Published in final edited form as:

Can J Exp Psychol. 2015 March ; 69(1): 136–155. doi:10.1037/cep0000037.

Working Memory and the Revision of Syntactic and Discourse Ambiguities

William S. Evans¹, David Caplan², Adam Ostrowski¹, Jennifer Michaud², Anthony Guarino³, and Gloria Waters¹

¹Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA

²Neuropsychology Laboratory, Department of Neurology, Massachusetts General Hospital, Boston, MA, USA.

³ Institute for Health Professions, Massachusetts General Hospital, Boston, MA, USA.

Abstract

Two hundred participants, 50 in each of four age ranges (19 – 29; 30 – 49, 50 – 69, 70 – 90) were tested for short term working memory, speed of processing and on-line processing of three types of sentences in which an initially assigned syntactic structure and/or semantic interpretation had to be revised. Self-paced reading times were longer for the segments which signaled the need for revision, and there were interactions of age and sentence type and of speed of processing and sentence type, but not of working memory and sentence type, on reading times for these segments. The results provide evidence that working memory does not support the processes that revise the structure and interpretation of sentences and discourse.

Introduction

This paper investigates the role of short-term working memory in comprehension. Short-term working memory is often conceived as an aspect of the memory system responsible for effortful, conscious, and controlled manipulation of active information. While some aspects of sentence comprehension may intuitively seem to require this type of deliberate effort (for example, making sense of a complex run-on sentence), most operations involved in typical on-line sentence comprehension are highly skilled, automatic, and unconscious. We have argued in previous work that skilled operations of this type do not rely on short-term working memory resources (e.g., Caplan and Waters, 1999, 2013). This paper explores a middle ground between these extremes of automatized processing: namely, the role of short-term working memory in contexts where the reader is required to revise an initial interpretation during on-line sentence comprehension.

We will begin with a brief overview of literature related to this topic. There is broad agreement among psychologists that one component of human memory is a short-term memory (STM) that maintains small amounts of item and order information for short

periods of time (Barroullet and Camos, 2012; Brown et al, 2007; Cowan, 2000; Henson, 1998; Jones, 1993; Jones and Macken, 1993; Jones et al 1993, 2006, 2007; Macken and Jones, 1995; Oberauer and Lewandowsky, 2011; Page and Norris, 1998; Salame and Baddeley, 1986; Verhaeghen and Basak, 2005). A longstanding question has been what role STM plays in cognitive functions. One of the points Crowder (1982) made in his famous critique of the concept of STM was that there was no convincing evidence for a functional role of STM, saying “short-term memory, as traditionally defined, is remarkably insensitive to almost everything (p. 305).”

Baddeley and Hitch (1974) addressed this issue in a series of studies in which they had participants perform memory, comprehension and reasoning tasks while retaining sequences of digits of various lengths. They found that concurrent retention of three-digit sequences had no effect on primary task performance, but that concurrent retention of six-digit led to small but discernable effects on these tasks. They argued that three digits could be retained in a specialized memory system (in their original terms, a “Phonological Loop”) that was not used in memory, comprehension and reasoning tasks, but that a “Central Executive” that was used in those tasks was needed to retain six digits and their order. They coined the term “working memory” (WM) to refer to the combination of the specialized system that retained three digits and the less specialized system that supported both retention of longer lists of items and other cognitive functions.

Since 1974, there has been extensive study of Baddeley's Working Memory model, which has resulted in many proposals regarding the nature of both the specialized system (Baddeley's Phonological Loop) and the more general-purpose component of ST-WM (Baddeley's Central Executive). We will not review this work here (for a critical review, see Caplan et al, 2012), but we will describe our use of terminology. Because we will later contrast this proposed “working memory” memory system with a different “working memory” system, we will use the term “short-term working memory” (ST-WM), rather than WM *tout court*, to refer to Baddeley's model. We will use Baddeley and Hitch's original term “Phonological Loop” to refer to the hypothesized specialized system that retains item and order information in phonological form aided by subvocal rehearsal – Baddeley has divided this component of the system into a “Phonological Store” and an “Articulatory Loop.” We will use Baddeley and Hitch's original term “Central Executive” to refer to the more general-purpose component of the system that retains information in a more abstract code.

As noted, the earliest work on ST-WM postulated a direct role for ST-WM in the process of language comprehension. There are several powerful intuitions that support this hypothesis. The time period over which information must be maintained and retrieved in language comprehension corresponds to the time scale in which ST-WM operates. Limitations in the ability to understand certain types of structures have been attributed to memory constraints (Miller and Chomsky, 1963), which in turn have been related to ST-WM capacity constraints (Lewis, 1996). Interestingly, despite his negative comments regarding possible roles for STM in cognition, Crowder (1982) suggested that “The role of short-term memory may really be in deciphering syntax. (p. 316).”

There is broad agreement that the Central Executive is critical to any role of ST-WM in supporting cognitive functions, including language comprehension. Daneman and Carpenter (1980) contrasted simple span tasks, which mostly reflected operations of the Phonological Loop, against a novel sentence span task which reflected operations of both the Phonological Loop and the Central Executive. They looked at correlations between performance on these tasks and a language comprehension task (the Verbal Scholastic Aptitude Test) and found much higher correlations for the sentence span than simple span tasks. This pattern was confirmed in a review (termed a “meta-analysis”) of 77 studies by Daneman and Merickle (1996). Just and Carpenter (1992) specifically related the Central Executive, and not the Phonological Loop, to comprehension. Many studies in the past 20 years have used complex span tasks, similar to Daneman and Carpenter's (1980) sentence span task, to measure the Central Executive or the entire ST-WM system in studies of the relation of ST-WM to comprehension.

When reviewing studies on the relationship between ST-WM and sentence comprehension, the comprehension operations which have been studied most extensively are “parsing” (the assignment of syntactic structure) and “interpretation” (determining aspects of sentence meaning such as agent, theme, and goal). Parsing and interpretation are natural operations to focus on both because they require relating words to one another over long distances (Lashley, 1952) and performing computations on the items held in memory, and are also because they are thought to be constrained by memory limitations. Several processes that contribute to performance on complex tests could contribute to maintaining information in memory in the process of parsing and interpretation, including the mechanisms that encode information in memory (Jones et al 2006, 2007; Neath, 2000; Salame and Baddeley, 1986), the stores in which information is retained in abstract multidimensional form (Page and Henson, 2001), the mechanisms that maintain serial order (Henson, 1998) and the processes that refresh representations (rehearsal: Baddeley et al, 1975; Service, 1998; and mental review: Cowan et al 2000). Similarly, mechanisms that impose limits on complex span performance, such as temporal decay (Barrouillet et al, 2004, 2007) and interference (Lewandowky and Oberauer, 2008), could set limits on parsing and interpretation. We cannot properly review the possible relations between these mechanisms and operations of the parser/interpreter here; for reviews, see Caplan et al (2012) and Caplan and Waters (2013).

