Journal of Experimental Psychology: General
2007, Vol. 136, No. 1, 64–81

0096-3445/07/$12.00 DOI: 10.1037/0096-3445.136.1.64

The Role of Working Memory in Syntactic Ambiguity Resolution: A Psychometric Approach

Benjamin Swets
Michigan State University

Timothy Desmet
Ghent University

David Z. Hambrick and Fernanda Ferreira
Michigan State University

In 2 studies, the authors used a combination of psychometric and experimental techniques to investigate the effects of domain-general and domain-specific working memory factors on offline decisions concerning attachment of an ambiguous relative clause. Both studies used English and Dutch stimuli presented to English- and Dutch-speaking participants, respectively. In Study 1, readers with low working memory spans were less likely to use recency strategies for disambiguation than were readers with high spans. This finding is inconsistent with predictions of locality- and resource-based accounts of attachment. Psychometric analyses showed that both domain-specific (verbal) and domain-general working memory accounted for the effect. Study 2 found support for the hypothesis that segmentation strategies imposed during silent reading can account for the counterintuitive relationship. Results suggest that readers with low spans have a greater tendency to break up large segments of text because of their limited working memory, leading to high attachment of the ambiguous relative clause.

Keywords: relative clause attachment preferences, working memory, psychometric, domain specificity, implicit prosody

To understand a sentence, readers and listeners must keep track of many pieces of information: the words they encounter, the order in which the words appear, the syntactic category of the words, the relations among the words, and so on. It is therefore no surprise that many psycholinguistic theories appeal to working memory to explain a variety of processing phenomena. This tendency to invoke working memory constraints goes back at least to Miller and Chomsky (1963), who argued that theories of linguistic competence described the knowledge that a speaker or hearer could in principle access during processing, whereas theories of linguistic performance had to take into account the way that the limits of the cognitive system lead to systematic patterns in performance. Miller and Chomsky's focus was the so-called center-embedded structure, which is grammatically licensed but virtually impossible to understand—as in, *The mouse the cat the dog hated chased got caught in the trap*. In particular, they argued that the reason for the comprehension failure is that each of the initial noun phrases had to be held in working memory until its appropriate predicate could be located and that more than two such noun phrases exceeded the capacity of most people's working memory. Numerous theoretical accounts of sentence parsing since have been based on the idea that complexity is increased by the number of partially processed but incomplete syntactic dependencies that the parser has to store in memory (Abney & Johnson, 1991; Chomsky & Miller, 1963; Gibson, 1991, 1998; Hakuta, 1981; Kimball, 1973; Lewis, 1993; MacWhinney, 1987; Miller & Chomsky, 1963; Miller & Isard, 1964; Pickering & Barry, 1991; Stabler, 1994; Yngve, 1960).

This early emphasis on the role of memory constraints on sentence processing influenced a large number of theoretical accounts of the system architecture. Notably included among such accounts is the garden-path model (e.g. Frazier, 1979), which holds that the decision principles used by the language system are a consequence of working memory constraints. Specifically, the garden-path model assumes that the system computes a single analysis because computing more than one syntactic structure at a time is burdensome and because not computing a structure immediately would force words to be held unanalyzed in memory, which also taxes working memory. The system's tendency to adopt the most minimal grammatical analysis is therefore not stipulated but is viewed as a consequence of the system's limited working memory capacity.

Sentence processing accounts that assume parallel rather than serial parsing architectures have also been strongly influenced by notions of working memory constraints. An example of a model along these lines is Gibson's (1998) syntactic prediction locality theory, which assumes that language processing is highly constrained by the finite amount of information that a processor can maintain and integrate over time (e.g., Chen, Gibson, & Wolf,

Benjamin Swets, David Z. Hambrick, and Fernanda Ferreira, Department of Psychology, Michigan State University; Timothy Desmet, Department of Experimental Psychology, Ghent University, Ghent, Belgium.

Fernanda Ferreira is now at the Department of Psychology, University of Edinburgh, Edinburgh, Scotland.

This work was supported by National Institute of Mental Health Grant MH-65310 awarded to Fernanda Ferreira. Timothy Desmet is a Postdoctoral Fellow of the Research Foundation—Flanders (FWO-Vlaanderen).

Correspondence concerning this article should be addressed to Benjamin Swets, who is now at the Department of Psychology, State University of New York, Stony Brook, NY 11794-2500. E-mail: bswets@notes.cc.sunysb.edu

2005; Gibson, Desmet, Watson, Grodner, & Ko, 2005; Grodner & Gibson, 2005). Lewis's (1996) parsing model also uses working memory constraints as a vital component of its architecture, but it focuses on the interference constraints that lead to difficulty in parsing multiple center-embeddings.

In other approaches to language processing, working memory constraints are invoked not to explain the architecture of the system but to capture the extent to which various sources of information can be integrated online. One well-known example of this approach assumes that individual differences in working memory capacity are associated with the extent to which people are influenced by different constraints on sentence processing (Just & Carpenter, 1992; MacDonald, Just, & Carpenter, 1992; Pearlmutter & MacDonald, 1995). Proponents of the garden-path model had argued that the component of the language system that builds syntactic structure—the parser—consults only grammatical information during its initial pass through a sentence or utterance (e.g., Frazier, 1979, 1987; Frazier & Fodor, 1978; Frazier & Rayner, 1982). A source of information such as discourse plausibility would affect only later reprocessing. In contrast, the approach that emphasizes individual differences in working memory capacity claims that people with high capacity can take advantage of constraints such as plausibility because they have the "room" to hold all the relevant information (e.g., Just & Carpenter, 1992; MacDonald et al., 1992; Pearlmutter & MacDonald, 1995).

The Present Study

Working memory is thus a central construct in a wide range of sentence processing theories. One phenomenon that has been attributed to working memory limitations is the preference of readers and listeners to connect incoming linguistic material to the most recently processed constituent. The sentence *John said that Mark will die yesterday* is difficult to understand because it violates this recency preference: The adverb *yesterday* must be attached to the nonrecent verb phrase headed by *said* instead of to the recent verb phrase headed by *will die*. The recency preference was introduced by Kimball (1973) as *right association* and revised slightly and named *late closure* in the garden-path model (Frazier, 1979). The importance of this general recency preference has also been recognized in theoretical proposals that are quite different from the garden-path model, such as Gibson's (1998) syntactic prediction locality theory and Stevenson's (1994) connectionist network model of sentence parsing.

It is assumed that the recency principle minimizes reliance on the limited working memory resources available for sentence processing because connecting incoming linguistic material to the most recently processed constituent minimizes the risk that this material will be lost from working memory through decay or interference and because recently processed material is typically more available than older material. Nevertheless, this assumption has seldom been tested with an individual differences methodology to assess whether people with smaller working memory capacities rely more heavily on recency. One exception is a study by Felser, Marinis, and Clahsen (2003), in which working memory capacity was assessed with a listening span task (Gaulin & Campbell, 1994) modified from the original Daneman and Carpenter (1980) reading span task for use with children. Six- to 7-year-old children's attachment preferences for sentences containing an ambiguous relative clause (e.g., *Someone shot the servant of the actress who was on the balcony*) were also assessed. In this structure, a relative clause (*who was on the balcony*) can either be attached high to the first noun phrase (N1, *the servant*) or low to the more recent noun phrase (N2, *the actress*). Using a self-paced listening task, Felser et al. found indeed that children with a low working memory span were more likely to interpret the relative clause as being about the recent noun phrase (N2) than were participants with a high working memory span.

A major goal of this study is to determine whether individual differences in working memory capacity predict relative clause attachment preference. If capacity affects attachment decisions and if it varies across individuals, then there should be individual differences in attachment preferences. We tested large samples of adults from two different language communities, Belgian Dutch (Flemish) and American English, and used a rigorous psychometric methodology.

We chose the relative clause attachment ambiguity for several reasons. First, it would be interesting to see whether the pattern shown in children's preferences and described earlier (Felser et al., 2003) extends to adults. More important, the recency principle makes very straightforward predictions for this specific syntactic ambiguity: Participants with low working memory spans should show a greater tendency to attach to N2 than should participants with high working memory spans. Also, the preference for one of the two alternative interpretations is close to 50% for most languages (e.g., 40% N1 in English and 63% N1 in Spanish reported in Cuetos & Mitchell, 1988; 55% N1 in Spanish reported in Igoa, Carreiras, & Meseguer, 1998; 60% N1 in Dutch reported in Desmet, Brysbaert, & De Baecke, 2002; and 55% N1 in Italian reported in De Vincenzi & Job, 1993). Perhaps because the preference is not strongly biased, it can easily be made to shift by varying the presence or absence of other linguistic constraints (Desmet, Brysbaert, & De Baecke, 2002; Desmet, De Baecke, Drieghe, Brysbaert, & Vonk, 2006; Desmet & Declercq, 2006; Gilboy, Sopena, Clifton, & Frazier, 1995; Thornton, MacDonald, & Gil, 1999; Traxler, Pickering, & Clifton, 1998). The relative clause ambiguity, therefore, represents an ideal structure with which to explore whether differences in working memory resources can influence the preference for recency-based attachments.

A second specific question of interest was whether individual differences in attachment preference arise from domain-specific or domain-general components of working memory, or both. The approach we present here is inspired by work (Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004) demonstrating that working memory tests designated to gauge proficiency in particular cognitive domains also measure domain-general working memory capacity. In a study by Kane and colleagues (2004), participants performed a battery of working memory tasks, including three verbal tasks and three nonverbal tasks. Two models were compared. The first model, a one-factor model, assumed a single, domain-general factor. This model fit the data well. The other model, a two-factor model with distinct but correlated domain-specific factors, fit the data slightly better than the one-factor model. But, importantly, the correlation between the two factors was very high ($r = .84$). This high correlation between the two domain-specific working memory factors implies the existence of a domain-general factor that acts alongside domain-specific ones.

These findings suggest an architecture of working memory consisting of at least three components: Specific factors that correspond to verbal and spatial processing as well as a general factor, which Kane and colleagues hypothesized corresponds to the ability to control attention.

At the other end of the domain-specificity spectrum are theories of memory (Kintsch, 1998) and psycholinguistics (MacDonald & Christiansen, 2002) holding that domain-general memory factors have little bearing on the operations of specific cognitive domains such as language. According to this perspective, the reason that individual differences in reading span correlate with success in language processing tasks is that the reading span assessment tool is itself a language processing task. For example, MacDonald and Christiansen (2002) argued that "the reading span task and its auditory equivalent, listening span, are basically measures of participants' abilities to do particular language-processing tasks" (p. 46). Similarly, in his construction–integration theory, Kintsch (1998) stated "To use a metaphor, it is not that good readers have a larger box to put things in for temporary storage, but that they are more skilled in putting things into long-term storage and retrieving them again" (pp. 239–240). These accounts are similar to theories assuming that individual differences in language processing are not necessarily associated with individual differences in other processing domains (Shah & Miyake, 1996).

