resource supply, which could place an upper limit on comprehension" (Just et al., 1996, p. 775), it is a puzzle where the linguistic representations are stored in the brain. For us, the answer is clear: our account does not maintain a distinction between declarative knowledge and the computation space in which knowledge is activated, and there is not a linguistic working memory capacity that is separate from linguistic processing. The activation patterns that are observed during language comprehension using imaging methods do show the locus of various kinds of linguistic processing, but they do not show the locus of a separate working memory system. The fact that Just et al. found more activation in the extrinsic load task than in the read-only condition is interesting, but it does not demand a distinct working memory.

### *Age-Related Differences*

A number of researchers have observed declines in linguistic performance in aging, though many other studies have failed to find reliable age-related differences (see Kemper, 1992; Light, 1992, for review). Just and Carpenter (1992) attributed age-related declines to reductions in working memory capacity in aged adults (pp. 129–130; see Waters & Caplan, 1996, p. 765). Our account of these data requires an exploration of the relationship between age and performance on working memory tasks as well as the relationship between age and language comprehension performance.

The age-related decline in performance on working memory measures is well attested (e.g., Hasher & Zacks, 1988; Salthouse, 1994), but there is considerable disagreement concerning the reason for this result. Whereas Just and Carpenter attribute performance declines to a shrinkage in the working memory capacity in aging, Hasher and Zacks (1988; Stoltzfus et al., 1996) have attributed it to a loss of ability to inhibit irrelevant information, resulting in noisier computations. Another alternative is provided by Salthouse (1994, 1996), who has observed that older adults typically exhibit slowed perceptual and cognitive processes in comparison with younger adults. Salthouse has attributed declines in performance on working memory tasks to this age-related slowing and

has reported that age predicts little or no variance in performance on working memory tasks after measures of processing speed have been taken into account (see also Fisk & Warr, 1996). Thus, Salthouse used the notion of processing speed, for which age-related declines have been independently demonstrated in perceptual tasks that have no memory component, to explain the age-related declines on working memory tasks, without positing a shrinkage in working memory capacity. This position, which is clearly compatible with our own, unifies the decline in perceptual and working memory tasks within one framework, whereas Just and Carpenter's claim for a shrinking working memory capacity in aging does not provide an explanation for the age-related decline on purely perceptual tasks.

The processing speed view can be readily extended to account for age-related differences in linguistic performance, as there is evidence that this biological constraint on processing efficiency affects speech interpretation processes (Wingfield, Poon, Lombardi, & Lowe, 1985). Moreover, older adults have been shown to have reduced abilities to use contextual information in ambiguity resolution and other linguistic tasks (Hess & Higgins, 1983; Micco

<sup>7</sup> Just et al.'s extremely concise reporting of their fMRI methods makes it difficult to evaluate their data precisely, but it should be noted that their findings appear to be open to a number of alternative interpretations, including ones that are compatible with Waters and Caplan's position. First, Just et al. assumed that Waters and Caplan's claim of a dissociation between the reading span task and language comprehension necessarily predicts different anatomical locations for the two processes, but this is not a necessary assumption. Many anatomical regions of the brain participate in more than one function, and Caplan (1992, 1994) himself has been a leader in noting the complicated relationship between location and function. Thus, it is not clear that Just et al.'s localization data have a straightforward relationship to Waters and Caplan's hypothesis. Second, Just et al. had their participants make a true–false judgment for each sentence in (apparently) both the read-only and the read-and-maintain conditions. Judging the truth of a sentence is exactly the kind of postinterpretative task that Waters and Caplan claim is governed by the conscious pool of resources that also governs performance on the reading span task. Thus in Waters and Caplan's view, it makes perfect sense for both the read-only and read-and-maintain conditions to activate the same regions, because both tasks require the conscious pool of resources. Third, the fact that true–false judgments and word recall were all performed silently makes it impossible to verify the extent to which participants were actually maintaining the words or comprehending the sentences. Finally, it appears that Just et al.'s scanning time in the read-and-maintain condition could include at least some of the recall portion of the task in addition to the reading and maintaining portion, and so it is not clear whether the increased Wernicke's activation in the read-and-maintain condition should be attributed to the need to maintain the sentence-final words or to the generation of prearticulatory codes during the silent recall period in this condition.

<sup>8</sup> This is not to say that the only difference between the two conditions is additional phonological activation in the read-and-maintain condition. That is, the nature of sentence processing is not necessarily identical in the presence or absence of an extrinsic load in that the pattern of activation over many levels of representations (lexical, syntactic, discourse, etc.) may be altered by the presence of competing phonological activation from the extrinsic load words. This possibility underscores the difficulty of using a subtractive methodology to compare two tasks that may differ in complicated ways (Sergent, 1994).

