THEORETICAL NOTES


Gloria S. Waters
McGill University

David Caplan
Massachusetts General Hospital

The authors review M. A. Just and P. A. Carpenter's (1992) "capacity" theory of sentence comprehension and argue that the data cited by Just and Carpenter in support of the theory are unconvincing and that the theory is insufficiently developed to explain or predict observed patterns of results. The article outlines an alternative to the capacity theory, according to which the unconscious, obligatory operations involved in assigning the syntactic structure of a sentence do not use the same working memory resource as that required for conscious, controlled verbally mediated processes.

Just and Carpenter (1992) presented a "capacity" theory of language processing that relates verbal working memory to sentence comprehension. Their theory claims that humans have a limited set of processing resources that can be devoted to verbal tasks. According to the theory, these processing resources can be measured by a working memory task—the "reading span" of Daneman and Carpenter (1980), in which a participant must recall the last word of each of a set of sentences that he or she has read or heard. Just and Carpenter claimed that these resources, also known as working-memory capacity, are used in understanding sentences. According to the capacity theory, this working-memory system is needed to compute syntactic structures and to maintain these structures in memory while additional computations are being carried out. Just and Carpenter's evidence for this theory is the alleged finding that individual differences in working-memory capacity are associated with differences in the efficiency of sentence processing.

Just and Carpenter have made an important contribution by emphasizing the need to develop detailed models of the resources required for sentence comprehension. In recent years, psycholinguists have primarily focused on the nature of the representations and computations that are involved in constructing syntactic structures (Altmann & Steedman, 1988; Frazier, 1990; Trueswell, Tanenhaus, & Garnsey, 1994), and there has been very little research on the processing-resource requirements of sentence comprehension. However, although the focus on the resource requirements of sentence comprehension is valuable, we believe that Just and Carpenter's capacity theory is incorrect in its claim that the processing resources used in sentence comprehension are the same as those involved in other verbally mediated tasks and that they can be measured by participants' performance on the Daneman and Carpenter (1980) reading-span task. Our concerns were triggered by the incompatibility of this claim with much neuropsychological data. The neuropsychological results are too extensive to be reviewed in detail here, but they can be summarized briefly as follows.

Many patients with extremely limited verbal short-term memory spans have been shown to be able to use a wide range of syntactic structures to determine the meaning of sentences (Martin, 1993; Waters, Caplan, & Hildebrandt, 1991; see Caplan & Waters, 1990, for review). In stroke victims, limitations in verbal short-term memory are often due to disturbances affecting phonological storage or rehearsal (see articles in Vallar & Shallice, 1990), so it is possible that the integrity of syntactic processing in these patients reflects the integrity of the "central executive" part of the verbal working-memory system in these patients. However, we have shown that patients with Alzheimer's disease, whose verbal short-term and working-memory deficits are due to central executive limitations, also have well-preserved abilities to use syntactic structure to determine sentence meaning (Rochon, Waters, & Caplan, 1994). In fact, these patients, who have Daneman and Carpenter reading spans of 0 to 1, retain the ability to use syntactic structure to determine sentence meaning even when their already limited verbal working memories are rendered even less available for sentence comprehension by the requirement that they maintain a random digit sequence equal to their span in memory (Waters, Caplan, & Rochon, 1995).

These studies examine the sentence processing capacities of patients with low working-memory spans. A second neuropsychological approach is to examine patients who have disorders affecting syntactic processing that are in part due to reduced...
processing resources. If the resources needed for syntactic processing are the same as those used to maintain a verbal memory load, these patients' syntactic processing should be significantly affected by an external memory load. However, we found that the effect of syntactic complexity is not greater under a digit load condition in a group of aphasic patients with disturbed syntactic processing in sentence comprehension (Caplan & Waters, in press).

These neuropsychological results provide evidence that the verbal memory system that is measured by span and by tests like that of Daneman and Carpenter (1980) is not required for many aspects of syntactic processing in sentence comprehension. This conclusion led us to review the evidence that Just and Carpenter (1992) provided in support of their theory. This review indicated (to us) that much of the empirical evidence presented by Just and Carpenter is unconvincing and that the capacity theory makes apparently contradictory predictions about certain aspects of the performance of high- and low-span participants. In this article we present the results of this review. We first discuss the empirical evidence said by Just and Carpenter to support the theory, then turn to the explanatory and predictive power of the capacity model, and then briefly discuss several lines of research that could provide evidence for the possible fractionation of verbal working memory.


Just and Carpenter (1992) present six domains in which they said the performances of different participant groups support the capacity model of sentence comprehension. These are (a) the modularity of syntactic processing, (b) the nature of processing of complex embeddings, (c) age-related differences in syntactic processing, (d) the nature of processing of syntactically ambiguous sentences, (e) the effect of an extrinsic memory load on syntactic processing, and (f) distance effects on the assignment of pronominal reference. We discuss each of these areas in turn, concentrating on the most important claims and findings.


Modularity refers to several properties that some researchers claim are found in the language-processing system, one of which is that language processors only have a limited range of inputs (Fodor, 1983). With respect to syntactic processing, a modular theory maintains that nonsyntactic information does not affect initial syntactic processing but only later processes that integrate syntactic and other types of representations. Just and Carpenter (1992) argued that the appearance of modularity in some experimental results is misleading and that the results of these experiments can best be explained by considering the consequences of different working-memory capacities of different individuals for whether or not syntactic and nonsyntactic information interact in sentence comprehension. The data that Just and Carpenter indicate are relevant to this claim come from participants' processing of "garden-path" sentences, such as Sentence 1.

1. The defendant examined by the lawyer shocked the jury.

Readers and listeners generally initially assign the defendant as the subject of examined in Sentence 1 and have considerable difficulty revising this assignment in favor of one in which the defendant is the subject of shocked and examined by the lawyer is a reduced relative clause. This phenomenon is known as being "led down the garden path" ("garden-pathed," for short). One piece of evidence that garden pathing takes place in Sentence 1 is that participants' initial eye fixations are longer on the phrase that reveals that Sentence 1 is a garden-path sentence (by the lawyer) than on the same phrase in unambiguous, non-garden-path sentences, such as Sentence 2 (Ferreira & Clifton, 1986).

