

## The Capacity Theory of Comprehension: New Frontiers of Evidence and Arguments

Marcel Adam Just, Patricia A. Carpenter, and Timothy A. Keller  
Carnegie Mellon University

A capacity theory of comprehension (M. A. Just & P. A. Carpenter, 1992) has provided an integrated account of several central aspects of sentence comprehension, such as the processing of syntactic ambiguity, complex embeddings, syntactic (non)modularity, and individual differences, in terms of the working-memory capacity for language. Some of the evidence supporting the theory is questioned by G. S. Waters and D. Caplan (1996a). This article identifies some of Waters and Caplan's errors about the empirical support in Just and Carpenter (1992), evaluates Waters and Caplan's alternative hypothesis, and presents the results of a new neuroimaging study that supports capacity theory and not Waters and Caplan's separate resources hypothesis.

This article has the dual goals of refuting some of Waters and Caplan's (1996a) incorrect descriptions concerning the empirical support for capacity theory, as well as pointing out the theoretical and empirical difficulties with Waters and Caplan's alternative hypothesis. This article has three sections: (a) a critique of Waters and Caplan's hypothesis, (b) a new functional Magnetic Resonance Imaging (fMRI) study of the reading-span task that supports the capacity theory and not Waters and Caplan's alternative hypothesis, and (c) a reply to some of Waters and Caplan's inaccuracies concerning the empirical support in Just and Carpenter (1992).

The capacity theory adopts the basic premise that thinking is resource limited (Kahneman, 1973) but proposes a specific model of how the constraint is applied within a particular cognitive architecture and examines the implications in the domain of language comprehension. The theory deals with the resources used to support language comprehension computations, not the phonological buffer/articulatory loop of Baddeley's (1992) theory.

### Critique of Waters and Caplan's (1996a) Hypothesis

Two problems with Waters and Caplan's (1996a) proposed division of working-memory resources is that it conflates two

types of criteria for the division, which are neither individually adequate nor mutually consistent, and it ignores disconfirming evidence. Waters and Caplan (1996a) propose one (*psycholinguistic*) resource pool for on-line psycholinguistic operations, including operations that go from the acoustic signal into a discourse representation: "Syntactic parsing is only one . . . Others include . . . lexical access, . . . determination of sentential semantic values such as thematic roles, and . . . topic and coherent coreference" (p. 770). Waters and Caplan contrast this with the *conscious pool* that supports conscious, controlled, and verbally mediated processes, such as explicit reasoning. Furthermore, Waters and Caplan claim that the reading-span task does not draw on the psycholinguistic pool and hence should be unrelated to individual differences in the processing supported by that pool.

The partition they propose breaks down because, as Waters and Caplan (1996a) concede, psycholinguistic resources must support the use of general, nonspecialized information: "There is much debate about the boundaries and the nature of the on-line language comprehension process, and this uncertainty could raise a number of problems for this [Waters and Caplan's] theory" (p. 770). The problems arise from evidence of effects of "non-linguistic factors, such as pragmatic expectations (Trueswell et al., 1994) and the frequency with which particular constructions occur in a language (MacDonald, Pearlmutter, & Seidenberg, 1994)" (p. 770). Waters and Caplan abandon the definition of a class of "psycholinguistic processes," saying ". . . even if the language-processing system is not encapsulated with respect to its input, it may be domain-specific in terms of its output" (p. 770). But if the processes supported by the psycholinguistic pool can take nonpsycholinguistic information as input and modify their behavior in response, the partition evaporates.

A second problem is that the partition conflates two criteria that often diverge. The two criteria, (a) the type of process (e.g., lexical access) and (b) a process's automaticity (obligatory or automatic vs. conscious or controlled), can diverge because the automaticity of a process varies with circumstances. Although lexical access for a short familiar word may be automatic, it

---

Marcel Adam Just, Patricia A. Carpenter, and Timothy A. Keller, Department of Psychology, Carnegie Mellon University.

This research was supported in part by the Office of Naval Research Contract N00014-96-1-0322, National Institute of Mental Health Grant MH-29617, and Research Scientist Awards MH-00661 and MH-00662. We thank Henk Haarmann, Jonathan King, and Akira Miyake for their comments.

The functional Magnetic Resonance Imaging data were collected with the consultation of Keith R. Thulborn, director of University of Pittsburgh Medical Center Magnetic Resonance Research Center, who contributed his magnetic resonance biophysics and neuroradiological expertise.

Correspondence concerning this article should be addressed to Marcel Adam Just, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. Electronic mail may be sent via Internet to just@psy.cmu.edu.

is not for a very infrequent but known word, which can take hundreds of milliseconds longer in normal reading (Just & Carpenter, 1980). So, which pool supports nonautomatic lexical access? Similarly, which pool supports the slow syntactic processing of a difficult center-embedded, object-relative sentence? A related question concerns the syntactic, lexical, thematic, and discourse processes in verbal problem-solving tasks. Are they the same processes as the ones that are automatic and draw on the psycholinguistic pool or are there two duplicate sets of such processes? The conflated criteria result in inconsistencies and incompleteness.