& Masson, 1992; Dagerman, MacDonald, & Harm, 2001).<sup>9</sup> Dagerman et al. argued that this reduction in context use can be traced to slowed processing in older adults. In two cross-modal naming studies, they investigated how older and younger adults disambiguated words such as *fires*, which have both a noun and a verb interpretation. They found that both young and elderly participants were able to use strong syntactic constraints that forced one or the other interpretation of the ambiguous word in a sentence but that only the younger participants used probabilistic semantic cues from an immediately preceding context word to aid ambiguity resolution. In off-line paper-and-pencil tasks, however, elderly and younger participants showed equal sensitivity to the probabilistic contextual information. This result is reminiscent of Pearlmutter and MacDonald's (1995) finding that young adult low-span participants showed sensitivity to subtle probabilistic constraints in off-line but not on-line tasks, whereas high-span participants were sensitive to these constraints in both types of task. Dagerman et al. hypothesized that the elderly participants' reduced abilities to use context stemmed from slowed computational processes, which affected their on-line performance, where the rate of the speech input was not under the participant's control, but not their off-line performance, where participants could read the written stimuli at their own pace. Dagerman et al. simulated the differences in the two age groups and the two tasks in a simple localist dynamical network, in which the speed of activation of information was manipulated; a fast activation rate simulated the young participants, and a slow rate simulated the older participants. The faster network showed evidence of early context use, whereas context use emerged later for the slower network, consistent with the behavioral data. The networks had no separate working memory system, yet they captured important age-related individual differences of the sort that Just and Carpenter attributed to an age-related decline in working memory capacity.

### General Discussion

In this article, we have advocated an alternative approach to understanding individual differences in language comprehension. Unlike the approaches of Just and Carpenter (1992) and Waters and Caplan (1996), ours does not sanction one or more working memories functionally separated from the representation of linguistic knowledge. Instead the representation and processing of language is part of the same (network) system. We next turn to two other issues: (a) the extent to which our account is a real alternative account of individual differences or merely a redescription of some of Just and Carpenter's and Waters and Caplan's ideas and (b) the relationship between our approach and some of the many other alternative accounts of working memory and individual differences.

#### *Real Alternative or Redescription?*

Just and Carpenter (1992, pp. 124, 144–145) noted that although they prefer an account of individual differences that incorporates variations in the size of working memory capacity across individuals, the data that they have cited are also largely compatible with an efficiency-of-processing account. If an efficiency-based account and a working memory account are truly interchangeable, then it is important to ask whether we have offered a

real alternative to Just and Carpenter's position or whether we have merely amplified a slightly different perspective that Just and Carpenter and colleagues have already noted as a possible interpretation of their data. We argue below that our position is distinct in important ways from both Just and Carpenter's and Waters and Caplan's positions.

#### *Differences From Just and Carpenter*

Just and Carpenter are correct that within the particular cognitive architecture they adopt, the efficiency and capacity approaches are interchangeable. Thus had they preferred an efficiency account, Just and Carpenter could have simulated high- and low-span readers by keeping working memory capacity constant and setting the activation parameters for productions and declarative knowledge in a different way for the high- and low-span simulations, such that the high-span simulation activated knowledge and completed productions more efficiently. In other words, efficiency (whether owing to experience or to biological factors) and capacity are independent parameters in Just and Carpenter's framework. In the connectionist-based efficiency account that we propose, however, experiential factors and biological factors are inextricably intertwined, and in simulations, the model as a whole (both its architecture and its training experience) constitutes the source of performance limitations. Our account therefore promotes the study of individual differences as the interaction between biological (architectural) variations and differences in exposure to relevant aspects of language, rather than as independent factors.

This approach makes different predictions in a number of domains that Just and Carpenter have presented. First, our approach relates individual differences in the Frequency  $\times$  Regularity interaction in word recognition to similar phenomena at the sentence and discourse levels. Just and Carpenter's account offers no such integration and in fact does not extend well to word recognition. Second, our interpretation of the nature of the reading and listening span tasks correctly predicts that phonological representation accuracy, measured in tasks without a memory component, should correlate well with performance on these span tasks. Just and Carpenter's account makes no such prediction. Finally, our account captures the fact that processing speed, again measured in tasks without a memory component, correlates with performance

<sup>9</sup> The existence of some reduction in performance with aging is interesting in light of the fact that older adults have more experience with language than do younger adults. Thus, the negative effects of perceptual or cognitive slowing appear to overwhelm any positive effects of additional experience. As we have suggested, the interaction between experience and biological factors is likely to be complicated. For example, elderly adults certainly have more experience with language than younger adults, but that experience, particularly recent experience, may not be helpful for performance on a psycholinguistic experiment. That is, the kinds of materials that an elderly retired person reads are likely to be very different from that of a college student, and the often difficult materials in psycholinguistic experiments may be more similar to a student's recent reading than a retired person's. A second complication is that the helpful effects of experience need not be linearly increasing throughout the life span, so that 50 years of avid reading may not be much better than 30 years' experience, whereas the difference between 10 and 30 years' experience could be substantial.

on working memory tasks, whereas Just and Carpenter's account does not make this prediction.