2. The defendant that was examined by the lawyer shocked the jury.

The critical issue regarding modularity turns on Ferreira and Clifton's (1986) finding that participants' fixation durations were also longer on the by phrase in sentences such as 3 than in sentences such as 4.

3. The evidence examined by the lawyer shocked the jury.

4. The evidence that was examined by the lawyer shocked the jury.

In Sentence 1, the defendant is animate, and therefore it is a plausible agent of the verb examined, whereas in Sentence 3, the evidence is inanimate, and therefore not a plausible agent of the verb examined. If participants are able to use the nonsyntactic pragmatic cue of animacy to influence the syntactic analysis of the sentence, the garden-path effect found for sentences such as 1 should not be seen for sentences such as 3. Ferreira and Clifton's (1986) finding that there is evidence that participants are garden-pathed in both Sentences 1 and 3 is consistent with some version of the modularity model.

Just and Carpenter (1992) repeated the Ferreira and Clifton (1986) experiment with two groups of participants—those with high and low working-memory capacity. They claim that the results indicate that the low-span participants performed like those in the Ferreira and Clifton study but that the high-span participants did not. The data on which this conclusion is based are presented in the left panel of Figure 1 (Just & Carpenter, 1992, Figure 1). Just and Carpenter point out that the figure shows that the high-span participants showed longer fixation times on the disambiguating by phrase in reduced relative clause sentences with animate first nouns like Sentence 1 than in reduced relative clause sentences with inanimate first nouns like Sentence 3, whereas this was not true for the low-span participants.

However, this is not the appropriate comparison to make if the object of the analysis is to see whether high- and low-span groups show garden-path effects for sentences that begin with both animate and inanimate nouns. The correct comparisons are the ones that Ferreira and Clifton (1986) undertook—those between reduced and un-reduced sentences with animate first nouns (Sentences 1 and 2) and between reduced and unreduced sentences with inanimate first nouns (Sentences 3 and 4). Comparing fixation durations on the by phrase in Sentence 1 versus 2 and Sentence 3 versus 4 will indicate whether participants are misled as to the initial syntactic assignment of the first noun when the relative-clause marker (who was/that was) is not pres-
ent compared with when it is present and whether the animacy of the first noun in the sentence affects this phenomenon. The Just and Carpenter (1992) data are replotted in Figure 2, which clearly shows that both high- and low-span participants are garden-pathed both for Sentence Type 1 compared with 2 (reduced vs. unreduced animate first noun) and for Sentence Type 3 compared with 4 (reduced vs. unreduced inanimate first noun). These data are consistent with the view that both high- and low-span participants show modularity in syntactic processing (see also Ferreira, 1992, for similar discussion). Just and Carpenter's (1992) discussion of these data is ambiguous. They say that

The modularity explanation does not fit this pattern of results, unless one postulated that the syntactic processing of low-span subjects is modular and the syntactic processing of high-span subjects is interactive. But modularity was construed as a hypothesis about a universal functional architecture, a construal that is violated by a finding of individual differences. (p. 128)

It appears that the term modularity is used here in its standard psycholinguistic sense to refer to the presence of the garden-path effect in both semantically constrained and semantically unconstrained sentences, which bears on whether a syntactic structure is constructed on the basis of syntactic category information even in the face of disconfirming semantic information. As just noted, in this narrow sense of the term modularity, there is evidence for modularity in both high- and low-span participants, and Just and Carpenter's (1992) suggestion that the two groups differ is inaccurate.

However, Just and Carpenter (1992) also write that "this pattern of results is most easily explained in terms of a capacity difference between the two groups of subjects, such that only the high span subjects have the capacity to take the pragmatic information into account" (p. 128). Although this observation does not bear on the issue of modularity as it is reflected in the absence of semantic influences on a syntactic garden-path effect, it does focus on another possible sense of the term modularity—namely, the use of semantic information in addition to syntactic information to determine sentence meaning. In this respect, the data do show that high- and low-span participants used noun animacy differently. Figures 1 and 2 show that the high-span participants responded faster to sentences with an inanimate first noun, regardless of whether they had reduced or unreduced relative clauses, whereas the low-span participants did not show differences as a function of noun animacy for sentences with either reduced or unreduced relative clauses.

This observation raises the question of why the high-span

---

1 In fact, these data are not definitive on the issue of modularity. There are many complex considerations about the stimuli in these experiments that make the interpretation of these results difficult (Trueswell, Tanenhaus, & Garnsey, 1994).
participants used the animacy cue differently from low-span participants. Just and Carpenter (1992) claim that high- but not low-span participants had enough processing resources to use the pragmatic cue of animacy as well as syntactic structure to understand these sentences. If this is the correct conclusion to draw, the results would bear on differences in sentence processing that are related to working-memory capacity, even if they do not show that high- and low-span participants differ with respect to the modularity of syntactic processing. They 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. However, this account is inconsistent with other results regarding the use of pragmatic information by high- and low-span participants in the comprehension of relative clauses (King & Just, 1991), for which low- but not high-span participants used pragmatic information to help assign sentence meaning. We shall consider this issue in our discussion of the predictive and explanatory power of the capacity theory later.


Previous research (Wanner & Maratsos, 1978; Waters, Caplan, & Hildebrandt, 1987) has shown that sentences such as 5 (known as object relatives) are more difficult to comprehend than sentences such as 6 (subject relatives).

5. The reporter that the senator attacked admitted the error.
6. The reporter that attacked the senator admitted the error.

Just and Carpenter (1992) claim that individuals with low working-memory capacity have more difficulty than high-span participants in comprehending object relatives but not subject relatives. They also claim that the increase in processing difficulty arises at the portion of the sentence that demands the most capacity.