A third problem is that one of the criteria, automaticity, is treated as a dichotomous variable, making it unable to accommodate the evidence that automatic and controlled processes are not categorically different but define a continuum, varying in speed, parallelism, and difficulty of inhibiting the process.

Finally, Waters and Caplan's (1996a) proposal often generates null predictions specifically, that there should be no individual differences in comprehension or that individual differences should not interact with some other variable. As support, Waters and Caplan cite their own failures to obtain some such effects, without a quantitative analysis indicating that the study had sufficient power. Equally important, Waters and Caplan dismiss reliable individual differences in on-line comprehension, as we discuss in the section Individual Differences in Sentence Processing.

In sum, Waters and Caplan's (1996a) proposal is internally inconsistent. The general concept of a privileged psycholinguistic class of processes has been discredited by evidence showing that sentence comprehension is influenced by factors such as frequency and pragmatic knowledge.

### Brain Imaging in the Reading-Span Task

In this section, we report a brain-imaging study that provides evidence that the reading-span task activates the same brain regions that are involved in sentence comprehension and that tests one of the differentiating predictions of Waters and Caplan's proposal and the capacity theory. One source of evidence for capacity theory is that individual differences in sentence comprehension are related to differences in working-memory capacity, which are typically measured with the reading-span task (Daneman & Carpenter, 1980). The reading-span task requires comprehending a series of sentences while maintaining some active representation of the preceding sentence-final words (without covert rehearsal). Using fMRI, we tested our proposal that this maintenance draws, in part, on the same resources as does sentence comprehension. In a read-only condition, participants silently read a sequence of sentences. In a read-and-maintain condition, they in addition maintained the sentence-final words. Our proposal makes two predictions. First, the two conditions, read-only and read-and-maintain, should activate some of the same brain areas involved in sentence comprehension. Second, the degree of activation in a brain area activated by both conditions should be greater in the read-and-maintain condition because the demand on the common resource pool should be greater.

By contrast, Waters and Caplan (1996a) repeatedly propose that the central processes in the reading-span task and normal

reading draw on two different resource pools (Waters & Caplan, 1996a, pp. 761, 769–770). Waters and Caplan state "The memory load that is imposed in the Daneman and Carpenter task is unrelated to the computations of that task [i.e., the sentence comprehension task]. . ." (p. 769). Waters and Caplan predict no overlap in the brain areas activated by the two task components, so the read-only condition should activate an area or areas associated with sentence processing; the read-and-maintain condition should activate different areas, presumably areas involved in the conscious processing of verbal information.

### Method

During the scanning,<sup>1</sup> participants silently read sets of sentences with instructions to read-only or read-and-maintain for each set. Following each set in the read-and-maintain condition, participants were given time to recall silently the sentence-final words, which they were told would be tested at the end of the session. Each set had either two or four sentences, which was crossed with the two conditions. Each sentence was verified as true or false to ensure comprehension and decrease opportunities to rehearse, and the next sentence in the set appeared immediately. A sample true sentence was "It is known that most people use their right hand to eat, write, or drink." A total of 16 sets was randomly intermixed with at least four occurrences of a fixation condition, in which the participant fixated an asterisk in the middle of the screen for 24 s without performing any task. (The fMRI scanning protocol was identical to that used in other fMRI research on comprehension: Just, Carpenter, Keller, Eddy, & Thulborn, in press). The data analysis was based on 2 three-slice anatomical regions: the left laterosuperior temporal gyrus (Wernicke's area) and the left inferior frontal gyrus (Broca's area).

### Results

Significant activation occurred in the two classical language areas, Wernicke's and Broca's, in both the read-only and the read-and-maintain conditions. Furthermore, in Wernicke's area, the number of activated voxels was greater for the read-and-maintain condition than for the read-only condition for all 4 participants (with means of 17 vs. 9.5 activated voxels, respectively). The number of activated voxels was similar in Broca's area for the two conditions (identical for 3 of 4 participants and no statistical difference overall). As predicted, sentence comprehension (read-only) and the reading-span task (read-and-maintain) activated the same cortical areas, and fur-

<sup>1</sup> The fMRI were gradient echo echo-planar MR images. The 4 right-handed, mid-span volunteers (1 woman) completed informed consent forms approved by the University of Pittsburgh Institution Review Board. The scanner was a 1.5 Tesla General Electric Medical Systems Signa MRI scanner at the University of Pittsburgh Medical Center. Images were acquired of each of seven adjacent axial planes every 1,500 ms using a whole-head coil for 2 participants and two bitemporal 5-inch (12.7 cm) general purpose surface coils for the other 2 participants. The scanning parameters were time to echo (TE) = 50 ms, flip angle = 90°, slice gap = 1 mm.

To prevent any condition from having an undue influence on the imaging results by virtue of its having a longer duration, the data-sampling period was matched by considering only the early portion of each set, *early* defined by the duration of their shortest condition for each participant.

thermore, the reading-span task activated Wernicke's area more than comprehension alone did.