Studies of the relation of ST-WM to these operations have used two approaches: 1) correlations of individual differences in ST-WM with behaviors that reflect syntactically-based comprehension; and 2) interference studies in which the effect of an ST-WM load on syntactically-based comprehension are measured. In our own studies and in several reviews of the literature, we have argued that studies of both types generally converge on a model of the role that ST-WM plays in these processes. We have argued that ST-WM is not involved in the initial, incremental processes of parsing and interpretation that generally result in successful comprehension, and that it instead plays a role in maintaining the product of these processes in memory for use in performing tasks (Waters and Caplan, 1996a; Caplan and Waters, 1999, 2012; Caplan et al, 2011; DeDe et al, 2004 Evans et al, 2011). In our reviews of this topic, we noted that many researchers have interpreted their results as providing support for the view that ST-WM supports incremental parsing and interpretation, but that

close reading of these papers reveals flaws in the arguments that led to this conclusion. Space does not allow us to re-review this literature in detail here, but we will illustrate it with a discussion of two results, which we believe are typical.

One process in which ST-WM might be involved is finding the noun that a reflexive pronoun refers to (the “antecedent” of the reflexive). Cunnings and Felser (2013) studied the role of ST-WM in this process by measuring eye fixations in reading sentences such as (1):

1a. Helen/James worked at the army hospital for many years. She/He noticed that the soldier had wounded himself . . .

b. Helen/James worked at the army hospital for many years. She/He noticed that the soldier had wounded herself . . .

Sixteen low ST-WM participants had longer first fixations and regression path durations from the reflexive (*himself, herself*) when it mismatched its “antecedent” (*the soldier*) in expected gender (1b) than when it did not (1a); sixteen high ST-WM participants did not show these effects. Both groups showed effects of gender bias of the antecedent on re-reading times. The authors argue that the early effects of gender bias of the antecedent on early eye fixations in low ST-WM participants were due to them “keep[ing] anaphoric dependencies as short as possible (p. 204)” due to their limited ST-WM. However, low ST-WM participants also showed effects of gender bias of the antecedent (*the soldier*) on early eye fixations in sentences such as (2), in which noun phrases intervened between the antecedent and the reflexive:

2. Helen has worked at the army hospital for many years. The soldier who she treated on the ward wounded herself . . .

The authors do not report this contrast directly, because their interest was in other issues (how the syntactic structural constraint on the possible antecedents of reflexives (“Principle A of binding theory”) guided parsing and interpretation (Sturt, 2003). However, their Figure 2 and discussion of other significant effects makes it clear that it was found. These results show that low ST-WM participants access more distant noun phrases early in processing reflexives, and leave the role of ST-WM in retrieving the antecedent of a reflexive unclear.

A second operation in which ST-WM might be involved is the process whereby verbs retrieve the noun phrases that fill their thematic role slots. A widely studied contrast is the verbs of English subject- and object-extracted relative clauses (*pushed* in (3)).

3. The boy who [pushed the girl]/[the girl pushed] accidentally fell

The verb of an object-relative (*the girl pushed*) makes greater demands on memory than the verb of a subject relative (*pushed the girl*), due to the need to retrieve two noun phrases at that point. The effect of ST-WM on on-line processing of object relative clauses has been explored in several studies (King and Just, 1991; MacDonald et al., 1992; MacDonald and Christiansen, 2002; Just and Carpenter, 1992; Waters and Caplan, 1996a; Caplan and Waters, 1999, 2002; Caplan et al, 2011; see Caplan and Waters, 1999, 2013, for reviews). The most detailed observations were made by Traxler et al. (2005). In a series of three experiments, and out of hundreds of eye fixation measures reported, only one – total reading

time in object relative clauses with inanimate subjects and animate head nouns -- showed differences associated with ST-WM capacity. This result was uninterpretable: the magnitude of the object- vs. subject-extraction difference was greater in the mid- than in the low-span group and not present in the high-span group. Traxler et al. concluded that the results argued against individual differences in ST-WM affecting the on-line processing of relative clauses.

Because of results such as these, we have suggested that ST-WM is not the memory system that supports on-line, incremental, successful parsing and interpretation. On the other hand, ST-WM often correlates with performance on tasks that measure comprehension (see DeDe et al, 2004, for one extensive study), leading us to suggest that it aids in maintaining the products of the on-line comprehension process in memory for later use in performing tasks. This model leaves a gap in the possible functional role of ST-WM: it does not specify whether ST-WM supports processes that revise initially assigned structures and meanings. This occurs under two circumstances. One is when skilled comprehension fails, when the structure of a sentence or a discourse is extremely complex; e.g., (4):

4. The boy who the girl who the baby hugged tripped fell.

A second is when comprehenders assign an interpretation to an ambiguity that later has to be revised.

These two cases differ in ways that may be important for the role of ST-WM. ST-WM is measured in tasks such as complex span that require conscious, controlled cognitive operations. Attempts to assign structure and meaning to sentences such as (4) intuitively involve many such operations, including repeating parts of the sentence, using formal rules of grammar to assign structure, etc. It would not be surprising to find that ST-WM supports the memory demands of such strategies. Re-analysis of ambiguities is a middle case – some re-analyses are easy and seem to occur without conscious consideration, while others are difficult. The research reported here investigates the role of ST-WM in revising ambiguities that are difficult and/or that intuitively seem to involve conscious processes.

There is a small literature on ST-WM and processing ambiguities. Effects of ST-WM have been reported on end-of-sentence measures such as interpretation preferences of ambiguous relative clause attachments (Swets et al., 2007) and comprehension of “early closure” sentences (Christianson et al., 2006; see below), but these results may reflect a role for ST-WM in the ability to retain the meaning of a presented sentence in memory, not in the memory operations of the comprehension process itself. We review studies that report on-line measures here.

Kemper et al. (2004, exp 2) studied the effects of both age and ST-WM on eye fixations in “reduced relative clause (RR)” sentences such as (5) and the corresponding “main verb (MV)” sentences (6):

5. RR: The experienced soldiers warned about the dangers conducted the midnight raid.

6. MV: The experienced soldiers warned about the dangers before the midnight raid.

Low- but not high-span participants made more regressions from the disambiguating region of both (5) and (6) (conducted the midnight raid/before the midnight raid) to the sentence subject. They also showed longer regression fixation durations and total reading times on the subject than in matched unambiguous sentences. The effect of ST-WM in (6) argues against a role of ST-WM in revising initially assigned interpretations, since (6) is disambiguated towards its preferred meaning and therefore does not require revision. The results are consistent with parallel processing of both meanings of these ambiguous sentences, and with maintaining both representations leading to a memory load that exceeded the ST-WM capacity of low WM participants. However, these results contrast with those of MacDonald et al (1992), who found no significant differences between low- and high-capacity participants' reading times for the final phrases of these sentences.