The evidence cited by Just and Carpenter (1992) in support of this claim is based on work by King and Just (1991, Experiment 1). King and Just measured word-by-word reading times for high- and low-span participants in a task in which they were presented with either one, two, or three sentences and in which half the final sentences were subject relatives and half were object relatives. Participants read the sentences using a self-paced reading procedure and then were required to recall the sentence-final words of the one, two, or three sentences in the set. Sentence comprehension was tested with a true–false probe question following the recall task. Reading times, recall of sentence-final word, and accuracy on the questions were recorded. Just and Carpenter point to the fact that the accuracy data showed that differences between the high- and low-span participants
were greatest on the object-relative sentences as support for the capacity theory. They also claim that the data for reading time support the theory, by showing that differences between high- and low-span participants were greatest in the syntactically critical area of the object-relative sentences. The reading-time data are not strong, however, because reading times included sentences that participants failed to understand (about 20% of the trials in the one-sentence condition), and no statistical analyses were reported to support the contention that low-span participants had longer reading times on object relatives than on subject relatives or that this increase in reading time was greatest on the region of the sentence where there is the greatest processing load.\(^2\)

Moreover, other results of King and Just's (1991) experiments raise questions about the role of a general-purpose working-memory in syntactic processing. King and Just divided their low-span participants into groups of comprehenders, whose overall comprehension was above chance, and noncomprehenders, whose overall comprehension was at chance. They argued that differences in sentence-comprehension accuracy between the low-span comprehenders and noncomprehenders were smaller for the easier subject relatives (noncomprehenders = 70%, comprehenders = 80%) than for the more difficult object relatives (noncomprehenders = 51%, comprehenders = 81%). Because the low-span comprehenders and noncomprehenders had roughly equivalent working-memory capacities, any difference in processing resources that made for different performances in the two groups must occur in a processing resource system that is not adequately measured by the Daneman and Carpenter (1980) task.

In addition, as with the use of pragmatic information by high-span participants discussed earlier, other results have failed to find that low-span participants have more difficulty with more complex sentences than high-span participants. MacDonald, Just, and Carpenter (1992) and Waters and Caplan (1996d) both found no differences in high- and low-span participants' abilities to process syntactically complex garden-path sentences (see Section 4 later). Thus, although the King and Just (1991) results provide some support for the capacity theory, when additional analyses of these results and additional experiments are examined, the evidence that low-span participants have more difficulty than high-span participants with syntactically more complex sentences is unconvincing.

3. Age-Related Differences in Syntactic Processing (Just & Carpenter, 1992, pp. 128–129)

Just and Carpenter (1992) claim that data from older adult participants provide evidence for a relationship between working-memory capacity and sentence-processing abilities. They cite three pieces of data in support of the capacity theory: (a) Older adults show greater deficits when they must make inferences that require integration across several sentences (Cohen, 1979); (b) older participants are more impaired than younger participants when they must repeat sentences whose syntax makes larger demands on working memory (Kemper, 1986); and (c) there are age-related differences in spontaneous production of complex syntactic forms (Kemper, 1988; Kynette & Kemper, 1986). None of these studies provide strong support for the capacity model.

The Cohen (1979) study suggests that a general-purpose working-memory system may be related to the ability to draw inferences. This finding is consistent with several other studies that suggest that a general purpose, verbal-working-memory system is involved in making inferences and other verbally mediated tasks (e.g., Kyllonen & Christal, 1990), but it does not bear on the role of a general purpose, verbal-working-memory system in structuring sentences syntactically.

There are several problems associated with citing the studies by Kemper (1986, 1988; Kynette & Kemper, 1986) as evidence that individual differences in verbal working memory make for differences in the efficiency of sentence comprehension. First, Kemper's studies involved sentence production, not sentence comprehension. Daneman and Green (1986) have argued that sentence production involves a separate resource system from sentence comprehension. Second, participants' working-memory capacity was not measured using a Daneman and Carpenter (1980) type of reading-span task in the Kemper (1986) study. (In fact, none of the studies cited by Just and Carpenter, 1992, in the section on aging measure participants' performance on the Daneman and Carpenter task.) Third, the measure of sentence production used in some of Kemper's studies (Kemper, 1988; Kynette & Kemper, 1986) was the frequency of production of complex and simple structures in conversational speech, which is too unconstrained a task to provide reliable data about a participants' ability to produce a particular structure (see discussion in Caplan, 1992, chap. 8, and Caplan & Hanna, in press). The Kemper (1986) study, using repetition, provided the strongest evidence that syntactic processing abilities decline with age. However, because participants' working memory was not measured using a standard test, it is far from clear that this decline occurred in participants with low working memories. Without these data, an interaction between age and sentence type does not provide clear support for the capacity theory.


Just and Carpenter (1992) claim that their theory accounts for differences in how participants process ambiguous sentences. Their claims are based on results reported by MacDonald, Just, and Carpenter (1992), who presented sentences such as 7–10 to high- and low-span participants in a self-paced reading task.

7. The experienced soldiers warned about the dangers before the midnight raid.
8. The experienced soldiers spoke about the dangers before the midnight raid.
9. The experienced soldiers warned about the dangers conducted the midnight raid.

\(^2\) King and Just (1991) stated that "only in the main verb sector of the object relative sentences is there a significant difference in reading times between high and low span readers" (p. 590). However, no statistics are presented to support this statement, and it is not clear whether this statement results from a significant Group × Sentence Type × Region interaction.
10. The experienced soldiers who were told about the dangers conducted the midnight raid.

Sentences 7 and 9 are locally ambiguous, the verb warned could either be the main verb of the sentence, as in the preferred interpretation (Sentence 7), or a participle, as in Sentence 9, which induces a garden-path effect identical to the one seen in Sentence 1 earlier. Sentences 8 and 10 are unambiguous control sentences for Sentences 7 and 9, respectively.