Of course the production system architecture that Just and Carpenter adopt could be made to accommodate the findings that we have discussed, as this architecture permits independent manipulations of capacity, production efficiency, representation frequency, and so forth. Crafting the model in this way, however, would fail to capture the essential interactions between architecture and experience that are a natural part of our account. The insight we have offered is that although there are individual differences in processing skill or *capacity*, capacity is not a primitive in the theory and cannot be independently manipulated in a simulation. Capacity instead emerges from other factors that are an intrinsic part of representing knowledge within connectionist networks.

### *Differences From Waters and Caplan*

Because Waters and Caplan's article focused far less on their own account of working memory than on Just and Carpenter's data, their position is not articulated fully enough for us to formulate a reply with any precision. It is nonetheless very clear how our view of language comprehension differs from Waters and Caplan's: They advocate a modular approach to language comprehension and we advocate an inherently interactive one. These two approaches make markedly different predictions that have been extensively documented in the literature, and we do not review them here (see Frazier, 1987; MacDonald et al., 1994). With particular regard to working memory, Waters and Caplan suggest that different kinds of linguistic processes each have their own working memory capacity; in particular they distinguish between *interpretative* and *postinterpretative* processes. It might initially appear that our view, in which activation at multiple levels of representation replaces a separate working memory, is a sort of implementation of that approach. Indeed, Martin (1995) suggested a similar view, (a) that there are modular linguistic processes, (b) that each has its own working memory capacity, and (c) that they might be implemented in a connectionist system. However, Martin's (1995) view of working memory as a "dedicated temporary storage capacity" (p. 626) does not translate well to existing connectionist models, including the models she cites, Dell (1986) and McClelland and Kawamoto (1986). Neither of these models has a separate "storage capacity" that can temporarily store intermediate processing results, and there is nothing in these models that can readily be ascribed the function of a separate working memory, independent of the representation and processing of linguistic knowledge.<sup>10</sup> Thus though Waters and Caplan (and R. C. Martin, 1995) talk in a general way about capacity being intertwined with processing, the inherently symbolic and modular account that Waters and Caplan adopt prevents them from realizing this approach in the way that we have here.

### *Relationship to Other Theories*

This commentary has of course focused on the positions of Just and Carpenter and Waters and Caplan, but the concept of working memory admits a huge variety of interpretations that are not encompassed by these two views (Richardson et al., 1996, survey many of the alternatives). Space constraints prohibit a detailed

discussion here, but our approach does have consequences for several other accounts of working memory and its putative role in cognitive processes.

First, our advocacy of the importance of experience draws on Ericsson and Kintsch's (1995) claims in their own working memory model. Reviewing evidence for individual differences in domains ranging from text comprehension to chess playing, these authors found that when an individual has a greater than average working memory capacity, this capacity is restricted to skilled activities, suggesting that superior performance on some working memory assessment stems from experience-based processing efficiency. They suggested that the data reviewed in Just and Carpenter fit equally well with an experience-based approach, and that "the total capacity does not differ between good and poor readers, but the processing efficiency of good readers is assumed to be higher, so that their effective working memory capacity is enlarged because they can use their resources better" (Ericsson & Kintsch, 1995, p. 229). Although we share their appreciation for the importance of experience, there are nonetheless clear differences between our approach and Ericsson and Kintsch's. The most important of these is that they, like Just and Carpenter and Waters and Caplan, postulate a working memory for temporary storage and processing, separated from the representation of long-term knowledge—despite having ready access to parts of it. In our account, the long-term knowledge of language is not functionally separated from the locus of processing, leading to many desirable consequences, as we have seen. Moreover, our model provides the first step toward an account of how increased processing capacity in skilled performance may be acquired through learning, whereas Ericsson and Kintsch do not discuss a learning component.

Second, our approach relates to the considerable debate concerning whether linguistic working memory is separate from other nonlinguistic working memory systems, such as spatial working memory (e.g., Goldman-Rakic, 1997; Shah & Miyake, 1996; Smith & Jonides, 1997). This debate obviously has echoes within the debate between Just and Carpenter and Waters and Caplan concerning whether there is more than one working memory capacity within the linguistic system. Similar to their debate, a major source of evidence in this broader controversy is whether various linguistic working memory tasks correlate with nonlinguistic ones and which working memory tasks correlate with measures of cognitive performance (e.g., Engle et al., 1992; Shah & Miyake, 1996). We have argued that linguistic working memory tasks are simply special kinds of language-processing tasks. The situation in nonlinguistic domains appears similar, so that spatial working memory tasks, for example, draw on the same skills as are used to navigate through space, and persons who perform well on these tasks are ones who have a high degree of spatial-processing skill, owing to some combination of biological and experiential factors. To the extent that the biology and experience underlying