Effects of ST-WM have also been reported in the on-line use of information about the plausibility of events in processing syntactic ambiguities. Pearlmutter and MacDonald (1995) studied sentences such as 7:

7. The soup cooked in the pot but was not ready to eat.

which are ambiguous at *cooked* (Transitive: *The soup cooked the vegetables*; Intransitive: *The soup cooked slowly*; Passive participle: *The soup cooked in the pot was delicious*). High-span, but not low-span, participants took longer to read the word *in* in ambiguous compared to unambiguous control sentences (*The soup bubbled in the pot . . .*) and the plausibility of *cooked* as a passive participle predicted high-span, but not low-span, participants' reading times at the word *in*. Long and Prat (2008) found that high-, but not mid- or low-, ST-WM participants had longer self-paced reading times on the prepositional phrase (*for the party*) in "reduced relative clause" sentences such as (8) in which the verb had a past tense compared to a past participle bias (*tossed* vs. *mixed*):

8. The salad tossed/mixed for the party looked delicious

However, this effect of ST-WM was eliminated when controlling for the plausibility of the sentence initial noun as the agent or the theme of the verb, indicating that ST-WM affected the ability to use plausibility information, not information about the statistical properties of lexical forms, to resolve ambiguities on-line. The effects in both these studies appear to reflect differences in the use of plausibility information in the initial assignment of the structure and meaning of these sentences, not in revision of these ambiguities. It may be that this relationship is due to the strong relation of ST-WM to measurements of "fluid intelligence" (Kane and Engle, 2004), and not in the ability of ST-WM to encode, retrieve, or integrate linguistic structure *per se*.

To summarize, the literature provides evidence for a role of ST-WM in utilizing plausibility information in determining the meaning of an ambiguous structure on-line. There are inconsistent results regarding individual differences in ST-WM affecting encoding and/or maintaining multiple interpretations of ambiguous reduced relative/main clause sentences. The role of ST-WM in revising ambiguities remains unclear.

In the studies reported here, we examined the role of ST-WM in ambiguity resolution, and also the effects of speed of processing and age on these processes. The reason for studying

speed of processing is that it is a major determinant of ST-WM capacity (Salthouse, 1996), and any effect of ST-WM on a function may be largely or entirely due to speed of processing, not the memory abilities measured in ST-WM tasks. Despite the strong relation of speed of processing to ST-WM, its role in comprehension has received little attention (see Caplan et al, 2011, for one study related to the current work) and its role in revising the interpretation of ambiguities has not been studied.

The inclusion of age greatly expands the range of individual differences in ST-WM capacity in our participants here because ST-WM declines with age. However, in addition to changes in ST-WM, other changes in processing ability may also occur during the aging process. Therefore, we also investigated the unique effects of aging on ambiguity resolution.

We studied the effects of ST-WM, age, and speed of processing in three linguistic contexts where initially-assigned structures and meaning need to be revised. Experiment 1 investigated the reassignment of the antecedent of a pronoun when world knowledge necessitates a revision. Experiment 2 investigated the “Repeated Name Penalty (RNP)” (Gordon et al., 1993; Gordon and Hendrick, 1998), in which a repeated name is linked to a previously mentioned individual, rather than introducing a new entity into the discourse. Experiment 3 examined early closure (EC) sentences, in which an initial set of thematic roles has to be revised, leading to the re-structuring of a previously assigned syntactic representation. These three cases sample a broad range of revisions, because they are triggered by pragmatic, discourse, and syntactic factors, respectively. In each instance, we asked whether the retrieval operations and changes to previously assigned syntactic and/or semantic structures were moderated by individual differences in ST-WM capacity, whether this moderating effect was mediated by differences in age or speed of processing, and also whether age and speed of processing produced moderating effects on these processes in their own right.

Our logic is as follows: if ST-WM is used in the reanalysis operations studied here, then ST-WM should show a moderating effect on on-line processing measures at points of revision. If the role of ST-WM is supportive or facilitative in these contexts, then better ST-WM capacity should be associated with reduced revision costs. Also, if ST-WM effects exist and are not due to differences in speed of processing, then any moderating effects of ST-WM should remain when controlling for speed of processing.

Methods

Participants

Two hundred and thirty-four participants enrolled in all experiments. To achieve equal representation across the age range, roughly equal numbers of participants were recruited in four age groups: 70 –90 years (N 59), 50 – 69 years (N 63), 30 – 49 years (N 53), and 18 – 29 years (N 59).

Thirty-four participants did not complete the study for scheduling reasons. The remaining 200 individuals (a total of 50 from each age group) participated in the current experiments, and were also those reported on in Caplan et al (2011). Participants were recruited from

Boston University, the Harvard Cooperative Program on Aging, and ads placed on Craigslist and local newspapers. Participants had no current evidence and/or history of central nervous system disease, substance abuse, learning disabilities, psychiatric illness, or serious medical illness, and were not taking medications known to have centrally acting effects. All participants had normal or corrected- to-normal vision. All participants completed the WAIS III Vocabulary Test and those over the age of 50 were screened for cognitive function using the Mini Mental State Exam (Folstein, Folstein, & McHugh, 1975), with all participants who completed testing performing above the cut score of 25/30.¹

Procedure

Assessment of working memory capacity

Participants completed three tests of working memory capacity: alphabet span, subtract 2 span, and sentence span. These tasks all have good internal consistency, and a composite measure based on these tasks (total sum of item scores) had good test-retest reliability and resulted in stable classification of participants into discrete working memory groups (Waters & Caplan, 2003). In alphabet span, participants repeated a series of monosyllabic, unrelated words after rearranging them in alphabetical order. In subtract 2 span, participants repeated a random sequence of digits after subtracting 2 from each. In sentence span (Waters and Caplan, 1996b), participants read sets of syntactically simple sentences (cleft subject) and decided whether or not each sentence was acceptable. After seeing all of the sentences in a set, they were required to recall the final words of all of the sentences in the set. The sentence final recall measure served as the principal measure of sentence span (reaction times and accuracy on the sentence processing component of the task were also measured).

There were five trials at each span size. Participants were required to repeat all of the items in a trial in the correct serial order to obtain credit for the trial. If they did not know what response to make in a given position, they were instructed to say “blank.” Correct items in the correct location were scored as correct. On the alphabet and subtract 2 span, testing began at span size 2 and continued through span size 8. On sentence span, testing began at span size 2 and continued and through span size 6 to 8, with continued administration of higher span sizes contingent upon accuracy on at least 3 out of 5 trials of the prior span size. The total number of items correctly recalled was recorded (item measure). Additional details are reported in Caplan et al. (2011).

Assessment of speed of processing

Participants completed four different measures of processing speed: digit copying, boxes, pattern comparison, and letter comparison. These tasks are frequently used in the aging literature and have been shown to have good reliability in young and old adult samples (Earles et al., 1997; Salthouse et al, 1991).

¹The current study and Caplan et al. (2011) were part of a series of studies of the same set of participants. Caplan et al. (2011) incorrectly reported that “only participants with normal hearing in the speech frequency range were included” in these studies. Participants over the age of 50 were screened for hearing using pure-tone audiometry at 500, 1000, 2000, and 4000 Hz. Over 80% of the 70 to 90 year old group failed to meet a screening cut-off for normal hearing of 30DB HL. With the exception of two of the ST-WM measures discussed below (alphabet span, subtract 2 span), however, all stimuli were presented visually, and all participants were included in this study.