MacDonald et al. (1992) analyzed their results separately for the main verb and participle structures. For the main verb structure, they found that high-capacity readers showed longer reading times than low-capacity readers for sentences such as 7. They claim this was due to high-span participants, but not low-span participants, having enough working-memory capacity to maintain multiple representations of sentences like 7 in mind, which slowed down processing. However, the increase in reading times seen for high-span participants did not occur where the capacity theory predicts—in the sector in which the ambiguous word occurred or on subsequent words—but rather only on the last word of the sentence. In addition, all three groups of participants made more errors on Sentence Type 7 than 8. This would be consistent with both low- and high-span participants maintaining both the main verb and participial structure in memory. We have also been unable to replicate MacDonald et al.'s finding of longer reading times for high-span participants than low-span participants on sentences like 7 using their materials (Waters & Caplan, 1996c; but see Pearlman & MacDonald, 1995, for a replication of the MacDonald et al., 1992, results and related discussion).

Turning to the unpreferred garden-path interpretations, MacDonald et al. (1992) found that participants took longer to read the final words and made more errors in answering questions about Sentence Type 9 than for any other sentence type, confirming the difficulty of these garden-path sentences. Contrary to their predictions, they found that low-capacity participants were not more impaired than high-capacity participants in answering questions about Sentence Type 9 and that both groups of participants had longer reading times on Sentence Type 9 than 10. Low-capacity participants made more errors on Sentence Type 9 than 10, and high-capacity participants took longer to read the words in the final phrases of Sentence Type 9 than 10, but these results were not statistically significant. In addition, the reading time (RT) data were based on a very small number of items per condition. Despite these weaknesses in the database, MacDonald et al. interpreted the pattern of results as being consistent with their theory. They argued that low-capacity participants made more errors on Sentence Type 9 because they did not have the unpreferred syntactic structure available at the end of the sentence, whereas high-span participants took longer to read the words in the final phrases of Sentence Type 9 because they were maintaining both the preferred and unpreferred syntactic structures throughout the sentence. However, this pattern of results could simply have reflected different speed-accuracy tradeoffs in the two groups, rather than inherent differences in their abilities to process sentences.

MacDonald et al.'s (1992) analysis assumes that both high- and low-capacity readers initially construct both syntactic representations in sentences like 9 and claims that low-capacity readers abandon the unpreferred representation sooner than do high-capacity readers. However, MacDonald et al. claimed that high-capacity readers are not capable of maintaining both syntactic structures indefinitely. If processing is examined late enough, high-capacity readers will have abandoned the unpreferred representation. At that point, high- and low-capacity readers should not differ in their treatment of these garden-path structures. Also, high-capacity participants should show different patterns of performance when processing is measured early and late.

MacDonald et al. (1992) tested these predictions by examining high- and low-capacity readers on sentences like 11 and 12 in which additional words were inserted in the phrase before the disambiguation and comparing them with their shorter counterparts, like Sentences 9 and 10.

11. The experienced soldiers warned about the surprise enemy guerrilla attacks the midnight raid.
12. The experienced soldiers who were warned about the surprise enemy guerrilla attacks the midnight raid.

They claimed that there was an increase in reading times for the last word of short sentences such as 9 but not long sentences such as 11 for high-capacity participants, but an analysis of variance (ANOVA) on the reading-time data showed that the four-way interaction between participant group, ambiguity, length, and region was only marginally significant by participants ($p < .07$). (No item statistics are presented, presumably because of the very small number of sentences in each condition.) The predicted difference between the error rates of high-capacity participants on the shorter sentences (9) and the longer sentences (11) was not found. These results provide at best weak support for the claim that garden path sentences become more difficult for high-capacity readers only up to a point at which they abandon the effort to maintain two syntactic structures in working memory. MacDonald et al. found no increase in errors for the low-capacity participants compared with high-capacity participants for sentences such as 11. This is consistent with their theory that high-capacity participants abandoned the unpreferred syntactic representation by the end of these long sentences. However, given that high- and low-span participants also did not have different error rates for the shorter garden-path sentences, such as 9, this result is also consistent with the simple view that high- and low-span participants do not process these sentences differently.

Overall, MacDonald et al.'s (1992) results provide little evidence to support the view that participants with different working-memory capacities perform differently in processing garden-path sentences. Other data also have failed to show differences between low- and high-span participants in processing garden-path sentences. Using a sentence acceptability judgment task with both whole-sentence visual presentation and three different rapid serial visual presentation (RSVP) rates, we failed to find differences between high- and low-span participants' abilities to comprehend three different types of garden-path structures, including the reduced relative clause structures tested by MacDonald et al. (Waters & Caplan, 1996d). As noted earlier, the failure to find reliable differences in the abilities of high- and low-span participants to process these syntactically complex sentences makes the King and Just (1991) find-
ing that there were group differences in processing object relative clauses hard to interpret.

5. The Effect of an Extrinsic Memory Load on Syntactic Processing (Just & Carpenter, 1992, pp. 132–133)

Just and Carpenter (1992) claim that support for their theory comes from studies in which participants are required to perform a sentence comprehension task while concurrently maintaining an extrinsic memory load. Just and Carpenter claim that the results of the King and Just (1991) experiment show that participants with low working-memory spans as measured on the Daneman and Carpenter (1980) test perform less well on more complex sentences and that this effect is greater under concurrent memory load conditions. As described earlier, King and Just had participants read object-relative and subject-relative sentences word by word in a self-paced task while retaining one, two, or three sentence-final words in memory. Just and Carpenter claim that low-span participants were more affected than high-span participants by the addition of larger external memory loads, particularly on syntactically more complex sentence structures.

However, this claim does not correspond to the data reported by King and Just (1991). In the recall task, King and Just reported significant interactions between group and size of memory load and between sentence type and size of memory load, but no interaction between group, sentence type, and memory load. In the comprehension probe results, there was a statistically significant main effect of group; the effect of sentence type was marginally significant, but none of the interactions involving memory load approached statistical significance. No statistics are presented to support the claim that low-span participants had longer reading times on object relatives than on subject relatives or that reading times increased to a greater extent in low- than in high-span participants for the more complex sentences under a greater memory load. Overall, the analyses reported in King and Just (1991) do not show the interactions between group, sentence type, and load that Just and Carpenter (1992) claim as support for the capacity theory.