<sup>10</sup> Of course, it is possible to construct a hybrid model that includes a separate working memory, and indeed some efforts have been made in this direction (e.g., Kwasny & Faisal, 1990), but this approach violates the predominant connectionist approach to language, and the separate working memory component in such hybrid models will suffer from the same shortcomings identified for Just and Carpenter's and Waters and Caplan's symbolic models here.

language comprehension skills and the skills used in navigation through space are similar, or to the extent that particular spatial and linguistic working memory tasks have similar components, performance on linguistic and spatial working memory tasks will correlate with one another. Thus, in our view, the presence or absence of correlations between two “working memory” tasks or between “working memory” and “processing” tasks in no way implies that there is a separate working memory capacity for any cognitive domain. Such results merely indicate that there are processing skills in various cognitive domains, and working memory tasks in these domains draw on some subset of these skills.

More speculatively, our approach may have an important relationship to theories of attention. Like working memory, *attention* is a controversial notion (indeed the two terms sometimes appear interchangeable). Many of the same questions posed in this article have their counterparts in the attention literature, including the extent to which attention can be viewed as separate from perceptual processing. For example, Behrmann, Zemel, and Mozer (1998) argued that object-based attention is not a separate entity but rather arises from an experience-based visual processing mechanism that groups certain visual features together. This view is clearly consistent with our own, holding out the possibility that for both early perceptual processes and higher cognitive functions, the attention, working memory, and computational capacity of a system may be unified with the system’s computational processes, rather than viewed as separate independent entities.

#### Postscript 1: Reply to Just and Varma (2002)

Just and Varma (2002) unsuccessfully tried to blur the architectural distinctions between our different approaches. On the one hand, they attempted to cast connectionist networks as symbolic, as in their statement that grammar is in the network connection weights, whereas capacity varies with noise in the input signal. This claim is demonstrably false; knowledge and capacity are both constrained by noise and connection weights in connectionist networks. On the other hand, Just and Varma tried to co-opt the connectionist approach by suggesting that “the functions of the resources and the productions are completely intertwined” (p. 55) in CC-READER, because each cannot work without the other. This observation misses the point: Virtually any multicomponent model has interacting components, but in connectionist networks, capacity and processing do not just interact, they are the same thing: They are jointly realized by the same architecture, and they cannot be independently manipulated. Just and Varma’s confusion about architectural distinctions cascaded into a failure to perceive the real functional consequences of these distinctions, resulting in vague statements about our approach not being “empirically distinguishable” from theirs (p. 57). In fact, our approach makes substantially different predictions, including about the relationship between word recognition and sentence processing and the relationship between frequency, regularity, and capacity. Just and Varma offered nothing concerning these points. Much of what remains of the “empirically distinguishable” problem rests not with us but in the large number of free parameters in their approach: With potential individual differences in resources, productions, knowledge, and strategic allocation among them, the falsifiability of Just and Varma’s model is questionable.

#### Postscript 2: Reply to Caplan and Waters (2002)

Caplan and Waters (2002) mistakenly feel that they differ with us when they end their reply with, “at the end of the day, domain-specific limitations on computational capacity will be left standing” (p. 73). However, we fully agree that capacity limits exist, that they are interesting, and that they stem from both innate and experiential components. Our differences with Caplan and Waters lie in how capacity is realized. For us, capacity is an intrinsic part of the language comprehension system, not a separately modulated resource. Similarly, domain-specific capacity is not an advantage for their account. Caplan and Waters use this term to mean that different processing domains each have their own capacity, but our proposal captures this idea much more naturally than theirs can. For Caplan and Waters, there are two separate working memories serving interpretive and postinterpretive domains. This and any other two-way division conflicts with abundant data concerning the interactive nature of comprehension processes. By insisting on modules, each with a separate working memory, Caplan and Waters risk a proliferation of little domains and working memories in order to characterize language comprehension. In our account, the various interacting subcomponents of comprehension will necessarily have capacity constraints, because computational capacity is an inherent feature of the models we advocate. Thus, our view accommodates both the domain specificity and the biological endowment that Caplan and Waters insist that we have overlooked, all without having to appeal to large numbers of separate working memories. At the end of the day, what’s left standing is the notion of verbal working memory as emergent from language processing, not outside it.