However, the view that individual differences in language comprehension are entirely domain specific ignores a large body of research showing that reading span reflects domain-general attentional or memory processes (Caplan & Waters, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Just & Varma, 2002; Kane et al., 2004; Roberts & Gibson, 2002). One major problem with domain-specific accounts of working memory constraints on sentence processing is that psycholinguistic individual differences studies have typically used only a single measure of span, namely the reading span task (Just & Carpenter, 1992; Pearlmutter & MacDonald, 1995; Waters & Caplan, 1996). It is therefore possible that measurement of domain-general working memory has always been conflated with the measurement of domain-specific working memory in psycholinguistic individual differences research, because unless at least two measures are used, it is impossible to separate the contributions of domain-specific and domain-general factors (see below). The use of just a single measure of working memory has perhaps led researchers to conclusions that verbal representations and processes alone account for individual differences in language processing.

To remedy these limitations of previous studies, we included a nonverbal task that measured *spatial* working memory (spatial span, Shah & Miyake, 1996) to begin to examine the possibility that domain-general aspects of cognition contribute to individual differences in language comprehension. Structural equation modeling (SEM) was used to measure the intercorrelation between verbal and spatial working memory; we then estimated the relative contributions of the emergent domain-general factor and the domain-specific verbal factor to relative clause attachment ambiguity resolution. We argue from our results that language comprehension is influenced by at least two working memory factors: one that reflects domain-general processing and one that reflects domain-specific processing.

An additional goal of the present study was to investigate cross-linguistic differences in relative clause attachment preferences. The universality of late closure was famously challenged by the findings of Cuetos and Mitchell (1988), who demonstrated that native speakers of Spanish do not attach relative clauses the same way that native English speakers do. Specifically, in preferring N2 attachment, English speakers showed the trends predicted by the garden-path model; however, Spanish speakers violated the predicted trends by showing a small but significant N1 preference. This finding challenged the garden-path model by showing that late closure is, in fact, not a universal strategy. Subsequent studies demonstrated an N1 preference in several other languages, including Dutch (e.g., Brysbaert & Mitchell, 1996; Desmet, De Baecke, & Brysbaert, 2002). Such work led Frazier and Clifton (1996) to modify the manner in which the garden-path theory treats relative clauses and other modifier phrases, resulting in the construal model of sentence processing. The current study addresses both cross-linguistic and individual differences in relative clause attachment by administering the same (translated) sentences to English-speaking and Dutch-speaking participants, which allows us to examine cross-linguistic differences more directly than in any previous experiments. Also, by comparing the cross-linguistic effects on relative clause attachment with the individual differences effects, we can assess the importance of cross-linguistic differences in attachment preference compared with the contribution of individual memory capacity.

The Psychometric Approach

Broadly stated, psychometric research involves observing relationships among variables that are naturally occurring, that is, that are not experimentally manipulated. An inherent limitation of this approach is captured by the phrase *correlation does not imply causation.* Nevertheless, it is less often recognized that causation can generally be assumed to imply correlation (see, e.g., Pearl, 2002). To illustrate, if a psychological theory predicts that success in two different tasks—for example, a working memory task and a language processing task—depends on the same process or structure, then measures reflecting performance in these tasks should correlate (Underwood, 1975). Testing for correlations can therefore be considered a powerful tool for falsifying theories. However, what can be concluded from any particular test for a correlation depends on two assumptions.

The first assumption is that the constructs of interest were adequately measured. *Reliability* refers to the consistency or accuracy of a measure and is the most basic requirement for psychometric research because the reliability of a measure limits the degree to which the measure can correlate with any other measure. Therefore, if a measure has low reliability, a researcher has no hope of observing a correlation between that measure and other measures, predicted or not. *Validity* refers to whether a measure reflects what it is supposed to reflect. A measure is said to be valid to the extent that it correlates moderately with other measures of the construct and weakly with measures of other constructs (Cronbach & Meehl, 1955) and predicts certain outcomes but not others. At the same time, the task impurity problem (e.g., Burgess, 1997) is the observation that no complex task measures only the construct it is supposed to measure. Psychometric research therefore tends to emphasize use of multiple measures (two or more) of intended constructs and statistical techniques such as SEM that isolate the construct-relevant variance common to the measures and cancel out other sources of variance (e.g., method variance, error variance).

The second assumption is that the sample was adequate. A common practice in many areas of research—including psycholinguistics—is to sample a small number of participants representing extreme levels of one or more key variables. The advantage of this approach is that a given level of statistical power can be achieved using fewer participants than is necessary in a continuous-variable design. However, extreme-groups designs result in inflated effect size estimates, which may lead to erroneous conclusions about the practical or theoretical significance of the results; they also preclude evaluating whether the relationship between two variables is linear across the full range of scores on the variables (see Preacher, Rucker, MacCallum, & Nicewander, 2005, for a discussion). Also, given small sample sizes, statistical power to detect potentially interesting effects may be low, and confidence intervals around observed effects may be large, permitting little confidence in replication of results (see Roberts & Gibson, 2002, and Traxler, Williams, Blozis, & Morris, 2005, for previous examples of psycholinguistic studies using continuous analyses).

There are a number of other issues that bear on the interpretation of correlational evidence. The most relevant of these to research on language processing concerns counterbalancing. The counterbalancing of independent variables is a practice so pervasive in experimental psychology that it is easy to understand why psycholinguists have been reluctant to do away with it in individual differences studies. However, an inherent feature of counterbalanced experimental designs is that each participant will have a different experience of the experiment: Participants will view different items in different conditions and different conditions in different orders. This variability in participants' experiences runs in direct opposition to the goal of individual differences studies, which is to explain as much of the variance due to individual differences as possible—while minimizing the variance due to task differences. Accordingly, the typical approach in psychometric research is to administer materials in the same order to all participants rather than to counterbalance the materials.

Overview of Studies

Taking into account the considerations just described, we conducted two studies to (a) test whether there is a relationship between working memory and relative clause attachment preferences and (b) illuminate the nature of the relationship. In each study, participants performed two working memory tasks, one verbal and one spatial, and a task designed to tap relative clause attachment preference. Both studies sampled from an English-speaking population and a Dutch-speaking population. The general preference for N1 versus N2 attachment differs across languages (Cuetos & Mitchell, 1988), and we wanted to make sure that our findings could be generalized to languages with opposing preferences. To preview, results from correlations and structural equation models showed that speakers with low working memory spans attach relative clauses to first-mentioned noun phrases, contrary to the predictions of all models that assume that recency is a strategy for maximizing the economical use of processing resources.

Study 1

We assessed verbal and nonverbal (spatial) working memory spans in both an English- and a Dutch-speaking group, and we used an offline question-answering measure to determine individuals' relative clause attachment preferences after they read ambiguous relative clause sentences. SEM was used to test the hypothesis that both domain-general and domain-specific working memory components predict attachment preferences for such sentences.

We predicted that because the language comprehension system is limited by the availability of processing resources, the strategies chosen will reflect its attempt to make the most economical use of those limited resources. A positive correlation between the tendency to attach relative clauses to N1 and working memory might indicate that recency is the most efficient strategy to make up for costs associated with storage in working memory (as Felser et al., 2003, found in their child population). A negative correlation would suggest a different strategy altogether. Alternatively, if none of the latent working memory variables account for any variance in attachment preference, we would conclude that working memory of all types is unrelated to the resolution of the relative clause attachment ambiguity.

Method

Participants

Two hundred forty-six participants were tested in Study 1. There were two samples. One sample consisted of students at Michigan State University who were native speakers of English ($n = 150$). These participants received partial credit in their introductory psychology courses in exchange for their participation. Three participants in this sample were excluded because they either did not see the correct version of the experiment or they did not correctly fill out their answer sheets. The other sample of participants tested in Study 1 were native speakers of Dutch and were students at Ghent University in Belgium ($n = 96$). The Belgian participants received €5 ($6) in exchange for their participation.

Materials and Procedure

Three tasks were administered in a single 50-min session to groups of participants. Groups ranged in size from 1 to 12 participants. Dutch-speaking participants were tested at Ghent University, and English-speaking participants were tested at Michigan State University.

Working memory tasks. Two different kinds of working memory tasks were used: reading span to measure verbal working memory and spatial span to measure nonverbal spatial working memory. Each task consisted of 36 items divided into eight trials, or sets. Single trials consisted of set sizes of 3, 4, 5, or 6 items. Trials of each set size appeared twice in each working memory task. Crucially, each participant viewed the same items in the same order for the reasons put forth in the introduction.

The 36 reading span items were modeled on the Daneman and Carpenter (1980) design and then modified in a manner similar to that described in Turner and Engle (1989). Each item consisted of a sentence presented above a single to-be-remembered word that was highlighted in red. Half of the sentences made sense, and the other half were semantically implausible or unlikely. A question mark appeared after each sentence to indicate to participants that a response to the sentence was desired (see Figure 1 for an example).

The session began with participants reading the instructions for the reading span task. Each participant received an answer packet. Two pages of the answer packet were used for each trial: For the reading span task, the

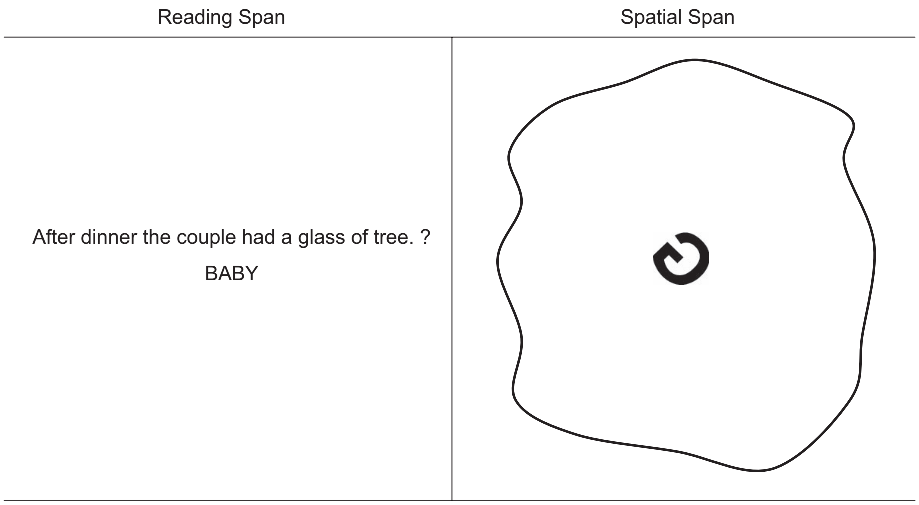

Figure 1. Item examples for the reading span and spatial span tasks in Studies 1 and 2.

first sheet for each trial showed six items for which participants circled *YES* or *NO* to indicate whether the sentences made sense (but all six were to be circled only for trials with set sizes of six). The second sheet showed six numbered blank lines in which to write in the to-be-remembered words. Each item in a given trial (the sentence/to-be-remembered word combination) was presented for 5 s. It was during this time that participants read the sentences, marked *YES* or *NO* on their answer sheets, and read the to-be-remembered word. After all of the items in a given set had been presented, the recall prompt (*???*) appeared. The duration of this prompt varied as a function of the set size of the trial: For each item in a set, the participants had 4 s. So, for set sizes of three, 12 s were allowed for recall; for set sizes of six, 24 s were allowed; and so on.