Aside from the King and Just (1991) results, Just and Carpenter (1992) also argue that data from Carpenter and Just (1988) support the position that low-span participants are less able to maintain an external memory load when the sentence comprehension task includes “more difficult” sentences. Carpenter and Just (1988) investigated the number of words that high-, medium-, and low-span participants recalled in a reading-span task similar to that of Daneman and Carpenter (1980) when the sentences were either “easy” or “difficult.” All participants performed the task at set sizes of 2, 3, and 4 under conditions in which the stimulus materials contained 0, 1, or all difficult sentences. High-span participants were only influenced by difficult sentences at Set Size 4, whereas medium-span participants were affected by difficult sentences at set sizes of 3 and 4. However, in the absence of a documented three-way interaction between span group, set size, and sentence type, it is impossible to know whether high- and medium-span participants were differentially affected at either Set Size 3 or 4 by sentence complexity. In addition, contrary to the predictions of the capacity theory, the performance of low-span participants was unaffected by sentence difficulty at any set size. Just and Carpenter claim that the low-span participants failed to show the predicted effect because their performance was essentially at floor level, but this was clearly not the case at Span Size 2, where they were able to recall at least 80% of the sentence-final words. Thus, again, the data provide no evidence for the capacity theory and may in fact contradict it.

6. Distance Effects (Just & Carpenter, 1992, pp. 133–134)

Just and Carpenter’s (1992) final claim is that working-memory capacity is an important determinant of the ability to find the antecedent for a pronoun. They cite data by Daneman and Carpenter (1980) as support for this claim. Daneman and Carpenter reported an effect of working-memory span on the ability to determine the antecedent of a pronoun. They presented passages in which a set of items was mentioned and in which a pronoun in the final sentence of the passage was related to one of the previous nouns that were separated from the pronoun by a variable number of sentences. Daneman and Carpenter found a strong correlation between a participant’s reading span and the distance in sentences over which they could find the correct antecedent of the pronoun.

Just and Carpenter (1992) cite this result as evidence for the role of working memory in the assignment of pronounal reference. We agree that this experiment provides evidence in favor of such a role for working memory; however, the question that remains is, What does the process of assignment of pronounal reference consist of? Just what aspects of language processing, as seen in the assignment of pronounal reference, are related to working memory?

A detailed description of how antecedents are assigned to pronouns is beyond the scope of this review. However, in brief, the relationship between pronouns and their antecedents is determined by three factors. First, the position of a pronoun and of noun phrases within the syntactic structure of a single sentence rules out certain noun phrases as possible antecedents for a pronoun (Chomsky, 1986; Reinhart, 1983). These linguistic structural constraints may operate as part of on-line processing of pronouns (Nicol & Swinney, 1989). Second, there is experimental evidence that pronouns in particular grammatical positions in one sentence are preferentially related to noun phrases in particular positions in an immediately preceding sentence (Gordon, Grosz, & Giliom, 1993; Sheldon, 1974). Both these factors may be considered to be “structural,” in the sense that a pronoun is related to a noun on the basis of aspects of sentence form, not the meaning of a discourse. In addition to these structural factors, a third factor—which we will call inferential—can ultimately determine how a pronoun will be interpreted in a discourse. For instance, in Sentence 13, anyone with knowledge of basketball will know that “he” must refer to “Bill” because taking a jump shot requires having the ball. This knowledge overrides the default interpretation of the pronoun as referring to “John,” which is generated from structural factors (Chang, 1980; Crawley, Stevenson, & Kleinman, 1990; Gordon & Searce, 1995; Sidner, 1979).

13. John passed the basketball to Bill, and he took a jump shot.
In all the discourses presented by Daneman and Carpenter (1980), a pronoun was related to a noun that occurred more than one sentence previously in the discourse. Because of this, structural factors are not likely to have been operative. In addition, because of the nature of the discourses, the determination of the antecedent of the pronoun is likely to have been made by a reasoning or inferencing process. Therefore, the correlation between working memory and the ability to assign the reference of pronouns reported in that study may reflect a role that working memory plays in inferencing (see also discussion of age effects, earlier).

The Predictive and Explanatory Power of Just and Carpenter’s (1992) Model

The preceding discussion reveals significant weaknesses in the empirical database that Just and Carpenter (1992) cite in support of the capacity theory. Our review of their article also indicated that the capacity theory is too underspecified to generate clear predictions or to provide explanations for several reported patterns of results. We shall discuss two studies cited by Just and Carpenter to illustrate the weaknesses in this theory.


Recall that Just and Carpenter (1992) found that high-span participants were influenced by the pragmatic cue of animacy in Sentences 1–4 (repeated here for ease of exposition), whereas low-span participants were not.

1. The defendant examined by the lawyer shocked the jury.
2. The defendant that was examined by the lawyer shocked the jury.
3. The evidence examined by the lawyer shocked the jury.
4. The evidence that was examined by the lawyer shocked the jury.

Although the results do not show that high- and low-span participants differ with respect to the modularity of syntactic processing (in the narrow sense of modularity), for the reasons discussed earlier, they do indicate that there are differences in the use of pragmatic information in sentence comprehension in groups of participants with different working-memory capacities. Does the capacity theory account for these differences?

As discussed earlier, Just and Carpenter (1992) indicate that these data show that high-span participants have sufficient working-memory capacity to use a pragmatic cue in sentence comprehension, whereas low-capacity participants do not. However, this is simply a post hoc interpretation of these results in terms of the theory. To explain these results, Just and Carpenter would need to formulate clear analyses of what the processing requirements are of using pragmatic and syntactic information in understanding a given sentence and of how the measurement of working-memory capacity is related to those requirements. In the absence of such analyses, the capacity theory can be invoked to account for a variety of results.