For digit copying, participants were presented with a sheet with 10 rows of 10 pairs of boxes with a digit in the top box and nothing in the bottom box. They were required to copy the digit from the top box to the one below. For boxes, participants were presented with a sheet with 10 rows of 10 three-sided squares, each with one side open, and were required to draw a line across the open side to make a closed box. For pattern comparison, participants were presented with 60 pairs of geometric patterns and were required to indicate whether the two members of each pair were the same or different. For letter comparison, participants were asked to decide whether pairs of letter strings 3, 6, or 9 letters in length were the same or different. A higher score indicates faster speed of processing (more items completed in the available time). Additional details are reported in Caplan et al. (2011).

Results

Relations between individual differences measures

Data regarding group performance on general psychological, ST-WM and speed of processing measures are reported in Caplan et al. (2011).

To determine whether the measures of ST-WM and speed of processing measured different capacities, confirmatory factor analyses were performed on the results of the ST-WM and speed of processing. Because of the strong correlations of subtract 2 span with several speed of processing measures, subtract 2 span was dropped from the analysis, leaving two measures of ST-WM. Based on low indicator reliability ($r = 0.23$), boxes was dropped from the analysis, leaving three measures of speed of processing. The resulting confirmatory factor analyses (Figure 1) showed good assignment of the ST-WM and speed of processing tests to different factors. For regression analyses, age was converted to a z score, and ST-WM and speed of processing scores were calculated for each subject as the average standardized item scores of the remaining relevant subtests of each type, giving each factor a mean of 0 and standard deviation of 1 while controlling for differences in the number of tests and total number of items between factors. These standardized summed item measures were used as factor scores in regression models based on recommendations that differential weightings (i.e., latent factors) generally provide little added analytical benefit in this approach (Pedhazur & Schmelkin, 2013, p.625-626).

Experiment 1

Materials

Experiment 1 was a phrase-by-phrase self-paced reading study that tested the ability of subjects to determine the antecedent of a pronoun through a reasoning process, based on Gordon and Searce (1995). Subjects read passages such as (10) and answered questions about them.

10)

P1/P2: Meg wanted Kate to look over/ some important papers./

P3/P4: Meg had to mail Kate/ the documents by Monday./

P5/P6: Unfortunately she/ never [sent/received] the documents./

P7/P8: As a result the whole deal/ fell behind schedule./

The critical segment is phrase 6 (P6), which was presented in a Continuation (“sent”) and a Shift (“received”) version. Prior to P6, syntactic prominence promotes the subject of the initial sentence (“Meg”) as the antecedent of “she” and this assignment of co-reference is maintained in the Continuation condition. However, in the Shift condition, situational constraints require that the anaphor “she” refer to “Kate.” This requires breaking the co-reference between “she” from “Meg” and instead retrieving “Kate” as the revised antecedent. The initial antecedent must also be demoted from its status as the preferred antecedent in this process, or else the pragmatically inappropriate interpretation would continue to be assigned. The difference between reading times in the Shift and Continuation conditions is therefore a measure of this additional processing (as well as a reflection of other potential operations such as recognizing that the original interpretation was implausible, making the decision to abandon it, etc.).

The stimuli in this experiment consisted of 3 practice passages, 24 experimental passages and 48 unrelated filler passages. All passages were four sentences in length, with each sentence split into two viewing segments. Passages were normed in a separate written completion study with 20 subjects to verify the presumed subject-biased interpretation of each passage prior to the critical region P6. Written responses were scored on their subject vs. object bias by 2 independent raters, and of 34 passages normed, the final 24 selected had a mean subject-bias interpretation of 89% (SD 7.6).

Passages were split evenly across two lists, with an equal number of each condition in each list. Each participant saw only one list, and presentation order was counterbalanced between this experiment and experiment 2, as it also included short passages and anaphor manipulation.

Procedure

Sentences were presented phrase by phrase (shown in (7)) on a laptop computer screen, with each sentence broken into two segments, using Linger self-paced reading software (Rohde, 2003; <http://tedlab.mit.edu/dr/Linger/>). Sentences were initially represented as a series of dashes, with each dash corresponding in length to an individual word in the sentence. Participants then paced their way through each stimulus by pressing a key to reveal each successive phrase. At the first key press, the dashes corresponding to the first phrase disappeared and were replaced by the words themselves. When they pressed the key again, the first phrase was replaced by dashes, and the second phrase was revealed.

Each passage was followed by a simple yes/no question presented as a whole sentence. Reaction times for key presses and accuracy on questions were recorded.

Statistical technique

Accuracy and reading time data were analyzed using linear mixed-effects models with crossed subject and item random effects. This class of models (also at times referred to as “hierarchical regression” or “multi-level regression”) has recently been gaining popularity in psycholinguistics research as an alternative to repeated measures-ANOVA. In the traditional

repeated-measures ANOVA approach (Clark, 1973), fixed effects of interest (e.g., the Shift vs. Continuation condition contrast in the current experiment) are evaluated in 2 separate models: one with by-subject random effects (F_1) and a second with by-item random effects (F_2). The first allows generalization from the subject sample to the general population, while the second allows generalization to the ‘population’ of items (in this case, all possible words or sentences with the same properties). The two models are then crossed via a statistic such as $\text{min-}F$ or merely treated as independent tests of significance ($F_1 \times F_2$; Barr et al, 2013). In contrast, mixed-effects models allow simultaneous modeling of these random subject and item effects and cross them in a single model, making both subject and item-specific adjustments to estimates at the level of individual trial data (Baayen et al, 2008). They possess a number of advantages: they are a form of generalized linear model, and therefore allow the use of continuous variables for fixed effects instead of requiring categorical grouping variables for measures such as ST-WM; they allow better handling of binary outcome variables via logistic regression techniques (Jaeger, 2008); they are more robust to missing data and unbalanced designs (Baayen, Davidson, & Bates, 2008); finally, they are claimed to be more powerful than traditional techniques when their random effects are appropriately specified (Barr et al, 2013).

Mixed-effects models were applied in the current study as follows: question accuracy data were evaluated using mixed logistic regression models (Jaeger, 2008), while reading time data were analyzed using linear mixed effect models (Baayen, Davidson, & Bates, 2008). In all cases, models were run with the “maximal” crossed random effects structure for subjects and items justified by the design (for discussion and recommendations, please see Barr et al., 2013). Subject effects were modeled using random intercepts and random slopes for condition, and where appropriate, random slopes for the interaction between condition and ST-WM, age, or speed of processing. Items effects were modeled with random intercepts.

Given the large sample size employed, p -values were derived from the relevant t and z scores by assuming an approximate normal distribution, and used a .05 criterion (Baayen et al, 2008; Levy et al, 2013). Analyses were conducted using R statistical software (R Development Core Team, 2008) and version 0.999375-37 of the lme4 package (Bates, 2005). For clarity, model specifications in lme4 R-style syntax are included in the appropriate tables.