The items used in the spatial span task were modified from the Shah and Miyake (1996) task of the same name. Individual items featured a letter in the center of a blob shape. The letter was either in a normal orientation or mirror-reversed—*G*, *F*, or *R* (see Figure 1 for an example)—and was rotated within the blob at one of eight possible orientations (0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°). There was a 50/50 split on the normal/mirror-reversed dimension within the items, and letters were equally likely to appear at all of the rotation orientations.

Timothy Desmet, a native Dutch speaker from Belgium who is also fluent in English, translated the English sentences and the to-be-remembered words from the Study 1 reading span task into Dutch. Items were kept in the same order as in the English task so that comparisons could be made between languages. Because Dutch uses the same letters as English, the spatial span task did not require translation.

After completing the reading span task, participants read instructions for the spatial span task. The first answer sheet for each trial showed six items for which participants circled *NORMAL* or *MIRROR* in response to whether the presented letter was normal or mirror-reversed (but all six were to be circled only for trials with set sizes of six). The second sheet showed six blank blobs in which to draw the directions in which the tops of the letters were pointing. Participants were instructed to look at the letter inside of a blob; to circle *NORMAL* if the letter was normal or *MIRROR* if it was mirror-reversed; and when they were prompted by a cue (again, *???*), to draw on the next sheet the directions in which the tops of the letters had been pointing in the order in which they appeared. The presentation of each letter-blob figure lasted for 5 s. It was during this time that participants studied the letters and marked *NORMAL* or *MIRROR* on their answer sheets. After all of the items in a given set had been presented, the recall prompt appeared. The duration of this prompt varied as a function of the set size of the trial, as in the reading span task.

The working memory tasks were scored in the following way. For each set, an item was scored as correct only if (a) the answer for the processing component was correct (e.g., for the sentence *After dinner the couple had a glass of tree* in the reading span task, *NO* would have to be circled) and (b) the word or direction for the memory component of each item was recalled in the correct serial position (e.g., for a set *CANDY–GUN–TREE* in the reading span task, *GUN* would have to be recalled in the second position to be counted as correct). The total number of correct items from the reading span task was taken as the indicator for verbal working memory, and the total number of correct items from the spatial span task was the indicator for spatial working memory.

Relative clause attachment task. Relative clause ambiguity sentences used in Studies 1 and 2 are shown in the Appendix. As described above, we considered it a top priority to ensure that each participant performed the same task so as to most accurately measure the degree to which participants differed from each other on the various behavioral dimensions. To this end, only a single list of items was created for the task in which we measured offline preferences for sentences containing relative clauses that ambiguously attach to either noun phrase within a complex noun phrase. The list consisted of 100 items, 20 of which were experimental. The remaining 80 items were filler trials. Each item consisted of a sentence, a question, and two possible answers to the question that participants responded to.

The 20 experimental sentences were modified from an experiment by Traxler et al. (1998). The sentence's subject began with a determiner, a head noun, and then a prepositional phrase modifier (e.g., *The maid of the princess*). The preposition *of* was always the head of this particular prepositional phrase. The rest of the subject consisted of a relative clause containing a reflexive noun that could refer to either constituent of the complex noun phrase subject (e.g., *who scratched herself*) and a prepositional phrase or adverbial modifier (*in public*). The subject was followed by the matrix verb phrase (e.g., *was terribly embarrassed*). An example sentence is listed below:

> *The maid* (N1) *of the princess* (N2) [*who scratched herself in public*] (relative clause) *was terribly embarrassed.*

For each experimental item, we created questions intended to measure participants' interpretations of the relative clause ambiguity (e.g., *Who scratched herself in public?*). The two possible response alternatives were N1 and N2, and these were presented one above the other. Half of the items presented the word corresponding to N1 above the word corresponding to N2, and the other half were shown in the opposite arrangement. We should note here that in this first study (English only), we did use a counterbalancing factor: whether the N1 or N2 appeared on top or on bottom for each item. If N1 appeared on top in one of the lists, it appeared on the bottom in the other list, and vice versa. We counterbalanced in this way to ensure that the location of answers in the multiple choice set did not bias responses. Because we found that this counterbalancing variable did not bias responses, the second study (which we describe in the next section) did not counterbalance the placement of the response alternatives.

There were four types of fillers: causation sentences (*Chuck delighted Janet because she had not expected the visit./Who had not expected the visit?/Janet–Chuck*), inference sentences (*Steve was found unconscious next to an empty fifth of cheap whiskey./Did Steve drink alcohol?/yes–no*), color sentences (*The leaves that turned brown fell off the trees in November./What color were the leaves?/brown–orange*), and location sentences (*The cheese that was next to the butter tasted awful./Where was the cheese?/next to the butter–next to the milk*). Correct answers were equally likely to be on top or on bottom, and this factor was counterbalanced across the two lists. Roughly two fifths of the filler sentences contained relative clauses, somewhat equally distributed between relative clauses that modified referents in subject position and in object position.

The attachment task required only one answer sheet that contained 100 items for which participants were to circle either *TOP*, in which case the top alternative was the best answer to the question about the presented sentence, or *BOTTOM*, meaning that the bottom alternative was a better answer. Participants read instructions to carefully read the sentences that were presented on the screen; to carefully read the question; and when the two possible answers came on the screen, to circle *TOP* or *BOTTOM*, depending on which was the best answer to the question. Each sentence was presented for 5 s, followed by the question presented for 3 s, after which a fixation cross appeared for 500 ms, leading finally to the multiple choice answers presented for 4 s. The next trial started immediately after the answers were presented.

Attachment for given sentences was straightforwardly assessed from participants' answers to the question about the relative clauses. The number of N1 attachments in the attachment task was divided by 20 to give a percentage N1 attachment figure.

The tasks were merged in a single PowerPoint file on a Dell laptop that was used to project the stimuli onto a large screen that all participants viewed. The session ended with a debriefing in which participants were informed about the purpose of the experiment.

For the Dutch sample, the sentences from the relative clause attachment task were translated into Dutch and kept in the same order, as described for the reading span task above.

Words on the answer sheets such as *Practice*, *Set 1a*, *Yes–No*, *Normal–Mirror, Top–Bottom,* and *Answer Sheet* were also translated into Dutch. The only difference between English and Dutch procedures was that some Dutch groups had as many as 15 people per session.

Results

Two steps were taken to screen the data for possible outliers. First, scores on any task that were more than 3.5 *SD* from the mean were defined as univariate outliers. Second, for each participant, Cook's *D* statistic (Cook, 1977) was computed by regressing the attachment preference variable onto the verbal and spatial working memory variables. Cook's *D* reflects the change in regression coefficients caused by deleting data for a given participant. Cook's *D* values exceeding 1 are considered indicative of possible outliers. By these criteria, there was no evidence for either univariate or multivariate outliers. Outside of these criteria, 2 participants in the English sample and 2 participants in the Dutch sample failed to score a single point in spatial span, so all data from those 4 participants were excluded from analysis.

Descriptive Statistics

Table 1 displays descriptive statistics for the working memory and attachment preference variables. Reading span scores were significantly lower in the English sample ($M = .49$, $SD = .17$) than in the Dutch sample ($M = .65$, $SD = .17$): $t(244) = -7.41$, $p < .01$. Differences in spatial span were nonsignificant. Furthermore, consistent with previous reports, the preference for N1 attachments (abbreviated as N1% in tables and text below) was slightly higher in the Dutch sample ($M = .56$, $SD = .30$) than in the English sample ($M = .47$, $SD = .31$): $t(244) = -2.12$, $p < .05$. Additional analyses showed that although the Dutch speakers showed an attachment preference marginally different from 50%, $t(95) = 1.88$, $p = .06$, the English N2 preference was not reliably different from 50%, $t(149) = 1.06$, $p = .29$.

Correlations and Reliabilities

Table 2 displays reliability estimates (i.e., coefficient alphas) and the correlations among the variables. Reliability was very high (αs $> .80$), indicating that all measures had excellent internal consistency reliability. Especially noteworthy are the reliability estimates for N1% (English $\alpha = .92$; Dutch $\alpha = .92$): Across sentences, participants were highly consistent in their attachment preferences; that is, if a given individual attached one item to N1, for instance, he or she tended to attach all items to N1. This consistency is unlikely to be due to syntactic priming, because recall that the 20 critical items were embedded in a list of 100 sentences and the experimental items were always separated by at least two fillers. Note also that reliability estimates were similar or identical in the Dutch and English samples. Therefore, differences in relations among the variables across samples cannot be attributed to differences in reliability.

Inspection of the correlations reveals two noteworthy findings. First, there were moderate positive correlations between reading span and spatial span (ps $< .01$; English, $r = .41$; Dutch, $r = .45$). Second, both working memory measures correlated significantly and negatively with N1%: verbal span (English, $r = -.34$; Dutch,

Table 1
Descriptive Statistics (Studies 1 and 2)

	English		Dutch				
Measure	*M*	*SD*	*M*	*SD*	*t*	*df*	*d*
Study 1							
Reading span	.49	.17	.65	.17	-7.41^{**}	244	.94
Spatial span	.56	.24	.50	.24	1.91	244	.25
N1%	.47	.31	.56	.30	-2.12^{*}	244	.29
Study 2							
Reading span	.47	.18	.60	.18	-6.11^{**}	290	.72
Spatial span	.50	.26	.49	.21	0.19	290	.04
N1%	.71	.25	.75	.19	-1.52	290	.17

Note. Study 1: English, $n = 150$; Dutch, $n = 96$. Study 2: English, $n = 166$; Dutch, $n = 126$. N1% = preference for attachment to the first noun phrase in the relative clause ambiguity sentences.
* $p < .05$. ** $p < .01$.

Table 2
Correlations and Reliability Estimates (Studies 1 and 2)

	Study 1						Study 2					
	English			Dutch			English			Dutch		
Measure	1	2	3	1	2	3	1	2	3	1	2	3
1. Reading Span	.83			.84			.84			.84		
2. Spatial Span	.41	.92		.45	.89		.34	.93		.47	.89	
3. N1%	−.34	−.21	.92	−.41	−.23	.92	−.22	−.06	.89	−.05	−.05	.92

Note. Values along the diagonal are coefficient alphas. For Study 1, correlations with an absolute magnitude greater than .16 (English sample) or .22 (Dutch sample) are statistically significant ($p < .05$). For Study 2, correlations with an absolute magnitude greater than .16 (English sample) or .18 (Dutch sample) are statistically significant ($p < .05$). N1% = preference for attachment to the first noun phrase in the relative clause ambiguity sentences.