With respect to the differences in the use of animacy cues by high- and low-span participants, the results reported by Just and Carpenter (1992) for Sentences 1–4 are opposite to effects claimed to be found in King and Just (1991, Experiment 2) regarding the effect of a different pragmatic cue—semantic plausibility—on high- and low-span participants’ comprehension. They studied the ability of high- and low-span participants to comprehend object-relative sentences that contained verbs that either did or did not provide strong pragmatic cues as to which of the two potential actors in the sentence was the agent of a given verb. Participants were tested on the four sentence types shown in 14–17, in which either both verbs were pragmatically biased (Sentence 14), only the relative clause verb was biased (Sentence 15), only the main clause verb was biased (Sentence 16), or neither verb was biased (Sentence 17).

14. The robber that the fireman rescued stole the jewelry.
15. The robber that the fireman rescued watched the program.
16. The robber that the fireman detained stole the jewelry.
17. The robber that the fireman detained watched the program.

Comprehension was tested with probes of either the relative clause or the main clause. Although none of the relevant interactions were significant, from visual inspection of the data, King and Just claimed that, in the relative clause condition, the comprehension of low-span but not high-span participants improved in the pragmatic bias condition.

If we accept King and Just’s (1991) analysis, the findings in this experiment directly contradict those of the Just and Carpenter (1992) experiment on garden-path sentences with respect to whether low- or high-span participants use pragmatic information in the assignment of sentence meaning. Nonetheless, both patterns of results can be “explained” by the capacity theory. In the case of garden-path sentences 1–4, Just and Carpenter claim that individuals with low working memories for language do not have sufficient working-memory capacity to keep both syntactic and pragmatic information activated. As a result, the sentence comprehension of low-span participants is not influenced by nonsyntactic information. In the object-relative case, King and Just (1991) claim that pragmatic cues relieve low-span participants of the need to process syntactic cues and so aid comprehension. Just and Carpenter are thus claiming that high-span participants have enough resources to use pragmatic information in a sentence like 4, The evidence that was examined by the lawyer shocked the jury; but not in a sentence like 14, The robber that the fireman rescued stole the jewelry; whereas low-span participants use pragmatic information in sentences like 14 but not 4.

On the surface, these claims appear ad hoc, post hoc, and inconsistent with one another. However, they need not be mutually inconsistent. All the patterns of performance described in the two experiments could result from different degrees of reduction in processing resources. This could be tested in a simulation of sentence comprehension, such as the CCREADER model developed by Just and Carpenter (1992). In CCREADER, both maintaining items in memory and constructing new structures requires processing resources. If more processing resources are required than are available at a given point in time, CCREADER can either simplify the structures it holds in memory or undertake fewer computations, or both. This increases error rates and time spent in processing certain regions of a sentence. Just and Carpenter report a variety of simulations in which the resources allocated to CCREADER are reduced, which they claim mimic the empirical data in different experi-
ments. However, Just and Carpenter do not present the results of simulations across the entire set of experiments. It is unclear whether the degree of reduction in processing resources that was used to simulate different span groups' performance on object-relative sentences would produce the observed pattern of results for the garden path stimuli.

CC-READER has the strength of providing a means of exploring the effect of reduction in processing resources on sentence comprehension. However, as it stands, CC-READER is only a model of sentence comprehension, not a model of the relation between general working-memory capacity and sentence comprehension. For it to become a model of this relationship, it is necessary to specify how a given span size is related to the amount of processing resources available in CC-READER. Once that is done, we can determine whether the model is capable of relating performance on the Daneman and Carpenter (1980) task to sentence comprehension performances across different tasks. This would go a considerable way toward allowing the capacity model to explain empirical results.

2. Predictive Power: Predicting Results Pertaining to Syntactic Ambiguity (Just & Carpenter, 1992, pp. 130–132)

The capacity theory is also imprecise in generating predictions. Consider again the experiments reported by MacDonald et al. (1992), in which participants were tested on Sentences 7–10:

7. The experienced soldiers warned about the dangers before the midnight raid.
8. The experienced soldiers spoke about the dangers before the midnight raid.
9. The experienced soldiers warned about the dangers conducted the midnight raid.
10. The experienced soldiers who were told about the dangers conducted the midnight raid.

MacDonald et al. generated predictions for the performance of high- and low-span participants on these sentences in the following way. According to MacDonald et al., all readers initially construct all possible syntactic structures that are compatible with any incoming sequence of words. The activation level of each representation depends on its frequency, syntactic complexity, pragmatic plausibility, and other factors. Readers with smaller working-memory capacities cannot maintain more than one syntactic representation in memory for more than a short period.

From these assumptions, MacDonald et al. (1992) made the counterintuitive prediction that high-capacity readers will show longer reading times than low-capacity readers in the interpretation of locally ambiguous sentences that are resolved in favor of the preferred syntactic structure (Sentence 7). This prediction is based on the idea that high-span but not low-span participants maintain both syntactic structures for these sentences in memory and that maintaining two representations in memory requires additional processing resources and adds to processing time. However, MacDonald et al. could have made the opposite prediction on the basis of the capacity model. They could have just as easily argued that high-span participants have sufficient capacity to maintain two interpretations without any difficulty, whereas low-span participants incur processing difficulties by trying to maintain two representations. On this view, low-span participants would have been expected to have more trouble with Sentence Type 7 than would high-span participants.

MacDonald et al.'s (1992) predictions for differences between low- and high-capacity readers with respect to the time needed to process garden-path structures such as Sentence 9 are not straightforward. They say that low-capacity readers require more time to process garden-path sentences than unambiguous sentences with identical syntactic structures because they need time to reconstruct the unpreferred syntactic structure that they have abandoned. High-capacity readers do not need to do this reconstruction, but they should the ongoing processing burden of maintaining more than one syntactic representation until a disambiguation occurs, which adds to the time they require to process these sentences. It is not clear which of these two factors—the need to reconstruct an unpreferred syntactic structure or the maintenance of several syntactic structures—has a greater effect on processing time, and therefore MacDonald et al. made no prediction about whether one type of reader will show longer reaction times than the other in interpreting garden-path sentences correctly.