Figure 1 shows significant activation relative to the fixation control in one slice in Wernicke's area (in the rectangle); the white voxels are those that were significantly activated above the activation level of the fixation condition,  $t > 5.75$  (Eddy, Fitzgerald, Genovese, Mockus, & Noll, 1996). The statistical analysis, contrasting the activation levels in the two conditions for those voxels in the region of interest that were activated in either condition (with voxels nested within participants), indicated that the activation amplitude was significantly higher for the read-and-maintain condition compared with the read-only condition in Wernicke's area,  $F(1, 73) = 20.76, p < .001$ .

The same type of quantitative analyses indicated no statistical difference in the activation amplitudes in Broca's area in the two conditions ( $F$  ratios  $< 1$ ), suggesting that the extra demand imposed by the maintenance aspect of the reading-span task (as compared with the read-only condition) did not increase the activation in Broca's area as it did in Wernicke's area. An emerging pattern is that Broca's area is a generator of a speech-based code of the kind used for verbal rehearsal (Awh et al., 1996), whereas the more posterior Wernicke's area (or the adjacent angular and supramarginal gyri) might be a site of verbal computation and storage.

The two cortical areas that were activated in this study overlap with areas that were found activated in other neuroimaging studies of auditory sentence comprehension (Binder et al., 1994) and reading comprehension (Just et al., in press). Also, the increase in activation in Wernicke's area with the demand is consistent with capacity theory and with an independent study of sentence comprehension (Just et al., in press). Third, Wernicke's area has been identified as the site of a "metabolic lesion" that is common to almost all aphasic patients. In two independent labs, 96% of all aphasic patients tested have a common site of metabolic impairment in the left temporal and temporal parietal regions; the impairment is a hypometabolism measured when the patient is at rest with Positron Emission Tomography using ( $F-^{18}$ )-fluorodeoxyglucose, which assesses glucose utilization, an index of neuronal activity. This common impairment is found in spite of wide variation in the clinical category, severity, and the locations of the aphasic patients' structural lesions (Karbe et al., 1989; Metter et al., 1990). Moreover, the degree of the metabolic impairment for an aphasic patient correlates with the degree of impairment in sentence comprehension (measured off-line by standardized tests), with the highest correlations for comprehension and hypometabolism of regions overlapping with the present ones (left latero-superior temporal and middle temporal gyri). The capacity theory provides a mechanism to explain the correlation, namely, that the measure of resting PET activation is an index of the size of the patients' potential resource supply, which could place an upper limit on comprehension.

Thus, the present study supports the argument that the maintenance aspect of the reading-span task draws on processes that overlap with those in sentence comprehension. The data are thus consistent with the numerous behavioral correlations between working-memory span and language comprehension found in a meta-analysis of over 30 published studies (Daneman & Merikle, 1994). The results support the capacity

theory, fail to support Waters and Caplan's (1996a) proposal and illustrate the potential of functional neuroimaging to inform issues at the cognitive level.

### Individual Differences in Sentence Processing

Individual differences in on-line sentence comprehension that correlate with the reading-span task provide one source of support for capacity theory. We briefly consider some of Waters and Caplan's (1996a) errors concerning this evidence.<sup>2</sup>

#### *Comprehending Complex Embedded Sentences*

Waters and Caplan (1996a) incorrectly describe and then dismiss the individual differences (King & Just, 1991) in the word-by-word reading times of high- and low-span individuals reading sentences that vary in structural complexity:

- (1) subject relative, "The reporter that attacked the senator admitted the error after the hearing."
- (2) object relative, "The reporter that the senator attacked admitted the error after the hearing."

Much of the difference in computational demand between the two sentences can be localized to the verb in the main clause (e.g., *admitted*). Figure 2 is reprinted to show (as King and Just, 1991, p. 589, reported) that the reading time on this word showed a reliable effect of sentence (object relatives taking longer than subject relatives),  $F(1, 32) = 23.99, p < .001$ ; a reliable effect of span,  $F(1, 32) = 4.99, p < .05$ ; and importantly, a reliable interaction, such that the low-span readers' 197 ms extra time for the object-relative sentences compared to the subject-relative sentences was greater than the high-span participants' 87 ms extra time, "the predicted interaction between reading span and sentence type, which is indeed reliable ( $F(1, 32) = 4.26, p < .05$ )" (p. 589).