In addition to standard p -values, we also report likelihood ratios (“bits of evidence”) for models that tested for interaction effects, per the methods of Lawrence and Klein (2013). These ratios are generated by calculating AIC-adjusted likelihoods for each model of interest and its ‘restricted’ counterpart (i.e., the same model run without the fixed effect interaction term), then calculating the likelihood ratio and converting to log-base-2 scale. Lawrence and Klein state that a positive ratio may be interpreted as evidence for the interaction, while a negative ratio is evidence for a null effect. While one of the benefits of this approach is that it provides an objective measure of evidence independent of categorical thresholds for significance, they cite rough interpretive guidelines provided by Royal (1997): per his recommendations, a ratio of ± 3 bits may be considered “fairly strong evidence”, whereas a ratio of ± 5 bits may be considered “strong evidence”².

Accuracy

Table 1 presents mean comprehension accuracy and region reading times by condition. We first examined the effects of condition, ST-WM, speed of processing, and age on question response accuracy (Table 2). Mixed logistic regression found no effects of condition on accuracy. Given the level of co-linearity between subject-level factors, we tested the relationship between continuous measures of ST-WM, speed of processing, age, and effects of condition in three separate models: each model included an interaction between an individual subject-level factor (e.g., ST-WM) and condition, controlling for the other subject-level factors (e.g., speed of processing and age). Therefore, these models each tested for the unique (Type III) moderating effect of a subject-level effect on condition. These models did not reveal any significant interactions between subject-level factors and condition on accuracy (Table 2). For comparison, we also ran a second set of interaction models without including other subject-level effects (e.g., testing the interaction of ST-WM and condition on accuracy without controlling for speed of processing and age). These models therefore tested the moderating effect of each subject-level effect without partialing out any shared covariance. In each case, model results for all interaction terms remained non-significant (models not reported), indicating that these null effects were not merely due to co-linearity.

The condition effect in the above models used standard treatment (“dummy variable”) coding, and therefore the simple effects may not be interpreted as main effects in the presence of an interaction term (Cohen et al., 2003). As a result, we also ran a main effects-only model without any interaction terms looking for effects of condition, ST-WM, speed of processing, and age; this model found a positive main effect only for speed of processing on accuracy (Table 2).

Reading Times

A floor was set at 150 ms for each phrase region, and outliers were trimmed at ± 2.5 standard deviations by subject and phrase after controlling for phrase length (affecting 3.1% of data overall). The remaining reading times on accurate trials were analyzed separately by region using linear mixed effects models. Analyses of the critical disambiguating P6 and the following P7 revealed significant main effects of condition (p 's $< .01$), with the Shift condition taking longer to read than the Continuation condition in both cases. There were no effects of condition in any other region (p 's $> .05$ in all models; not reported).

Relationships between subject-level factors and condition on reading times for P6 were analyzed as in the accuracy analyses, but using linear instead of logistic regression due to the continuous outcome variable. The main effects-only models revealed negative main effects for ST-WM and speed of processing, and a nonsignificant positive effect of age. Again, interaction models were run crossing one subject-level (e.g., ST-WM) with condition and controlling for the others (Table 2, Figure 2). These models found the following: there was a significant interaction between condition and age, such that the effect of condition

²Royal (1997) recommended thinking about likelihood ratios in log-base-2 form due to their interpretive appeal: the “strong” ratio of 5 bits is equivalent to flipping a coin and having it come up heads 5 times in a row (odds of 32 to 1, given a balanced 2-sided coin).

increased as age increased (i.e., older individuals tended to show longer reading times in the Shift compared to Continuation condition). There was also a significant interaction between condition and speed of processing, such that the effect of condition was reduced as speed of processing increased. The interaction between ST-WM and condition was not significant ($p = .95$). The log-base-2 likelihood ratio for the interaction of ST-WM and condition was -2.88 (in other words: given the observed data, a null interaction effect is approximately 7.4 times more likely than a significant one). A separate set of interaction models run without controlling for other subject-level covariates showed the same pattern of significance for all 3 interactions (not reported).

Discussion

Results from reading time measurements replicate the findings of Gordon and Searce (1995), in that reading times increased when semantic constraints forced reanalysis of a pronoun from a syntactically prominent to a non-prominent antecedent.

Faster speed of processing was associated with improved accuracy. The main effect of speed of processing and not ST-WM on accuracy suggests that effects of ST-WM on comprehension accuracy reported in the literature may in part be due to speed of processing in some cases.

The self-paced reading data show expected main effects of ST-WM and speed of processing. The absence of a main effect of age indicates that across these sentence types, effects of age are attributable to related differences in ST-WM and speed of processing. Interactions of sentence type and the factors of ST-WM, speed of processing and age in the self-paced reading times speak to whether these factors influence the operations found in the Shift condition – detaching the antecedent from the pronoun, demoting the original antecedent as the preferred antecedent, finding a different antecedent and attaching it to the pronoun. We begin with consideration of the results for ST-WM.

ST-WM did not interact with sentence condition, and the likelihood ratio supports this null effect. This null effect is unlikely to be due to inadequate power, as the variance in ST-WM was similar to that of speed of processing and age, and only the estimated effect size was smaller. It is also not likely to be due to ceiling or floor effects. The possibility that it is due to the wide age range of participants in these studies also seems unlikely; on-line effects predicted by the ST-WM hypothesis in young individuals should only increase with age. This null effect may be related to the current use of continuous samples: the more common approach of employing extreme-groups analyses has been criticized for magnifying statistical significance by inflating effect sizes (Preacher et al., 2005).

The interaction between speed of processing and condition points to a benefit of faster general speed of processing on the aspects of processing found in the shift condition. This suggests that measurements of speed of processing made in the lab are not material- and task-specific, but capture a more broadly applicable aspect of cognitive function.

The interaction between age and condition indicates that the aspects of processing found in the Shift condition become less efficient with age. Controlling for ST-WM or speed of

processing did not eliminate the effect of age on Shift reading times, indicating that the age-related decline in the efficiency of these processes is not due to age-related declines in either ST-WM or speed of processing. A candidate for a process that could be affected by age above and beyond age-related declines in speed of processing or ST-WM is the ability to demote the initial antecedent from its position of prominence in the memory set that is contacted by the retrieval cue set up by the pronoun. This process could be affected by a lessened ability of older individuals to inhibit representations that have been activated (Hasher et al., 2007).

Experiment 2

Materials

Experiment 2 examined the effects of age, ST-WM and speed of processing on the repeated name penalty. Proper nouns (names) are generally used to introduce new characters into a discourse; using a proper noun to refer to a character already mentioned in a discourse therefore requires a re-analysis of their reference. The revision needed in this case differs from that examined in Experiment 1 in two respects. First, the error signal that triggers the need for a revision is a due to a violation of the conventions of discourse, not possible events in the real world. The revision is therefore a result of more “linguistic”-type knowledge than in Experiment 1. Second, the revision itself requires removing an entity from the discourse, not re-ranking possible referents of an anaphoric phrase. If the effects of age seen in Experiment 1 are due to demoting an initial antecedent from its position of prominence, they would not be expected to be seen in Experiment 2, where this process does not take place.

The materials in this task were based on the work of Gordon et al. (1993), and consisted of 36 passages each 3 sentences in length, as in (11). In the first sentence of each passage, two characters were introduced by name. In the second sentence, the references to both characters were systematically varied between name and pronoun, resulting in four versions: Name-name, pronoun-pronoun, pronoun-name, or name- pronoun. The third sentence of each passage was a coherent continuation that did not mention either character, and each passage was followed by a yes-no question.