#### References

- Anderson, J. A., Kline, P., & Lewis, C. (1977). A production system model of language processing. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 271–312). Hillsdale, NJ: Erlbaum.
- Baddeley, A. (1986). *Working memory*. Oxford, England: Oxford University Press.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–90). New York: Academic Press.
- Behrmann, M., Zemel, R. S., & Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal participants and a computational model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1011–1036.
- Bertelson, P. (Ed.). (1987). *The onset of literacy*. Cambridge, MA: MIT Press.
- Bever, T. G. (1970). The cognitive basis for linguistic structure. In J. R. Hayes (Ed.), *Cognitive development of language* (pp. 279–362). New York: Wiley.
- Bock, J. K. (1987). An effect of the accessibility of word forms on sentence structure. *Journal of Memory and Language*, *26*, 119–137.
- Bowey, J. A., Cain, M. T., & Ryan, S. M. (1992). A reading-level design study of phonological skills underlying fourth-grade children’s word reading difficulties. *Child Development*, *63*, 999–1011.
- Caplan, D. (1992). *Language: Structure, processing, and disorders*. Cambridge, MA: MIT Press.
- Caplan, D. (1994). Language and the brain. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1023–1053). New York: Academic Press.



- Caplan, D., & Hildebrandt, N. (1988). *Disorders of syntactic comprehension*. Cambridge, MA: MIT Press.
- Caplan, D., & Waters, G. S. (1995). Aphasic disorders of syntactic comprehension and working memory capacity. *Cognitive Neuropsychology*, *12*, 638–649.
- Caplan, D., & Waters, G. S. (1999a). Age effects on the functional neuroanatomy of syntactic processing in sentence comprehension. In S. Kemper & R. Kliegel (Eds.), *Constraints on language: Aging, grammar, and memory*. Boston: Kluwer.
- Caplan, D., & Waters, G. S. (1999b). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*, 77–94.
- Caplan, D., & Waters, G. S. (2002). Working memory and connectionist models of parsing: A response to MacDonald and Christiansen (2002). *Psychological Review*, *109*, 66–74.
- Carpenter, P. A., Miyake, A., & Just, M. A. (1994). *Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging*. San Diego, CA: Academic Press.
- Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (1999, March). *Structural priming as implicit learning: A comparison of models of sentence production*. Talk presented at the 12th Annual CUNY Conference on Human Sentence Processing, New York.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Christiansen, M. H. (1992). The (non) necessity of recursion in natural language processing. In *Proceedings of the 14th Annual Cognitive Science Society Conference* (pp. 665–670). Hillsdale, NJ: Erlbaum.
- Christiansen, M. H. (2001). *Intrinsic constraints on the processing of recursive sentence structure*. Manuscript in preparation.
- Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*, 157–205.
- Christiansen, M. H., & MacDonald, M. C. (1999). Fractionating linguistic working memory: Even in pebbles, it's still a soupstone. *Behavioral and Brain Sciences*, *22*, 97–98.
- Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, *55*, 75–83.
- Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, *55*, 429–432.
- Dagerman, K. S., MacDonald, M. C., & Harm, M. (2001). *Aging and the use of context in ambiguity resolution*. Manuscript in preparation.
- Daneman, M. E. (1988). Word knowledge and reading skill. In G. E. MacKinnon & M. E. Daneman (Eds.), *Reading research: Advances in theory and practice* (pp. 145–175). San Diego, CA: Academic Press.
- Daneman, M. E., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.
- Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.
- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*, 123–147.
- Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*, 287–314.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in normal and aphasic speakers. *Psychological Review*, *104*, 801–838.
- Ehrlich, S. F., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning & Verbal Behavior*, *22*, 75–87.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.
- Elman, J. L. (1992). *Tlearn simulator* [Computer software]. Available: <http://crl.ucsd.edu/innate/tlearn.html>
- Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 972–992.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.
- Fisk, J. E., & Warr, P. (1996). Age and working memory: The role of perceptual speed, the central executive, and the phonological loop. *Psychology and Aging*, *11*, 316–323.
- Fodor, J. A. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- Folk, J. R., & Morris, R. K. (1995). Multiple lexical codes in reading: Evidence from eye movements, naming time, and oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1412–1429.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Eds.), *Attention and performance XII: The psychology of reading* (pp. 559–586). Hillsdale, NJ: Erlbaum.
- Gibson, E., & Ko, K. (1998, September). *An integration-based theory of computational resources in sentence comprehension*. Talk presented at the Fourth Architectures and Mechanisms in Language Processing Conference, University of Freiburg, Freiburg, Germany.
- Givón, T. (1976). Topic, pronoun, and grammatical agreement. In C. N. Li (Ed.), *Subject and topic* (pp. 151–188). New York: Academic Press.
- Goldman-Rakic, P. S. (1997, April 10). Space and time in the mental universe. *Nature*, *386*, 559–560.
- Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*, 311–347.
- Gottardo, A., Stanovich, K. E., & Siegel, L. S. (1996). The relationships between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third-grade children. *Journal of Experimental Child Psychology*, *63*, 563–582.
- Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language*, *59*, 267–333.
- 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.
- Harm, M., & Seidenberg, M. S. (1999). Reading acquisition, phonology, and dyslexia: Insights from a connectionist model. *Psychological Review*, *106*, 491–528.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 193–225). San Diego, CA: Academic Press.
- Hayes, D. P. (1988). Speaking and writing: Distinct patterns of word choice. *Journal of Memory and Language*, *27*, 572–585.
- Hess, T. M., & Higgins, J. M. (1983). Context utilization in young and old adults. *Journal of Gerontology*, *38*, 65–71.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: The University of Michigan Press.
- Holmes, V. M., & O'Reagan, J. K. (1981). Eye fixation patterns during the reading of relative clause sentences. *Journal of Verbal Learning and Verbal Behavior*, *20*, 417–430.
- Jared, D., & Seidenberg, M. S. (1991). Does word recognition proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General*, *120*, 358–394.
- Juliano, C., & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research*, *23*, 459–471.
- Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Newton, MA: Allyn & Bacon.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *98*, 122–149.
- Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory