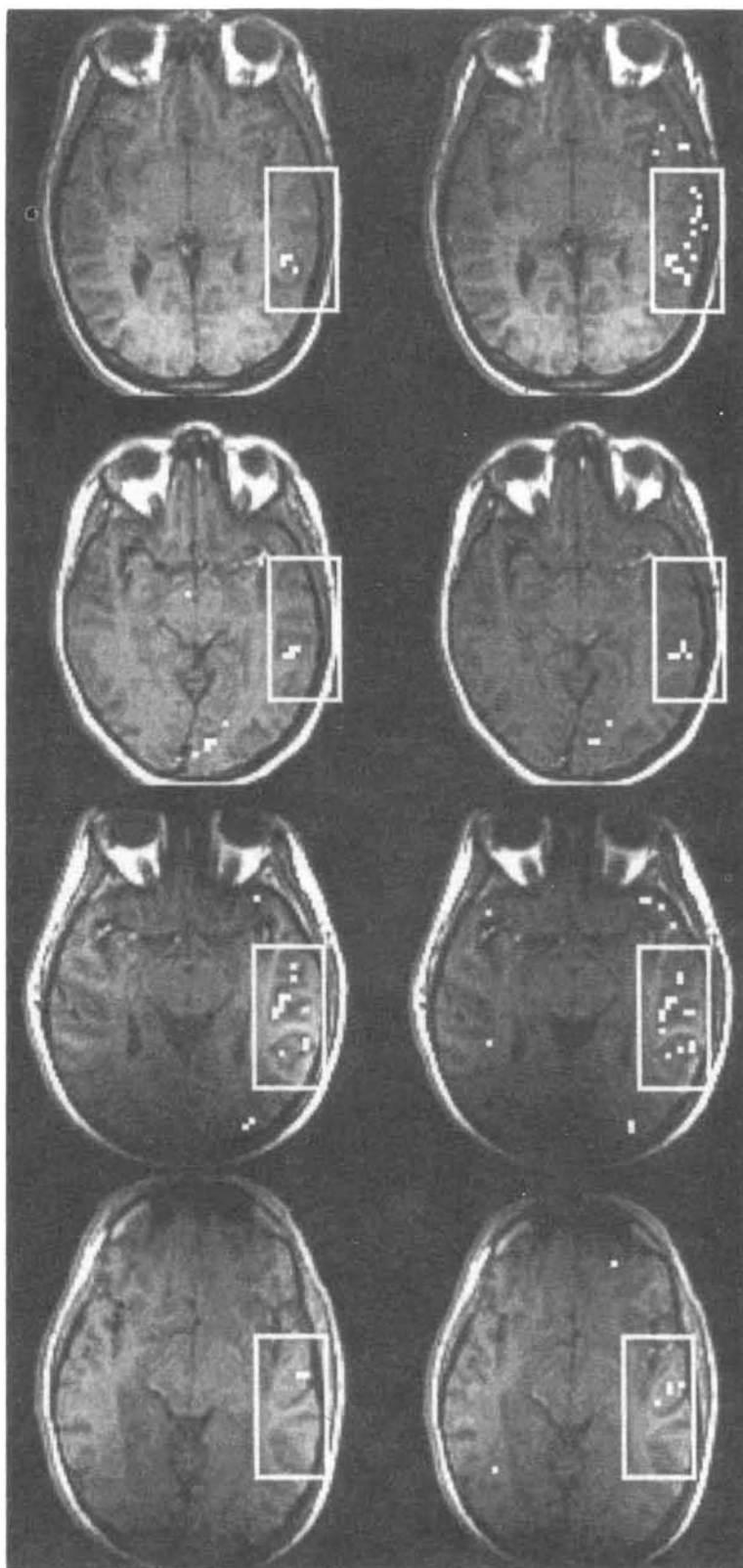
In the face of these reported results, Waters and Caplan (1996a) conclude that "no statistical analyses were reported to support the contention that low-span participants had longer reading times on object relatives than on subject relatives" (pp. 764-765). The reported results support precisely that contention. This fundamental error of Waters and Caplan undermines a central basis for their criticism.

Waters and Caplan (1996a, p. 765) repeatedly refer to these results as unconvincing and make other claims about the data. Waters and Caplan (p. 764) object that our analysis of Experiment 1 (King & Just, 1991) included sentences on which the reader subsequently made errors to comprehension-probe questions. However, as the original report stated, the results did not substantially change when such trials were excluded. Finally, the central result (that low-span readers took longer than high-span readers on the main-clause verb of object-relative sentences) was replicated in Experiment 2 of King and Just (1991), which

<sup>2</sup> This journal's current policy of requesting a review from the authors of a critiqued article was not in force when Waters and Caplan's (1996a) manuscript was accepted, so we had no opportunity to point out the errors in Waters and Caplan's manuscript to the editor prior to its acceptance.