$r = -.41$) and spatial span (English, $r = -.21$; Dutch, $r = -.23$). However, this negative correlation was somewhat stronger for reading span than for spatial span. A one-tailed test for differences in correlations confirmed this difference in predictive strength for Dutch (rs $= -.41$ vs. $-.23$), $t(93) = 1.81$, $p < .05$. This difference was not significant for English (rs $= -.34$ vs. $-.21$), $t(14) = 1.54$, $p = .06$.

To illustrate the above trends, we created groups of approximately equal size representing three levels of working memory span (low, medium, and high). (Note that we created these groups solely for the purpose of illustration but treated the variables as continuous in all statistical analyses.) As shown in Figure 2, participants with low reading spans showed a greater preference for N1 attachment than did participants with high spans. This result is contrary to the predictions of models that assume that locality preferences arise because of constraints on working memory. Instead, we found that those with the smallest working memory spans preferred the more distant attachment site, whereas those with high spans preferred the more recent site.

Effect Sizes

Effect sizes of cross-language differences in attachment preference and individual differences in attachment preference (computed with scores on reading span) were computed with Cohen's *d*. Note that the cross-language difference was considerably smaller than the within-language differences related to reading span. The effect size (Cohen's *d*) for the cross-language difference was .29. By Cohen's (1988) standards, this is considered a small effect. On the other hand, the within-language effects of reading span were .72 in the English sample and .90 in the Dutch sample, and these values are conventionally viewed as representing large effects.

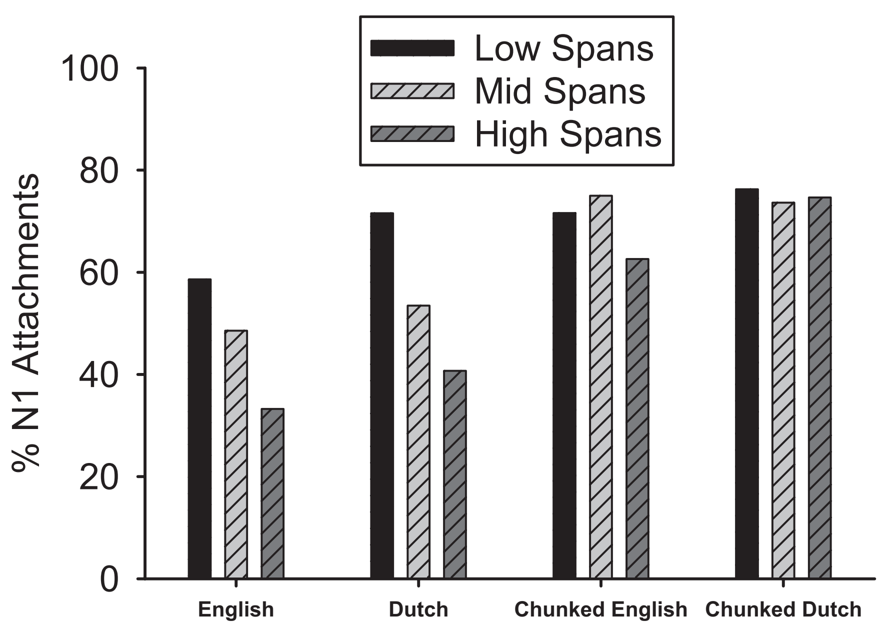

Figure 2. N1 attachments as a function of reading span group in Studies 1 and 2. The English and Dutch bars show results from Study 1, and the Chunked English and Chunked Dutch bars show results from Study 2. N1 = the first noun phrase in the relative clause ambiguity sentences.

Thus, although cross-linguistic differences are theoretically interesting in psycholinguistics, they are not nearly as robust as the individual differences that may be observed within a homogeneous language community.

SEM

We have found a relationship between working memory span and attachment preference, such that participants with low spans tend to prefer N1 attachment to a greater extent than do participants with high spans, and this relationship was stronger for reading span than for spatial span. At the same time, reading span and spatial span correlated positively and moderately, consistent with previous research (e.g., Kane et al., 2004) and with the existence of a domain-general working memory factor. That is, shared variance between the span measures was approximately 17% in the English sample ($.41^2 \times 100 = 16.8\%$) and approximately 20% in the Dutch sample ($.45^2 \times 100 = 20.1\%$). The question we address now is, to what extent did the domain-general factor implied by this shared variance contribute to individual differences in attachment preference, beyond the contribution of variance specific to reading span.

We performed structural equation analyses to address this question. For use in these analyses, we created four subscores for each task. Two subscores were the sum of scores for trials with set sizes of three and five; the other two were the sum of scores for trials with set sizes of four and six. (Again, set sizes ranged from three to six in each working memory task, with two trials at each set size, for a total of eight trials.) There also were four subscores for attachment preference, each reflecting the number of N1 attachments for five trials, as follows: ATT1 (Trials 1, 5, 9, 13, and 17); ATT2 (Trials 2, 6, 10, 14, and 18); ATT3 (Trials 3, 7, 11, 15, and 19); and ATT4 (Trials 4, 8, 12, 16, and 20).

For all analyses, we used covariance matrices as input but present standardized solutions. Model fit is characterized in terms of a number of commonly reported fit statistics, and we scaled the latent variables by fixing the factor loading of one variable per construct to 1.0. The χ^2 statistic reflects whether there was a significant difference between the reproduced and observed covariance matrices. Therefore, nonsignificant χ^2 values are desirable. However, when moderate to large sample sizes are used, even slight differences in the observed and reproduced matrices can result in significant χ^2s. We report a number of other fit statistics that are less sensitive to sample size. The confirmatory fit index (CFI) and normed fit index (NFI) reflect the proportion of the observed covariance matrix explained by the model. The root-mean-square error of approximation (RMSEA) reflects the average squared difference between the observed and reproduced covariances. CFI and NFI values greater than .90 and RMSEA values less than .08 indicate acceptable fit (see Kline, 1998).

Confirmatory Factor Analyses (CFAs)

First, we performed CFAs to establish measurement models for the variables, with a separate analysis for each sample. As illustrated in Figure 3, we considered two possible models for the working memory variables. Results concerning model fits are displayed in Table 3. Model 1 included a domain-general factor (WM_G), with loadings from all working memory variables (see Figure 3A). For both samples, model fit was poor: English, $\chi^2(20, N = 150) = 101.56$, CFI = .81, NFI = .78, RMSEA = .17; Dutch, $\chi^2(20, N = 96) = 57.93$, CFI = .84, NFI = .78, RMSEA = .14. Model 2 included a domain-general factor but also a verbal factor (WM_V) with loadings from the four reading span variables (see Figure 3B). For both samples, this model provided an acceptable fit to the data: English, $\chi^2(16, N = 150) = 22.11$, CFI = .99, NFI = .95, RMSEA = .05; Dutch, $\chi^2(16, N = 96) = 16.28$, CFI = 1.00, NFI = .94, RMSEA = .01. Furthermore, improvement in model fit over Model 1 was statistically significant: English, $\Delta\chi^2(4, N = 150) = 79.45$, $p < .01$; Dutch, $\Delta\chi^2(4, N = 96) = 41.65$, $p < .01$. Therefore, we considered Model 2 to be the best representation of the working memory variables.[1] (As shown in Table 4, the reading span variables tended to have stronger loadings on the WM_V factor than on the WM_G factor; however, the

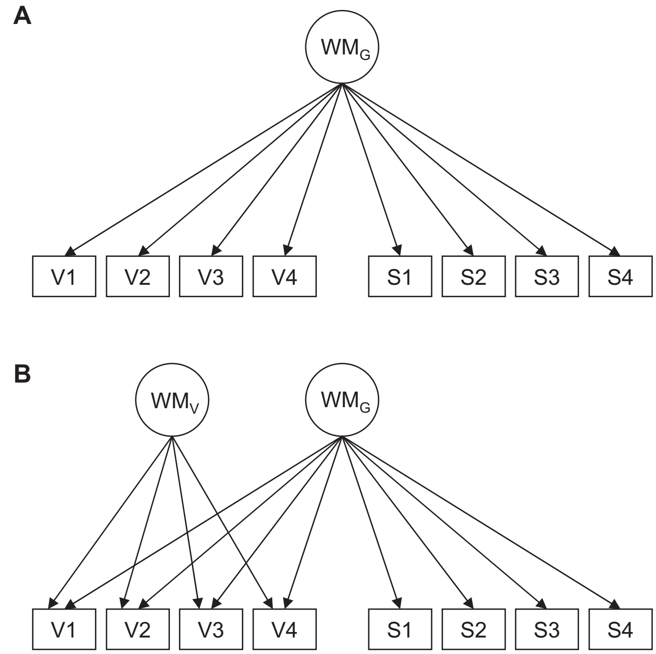

Figure 3. Predictor-side measurement models tested in Studies 1 and 2. A: Model including a domain-general working memory factor (WM_G), with loadings from all working memory variables. B: Model including WM_G and a verbal working memory factor (WM_V), with loadings from the four reading span variables. V1–V4 = reading span variables; S1–S4 = spatial span variables.

[1] Although we had no reason to expect effects of a spatial working memory factor (WM_S) on attachment preference, for each of the four independent data sets reported in this article (English and Dutch in Study 1 and Study 2), we tested a model that included three factors—WM_G, WM_V, and WM_S. (Each variable loaded onto WM_G, with the four reading span variables loading onto WM_V and the four spatial span variables loading onto WM_S.) In each case, the model was identified but the solution was not admissible (i.e., it did not converge). Therefore, we do not report fit statistics for these models. Consistent with previous research in which WM_G and WM_S factors have been shown to correlate very highly (e.g., Kane et al., 2004), inspection of the output from each model suggested that this was probably because there was a high degree of multicollinearity between WM_G and WM_S in our data sets.

Table 3
Fit Statistics for Predictor-Side Measurement Models (Study 1)

Model	$\chi^2(df)$	CFI	NFI	RMSEA	$\Delta\chi^2(df)$
English (N = 150)					
1: WM_G	101.56 (20)	.81	.78	.17	
2: $WM_G + WM_V$	22.11 (16)	.99	.95	.05	79.45 (4)
Dutch (N = 96)					
1: WM_G	57.93 (20)	.84	.78	.14	
2: $WM_G + WM_V$	16.28 (16)	1.00	.94	.01	41.65 (4)
Equivalence analyses (N = 246)					
Predictor Side					
E1: No constraints	38.40 (32)	.99	.95	.03	
E2: Factor loadings equal	54.36 (42)	.98	.93	.04	15.96 (10)
E3: Factor loadings and variances equal	56.24 (44)	.98	.92	.03	1.88 (2)
Criterion Side					
E1: No constraints	4.90 (4)	1.00	.99	.03	
E2: Factor loadings equal	7.02 (7)	1.00	.99	.00	2.12 (3)
E3: Factor loadings and variances equal	7.16 (8)	1.00	.99	.00	0.14 (1)

Note. CFI = confirmatory fit index; NFI = normal fit index; RMSEA = root-mean-square error of approximation; WM_G = domain-general working memory factor; WM_V = verbal working memory factor.

loadings on WM_G were all statistically significant.) For attachment preference, a one-factor model (see Table 5) provided an excellent fit to the data: English, $\chi^2(2, N = 150) = 1.55$, CFI = 1.00, NFI = 1.00, RMSEA = .00; Dutch, $\chi^2(2, N = 96) = 3.35$, CFI = 1.00, NFI = .99, RMSEA = .08.