- of comprehension: New frontiers of evidence and arguments. *Psychological Review*, 103, 773–780.
- Just, M. A., & Thibadeau, R. H. (1984). Developing a computer model of reading times. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 349–364). Hillsdale, NJ: Erlbaum.
- Just, M. A., & Varma, S. (2002). A hybrid architecture for working memory: Reply to MacDonald and Christiansen (2002). *Psychological Review*, 109, 55–65.
- Keir, J. A., & Duffy, S. A. (1999). The role of phonology in error recovery during reading comprehension. *Abstracts of the Psychonomic Society*, 4, 70.
- Kemper, S. (1992). Language and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 213–270). Hillsdale, NJ: Erlbaum.
- 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. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.
- Kwasny, S. C., & Faisal, K. A. (1990). Connectionism and determinism in a syntactic parser. *Connection Science*, 2, 63–82.
- Leather, C., & Henry, L. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology*, 58, 88–111.
- Light, L. L. (1992). The organization of memory in old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 111–166). Hillsdale, NJ: Erlbaum.
- Long, D. L., Oppy, B. J., & Seely, M. R. (1997). Individual differences in readers' sentence- and text-level representations. *Journal of Memory and Language*, 36, 129–145.
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24, 56–98.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Mann, V. A., Cowin, E., & Schoenheimer, J. (1989). Phonological processing, language comprehension, and reading ability. *Journal of Learning Disabilities*, 22, 76–89.
- Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities*, 17, 592–599.
- Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, 24, 641–682.
- Martin, R. C. (1995). Working memory doesn't work: A critique of Miyake et al.'s capacity theory of aphasic comprehension deficits. *Cognitive Neuropsychology*, 12, 623–636.
- Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, 33, 83–111.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McClelland, J. L., & Kawamoto, A. (1986). Mechanisms of sentence processing: Assigning roles to constituents. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processes: Explorations in the microstructure of cognition* (Vol. 2, pp. 272–325). Cambridge, MA: MIT Press.
- McCutchen, D., Dibble, E., & Blount, M. M. (1994). Phonemic effects in reading-comprehension and text memory. *Applied Cognitive Psychology*, 8, 597–611.
- Micco, A., & Masson, M. J. (1992). Age-related differences in the specificity of verbal encoding. *Memory & Cognition*, 20, 244–253.
- Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorder: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology*, 11, 671–717.
- Miyake, A., Carpenter, P. A., & Just, M. A. (1995). Reduced resources and specific impairments in normal and aphasic sentence comprehension. *Cognitive Neuropsychology*, 12, 651–679.
- Molfese, D. L., & Molfese, V. J. (1997). Discrimination of language skills at five years of age using event-related potentials recorded at birth. *Developmental Neuropsychology*, 13, 135–156.
- Montgomery, J. W. (1995). Sentence comprehension in children with specific language impairment: The role of phonological working memory. *Journal of Speech & Hearing Research*, 38, 187–199.
- Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review*, 104, 686–713.
- Navon, D. (1984). Resources—A theoretical soup stone? *Psychological Review*, 91, 216–234.
- Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4(2), 135–183.
- Patterson, K. E., Graham, N., & Hodges, J. R. (1994). The impact of memory loss on phonological representations. *Cognitive Neuroscience*, 6, 57–69.
- Patterson, K. E., Seidenberg, M., & McClelland, J. (1989). Dyslexia in a distributed, developmental model of word recognition. In R. Morris (Ed.), *Parallel distributed processing*. Oxford, England: Oxford University Press.
- Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, 34, 521–542.
- Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, 12, 765–805.
- Plaut, D. C., McClelland, J. L., Seidenberg, M., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115.
- Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 148–162.
- Richardson, J. T. E., Engle, R. W., Hasher, L., Logie, R. H., Stoltzfus, E. R., & Zacks, R. T. (1996). *Working memory and human cognition*. Oxford, England: Oxford University Press.
- Roth, F. P. (1984). Accelerating language learning in young children. *Child Language*, 11, 89–107.
- Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processes: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.
- Salthouse, T. A. (1994). Aging associations: Influence of speed on adult age differences in associative learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1486–1503.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 10, 403–428.
- Sanford, A. J., & Garrod, S. (1981). *Understanding written language: Explorations in comprehension beyond the sentence*. Chichester, England: Wiley.
- Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, 19, 1–30.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Sergent, J. (1994). Cognitive and neural structures in face processing. In A. Kertesz (Ed.), *Localization and neuroimaging in neuropsychology*:

- Foundations of neuropsychology* (pp. 473–494). San Diego, CA: Academic Press.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 25, 4–27.
- Shankweiler, D., Crain, S., Brady, S., & Macaruso, P. (1992). Identifying the causes of reading disability. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 275–305). Hillsdale, NJ: Erlbaum.
- Shankweiler, D., Crain, S., Katz, L., Fowler, A., Liberman, A., Brady, S., et al. (1995). Cognitive profiles of reading-disabled children: Comparison of language skills in phonology, morphology, and syntax. *Psychological Science*, 6, 149–156.
- Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, 33, 5–42.
- St. John, M., & Gernsbacher, M. A. (1998). Learning and losing syntax: Practice makes perfect and frequency builds fortitude. In A. F. Healy & L. E. Bourne Jr. (Eds.), *Foreign language learning: Psycholinguistic experiments on training and retention* (pp. 231–255). Mahwah, NJ: Erlbaum.
- Stanovich, K. E. (1992). Speculations on the causes and consequences of individual differences in early reading acquisition. In L. C. Ehri & P. B. Gough (Eds.), *Reading acquisition* (pp. 307–342). Hillsdale, NJ: Erlbaum.
- Stoltzfus, E. R., Hasher, L., & Zacks, R. T. (1996). Working memory and aging: Current status of the inhibitory view. In J. T. E. Richardson, W. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, & R. T. Zacks (Eds.), *Working memory and human cognition* (pp. 66–88). Oxford, England: Oxford University Press.
- Tabor, W., Juliano, C., & Tanenhaus (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, 12, 211–271.
- Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication*. San Diego, CA: Academic Press.
- Thibadeau, R. H., Just, M. A., & Carpenter, P. A. (1982). A model of the time course and content of reading. *Cognitive Science*, 6, 157–203.
- Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73, 89–134.
- Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound and reading. *Memory & Cognition*, 15, 181–198.
- Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, 103, 761–772.
- Waters, G., Caplan, D., & Hildebrandt, N. (1991). On the structure of verbal short-term memory and its functional role in sentence comprehension: Evidence from neuropsychology. *Cognitive Neuropsychology*, 8, 81–126.
- Weckerly, J., & Elman, J. (1992). A PDP approach to processing center-embedded sentences. In *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society* (pp. 414–419). Hillsdale, NJ: Erlbaum.
- Wingfield, A., Poon, L. W., Lombardi, L., & Lowe, D. (1985). Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time. *Journal of Gerontology*, 40, 579–585.

## Appendix

The probabilistic context-free grammar used in the simulations is presented in Table A1. The grammar resembles that used by Elman (1990), save that we included past tense verbs, a determiner, and associated explicit probabilities with each rule expansion (as indicated by the numbers in parentheses in Table A1). All nouns were equibiased as were the verbs, with the exception that the three forms of the optionally transitive verbs *phone* and *understand* occurred about twice as often as other verbs because they were allowed in both V(i) and V(t) constructions. No semantic constraints were imposed on the grammar. The grammar produced sentences such as

Table A1  
*The Probabilistic Context-Free Grammar Used to Generate the Training Corpora*

---

S	→ NP VP “.” (1.0)
NP	→ det N (.95) det N rel (.05)
VP	→ V(i) (.5) V(t) (.5)
rel	→ <i>that</i> VP (.5)  <i>that</i> NP V(t) (.5)
N	→ { <i>lawyer, lawyers, senator, senators, reporter, reporters, banker, bankers, judge, judges</i> }
V(i)	→ { <i>lies, lie, lied, hesitates, hesitate, hesitated, phones, phone, phoned, understands, understand, understood</i> }
V(t)	→ { <i>praises, praise, praised, attacks, attack, attacked, phones, phone, phoned, understands, understand, understood</i> }
det	→ { <i>the</i> }

---

*Note.* Values in parentheses indicate the probability of a particular rule expansion. S = sentence; NP = noun phrase; VP = verb phrase; rel = relative clause; N = noun; V(i) = intransitive verb; V(t) = transitive verb; det = determiner.