Thus, the capacity model could have predicted either faster or slower response times for high-span participants both for garden-path sentences and for sentences that are resolved in favor of the preferred interpretation (see also Gathercole & Baddeley, 1993, pp. 222–226). The reason why the model is overly flexible is because it contains too many free parameters. The point at which the processing load associated with sentence processing becomes too great for a participant with a given span, the point at which a participant drops a representation because of limitations in processing resources, and the processes that arise after a representation has been dropped are all inadequately specified in the current version of CC-READER. Future development of these aspects of CC-READER would help to relate it to general purpose, working-memory capacity.

General Discussion

Just and Carpenter (1992) argue that individual differences in working-memory capacity, measured on the Daneman and Carpenter (1980) reading-span task, are related to individual differences in sentence processing. Our review of the evidence they provide in support of their capacity model indicates that it rests on an unconvincing empirical foundation. In addition to the problems with the empirical support for this model, there are theoretical reasons for questioning the view that sentence processing uses the same resource pool that is measured by the Daneman and Carpenter task. The Daneman and Carpenter task requires conscious retrieval of items held in memory, to an extent not found in processing sentence structure (Marslen-Wilson, 1975; Marslen-Wilson & Welsh, 1978; Tyler & Marslen-Wilson, 1977). Second, the memory load that is imposed in the Daneman and Carpenter task is unrelated to the computations of that task, whereas the verbal material that is stored in sentence processing is relevant to the ongoing computations required by that task. The Daneman and Carpenter task and the
sentence-comprehension process are thus qualitatively different in ways that add a theoretical impetus to raising the question of whether the former uses and accurately measures the processing resources that are used in the latter.

One alternative to the capacity model is the hypothesis that there are specializations within the verbal-processing resource system for different verbally mediated tasks. Our working hypothesis is that one resource pool is used in obligatory, on-line psycholinguistic operations in the comprehension process and another in controlled, verbally mediated tasks (Caplan & Waters, 1990; Waters et al., 1995; Waters & Caplan, 1996a). We suggest that “obligatory, on-line psycholinguistic operations in the comprehension process” consist of those operations that transform the acoustic signal into a preferred, discourse-coherent, semantic representation. Syntactic parsing is one of a set of operations that achieve this function. Others include acoustic-phonetic conversion, lexical access, assignment of intonational contours, determination of sentential semantic values such as thematic roles, and determination of discourse-level semantic values such as topic and coherent coreference. An intuitive way to think of the system that we are trying to delineate is as the one that yields the meanings of utterances that listeners extract from normal, everyday conversations. The domain of on-line language comprehension contrasts with conscious, controlled, and verbally mediated processes, such as the deliberate search through semantic memory for a piece of information, explicit reasoning, and other tasks, which we postulate utilize a different set of resources.3

There are several ways in which the theory of separate language-processing resources can be experimentally distinguished from the capacity model. One is Just and Carpenter's (1992) approach to the issue—the investigation of the relationship of individual differences in verbal working memory to sentence-processing efficiency. The capacity theory predicts that having a low working-memory capacity will reduce the resources available for sentence processing and make it less efficient; the theory of separate language-processing resources predicts that performance on general verbal working-memory tasks will not predict language-processing efficiency. The many instances in which experimental evidence has failed to reveal differences in the efficiency of sentence processing in low- and high-span participants speak against the capacity model and are consistent with the theory of separate language-processing resources.

A second approach to the issue is to investigate the pattern of mutual interference (or noninterference) of verbally mediated tasks. The capacity model claims that verbal memory loads that are imposed external to the comprehension task (such as a concurrent digit-span task) and sentence comprehension draw on the same pool of resources, and therefore it predicts that there will be interference between the two. The theory of separate language-processing resources claims that “comprehension-external” and “comprehension-internal” factors draw on different resource pools, and therefore it predicts main effects of each but no interference between the two.

Relevant data come from experiments in which participants process sentences that vary in their syntactic structure under task conditions that vary in their memory load. Our review of such studies indicates that whether a concurrent memory load interferes with syntactic processing depends on the nature of the recall and sentence-processing tasks. The largest experience is with the combination of a sentence comprehension and a word- or digit-recall task. When the stimuli in both the sentence and recall tasks have been presented in an uninterrupted fashion, as when the sentence-processing task is a sentence–picture matching task and a digit load is presented prior to each sentence, no interference between these two tasks has been found (Caplan & Waters, in press; Waters et al., 1995; Waters & Caplan, 1996a).4 On the other hand, when the presentation of the stimuli in one task has been interrupted by the presentation of a stimulus relevant to the second, as when the sentence is interrupted by a series of words while it is being presented (Wanner & Maratos, 1978), or the presentation of the to-be-recalled items is staggered across the sentence task (as in sentence-final-word recall; King & Just, 1991; Waters & Caplan, 1996b), a digit load of sufficient size has interfered with processing syntactically complex sentences more than with processing syntactically simple sentences. An important point to note is that the tasks in which an interaction between digit load and syntactic complexity has not been found are not simply so easy that this interaction cannot be detected, because participants performed below ceiling on both the recall task and on at least some sentences types in which the effects of syntactic complexity could be examined (Caplan & Waters, in press; Waters et al., 1995; Waters & Caplan, 1996a). The pattern of results thus suggests that retaining a sequence of digits and struc-

---

3 There is much debate about the boundaries and the nature of the online language comprehension process, and this uncertainty could raise a number of problems for this theory. For instance, some recent work suggests that on-line processing is influenced by nonlinguistic factors such as pragmatic expectations (Trueswell et al., 1994) and the frequency with which particular constructions occur in a language (MacDonald, Pearlmutter, & Seidenberg, 1994). This raises the theoretical question of how a language processor that is not informationally encapsulated can use a processing resource system that is separate from that used in other verbally mediated tasks. Our answer to this question is that, even if the language-processing system is not encapsulated with respect to its input, it may be domain-specific in terms of its output. It computes lexical forms and their meanings, syntactic and prosodic structures and propositional meaning, and coherent discourse structures. The comprehension process may recruit a dedicated resource pool to compute its output regardless of what its input is. Nor does the fact that there are uncertainties about whether certain comprehension processes take place on-line vitiate the theory we have outlined or make it untestable. Despite areas of uncertainty, there is a large set of operations that clearly fall within the postulated on-line comprehension system and an equally large set of operations that fall outside it. For instance, constructing a simple syntactic structure, such as the main verb reading of the sentence types used by MacDonald et al. (The policemen fired into the empty lot), falls within the system, whereas consciously searching semantic memory for the referent of the object of the sentence John recognized the only member of the cat family that spontaneously swims does not.