Read-Only

Read-and-Maintain



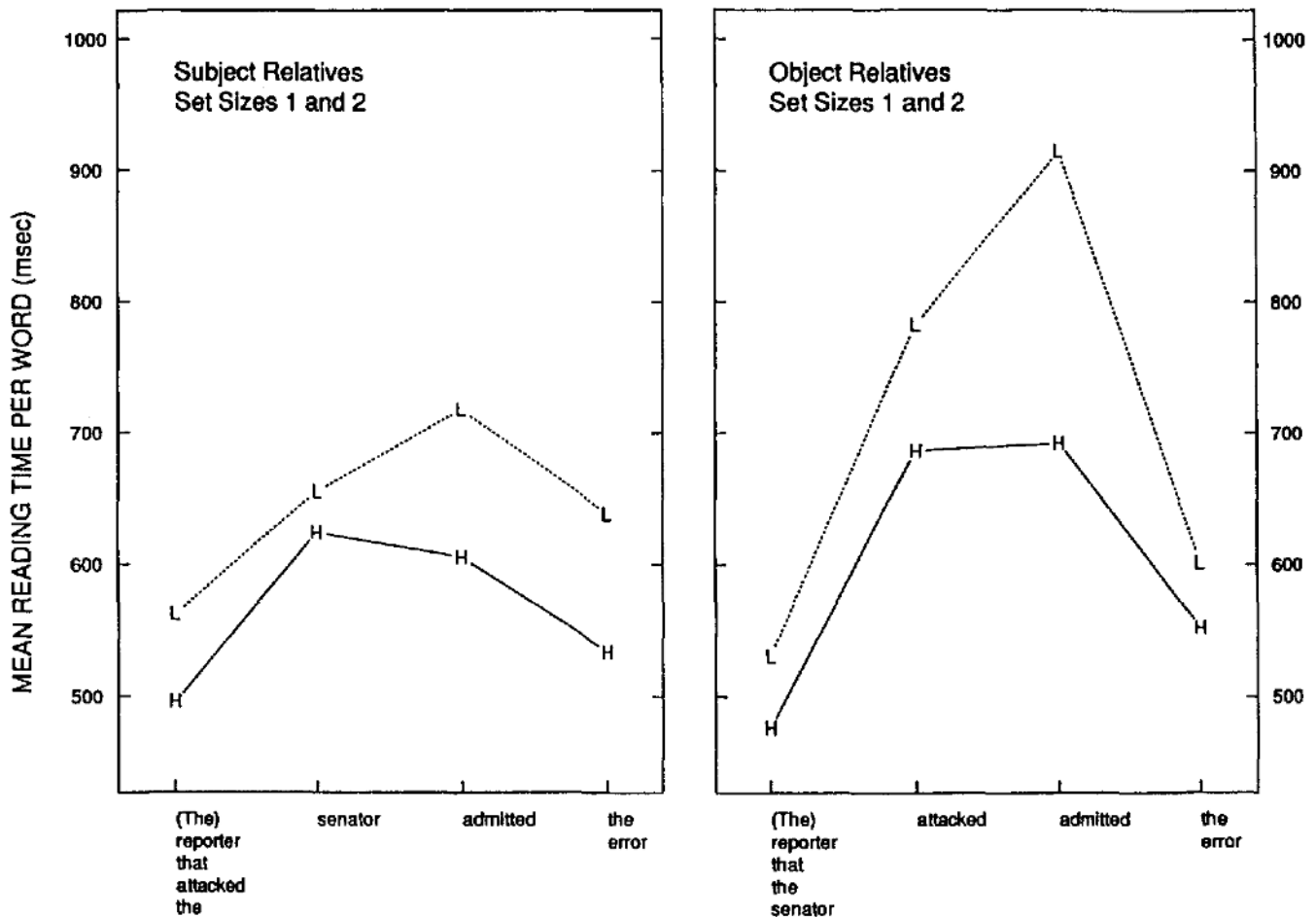


Figure 2. Mean reading time per word for successive words for the high-span (H) and low-span (L) readers for subject-relative sentences (left) and object-relative sentences (right). These reading times show the significantly greater reading time for low-span readers than for high-span readers on the more demanding object-relative compared with the subject-relative construction. Reprinted from "Individual Differences in Syntactic Processing: The Role of Working Memory," by J. King and M. A. Just, 1991, *Journal of Memory and Language*, 30, p. 589. Copyright 1991 by Academic Press. Reprinted with permission.

is additional evidence of the finding's reliability and of its generalization to independent groups of high- and low-span readers and to different sentence materials.

The model also guided the interpretation of a difference among low-span readers, some of whom answered the comprehension probes significantly above chance level (comprehenders) and others who responded at chance (noncomprehenders). Comprehenders showed more marked effects of

computational demand on the two verbs, whereas noncomprehenders did not. Waters and Caplan (1996a, p. 765) suggest this within-group effect is outside of the model. On the contrary, CC READER (Capacity-Constrained Reader model) has an inherent mechanism to account for differential strategic resource allocation (trading off speed for accuracy; Just & Carpenter, 1992). It is to the model's credit that it accounted for this speed-accuracy tradeoff in comprehension with such detail.

Figure 1. (opposite). Functional magnetic resonance images of four participants (arranged in four rows). The left-hand column of images is from the read-only condition, and the right column is from the read-and-maintain condition. The left hemisphere is depicted on the right side of the image. The white rectangles highlight one slice in Wernicke's area, indicating that the  $3 \times 3 \times 5$  mm volume of cortical tissue is significantly activated relative to the fixation condition. The functional images are superimposed on a structural magnetic resonance image of the same location. The depicted slices also show activation in the visual cortex (around the 6-o'clock position) associated with the visual processing in reading. These slices depict hardly any of the activation in Broca's area, which is more superior.