*The lawyer lied.*

*The senator understands the bankers.*

*The reporters praise the judge that attacked the senator.*

*The reporter that attacked the senator praised the judge.*

*The reporter that the senator attacked praised the judge.*

*The banker that the lawyer that the judge phoned understood hesitated.*

The grammar in Table A1 was used to generate 10 different training corpora, each containing 10,000 sentences, to simulate the fact that language learners are not exposed to the exact same input. Because the generation of sentences was probabilistic given the constraints imposed by the grammar, the number of words in each corpus varied. To provide all nets with exposure to the same number of words, we set an epoch to correspond to 55,000 words (independent of the actual number of words in a given corpus).

The SRNs used in the simulations had 31 input and output units (see Figure 1), one for each of the 30 words and one designating the end of a sentence, as well as 60 hidden and 60 context units. The weights of all nets were initialized randomly within the interval [−.15, .15] prior to training. We gave each network in a set different random starting values to simulate individual differences in initial conditions prior to learning. Thus, the simulations involved a set of 10 SRNs, each with different initial weights and each trained on a different corpus. The learning rate was set to .1, and no momentum was used. Cross-entropy was used to calculate the error used by the back-propagation algorithm. The simulations were conducted using Elman’s (1992) Tlearn simulator available from Center for Research on Language at the University of California, San Diego.

To evaluate the extent to which a network has learned a grammar after training, performance on a set of test sentences is measured. For each word in the test sentences, a trained network should accurately "predict" the next possible words in the sentence; that is, it should activate all and only the words that produce grammatical continuations of that sentence. Moreover, the degree of activation of grammatical continuations should correspond to the probability of those continuations in the training set. Our measure of network performance, grammatical prediction error (GPE, Christiansen, 2001; Christiansen & Chater, 1999) assesses all of these facets of network performance. It takes hits ( $H$ ), false alarms ( $F$ ), misses ( $M$ ), and correct rejections into account. The GPE for predicting a particular word was calculated using

$$\text{GPE} = 1 - \frac{H}{H + F + M},$$

where  $H$  and  $F$  consisted of the accumulated activations of the set of units,  $G$ , that were grammatical given the grammar and the set of activated ungrammatical units,  $U$ , respectively:

$$H = \sum_{i \in G} u_i \quad F = \sum_{i \in U} u_i$$

$M$  was calculated as the sum of a proportion of the total activation determined by the correct units' lower-than-expected values:

$$M = \sum_{i \in G} (H + F)m_i,$$

where the potential missing activation,  $m_i$ , for a grammatical unit was calculated as the (positive) discrepancy between that unit's theoretical target activation,  $t_i$ , derived from the probabilistic context-free grammar, and its actual activation,  $u_i$ :

$$m_i = \begin{cases} 0 & \text{if } t_i - u_i \leq 0 \\ t_i - u_i & \text{otherwise.} \end{cases}$$

This calculation is based on the notion that misses are a product of misplaced activation and that they therefore should be calculated as a (misplaced) proportion of the total activation. Note also that implicit in the misses is a penalty for overestimates ( $u_i > t_i$ ) because too much activation in one place results in too little activation in some other place (given that the cross-entropy error training leads to output activations that typically sum to 1).

Thus construed, the GPE provides a measure of how much of the activation for a given item has been placed correctly according to the grammar ( $H$ ) in proportion to the total amount of activation ( $H + F$ ) and the penalty for not activating grammatical items sufficiently ( $M$ ). Although not an explicit part of the above equation, correct rejections are also taken into account under the assumption that they correspond to zero activation for units that are ungrammatical given previous context. The GPE is a very conservative measure of performance; to obtain a perfect GPE score of 0, the network must not only predict all grammatical next items and no ungrammatical ones, but it must also correctly scale those activations according to the probabilistic constraints on lexical classes found in the grammar.

The GPE for an individual word reflects the difficulty that the SRN experienced for that word given the previous sentential context and can be mapped qualitatively onto word reading times, with low GPE values reflecting a prediction for short reading times and high values indicating long predicted reading times. The average GPE across a whole sentence expresses the difficulty that the SRN experienced across the sentence as a whole, and has been found to map onto sentence grammaticality ratings (Christiansen, 2001; Christiansen & Chater, 1999), with low average GPE scores indicating high goodness ratings and high scores reflecting low ratings.

Received September 18, 1997

Revision received April 19, 2000

Accepted April 21, 2000 ■

### Wanted: Your Old Issues!

As APA continues its efforts to digitize journal issues for the PsycARTICLES database, we are finding that older issues are increasingly unavailable in our inventory. We are turning to our long-time subscribers for assistance. If you would like to donate any back issues toward this effort (preceding 1982), please get in touch with us at [journals@apa.org](mailto:journals@apa.org) and specify the journal titles, volumes, and issue numbers that you would like us to take off your hands.