11)

S1: Sue gave Liz a pet hamster.

S2: [**Sue/She**] told [**Liz/her**] that hamsters eat vegetable scraps and live about four years.

S3: Giving a pet as a gift can be an imposition.

Gordon and his colleagues (Gordon et al., 1993; Gordon and Hendrick, 1998; Camblin et al., 2007) found longer reading times for proper nouns compared to pronouns in subject position of *S2*, which they explained in the context of Centering Theory (Grosz et al., 1995, Joshi et al., 2006). According to Centering Theory, local discourse coherence is maintained via the mechanisms of forward-looking and backward-looking centers. Forward-looking centers are ranked on the basis of a variety of features, one of which is their grammatical role; discourse representation theory (Kamp and Reyle, 1993) maintains that subjects outrank objects.

Centering Theory maintains that pronouns are backward-looking centers that begin their search for an appropriate antecedent starting with the most prominent forward-looking center. Gordon and his colleagues argued that names initially introduce new discourse entities, and that establishing an anaphoric reference for a repeated name involves removing the entity introduced into the discourse representation upon encountering a name, and finding an antecedent for the repeated name. The search for an antecedent begins with the least prominent discourse entity and eventuates in finding the highest ranked forward-looking center.

Procedure

Passages were presented sentence-by-sentence on a laptop computer screen using Psyscope experimental software (Cohen et al., 1993). Participants pressed a button on a button box to see each successive sentence in the passage, and in response to the yes-no question. Reading times for each sentence, and question accuracy for each passage were recorded. The stimuli were split into 4 lists, such that only one version of a passage appeared in each list, and each list contained 9 passages of each condition. In addition to the experimental stimuli, 4 practice passages and 72 filler passages were included for a total of 112 passages in each list, and each participant saw only one list. Practice and filler passages were also each three sentences in length and followed by a question, but differed in structure and content from the experimental stimuli (e.g., contained only one character, contained various constructions, used personal pronouns, etc). Lists were presented in a pseudo-randomized order unique to each participant, such that no more than 3 fillers or 3 experimental stimuli appeared in a row.

Results

All analyses were conducted as in experiment 1, with the exception of the following changes: in this experiment, the critical sentence condition contrast was tested by looking at 2-way interactions between referent type (name vs. pronoun) and sentence position (subject, object), and models investigating the role of subject-level factors therefore employed 3-way interactions. Therefore, subject effects were modeled using random intercepts and random slopes for this 2-way interaction, and where appropriate, random slopes for the 3-way interaction between sentence-level effects and ST-WM, age, or speed of processing.

Given the complexity of these random effect structures, in some instances these models were unable to converge on a unique model specification (generally in cases where very small fixed effects were fit to binomial data). In such cases, simpler models with crossed random subject and item intercepts were run and reported, with these alternate random effect structures provided explicitly in Table 4. Significant result in these models should be interpreted with some caution, however, as this class of model specification has been criticized for being anti-conservative in certain contexts (Barr et al, 2013).

Accuracy

Table 3 presents mean comprehension accuracy and sentence reading times by condition. There were no main effects of or interactions between referent type and sentence position. A

main effects-only model on the effects of ST-WM, speed of processing, age, and condition found positive main effects for ST-WM and Age, but not speed of processing. There were no interactions between age, ST-WM or speed of processing and condition in models with subject-level covariates (Table 4) or in a second set of models run without these covariates (all p 's $> .05$; not reported).

Reading Time

A floor was set at 800 ms for each whole-sentence response time, and outliers were trimmed at ± 2.5 standard deviations by subject and sentence after controlling for total sentence length in characters (affecting 2% of data overall).

Reading times were analyzed separately by sentence. Table 4 presents model results for the critical sentence 2. There were significant main effects for referent type in both the subject and object positions, with repeated names taking longer to read than pronouns, with no interaction between referent type and position. Examination of sentences 1 and 3 did not reveal any effects of condition (p 's $> .05$; not reported). A main effects-only model on sentence 2 showed significant effects of ST-WM and speed of processing but not of age when controlling for other factors, with higher ST-WM and speed of processing associated with shorter reading times overall. There were no significant 2-way interactions involving referent type in subject position and either age, ST-WM or speed of processing. The log-base-2 likelihood ratio for the interaction of ST-WM and condition was -3.96. There were no significant 3-way interactions involving referent type in subject and object position and either age, ST-WM or speed of processing in any models with subject-level covariates (Table 4, figure 3) or in a second set of models run without these covariates (not reported; all p 's $> .05$).

Discussion

The increased accuracy with ST-WM is expected. The increased accuracy with age may be due to a more conservative decision making process in aging, which has been documented in two-alternative forced-choice tasks such as the verification task used here (Ratcliff et al. 2006, 2010).

The reading time data show a repeated name penalty in object as well as subject position. Previous work (e.g., Gordon et al, 1993; Kennison and Gordon, 1997) has found repeated name effects in subject but not object position. One potential explanation for this difference could be power considerations: the studies by Gordon and his colleagues each had 12-20 subjects while the current study had 200. However, the current study used mixed effect models, while previous work used by-subjects and by-items ANOVAs. To rule out methodological differences, we also ran by-subjects and by-items repeated measures ANOVAs for the critical sentence 2: these models showed significant main effects for subject position ($F_1(1, 199) = 30.1, p < .0001$; $F_2(1, 35) = 19.1, p = .0001$), and for object position in the by-subjects analysis ($F_1(1, 199) = 11.4, p < .001$; $F_2(1, 35) = 3.5, p = .07$), and no interaction for reference type and sentence position for either the subjects or items analyses ($F_1(1, 199) = 0.1, p = .8$; $F_2(1, 35) = .6, p = .44$). The possibility that the lack of interaction effect in our study was due to age differences in our sample not present in

Gordon (1993) and Kennison and Gordon (1997) is unlikely, because an age effect would have been manifest in a 3-way interaction between the sentence conditions and age for this experiment, which we did not find. As a check, we ran the multi-level models on the college-aged subset of participants ($n=50$), with the same null interaction findings. The repeated name penalty in object position thus appears to be real. This is consistent with the principles of Centering Theory applying to pronominal and nominal anaphors in object as well as subject position.

There were no interactions of age, speed of processing or ST-WM and referent type. As with the results of experiment 1, the absence of an effect of ST-WM is not likely to be due to inadequate power, ceiling or floor effects, or inclusion of older participants. The likelihood ratio provides evidence for the null effect. This suggests that the operations required to establish the anaphoric relation of the repeated name to its antecedent do not employ the memory system measured by ST-WM tasks, as was the case for operations needed to re-assign the antecedent of a pronoun on pragmatic grounds in Experiment 1.

Though the operations in experiments 1 and 2 both involve revising the antecedent of an anaphor, they differ in the two studies. Establishing the antecedent of a repeated anaphoric name involves deleting an entity introduced into the discourse, while the “shift” case does not involve deleting an entity in the discourse, but requires demoting prominence or changing the criteria of the search. The repeated name case involves searching for an antecedent beginning with entity that is ranked lowest in the discourse at the time of the search, while the “shift” case involves searching for the entity that is ranked highest in the discourse at the time of the search. The results of experiments 1 and 2 thus indicate that ST-WM does not support a variety of operations and search processes involved in revision of the establishment of the referents of anaphoric expressions.