In sum, the reliable individual differences on the more demanding parts of the sentence, differences that correlate with the reading-span task, constitute support for the capacity theory and evidence against Waters and Caplan's (1996a) proposal.

### *Animacy and Modularity*

Waters and Caplan (1996a, pp. 762–764) discuss a study that examined individual differences in the use of semantic cues (animate or inanimate first nouns) in resolving a syntactic ambiguity, such as:

2a (vs. 2b) reduced relative clause "The defendant (vs. evidence) examined by the lawyer shocked the jury."

3a (vs. 3b) unreduced relative clause, "The defendant (vs. evidence) that was examined by the lawyer shocked the jury."

The study examined whether first-pass syntactic processing is modular in the sense that a semantic cue to the relative-clause interpretation (the inanimacy of *evidence*, which should disqualify *evidence* as the agent of *examined*) might fail to penetrate the initial syntactic analysis (Just & Carpenter, 1992, pp. 125–128). The measure of cue effectiveness is a decrease in the first-pass gaze duration on the disambiguating *by* phrase (e.g., *by the lawyer*), which resolves the ambiguity in favor of the relative-clause interpretation.

Our results are best understood by considering how the parsing of these sentences is influenced by two cues that signal the relative-clause interpretation: (a) the inanimacy of the first noun and (b) the presence or absence of the syntactic marker (*that was*). Neither cue alone, not even the syntactic cue, entirely eliminates the ambiguity, as indicated by the increased facilitation on the disambiguation for the high-span participants when both cues were present as compared with when only a single cue was present. For the high-span readers, there was facilitation from either type of cue alone and (additively) from both cues together. A key point is that the facilitation was not global but a specific decrease on the disambiguation. For the low-span readers, only the syntactic cue provided facilitation, suggesting that semantic information may be gated out by a resource constraint rather than an architectural boundary of a syntactic module. Waters and Caplan's (1996a) argument fails to acknowledge that the effect of the semantic cue was specific to the disambiguation and that it differed for the high- and low-span groups. It is the specificity of the effect and its interaction with working-memory capacity that is incompatible with Waters and Caplan's position.

A related but separate point is that Waters and Caplan (1996a) claim we made inappropriate comparisons to test the modularity hypothesis. However, Ferreira and Clifton (1986) proposed precisely the contrast that we tested: "If this semantic information [the inanimacy cue] can be used to guide the analysis of these sentences, the difficulty of (2b) should be reduced or eliminated. Specifically, subjects should be faster to read regions *c* [the *by* phrase] and *c* + 1 [the verb] for sentence (2b) [inanimate reduced] than (2a) [animate reduced]" (p. 352).

Finally, Waters and Caplan (1996a) attempt to rescue modularity by proposing a narrow definition of modularity, which actually tests a different issue of whether a *syntactic* cue can

influence the *syntactic* process in the face of inconsistent semantic information, that is, whether syntax penetrates semantics. In summary, the capacity theory led to new findings and cast a new light on syntactic modularity.

### *Sensitivity to Pragmatic and Inanimacy Cues*

Waters and Caplan (1996a, pp. 763–764, 765, 768) incorrectly claim a contradiction between the results of two studies, one involving noun inanimacy cues (described earlier) and one involving pragmatic noun-verb relations in sentences such as "The robber that the fireman rescued stole the jewelry" (King & Just, 1991, Experiment 2). In this study, both low- and high-span readers showed a similar slight, marginally reliable facilitation (25 ms) in the word-by-word reading times on the main verbs compared with reading times on sentences without pragmatic noun-verb relations ( $p < .10$ ). This correct characterization is entirely consistent with the results of the noun inanimacy study.

Waters and Caplan's (1996a) claim of inconsistency is partially based on a confounded comparison between the first-pass gaze durations of the low-span readers in the inanimacy study and their error rates to verification probes in the pragmatic relations study. In King and Just (1991), low-span readers made fewer errors in verifying the subsequent probe involving such pragmatic relations, suggesting that they used the pragmatic relations (e.g., *firemen rescue*, *robbers steal*), an effect that could have occurred long after first-pass reading. High-span readers' accuracy was high regardless of pragmatic cues, and the interaction of span and pragmatic cue was not reliable,  $F(3, 138) = 1.84, p > .1$ . Waters and Caplan's comparison of question-answering errors from one study and first-pass gaze durations from another study also confounds other important differences between the two studies, including different types of cues and sentence constructions. Moreover, the flaw in comparing cue effectiveness across unequal time ranges was made apparent by Ni, Crain, and Shankweiler (in press), who found that high-span readers showed sensitivity to pragmatic cues in their first-pass gaze durations, whereas low-span readers showed it later, in their subsequent regressions. In conclusion, when the two studies are accurately characterized, the results are compatible.