4 We are aware of only one report of an interaction of digit load with syntactic structure that does not involve interrupting one of two tasks—the written whole-sentence acceptability judgment task reported by Waters et al. (1987). The interaction was only found in the analyses by participants, not items, which raised questions about its reliability. However, the issue of what task combinations produce this interaction remains an important one to explore.
turing a sentence syntactically do not compete for the same resource pool, but rather that the attentional demands associated with interrupting each task interfere with participants' abilities to structure sentences syntactically or to use that structure to assign sentence meaning.

The two approaches—the effects of individual differences and of external load on language-processing efficiency—can be combined. The capacity theory asserts that comprehending more complex sentences, maintaining a larger digit load in memory, and producing a higher score on the Daneman and Carpenter (1980) working-memory task all require more processing resources from the same pool. Thus it predicts that the expected impairment of low-span participants on more complex sentences will be exacerbated by a concurrent memory load. In contrast, the theory of separate language-processing resources holds that maintaining a digit load in memory and performing the Daneman and Carpenter working-memory task call on one resource pool and comprehending sentences calls on another. Therefore, this model predicts that low-span participants should perform less well under conditions of increased verbal memory load, but it does not predict that this effect should be greater for syntactically complex sentences. Again, the available data favor the theory of separate language-processing resources. There are many instances in the literature reviewed earlier in which low-span participants were more impaired than high-span participants by an external memory load (e.g., King & Just, 1991; Waters et al., 1995). There are also many instances in which low-span participants have been more affected than high-span participants when other task demands increased, such as when sentences were presented under RSVP conditions (Miyake, Carpenter, & Just, 1994; Waters & Caplan, 1996d). These findings are consistent with the view that performing the reading-span task and maintaining a verbal memory load, or dealing with other task demands, use the same resource system. However, strikingly, there are no instances in which an extra processing load differentially decreased low-span participants' performance on syntactically more complex sentences. This is consistent with the view that syntactic processing does not compete for the resources that are recruited by these other task features.

All these sets of results are not definitive for many reasons. The failure to find two-way interactions between syntactic complexity and memory load, or between span group and syntactic complexity, or three-way interactions between span group, load, and syntactic complexity could be due to inadequate experimental power, floor and ceiling effects altering the expected patterns of factor interactions, or other causes. However, many of these possible reasons for the absence of these interactions are unlikely. In particular, many experiments have had sufficient power to generate the main effects and interactions that are predicted by the theory of separate language-processing resources but not the interactions of span group, load, and syntactic complexity predicted by the capacity theory and, as noted earlier, have examined participants in whom ceiling-and-floor effects were not present on at least some sentence types (Caplan & Waters, in press; King & Just, 1991; Miyake et al., 1994; Waters et al., 1995; Waters & Caplan, 1996a, 1996d).

In summary, we suggest that a number of lines of research favor a model in which the processing resources that are used in unconscious, on-line language comprehension are at least partially separate from those used in controlled, verbally mediated functions. Whether this is the correct division, whether verbal processing resources are fractionated along some other lines, whether either of these putative resource systems should be further divided, how such a specialization might develop, and other related issues remain open questions.

We are indebted to Robert Bjork for pointing out that assumptions that are made about floor and ceiling effects in participants' working-memory capacities can greatly affect the patterns of factor interactions that a model predicts. For instance, in a simple case, if the combination of task and sentence-processing demands is less than the working-memory resources available to even low-span participants, ceiling effects would be expected to eliminate interactions between load, sentence type, and span group. Much more complex patterns of results could be seen. Consider the capacity theory's prediction that high-span participants' processing of more complex sentences will be less affected than that of low-span participants by an external load. Ceiling effects can alter this prediction. If high-span participants have excess working-memory capacity when dealing with simple sentences in a no-load condition and are at ceiling when sentences become more complex or when a load is imposed, a shift from a simple-sentence-no-load condition to either a complex-sentence-no load or simple-sentence-load condition may have little or no effect on their performance, whereas a shift from either of these conditions to a complex-sentence-load condition may have a substantial detrimental effect. If low-span participants are at their working-memory ceiling in the simple-sentence-no-load condition and are short of working-memory capacity in the others, a shift from the simple-sentence-no-load to either the complex-sentence-no load or simple-sentence-load condition may have large or larger effect than a shift from one of these two conditions to the complex-sentence-load condition. Under these conditions, the high-span participants, but not the low-span participants, might be more affected on the harder sentences by a larger memory load—a three-way interaction that has the opposite pattern of terms from that predicted earlier. Our discussion of the predictions and explanations offered by King and Just (1991), MacDonald et al. (1992), and Just and Carpenter (1992) indicates that they frequently rely on the assumed presence of these types of ceiling (or floor) effects. The point we wish to make is not that such assumptions are incorrect or that this is an issue only for Just and Carpenter's capacity theory, but rather that they may play crucial roles in generating predictions. It is therefore important either to avoid ceiling and floor effects in experimental results or to provide empirical evidence that they are present if their presumed existence is an implicit factor in generating predictions and explaining results.

References


Received May 17, 1994
Revision received October 9, 1995
Accepted October 18, 1995