There was no interaction of age and anaphor type. As mentioned above, we suggested that the reason for the effect of age on revision of antecedent assignment in experiment 1 may have been an age-related decline in inhibitory function, which could have affected the ability to demote an entity in the discourse to a lower position. As noted, the operations involved in establishing the antecedent of the anaphoric repeated noun involve removal of a representation from the discourse, not demotion of an entity. This may account for the absence of an interaction of age and anaphor type in experiment 2.

There was no interaction of speed of processing and anaphor type. This suggests that variation in speed of processing has limited effect on the operations involved in revising the reference of a repeated name. The contrast with the result in experiment 1 point to differences in the way speed of processing affects different operations in the comprehension process, which remain to be explored.

Experiment 3

Materials

Experiment 3 examined the effects of age, ST-WM and speed of processing on the interpretation of “early closure (EC)” sentences (12), compared to “late closure (LC)” sentences (13).

12. EC: After the hunter shot the deer ran away

13. LC: After the hunter shot the deer it ran away

In (12) and (13), *the deer* is initially attached as the object of *shot*. In (12), when the disambiguating segment (*ran away*) is encountered, *the deer* must be re-analyzed as the subject of the matrix clause. This revision is a “non-monotonic” re-analysis, in which a constituent is attached higher in the second structure that is constructed than in the first. It has been suggested that nonmonotonic re-analysis leads to difficult garden path effects (Sturt and Crocker, 1996).

Re-analysis of EC sentences differs from the re-analyses in both Experiments 1 and 2 because it requires revision of syntactic structure, not simply revision of assignment of reference as is required in experiments 1 and 2. It shares some features with the previous experiments. The need for re-analysis is triggered by a syntactic factor – the fact that the main verb requires a subject – and therefore has similarities to the re-analysis required in Experiment 2, which is also triggered by an anomaly at a different linguistic level. The required syntactic revision involves raising a noun phrase (*the deer*) from a lower syntactic position (object of the verb of an adverbial clause) to a higher syntactic position (subject of the main verb), which has similarities to the re-ordering of the ranking of possible antecedents of a pronoun in Experiment 1. If our account of the reason for the effect of age in Experiment 1 is correct, we expect to see effects of age on this reanalysis as well.

There are data on the effect of age and ST-WM on processing of EC sentences. Christianson et al. (2006) studied the effects of age and ST-WM on interpreting (12) meaning as *the hunter shot the deer*. They varied the nature of the verb, presenting sentences with optionally transitive verbs (*shot*) and verbs such as *dress* which, when used as an intransitive, convey the “reflexive” meaning that the sentence subject is the verb theme (e.g., that *Anna dressed* means that *Anna dressed herself*). They found that older participants were less likely than young participants to endorse *Anna dressed herself* as the meaning of *After Anna dressed the girl left*. There was a significant positive correlation between ST-WM and rate of endorsement of the reflexive interpretation of ambiguous sentences in older subjects. The authors argued that these results indicate that endorsing a reflexive meaning requires re-analyzing the verbatim form of the sentence, and concluded that this process involves ST-WM, which declines with age. However, as noted above and as the authors say, the memory operations that were implicated are ones that were triggered by the need to answer a question about the sentence and therefore involve a memory system that may not be used when revisions are made on-line at the point at which an interpretation of an ambiguity is recognized as requiring revision.

In the present study, participants read sentences such as (12) and (13) in word-by-word self-paced reading and answered questions about them. The stimuli consisted of 36 sentence pairs, based on Ferreira and Henderson (1998). The first sentence of each pair was an early closure (EC) sentence, in which the initially preferred syntactic parse required reanalysis. The second sentence of each pair was a locally ambiguous late closure (LC) sentence that did not require reanalysis. Both sentences were identical in early regions.

Procedure

Sentences were presented using the self-paced reading methods in experiment 1, except that in this case, each key press revealed a single word in a sentence instead of a phrase in a short passage. On one third of trials, the sentence was followed by a simple yes/no question presented as a whole sentence. Reaction times for key presses and accuracy on questions were recorded.

Stimuli pairs were divided into two lists, such that each list contained 18 sentences of each condition, and the sentences of each pair were presented in separate lists. In addition to the 36 experimental stimuli, 100 filler sentences (including the relative clause/ sentential complement sentences presented in Caplan et al., 2011) were included in each list. Sentences were presented in the same pseudo-randomized order within a list for all participants, such that no more than three experimental items were seen in a row. Each participant saw both lists of stimuli, and both lists were presented during separate testing sessions at least two weeks apart. The presentation order of each list was counterbalanced across participants.

Results

Data were analyzed as in experiment 1, but with the following changes: since all versions of all items were seen by each subject, both subject and item effects were modeled using random intercepts, random slopes for effects of condition, and random slopes for interactions between condition and subject-level factors where applicable.

Accuracy

Table 5 presents mean comprehension question accuracy and word reading times by condition. Mixed logistic regression revealed no effects of condition on accuracy. A main effects-only model revealed positive main effects for age on accuracy, and no effect of ST-WM or speed of processing. There were no interactions between sentence and subject-level effects in either models with subject-level covariates (Table 6) or in a second set of models run without covariates (not reported; p 's > .05).

Reading Times

A floor was set at 150 ms for each word, and outliers were trimmed at ± 3 standard deviations by subject and word after controlling for word length in characters and presentation order (affecting 2.6% of data overall). Inaccurate trials were also discarded prior to analysis. Given the fact that sentences were presented to all participants using a single pseudo-randomized order, centered item presentation order was included as a

covariate in each model to control for potential “speed-up” effects on items occurring later in the experiment (Farmer et al., 2011).

There was an effect of condition on reading times for the disambiguating segment V2, with longer reading times for EC compared to LC versions of the sentences, and at no other region (not reported; p 's $> .19$). Effects of subject-level factors focused on the critical region, V2. A main effects-only model on this region found significant negative effects of ST-WM and speed of processing, and a significant positive effect of age (p 's $< .05$). There were interactions between age and condition and between speed of processing and condition at V2 in the models with subject-level covariates (Table 6, figure 4). In these interactions, condition effects decreased with faster speed of processing, and increased with advancing age. There was no significant interaction between ST-WM and condition ($p = .22$). The log-base-2 likelihood ratio for the interaction of ST-WM and condition was $-.72$. A second set of interaction models without subject-level covariates showed the same pattern of significance for all 3 interactions (not reported).

Discussion

The finding that subjects were equally accurate answering questions about late closure and early closure sentences may reflect the nature of the comprehension questions: on re-examination, we determined that late closure questions often asked for verification of more complex thematic roles and may have been harder than early closure questions (e.g., a typical late closure question was “*Do the Indians trap the coyote?*” and a typical early closure question was, “*Does the coyote hunt?*”). As in experiment 2, accuracy again increased with age. In this case, older subjects also spent longer reading overall, which also supports the claim that they were more conservative in their decision making.