### *Processing Syntactic Ambiguity*

The capacity theory led to the prediction that high-span readers are more able than low-span readers to maintain two interpretations of a structural ambiguity (Just & Carpenter, 1992; MacDonald, Just, & Carpenter, 1992). Support came from a reading-time study for main-verb sentences, which are temporarily ambiguous if the verb can be a reduced relative (e.g., *warned*) compared with an unambiguous verb (e.g., *spoke*):

7 (vs. 8) main verb, "The experienced soldiers warned (vs. spoke) about the dangers before the midnight raid."

High-span readers took longer reading ambiguous main-verb sentences than their unambiguous counterparts, immediately after the ambiguous verb and particularly at the end of the sentence, reflecting the processing cost of maintaining the multiple

interpretations. Waters and Caplan (1996a) give an inaccurate description of the results, claiming "the increase in reading times seen for high-span participants did not occur where the capacity theory predicts . . ." (p. 766). They also describe the reported statistical analyses inaccurately, and they ignore a meta-analysis that showed that high-span readers reliably take longer immediately after the ambiguity, as well as at the end of the sentence, results that are consistent with the theory. The low-span readers abandoned the alternative sooner and showed almost no time differences between sentences 7 and 8 at the end of the sentence, resulting in a reliable interaction of span and ambiguity in all three studies. Waters and Caplan also imply the results are unreliable, citing a study of their own that is "in preparation" (p. 766), when in fact there are four published studies of the effect (MacDonald et al., 1992, Experiments 1, 1b, and 3; Pearlmutter & MacDonald, 1995).

Reading span also correlated with localized reading-time differences for sentences that contained relative clauses that were either temporarily ambiguous or unambiguous. These sentences were included mainly as controls, but the general pattern of data was consistent with the model. Waters and Caplan (1996a, p. 766) dismiss these effects. They claim (p. 766) that the high-capacity participants' longer times for the ambiguous sentences than the unambiguous ones were not statistically significant, ignoring the reliability of the effect in two studies and meta-analyses across participants and materials. They also propose a speed-accuracy trade-off account, which was evaluated and rejected by MacDonald et al. (1992, p. 69) because the differences in reading times between span groups were localized rather than general.

#### *Working Memory Capacity in the Elderly*

There is evidence that age-related decrements in language comprehension are, in part, mediated by a reduction in working-memory capacity (e.g., Craik, Morris, & Gick, 1990). Waters and Caplan (1996a) incorrectly characterize our claims concerning three cited studies and ignore a body of evidence showing age-correlated impairments on syntactically more complex sentences (Kemper & Anagnopoulos, 1993; also Carpenter, Miyake, & Just, 1994).

#### *Neuropsychological Evidence*

Far from being incompatible with neuropsychological findings (as Waters and Caplan suggest), the capacity theory provides a precise hypothesis about aphasic comprehension deficits, namely, that individuals with aphasia share a common deficit, a reduction in the resources of the working-memory system that subserves sentence comprehension (see Miyake, Carpenter, & Just, 1994, 1995). A computational model instantiating this hypothesis, in which the resources supporting comprehension are pathologically reduced, demonstrates partial, aphasic-like comprehension across different sentence types that correlates with error rates for individuals with aphasia (Haarmann, Just, & Carpenter, in press). Moreover, in a study of diverse aphasic patients, in which the listening-span task was modified by changing the response mode from word recall to picture selection, the listening span for individuals with aphasia

correlated highly (.82) with an independent measure of silent text comprehension (Caspari, Parkinson, LaPointe, & Katz, 1994). Thus, the theory provides an account of the comprehension performance of individuals with aphasia along several dimensions, such as brain imaging, computational modeling, and individual differences in comprehension among aphasic patients.

#### *Statistical Interactions, Predictions, and Models*

Although statistical interactions can be a powerful tool, for sensible interpretation researchers need to specify the relationship between a theoretical model, the experimental design variable, and the response scale (Bogartz, 1976). These issues are not given weight when Waters and Caplan freely switch among experimental paradigms and response measures when predicting interactions. Waters and Caplan (1996a) cite Waters and Caplan (1996b) as failing to replicate MacDonald et al. (1992). However, MacDonald et al.'s study was a sentence comprehension task measuring word-by-word reading time, whereas Waters and Caplan's (1996b) study involved acceptability judgments, using whole sentence presentation, Rapid Sequential Presentation (RSVP), and an A' measure for which Waters and Caplan have no model.

The lack of a careful task analysis also undermines Waters and Caplan's (1996a) claims concerning load tasks. According to Waters and Caplan, our theory predicts that load should always interact with sentence complexity. This is not true because our theory takes into account the type and amount of demand relative to working-memory span. Several studies show, for example, that if a load is encoded and maintained in a peripheral buffer (i.e., the phonological buffer/articulatory loop), it does not interact with sentence complexity (Craik et al., 1990). Such peripheral storage may be encouraged by paradigms that allow the load to be previewed, preorganized, or rehearsed prior to the presentation of the sentence; these are characteristics of the paradigms that Waters and Caplan frequently use. Not only are the task's characteristics important but also the individual's processes in trying to cope with the demand. Thus, to determine when and whether interactions should occur, one must prevent alternative strategies between groups that may impact differentially on their comprehension time, accuracy, and load recall (King & Just, 1991, pp. 588-589).