Tables 4 and 5 display the predictor-side and criterion-side measurement models by sample (English vs. Dutch). As shown, factor loadings were very similar. We performed an additional series of CFAs as a formal test of the equivalence of the measurement models in the Dutch and English samples (i.e., factorial invariance; see the bottom portion of Table 3). For both predictor-side and criterion-side models, progressively more restrictive models (a) allowed the samples to have separate loadings on the factor or factors (E1), (b) constrained factor loadings to be equivalent (E2), and (c) constrained factor loadings and factor variances to be equivalent (E3). As shown in the bottom portion of Table 3, the constraints resulted in no significant loss in model fit relative to the model with no constraints (E1). Therefore, there was evidence for factorial invariance: Across samples, the same variables related to the factors (configural invariance), and factor loadings did not differ (metric invariance).

Structural Equation Analyses

Having identified the factors, we performed path analyses to estimate the relative contributions of WM_G and WM_V to attachment preference. Results are displayed in Figure 4. For both samples, model fit was excellent: English. $\chi^2(48, N = 150) = 44.47$, CFI = 1.00, NFI = .96, RMSEA = .00; Dutch, $\chi^2(48, N = 96) = 66.36$, CFI = .97, NFI = .89, RMSEA = .06. As expected

Table 4
Loadings of Working Memory Variables on Latent Factors (Studies 1 and 2)

	Study 1				Study 2			
	English		Dutch		English		Dutch	
Variable	WM_G	WM_V	WM_G	WM_V	WM_G	WM_V	WM_G	WM_V
V1	.39	.34	.31	.59	.36	.53	.45	.58
V2	.33	.55	.35	.47	.27	.64	.38	.69
V3	.29	.74	.45	.46	.25	.59	.50	.59
V4	.31	.54	.33	.61	.23	.70	.37	.50
S1	.81		.77		.85		.70	
S2	.78		.76		.82		.76	
S3	.77		.74		.79		.63	
S4	.81		.76		.84		.73	

Note. WM_G = domain-general working memory factor; WM_V = verbal working memory factor; V = verbal; S = spatial.

Table 5
Criterion-Side Measurement Models (Studies 1 and 2)

	Study 1		Study 2	
Variable	English	Dutch	English	Dutch
ATT1	.84	.83	.81	.59
ATT2	.89	.83	.79	.67
ATT3	.87	.90	.88	.82
ATT4	.93	.87	.83	.77

Note. ATT1 = number of attachments to the first noun phrase (N1) in Trials 1, 5, 9, 13, and 17; ATT2 = number of N1 attachments in Trials 2, 6, 10, 14, and 18; ATT3 = number of N1 attachments in Trials 3, 7, 11, 15, and 19; ATT4 = number of N1 attachments in Trials 4, 8, 12, 16, and 20.

on the basis of the zero-order correlations, there were negative effects of WM_V on attachment preference (*p*s $< .01$; English, −.33; Dutch, −.41). High levels of WM_V were associated with low levels of N1%. However, there also were negative effects of WM_G (*p*s $< .05$; English, −.23; Dutch, −.27). Thus, there was an effect of WM_G on attachment preference, independent of the effect of WM_V.

We conducted additional analyses to address two issues. The first question was whether the effect of WM_V on attachment preference was greater than that of WM_G. To answer this question, we compared the model reported above with one in which these effects were constrained to be equal. Loss of model fit was statistically significant for the English sample, $\Delta\chi^2(1, N = 150) = 6.50$, $p < .05$, but not for the Dutch sample, $\Delta\chi^2(1, N = 96) = 2.04$, $p = .15$. Thus, for the English sample, the effect of WM_V on attachment preference was stronger than that of WM_G. This was not the case for the Dutch sample, although it is possible that a smaller sample size resulted in the loss of sufficient power to observe a significant difference. The second question was whether there were differences across samples in effects of WM_G and WM_V on attachment preference. Was, for example, the effect of WM_G on attachment preference greater for Dutch (−.41) than for English (−.33)? To answer this question, we constrained the paths leading from WM_G and WM_V to attachment preference to be equal in the two samples. Loss of fit relative to a model in which these paths were allowed to vary was nonsignificant, $\Delta\chi^2(2, N = 246) = 0.94$, $p = .63$. We therefore conclude that there were no differences in predictive relations in the English and Dutch samples.

Discussion

Several noteworthy findings were uncovered in Study 1. First, we found that the well-reported cross-linguistic differences in attachment preferences between English and Dutch also held for our samples: Dutch speakers tended to attach more often to N1 than English speakers did. Second, we found a relationship between working memory span and attachment preference that is not predicted by accounts that regard recency as a resource-preserving strategy. An explanation for this finding is deferred until we describe Study 2.

Note that the relationship between working memory and attachment preference was much larger than the cross-linguistic effect: The individual differences were roughly three times larger than the cross-linguistic differences. Because studies that have shown cross-linguistic differences in attachment preference never controlled for this substantial variation in individual working memory differences in their between-groups comparisons, it is likely that these differences have been overinterpreted as evidence against universal late closure strategies. However, the finding of large individual differences in itself could be viewed as strong evidence against the universality of late closure: Individuals clearly differ in the extent to which they use it, regardless of whether their native language has an independent effect.

Finally, both domain-general and domain-specific working memory were related to relative clause attachment preference. This result is consistent with recent evidence that working memory measures of all sorts have a substantial domain-general component (Engle et al., 1999; Kane et al., 2004) as well as with views holding that language processing can be affected by memory components that are both unique to language and shared with other parts of the cognitive architecture (Baddeley, 1986, 2000). However, this result is inconsistent with the view of working memory as being a language-specific factor (MacDonald & Christiansen,

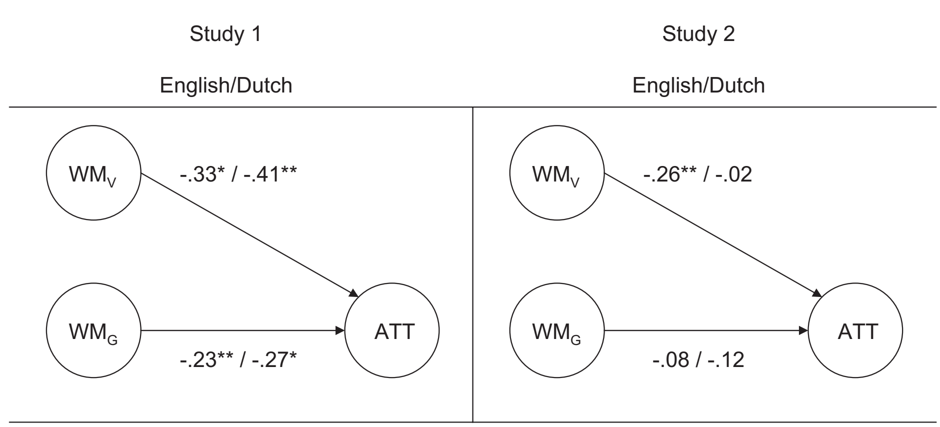

Figure 4. Path models for Studies 1 and 2. Values on the left of the slashes are for English. Values on the right of the slashes are for Dutch. WM_V = verbal working memory factor; WM_G= domain-general working memory factor; ATT = attachment to the first noun phrase in the relative clause ambiguity sentences.

2002) and the view of a purely domain-specific working memory (Shah & Miyake, 1996). We address the implications of our findings for such accounts of domain specificity in the General Discussion.

Study 2

Study 1 showed that for both English and Dutch speakers, the decision about which constituent of a complex noun phrase is modified by an ambiguous relative clause is influenced by the availability of working memory resources. The obvious remaining question is why speakers with smaller spans attach relative clauses to the more distant site.

To account for this finding, one might invoke the existence of a processing principle such as predicate proximity (Gibson, Pearlmutter, Canseco-Gonzalez, & Hickok, 1996) or relativized relevance (Frazier & Clifton, 1996). According to these principles, N1 is a more accessible discourse entity than N2 because of its prominence as an argument of the main verb. Given that Gibson et al. (1996) have already hypothesized that predicate proximity is a constraint that varies in strength across languages, one might also propose that it varies in strength across span types. So, whereas people with low spans might have a strong bias to attach to the more accessible N1, people with high spans might be better able to keep track of additional discourse entities (apart from the most accessible one) and, therefore, on average have a lower attachment bias. This account could explain why participants with low spans have an overall preference to attach high and why working memory capacity is inversely related with attachment preference. However, the appeal to prominence factors ultimately fails to account for the overall tendency of participants with high spans to attach low. Although participants with high spans would on average be less likely to attach high than would participants with low spans, they would be unlikely to have an overall preference to attach to the less accessible entity over the more accessible entity.

Alternatively, it could be argued that people with low spans prefer N1 attachment because they are less sensitive to the infelicity of multiple modification. Because N1 is already modified by N2, at a discourse level it is infelicitous to modify it even further (see Altmann & Steedman, 1988, for a discussion of processing difficulty related to invoking additional contrast sets when modifying a noun phrase, and see Thornton et al., 1999, for evidence that modified noun phrases are less likely to receive further modification). People with high spans might be sensitive to this Gricean principle and therefore have a preference to attach low. People with low spans might be less sensitive to this principle and therefore, on average, be less likely to attach low. Again, this account fails to explain the overall high attachment of people with low spans: They might be less sensitive to this principle and have a higher attachment preference than people with high spans, but it is unlikely that they have an overall preference for the least felicitous interpretation over the most felicitous interpretation.

Another, more viable, explanation is based on the concept of implicit prosody (Fodor, 1998, 2002). According to the implicit prosody hypothesis, silent prosody affects syntactic decisions made during reading. As one reads a sentence, the way a sentence gets parceled into prosodic units by an internal voice can be used as discriminatory information when syntactic and semantic information leave structural input ambiguous. Fodor (1998, 2002) has used this theory as an explanation for cross-linguistic differences in relative clause attachments: Whereas speakers of languages such as Dutch tend to put a prosodic break between the complex noun phrase and the relative clause in sentences that are similar to our critical stimuli, English omits this break and tends to place one after the entire relative clause (see also Jun, 2003). A prosodic break before the relative clause has been shown to lead to higher attachment decisions in studies when explicit prosody is manipulated directly in spoken sentences (Carlson, Clifton, & Frazier, 2001), and cross-linguistic studies have found that languages indeed show overall attachment preferences that fall in line with their default prosodic breaking patterns (Jun, 2003). The prosodic break is assumed to induce N1 preferences because a prosodic break before a relative clause could be interpreted “as marking a structural discontinuity in the syntactic tree” (Fodor, 2002, p. 4). This interpretation would then lead to the formation of a tree in which the entire complex noun phrase is modified by the relative clause rather than just N2.

An intriguing possibility is that this prosodic parceling mediates not only cross-linguistic differences but also the individual differences factor found here. Perhaps people with high spans “chunk” more material together as they read silently than do people with low spans. This tendency would increase the chances that the complex noun phrase and the relative clause would be grouped together with no break, leading to the tendency to attach low (Carlson et al., 2001). On the other hand, people with low spans may insert a break between the complex noun phrase and the relative clause, a strategy that would be associated with a tendency to attach relative clauses more often to N1.

In Study 2, we tested the chunking hypothesis by forcing all participants to parcel the complex noun phrase–relative clause units into two units. Specifically, we presented each of the Study 1 sentences in three separate pieces: first, the complex noun phrase (*The maid of the princess*), then the relative clause with a modifying prepositional phrase (*who scratched herself in public*), and finally the matrix verb phrase (*was terribly embarrassed*). If the individual differences in attachment preferences observed in Study 1 were due to differences in the way people internally chunk the constituents while reading silently, then removing variance along that dimension should remove or drastically reduce the variance in attachment preferences, and all participants should attach more like the participants with low spans did in Study 1.

Note that the two alternative hypotheses presented above make crucially different predictions compared with the chunking hypothesis. Given that the presentation format does not change the prominence of N1 and N2 (N1 is still an argument of the main verb and should still be more accessible than N2) or the infelicity of the N1 modification (a high attachment interpretation still modifies N1 twice), these alternative explanations predict that the pattern of results in Study 2 should be identical to the pattern of results observed in Experiment 1.

Method

Participants

In this study, 292 participants from English (n = 166) and Dutch (n = 126) samples were recruited in the same manner as in Study 1.

Materials and Procedure

The materials were the same as those used in Study 1. The major procedural difference was as follows: Rather than presenting the sentence in its entirety for 5 s, we presented it noncumulatively in three segments for 2 s each, for a total of 6 s. This presentation method was used for both experimental and filler sentences. We chose to present segments for 2 s because pilot testing revealed that participants would sometimes fail to apprehend the content if segments were presented for a duration less than 2 s. Because the potential antecedents for the reflexive pronoun in the second segment were located in the first, it was important that participants process both segments long enough to be able to establish some type of coreference relation. All experimental sentences were divided in the way described in the Study 2 introduction (*The maid of the princess/who scratched herself in public/was terribly embarrassed*). Filler sentences were divided at syntactically and semantically appropriate points to ensure that the presentation segments were roughly equal in size and that they corresponded to linguistic units.

Results

Applying the data screening criteria from Study 1, no cases were excluded as outliers. However, 4 participants in the current study (all from the English sample) failed to score a single point in the spatial span task, so all data from these participants were excluded from analysis.

Descriptive Statistics

Table 1 displays descriptive statistics for the working memory and attachment preference variables. As in Study 1, reading span scores were significantly higher in the Dutch sample ($M = .60$, $SD = .18$) than in the English sample ($M = .47$, $SD = .18$), $t(290) = 6.11$, $p < .01$. However, unlike in Study 1, the difference in percentage N1 attachments between English and Dutch was not significant (English, $M = .71$, $SD = .25$, vs. Dutch, $M = .75$, $SD = .19$), $t(290) = -1.52$, $p = .13$.

Correlations and Reliabilities

Table 2 displays correlations among the variables and reliability estimates (i.e., coefficient alphas). As in Study 1, reliability estimates were very high, including for attachment preference (English, $\alpha = .89$; Dutch, $\alpha = .92$). Thus, participants were again highly consistent in attachment preference. Furthermore, there were moderate and positive correlations between reading span and spatial span (English, $r = .34$; Dutch, $r = .47$). However, unlike in Study 1, correlations of attachment preference with the span measures were weak in both English and Dutch samples (reading span, rs $= -.22$ for English and $-.05$ for Dutch; spatial span, rs $= -.06$ for English and $-.05$ for Dutch). A one-tailed test for differences in correlations revealed a difference in predictive strength for English (rs $= -.22$ vs. $-.06$), $t(163) = 1.82$, $p < .05$, but not for Dutch (rs $= -.05$ vs. $-.05$, *ns*).

Figure 2 shows attachment tendencies broken down by reading span group. (As before, data were not grouped for statistical analyses.) The results of Study 1 are shown on the left side of the figure and the results of Study 2 on the right to facilitate comparisons. The figure shows clearly that in Study 2, both language groups and all three reading span clusters tended to attach the relative clause high, to N1.

SEM

Item parcels for both the working memory and attachment preference tasks were created as described in the results section for Study 1.

CFAs

Using a separate sample for each language (English, $n = 166$; Dutch, $n = 126$), we tested the two models for the working memory variables, as illustrated in Figure 3. Replicating the results of Study 1, the model containing domain-general (WM_G) and verbal (WM_V) factors provided the best fit to the data: English, $\chi^2(16, N = 166) = 15.10$, CFI = 1.00, NFI = .97, RMSEA = .00; Dutch, $\chi^2(16, N = 126) = 17.50$, CFI = 1.00, NFI = .95, RMSEA = .03. (See Table 6 for fit statistics and model comparisons.) And again, a one-factor attachment preference model provided an excellent fit to the data: English, $\chi^2(2, N = 166) = 0.49$, CFI = 1.00, NFI = 1.00, RMSEA = .00; Dutch, $\chi^2(2, N = 126) = 0.15$, CFI = 1.00, NFI = 1.00, RMSEA = .00. Tables 4 and 5 display the predictor-side and criterion-side measurement models by sample (English vs. Dutch). There was some evidence for factorial invariance across the samples (see Table 6). However, as can be seen in Tables 4 and 5, the factor loadings were similar across the two samples, and for both the predictor variables and the criterion variables, the most restrictive measurement model (i.e., Model E3) provided a good fit to the data, indicating that the degree of factorial invariance was quite small.

Path Analyses

Figure 4 displays a path model, with WM_G and WM_V as predictors of attachment preference. For both samples, model fit was excellent: English, $\chi^2(48, N = 166) = 49.13$, CFI = 1.00, NFI = .95, RMSEA = .01; Dutch, $\chi^2(48, N = 126) = 54.42$, CFI = .99, NFI = .90, RMSEA = .03. Let us first consider the path results for the English sample. As in Study 1, there was a negative effect of WM_V on attachment preference ($-.26$, $p < .01$). By contrast, the effect of WM_G ($-.08$) was nonsignificant. Thus, for the English sample, there was no effect of WM_G above and beyond the effect of WM_V. A somewhat different pattern of results emerged for the Dutch sample. Specifically, effects of both WM_G ($-.12$) and WM_V ($-.02$) were near zero and nonsignificant.

We conducted additional analyses to address two issues. The first was whether the effect of WM_V on attachment preference was greater than that of WM_G. To answer this question, we compared the model reported above with one in which these path effects were constrained to be equal. In the English sample, loss of model fit was statistically significant, but for the Dutch sample, it was not: English, $\Delta\chi^2(1, N = 166) = 4.80$, $p < .05$; Dutch, $\Delta\chi^2(1, N = 126) = 0.20$, $p = .65$. Thus, for the English sample, the effect of WM_V on attachment preference was stronger than that of WM_G, but both effects were equally nonsignificant in the Dutch sample. The second question was whether there were differences across samples in effects of WM_G and WM_V on attachment preference. For example, was the effect of WM_V on attachment preference greater for English ($-.26$) than for Dutch ($-.02$)? To answer this question, we constrained the paths leading from WM_G and WM_V to attachment preference to be equal in the two samples. Surpris-

Table 6
Fit Statistics for Predictor-Side Measurement Models (Study 2)

Model	$\chi^2(df)$	CFI	NFI	RMSEA	$\Delta\chi^2(df)$
		English (N = 166)			
1: WM_G	141.16 (20)	.78	.76	.19	
2: $WM_G + WM_V$	15.10 (16)	1.00	.97	.00	126.06 (4)**
		Dutch (N = 126)			
1: WM_G	86.78 (20)	.80	.76	.16	
2: $WM_G + WM_V$	17.50 (16)	1.00	.95	.03	69.28 (4)**
		Equivalence analyses (N = 292)			
Predictor Side					
E1: No constraints	32.61 (32)	1.00	.97	.01	
E2: Factor loadings equal	54.28 (42)	.99	.94	.03	21.67 (10)*
E3: Factor loadings and variances equal	60.37 (44)	.98	.94	.04	6.09 (2)*
Criterion Side					
E1: No constraints	0.65 (4)	1.00	1.00	.00	
E2: Factor loadings equal	7.02 (7)	1.00	.99	.00	6.37 (3)
E3: Factor loadings and variances equal	17.50 (8)	.98	.97	.06	10.48 (1)*

Note. CFI = confirmatory fit index; NFI = normed fit index; RMSEA = root-mean-square error of approximation; WM_G = domain-general working memory factor; WM_V = verbal working memory factor.
* $p < .05$. ** $p < .01$.

ingly, loss of fit relative to a model in which these paths were allowed to vary was nonsignificant, $\Delta\chi^2(2, N = 292) = 4.41$, $p = .11$. Thus, we conclude that there were no differences in predictive relations in the English and Dutch samples, despite the fact that it appears that English and Dutch have different predictive relationships, especially between the significance of the verbal factors on relative clause attachment. Hence, the upcoming analyses in which Studies 1 and 2 are compared become less complex.

Discussion

Briefly, the presentation format of the ambiguous relative clause sentences in Study 2 seemed to reduce the role of working memory in determining attachment preference and resulted in a greater tendency to attach to N1 in both languages. In the following section, we assess these effects statistically and discuss their implications.

Analyses Comparing Study 1 and Study 2

Table 1 displays the descriptive statistics for a combined sample including Studies 1 and 2. Here, it can be seen that although there were cross-linguistic differences in attachment preference between English and Dutch, the difference was not as great when sentences were presented in segments. That is, the English–Dutch difference in N1% was statistically significant in Study 1 but not in Study 2. It can also be seen that the percentage N1 attachments tended to be higher in Study 2 than in Study 1.

To see whether these differences in attachment preference were statistically significant, we performed an analysis of variance, with language (English vs. Dutch) and presentation format (regular vs. chunked) as between-subjects factors. Results revealed main effects of both language, $F(1, 534) = 7.18$, $p < .01$, and presentation format, $F(1, 534) = 81.78$, $p < .01$, with no interaction ($F < 1$). Therefore, we can conclude that cross-linguistic differences in attachment preferences do exist, that visual segmentation of the sentences encourages N1 attachments, and that Dutch and English speakers' attachment preferences were equally affected by presentation format.