#### *Computational Modeling and Theoretical Prediction*

Waters and Caplan (1996a) raise miscellaneous objections to the computational modeling. For example, they question whether the simulations (in Just & Carpenter, 1992) used different parameters to simulate different studies. However, as stated originally (Just & Carpenter, 1992, p. 137), the activation quotas were constant for each group across the simulations. Also, Waters and Caplan criticize the CC READER model for having too many free parameters, suggesting their lack of understanding of computational model development and evaluation. The theory and the 3CAPS (the Capacity Constrained Concurrent Activation-based Production System) architecture together provide a toolkit for building a family of related models for particular tasks, based on a set of common principles of

cognitive architecture, providing far more constraint than most other accounts of sentence processing, particularly the bare outline proposed by Waters and Caplan. The degree of prediction versus post hoc explanation that a model provides typically increases from the early to the late studies in a series of experiments on a given topic. Such theories are not evaluated by counting their parameters but primarily by competitive argumentation.

### Conclusions

A theory is foremost an engine of understanding. The capacity theory provides insights into individual differences in sentence comprehension and has also evolved to account for comprehension by individuals with aphasia and, more recently, brain imaging. The theory is not only up to the challenge of competitive analysis but also continues to expand the horizons of cognitive research.

### References

- Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppel, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science*, 7, 25–31.
- Baddeley, A. (1992). Working memory. *Science*, 255, 556–559.
- Binder, J. R., Rao, S. M., Hammeke, T. A., Yetkin, F. Z., Jesmanowicz, A., Bandettini, P. A., Wong, E. C., Estkowski, L. D., Goldstein, M. D., Haughton, V. M., & Hyde, J. S. (1994). Functional magnetic resonance imaging of human auditory cortex. *Annals of Neurology*, 35, 662–672.
- Bogartz, R. S. (1976). On the meaning of statistical interactions. *Journal of Experimental Child Psychology*, 22, 178–183.
- Carpenter, P. A., Miyake, A., & Just, M. A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1075–1122). San Diego, CA: Academic Press.
- Caspari, I., Parkinson, S. R., LaPointe, L. L., & Katz, R. C. (1994). *Working memory and aphasia*. Paper presented at the Clinical Aphasiology Conference, Traverse City, MI.
- Craik, F. I. M., Morris, R. G., & Gick, M. L. (1990). Adult age differences in working memory. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 247–267). Cambridge, England: Cambridge University Press.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Daneman, M., & Merikle, P. M. (1994). *Working memory and language comprehension: A meta-analysis*. Paper presented at the International Conference on Working Memory, Magdalene College, Cambridge, United Kingdom.
- Eddy, W. F., Fitzgerald, M., Genovese, C. R., Mockus, A., & Noll, D. C. (1996). Functional image analysis software—Computational olo. In A. Prat (Ed.), *Proceedings in computational statistics*. Heidelberg, Germany: Physica-Verlag.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348–368.
- Haarmann, H., Just, M. A., & Carpenter, P. A. (in press). Aphasic sentence comprehension as a resource deficit: A computational approach. *Brain and Language*.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (in press). Brain activation modulated by sentence comprehension. *Science*.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Karbe, H., Herholz, K., Szekely, B., Pawlik, G., Wienhard, K., & Heiss, W. D. (1989). Regional metabolic correlates of token test results in cortical and subcortical left hemispheric infarction. *Neurology*, 39, 1083–1088.
- Kemper, S., & Anagnopoulos, C. (1993). Adult use of discourse constraints on syntactic processing. In J. Cerella, J. Rybash, W. Hoyer, & M. L. Commons (Eds.), *Adult information processing: Limits on loss* (pp. 489–507). San Diego, CA: Academic Press.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580–602.
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24, 56–98.
- Metter, E. J., Hanson, W. R., Jackson, C. A., Kempler, D., Van Lancker, D., Mazziotta, J. C., & Phelps, M. E. (1990). Temporal cortex in aphasia: Evidence from positron emission tomography. *Archives of Neurology*, 47, 1235–1238.
- Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: 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.
- Ni, W., Crain, S., & Shankweiler, D. (in press). Sidestepping garden paths: Assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. *Language and Cognitive Processes*.
- Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, 34, 521–542.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influence on syntactic processing: Use of thematic information in syntactic disambiguation. *Journal of Memory and Language*, 33, 265–312.
- Waters, G. S., & Caplan, D. (1996a). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, 103, 761–772.
- Waters, G. S., & Caplan, D. (1996b). Processing resource capacity and the comprehension of garden path sentences. *Memory and Cognition*, 24, 342–355.

Received December 20, 1995

Revision received June 21, 1996

Accepted July 5, 1996 ■