A comparison of the results of structural equation analyses for Studies 1 and 2 (each set of results is presented in Figure 4) suggests that the relationships between the two types of working memory (WM_G and WM_V) and attachment preferences are reduced when the ambiguous sentences are presented in segments. In English, the relationship between WM_G and attachment preference that was found to be significant in Study 1 was not significant in Study 2, and the WM_V effect found for English participants in Study 1 remained significant in Study 2, although it was smaller. Reductions in the statistical significance of the paths for the Dutch samples also appear more pronounced. Specifically, both WM_G and WM_V went from significant in Study 1 to nonsignificant in Study 2. It therefore appears that there is a moderating effect of presentation format on the effects of working memory on attachment preference: Although WM_G and WM_V explain significant amounts of attachment preference variance in the whole visual sentence presentation format, the amount of variance they explain is reduced when the presentation format is chunked.

We performed a multiple-groups analysis using SEM techniques to provide statistical support for this conclusion as well as to reinforce our earlier claims of configural and metric invariance across samples. Two models were compared for both English and Dutch. In the first model, the paths from WM_V and WM_G to relative clause attachment were freely estimated. In the second model, the path from WM_V to relative clause attachment was constrained to be equal across Study 1 and Study 2 samples. The WM_G-attachment path was also constrained to be equal across samples. A significant reduction in model fit from Model 1 to Model 2 would indicate that the two working memory factors explain different amounts of variance across the text presentation formats. For both English and Dutch, constraining the models in this manner did result in significant losses in model fit: English, $\Delta\chi^2(2, N = 316) = 6.09, p < .05$; Dutch, $\Delta\chi^2(2, N = 222) = 10.55, p < .05$. We conclude that presenting the sentences in a segmented format significantly reduces the extent to which working memory factors explain variance in attachment preference.

General Discussion

The present work had two primary aims. The first was to examine whether individual differences in working memory resources affect decisions about how to attach an ambiguous constituent. The second aim was to examine the nature of this relationship—especially with regard to its directionality and the domain generality versus specificity of working memory constraints on sentence comprehension. We discuss the implications of our results for both issues in turn.

Working Memory and Syntactic Ambiguity Resolution

Concerning the first aim, we investigated the role of working memory in resolving the relative clause attachment ambiguity (e.g., *The maid of the princess who scratched herself in public*). This ambiguity is well suited for this research question because the two alternative interpretations have similar phrase structures, essentially, noun phrase → noun phrase + relative clause; that is, the larger noun phrase contains a smaller noun phrase and a relative clause, and the difference is in where the relative clause attaches inside the smaller noun phrase. Therefore, any effects of late closure or recency (e.g., Frazier, 1979; Gibson, 1998; Kimball, 1973), which have been motivated on the basis of limitations on working memory resources, should not be masked by the influence of other structure building constraints (e.g., minimal attachment, which favors structural simplicity). Study 1 showed a clear relationship between working memory and relative clause attachment preferences. However, the direction of this relationship was the opposite of what was predicted: Participants with lower working memory spans attached ambiguous relative clauses to the distant attachment site more often than did participants with higher working memory spans.

Although this relationship between working memory and the recency preference is unexpected (e.g., Frazier, 1979; Gibson, 1998; Kimball, 1973), it is clearly not spurious. First, the result is based on a sample that can be described as huge ($N = 246$) compared with sample sizes in most psycholinguistic experiments. Moreover, the finding was shown independently in two different languages. Our observation that the pattern is identical in these particular languages is important because English has been considered to be an N2 language (Cuetos & Mitchell, 1988) and Dutch an N1 language (Brysbaert & Mitchell, 1996; Desmet, De Baecke, & Brysbaert, 2002). The conclusion that N1 and N2 languages exist is mainly based on comparing different studies using entirely different materials. Our study is the first to use close translations of the same materials and it confirms this cross-linguistic difference. Taking into account the large sample size and the replication in two languages with a different baseline preference, we believe the negative relationship between working memory resources and the recency preference should be taken seriously.

Hence, this study finds evidence that individual differences in working memory are associated with different language comprehension strategies. However, the direction of the influence is not predicted by major accounts of cognitive resource constraints on sentence comprehension. In a second study, we tried to provide an explanation for this counterintuitive finding. Given the different pattern of findings in Study 2 compared with Study 1, we can rule out an explanation based on predicate proximity (Gibson et al., 1996) or relativized relevance (Frazier & Clifton, 1996), which would claim that people with low spans are more sensitive to the prominence of N1 because of its role as argument of the main verb. We can also rule out the explanation holding that people with high spans are more sensitive to multiple modification, on the basis of referential theory (Altmann & Steedman, 1988) and the finding that people prefer not to modify a noun phrase that has been modified before (Thornton et al., 1999). These two accounts incorrectly predict the same pattern of results in Study 1 and Study 2 because the presentation format is unlikely to have changed these factors.

On the basis of the results of Study 2, we believe the most reasonable explanation of the relation between working memory capacity and attachment preference is that people with low working memory spans may tend to insert an implicit prosodic break between the complex noun phrase and the relative clause, leading to N1 attachment, whereas people with high working memory spans may tend to leave out such breaks, leading to N2 attachment. This explanation is supported by two pieces of evidence. First, when the sentences were presented in three segments (Study 2), forcing participants to put a prosodic break between N2 and the relative clause, the average N1 attachment percentage was much higher than when the sentences were presented in full (Study 1): 73% versus 53% N1 attachments averaged across languages. Second, the individual differences in attachment preferences when everyone was free to choose the prosodic pattern with which to read the sentences (Study 1) were all but eliminated when segmentation encouraged everyone to insert a prosodic break (Study 2). One could thus describe the effect of inserting breaks between the complex noun phrase and relative clause as analogous to forcing everyone to read the way individuals with low spans do; that is, everyone ended up attaching high, to N1.

So far, we have based the underlying mechanism for the effect of forcing participants to break between N2 and the relative clause on the implicit prosody hypothesis (Fodor, 1998; Jun, 2003). Indeed, silent reading has been suggested to involve implicit prosody, and such implicit prosodic representations have been shown to influence syntactic decisions (Bader, 1998; Fodor, 1998, 2002; Jun, 2003), particularly the presence of prosodic boundaries between syntactic constituents. However, an alternative underlying mechanism appeals to the sausage machine model of parsing (Frazier & Fodor, 1978) by suggesting that the effects of presentation format arose from differences in the way phrases are packaged together by people with different memory spans. When the sausage machine model was proposed, it was largely a reaction to comprehension strategy theories that did not take into account either human computational constraints or time limits on processing (Fodor, Bever, & Garrett, 1974; Kimball, 1973). Such models accounted for the processing of sentences by proposing that the parser has a discrete set of strategies at its disposal to parse the meanings of sentences. Frazier and Fodor (1978) argued that those strategies were fine in principle but that limits of human memory capacity meant that the strategies could not apply to as much input as those models proposed—certainly not in a limited amount of time. The alternative they proposed, the so-called *sausage machine*, is a two-stage model. The first stage, the preliminary phrase packager (or PPP), groups some number of constituents together. This number is supposed to be constrained by the amount of short-term memory humans have at their disposal. This phrasal or clausal unit is then shunted off to the second processing stage, the sentence structure supervisor, which computes the full syntactic analysis. One possibility, then, is that the chunking effects observed in the present study approximate what the PPP does for the sausage machine and that the amount of material that gets packaged together for processing is subject to individual differences. Once this packaging is done, the prosodic breaks that are created between each package influence the final attachment decision. Planned future studies will help determine whether this indirect account of working memory–prosodic effects is superior to the direct prosodic influences of working memory constraints.

Even though the prosodic chunking account of the effects of working memory on relative clause attachment is well supported by the data, there are some aspects of the design that might obscure this interpretation. First, the comparison between full and chunked presentation formats was necessarily a between-studies comparison. However, we do not deem this comparison to be particularly damaging given the rigorous reliability analyses that we used to make sure that the two studies showed equivalence across samples. Further complicating the cross-study comparison, however, is the fact that readers in Study 1 had only 5 s to read each sentence, whereas the readers in Study 2 had a total of 6 s. Given this difference, it could be argued that the extra second allowed readers with low spans, who often read more slowly (King & Just, 1991), to make up for reading-speed differences that may have been more fully exploited within the shorter Study 1 reading window. Although we have no reading time data, it does not seem that this line of reasoning can account for the Study 2 effects. This extra-time hypothesis predicts that it is the performance of participants with low spans that should have changed in Study 2, but in fact what we observed is that participants with high spans performed differently—specifically, they switched from N2 to N1 attachments. We therefore conclude that the chunking account is better able to accommodate our pattern of results.

Second, our results may have arisen only because we measured offline attachment preferences instead of online attachment preferences. It could be that working memory factors explained significant amounts of variance in attachment preferences in the direction we observed because the offline questions probing attachment preference allowed some postinterpretive strategies to differentially arise across span types (Waters & Caplan, 1996, 2004). This possibility could help explain why our results differed from those of Felser et al. (2003), in which children performing online tasks showed the opposite pattern of preferences on the basis of working memory span (i.e., children with low spans were more likely to attach low than were children with high spans). In fact, Traxler (in press) has found that adults with high spans performing online resolution tasks prefer high rather than low attachment. However, further exploration of this issue should be a focus of future research. Indeed, an interesting project would be to evaluate whether attachment preferences are similar in both online and offline tasks and to assess whether individual differences in working memory capacity affect performance on the two types of tasks similarly.

The Domain Specificity of Working Memory in Language Processing

The second aim of this study was to find out whether the influence of working memory resources on syntactic ambiguity resolution is domain specific or domain general. CFAs showed that the two working memory span tasks (the reading span task and the spatial span task) reflected two sorts of working memory resources: a general working memory factor, WM_G, which explained variance common to both tasks, and a verbal working memory factor, WM_V, which explained variance unique to the reading span task. A statistical model in Study 1 confirmed that both factors explained a significant amount of variance in the latent variable reflecting relative clause attachment preference. In other words, syntactic ambiguity resolution recruits the services of at least two types of working memory: domain-general working memory and domain-specific verbal working memory.

The model we have proposed has been left underspecified with regard to what it means exactly for working memory components to be domain general versus domain specific. The present study is the first to correlate a measure of syntactic ambiguity resolution with two span tasks reflecting different working memory domains and to use SEM to detect the contribution of domain-general and domain-specific working memory resources in sentence processing. A key finding of this study that should be of interest to researchers in psycholinguistics and working memory is that a general factor was observed to explain significant variance within a measure of ambiguity resolution. This is a problem for many previous models of the role of working memory constraints in sentence processing. One influential model, for instance, postulates the use of two verbal working memory systems for human sentence processing (Waters & Caplan, 1996). Subsequent studies using dual-task paradigms have shown interference patterns implicating a domain-general working memory pool that constrains language processing both offline (Gordon, Hendrick, & Levine, 2002) and online (Federenko, Gibson, & Rohde, 2006). Such dual-task results are more in line with Just and Carpenter's (1992) account, holding that language processing is bounded by the limits of a single pool of domain-general working memory resources. A major theoretical contribution of the present study is to demonstrate that a domain-general working memory factor operates in concert with a domain-specific factor to influence language comprehension processes.

Some recent theories (Christiansen & MacDonald, 1999; Farmer, Christiansen, & Kemtes, 2005; MacDonald & Christiansen, 2002) assume that a factor related to experience explains most if not all interesting individual differences in language comprehension. The most explicit statement of this hypothesis (MacDonald & Christiansen, 2002) claims that individual differences in language comprehension have roots in the interaction of biology and language-specific experience, eliminating the need to postulate any reference at all to working memory capacities. But an experience-based account of working memory and ambiguity resolution predicts that speakers of N1 attachment languages such as Dutch (Desmet, Brysbaert, & De Baecke, 2002; Desmet et al., 2006) with high spans will show preferences consistent with the statistics of their own language and hence should prefer N1 attachment. Study 1 of the present work found evidence against this experience account of WM_V, because Dutch participants with high spans did not show an N1 preference (see Figure 2). Thus, the Dutch speakers with presumably the greatest amount of experience attaching these types of relative clauses high, to N1 (Desmet et al., 2006)—namely, people with high working memory spans (Farmer et al., 2005; MacDonald & Christiansen, 2002)—preferred the less frequently experienced low attachment site (N2). Moreover, even though English-speaking participants with high spans followed the more common N2 attachment for their language, an appeal to experience has trouble explaining the finding that participants with low spans attached to N1 rather than dividing their attachment decisions more evenly between N1 and N2. It seems, therefore, that the experience-based account of individual differences in language-specific attachment preferences cannot adequately account for the entire range of results that we observed. Hence, we favor the view that the WM_V factor found in Study 1 reflects verbal working memory capacity and not merely amount of language experience.

Conclusions

The present study maximized the opportunity to discover the variance that can be attributed to individual differences by testing a large sample of participants with a wide range of working memory scores, by analyzing results using continuous rather than categorical statistical procedures, and by minimizing the number of variable manipulations to ensure that each participant went through the same experience. This study also demonstrates the power of combining experimental and psychometric approaches to address questions of cognitive architecture (Cronbach, 1957) in psycholinguistic research. The manipulation of a single variable—text format of the relative clause sentences—across two separate psychometric studies allowed us to make significant conclusions about both language processing and the nature of working memory. The first conclusion is that language comprehension is bounded by the limits of working memory capacity and that if recency is a working memory-preserving strategy, it is covered up in this study by a different sort of chunking strategy that takes place during silent reading. Second, working memory constraints on sentence comprehension have domain-general and domain-specific components. Third, working memory constrains informational chunking in parsing. An attractive explanation for this phenomenon is that working memory and chunking strategies interact during silent reading to bias attachment decisions. And finally, because we showed that cross-linguistic differences in attachment preferences are very small compared with individual differences in attachment preferences, we use this opportunity to remind researchers that the differences between individuals can be at least as interesting and informative as the differences between groups of people.

References

Abney, S. P., & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research, 20,* 233–250.

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30,* 191–238.

Baddeley, A. D. (1986). *Working memory.* London: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. D. Fodor & F. Ferreira (Eds.), *Studies in theoretical psycholinguistics*: *Vol. 21. Reanalysis in sentence processing* (pp. 1–46). Dordrecht, the Netherlands: Kluwer Academic.

Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentences parsing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*(A), 664–695.

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81–116). Hove, England: Psychology Press.

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

Carlson, K., Clifton, C., Jr., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language, 45,* 58–81.

Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language, 52,* 144–169.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 269–321). New York: Wiley.

Christiansen, M. H., & MacDonald, M. C. (1999). Fractionating linguistic working memory: Even in pebbles, it's still a soupstone. *Behavioral and Brain Sciences, 22,* 97–98.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics, 19,* 133–169.

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

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition, 30,* 73–105.

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

Desmet, T., Brysbaert, M., & De Baecke, C. (2002). The correspondence between sentence production and corpus frequencies in modifier attachment. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55*(A), 879–896.

Desmet, T., De Baecke, C., & Brysbaert, M. (2002). The influence of referential discourse context on modifier attachment in Dutch. *Memory & Cognition, 30,* 150–157.

Desmet, T., De Baecke, C., Drieghe, D., Brysbaert, M., & Vonk, W. (2006). Relative clause attachment in Dutch: On-line comprehension corresponds to corpus frequencies when lexical variables are taken into account. *Language and Cognitive Processes, 21,* 453–485.

Desmet, T., & Declercq, M. (2006). Cross-linguistic priming of syntactic hierarchical configuration information. *Journal of Memory and Language, 54,* 610–632.

De Vincenzi, M., & Job, R. (1993). Some observations on the universality of the late-closure strategy. *Journal of Psycholinguistic Research, 22,* 186–206.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128,* 309–331.

Farmer, T. A., Christiansen, M. H., & Kemtes, K. A. (2005). Sentence processing in context: The impact of experience on individual differences. In *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (CD-ROM). Mahwah, NJ: Erlbaum.

Federenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language, 54,* 541–553.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition: A Journal of Developmental Linguistics, 11,* 127–163.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27,* 285–319.

Fodor, J. D. (2002, April). Psycholinguistics cannot escape prosody. In *Proceedings of the Speech Prosody 2002 Conference*. Available from http://aune.lpl.univ-aix.fr/sp2002/papers.htm

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar.* New York: McGraw-Hill.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* Unpublished doctoral dissertation, University of Connecticut.

Frazier, L. (1987). Theories of sentence processing. In L. J. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. 291–307). Cambridge, MA: MIT Press.

Frazier, L., & Clifton, C. (1996). *Construal.* Cambridge, MA: MIT Press.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6,* 291–325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178–210.

Gaulin, C., & Campbell, T. (1994). Procedure for assessing verbal working memory in normal school age children: Some preliminary evidence. *Perceptual and Motor Skill, 79,* 55–64.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68,* 1–76.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitation and processing breakdown.* Unpublished doctoral thesis, Carnegie Mellon University, Pittsburgh, PA.

Gibson, E., Desmet, T., Watson, D., Grodner, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics, 16,* 313–353.

Gibson, E., Pearlmutter, N. J., Canseco-Gonzalez, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition, 59,* 23–59.

Gilboy, E., Sopena, J. M., Clifton, C., & Frazier, L. (1995). Argument structure and association preference in Spanish and English complex NPs. *Cognition, 54,* 131–167.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory load interference in syntactic processing. *Psychological Science, 13,* 425–430.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science, 29,* 261–290.

Hakuta, K. (1981). Grammatical description versus configurational arrangement in language acquisition: The case of relative clauses in Japanese. *Cognition, 9,* 197–236.

Igoa, J. M., Carreiras, M., & Meseguer, E. (1998). A study on late closure in Spanish: Principle-grounded versus frequency-based accounts of attachment preferences. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 51*(A), 561–592.

Jun, S. (2003). Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research, 32,* 219–249.

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., & Varma, S. (2002). A hybrid architecture for working memory: Reply to MacDonald and Christiansen (2002). *Psychological Review, 109,* 55–65.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189–217.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition, 2,* 15–47.

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

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge, England: Cambridge University Press.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford Press.

Lewis, R. L. (1993). *An architecturally-based theory of human sentence comprehension* [Doctoral thesis, Carnegie Mellon University, Pittsburgh, PA]. Available from http://www-personal.umich.edu/~rickl/Documents/thesis.pdf

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research, 25,* 93–115.

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

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24,* 56–98.

MacWhinney, B. (1987). Toward a psycholinguistically plausible parser. In S. Thomason (Ed.), *ESCOL 1986* (pp. 1–8). Columbus: Ohio State University.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 269–321). New York: Wiley.

Miller, G. A., & Isard, S. (1964). Free recall of self-embedded English sentences. *Information and Control, 7,* 292–303.

Pearl, J. (2002). *Causality: Models, reasoning, and inference.* New York: Cambridge University Press.

Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language, 34,* 521–542.

Pickering, M. J., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes, 6,* 229–259.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods, 10,* 178–192.

Roberts, R., & Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research, 31,* 573–598.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125,* 4–27.

Stabler, E. (1994). The finite connectivity of linguistic structure. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 303–336). Mahwah, NJ: Erlbaum.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation, *Journal of Psycholinguistic Research, 23,* 295–322.

Thornton, R., MacDonald, M. C., & Gil, M. (1999). Pragmatic constraint on the interpretation of complex noun phrases in Spanish and English. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1347–1365.

Traxler, M. J. (in press). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory and Cognition.*

Traxler, M. J., Pickering, M., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39,* 558–592.

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language, 53,* 204–224.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154.

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist, 30,* 128–134.

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

Waters, G. S., & Caplan, D. (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 57*(A), 129–163.

Yngve, V. H. (1960). A model and a hypothesis for language structure. *Proceedings of the American Philosophical Society, 104,* 444–466.

Appendix

Relative Clause Ambiguity Sentences (Studies 1 and 2)

The sister of the actress who shot herself on the balcony was under investigation.

The uncle of the fireman who criticized himself far too often was painting the bedroom.

The companion of the salesman who amused himself quite a bit was writing a letter to the editor.

The mother of the bride who embarrassed herself at the reception was complaining to the vicar.

The uncle of the bishop who injured himself last summer was concerned about the infection.

The brother of the mayor who complimented himself constantly bothered the reporter.

The niece of the waitress who hurt herself on the broken glass was shocked by the accident.

The daughter of the seamstress who entertained herself most evenings was reading a book.

The sister of the schoolgirl who burned herself the other day was usually very careful.

The aunt of the nun who lost herself in thought was disturbed by the noise.

The father of the surgeon who made a fool of himself at the party was greatly embarrassed.

The maid of the princess who scratched herself in public was terribly embarrassed.

The great-uncle of the policeman who treated himself after the accident was watching the news.

The mother of the prostitute who killed herself last summer had lived in Wales.

The sister of the beautician who cut herself with the old scissors phoned for a doctor.

The uncle of the general who sacrificed himself for the cause was the subject of the biography.

The grandmother of the heiress who bankrupted herself last year still made risky investments.

The nephew of the fisherman who drowned himself in the ocean didn't know about the tricky current.

The neighbor of the actor who hated himself for lying left town in a hurry.

The brother of the count who crippled himself by falling off a horse took a long time to get over it.

Received January 10, 2006
Revision received July 13, 2006
Accepted July 19, 2006 ■