The interaction of age and condition on reading times indicates that the operations involved in recovering from the misanalysis of an early closure sentence are affected by age. Since these effects were independent of ST-WM and speed of processing, they point to factors other than ST-WM and speed of processing that affect processing syntactic ambiguities in age. As noted, the operations needed to resolve the ambiguity in early closures sentences include attaching the ambiguous noun phrase to a higher level in a syntactic tree than had been the case in the initial parse. This operation inverts level of attachment of this noun phrase and the adverbial clause verb in the syntactic structure. It therefore may have been affected by an age-related decline in inhibitory abilities.

Speed of processing interacted with condition, suggesting that, as in experiment 1, the tests that assess speed of processing measure a cognitive capacity which applies beyond the specific tasks in which it is assessed.

Again, ST-WM did not interact with condition, although the likelihood ratio provides weaker support for the null effect in this case compared to results in experiments 1 and 2. As those experiments, the absence of this interaction is unlikely to be due to inadequate power, ceiling or floor effects, or the inclusion of older participants. As in experiments 1 and 2, the absence of an interaction points to the lack of ST-WM effects on the efficiency and speed of

the operations involved in recovering from EC garden paths. The results of experiment 3 suggest that the Christianson et al. (2006) results reflect the use of ST-WM to review sentences when a question is asked about their meaning, not the memory system used in assigning structure and meaning, as the authors of that paper suggested may be the case.

General Discussion

The results of this study show effects of age and speed of processing on self paced reading times for segments of sentences that trigger a revision of an initially assigned grammatical structure and discourse congruent interpretation. They are consistent in not finding interaction effects of ST-WM on reading times for these segments. We begin by discussing the operations that are involved in these revisions, and then turn to the implications of the results for ST-WM, speed of processing and age. We will focus our discussion on ST-WM, which is the topic of this Special Issue, and discuss speed of processing and age only briefly.

We have characterized the operations that are required in experimental compared to baseline sentences as revisions of initially assigned structures. This description is appropriate regardless of the degree to which multiple structures and interpretations are constructed and maintained in parallel. If the parser/interpreter operates serially, initially constructing only one representation in ambiguous sentences and discourse, the revision process consists of discarding or changing the representation that is maintained and replacing it with another. If it operates in a ranked parallel mode, constructing several representations in ambiguous sentences and discourse and ranking them for their likelihood as the preferred interpretation, the process consists of re-ordering the rankings of representations. Regardless, the operations in each experiment require accessing representations in memory and performing additional computations on those representations, making them good candidates for relying upon ST-WM resources.

The absence of ST-WM effects is therefore of interest. Though it is a null effect, it was found with a large sample size, across the adult lifespan, in studies without ceiling and floor effects. We conclude that any effects of ST-WM on the processes studied here is very small, if it exists at all. This is supported by negative likelihood ratios in the experiments. It is also supported by the fact that although ST-WM shared significant covariance with both age and speed of processing, and both age and speed of processing showed moderating sentence effects, ST-WM still failed to show moderating effects even in models that did not control for these covariates.

The studied operations -- revising the antecedent of a pronoun, finding the antecedent of an anaphoric repeated noun, and assigning the structure and meaning of an early closure sentence -- involve a wide range of syntactic and structures interpretive operations. One of them -- comprehension of early closure sentences -- is known to be difficult and meets criteria for difficult revisions (non-monotonic parsing). The absence of an effect of ST-WM therefore leads to the conclusion that ST-WM capacity does not support a wide range of revision operations. These results extend the range of operations for which effects of ST-WM have either not been found or have only been found intermittently.

The conclusion that ST-WM does not support most processing that revises an initially assigned structure and discourse-preferred meaning is consistent with eye fixation data, which strongly suggest that most text review is not guided by verbal memory. Regressive eye movements overwhelmingly fixate on words immediately preceding the current fixation and on visually salient aspects of texts such as line edges and sentence beginnings marked by punctuation (von der Malsburg and Vasishth, 2011). In rare cases, regressions directly fixate words and phrases that are relevant to the clarification of the structure and meaning of the sentence (Murray and Kennedy, 1988; Frazier and Rayner, 1982; Inhoff & Weger, 2005; Weger & Inhoff, 2007; Meseguer et al, 2002; Mitchell et al, 2008; von der Malsburg and Vasishth, 2011). These “targeted regressions” have been taken as evidence that there is a representation in memory of both the words in a sentence and their spatial co-ordinates, but the memory system in which these representations are maintained is not clear. Although targeted regressions may rely to some extent on verbatim sentence representations that could be stored in ST-WM (Marslen Wilson and Tyler, 1976), it is more likely they rely on reconstructing sentence content from conceptual representations, which has been shown to support most of sentence recall (Potter and Lombardi, 1990; Lombardi and Potter, 1992). These conceptual representation could be directly linked to the spatial co-ordinates of words in a visual memory system, without any role for verbal ST-WM. The self-paced reading results here used a disappearing “moving window”, and therefore suggest that even when the visual record of a sentence or discourse is gone, review of previously presented sentence and discourse material still does not – or need not -- make use of ST-WM.

If this is the case, what memory system do parsing and interpretation, including revision processes, use? Two other proposed memory systems that might support these functions are short term semantic memory (Martin and He, 2004) and long term working memory (LT-WM: Ericsson and Kintsch, 1995). The role of short term semantic memory appears limited to maintaining information that has not yet been incorporated into higher-level syntactic and semantic units, such as strings of pre-nominal adjectives (*the big, black, slow moving ...*). LT-WM, however, which was postulated to account for memory phenomena in skilled tasks and after training, has many properties that equip it for this function. These properties include an unconscious mode of operation, the ability to retain very large amounts of information (e.g., waiters’ ability to remember orders for many diners), domain-specificity of memory ability (skilled waiters do not have superior memory for chess moves), and considerable resistance to interference. The principle differences between the mechanisms that support LT-WM and ST-WM are the complex chunking of stored information and the use of domain-specific retrieval cues in LT-WM. As Caplan and Waters (2013) pointed out, language comprehension has many of these features – unconscious operation; the ability to retain very large amounts of information; domain-specificity of knowledge; complex, domain-specific chunking of representations in storage; resistance to interference; use of domain- specific retrieval cues. Caplan and Waters (2013) suggested that LT-WM supported initial, incrementally successful comprehension. We suggest it supports revisions of assignment of ambiguities as well.

Following Baddeley and many others, we have characterized ST-WM as a short duration memory system that retains item and order information in the service of performing a task.

Tests that measure ST-WM assess these functions, but also require many other cognitive abilities (e.g., dividing attention). Engle and his colleague have emphasized the extent to which they correlated with tests of “fluid intelligence” and proposed that they measure “executive attention” – the ability to direct attention towards task-relevant representations and operations in the face of decay of and interference with task set (e.g., Engle and Kane, 2004). To the extent that this is the case, the results here speak to the role of executive attention in directing the comprehension process. The finding that reading times were faster for all sentences in individuals with higher ST-WM is consistent with the view that executive attention directs attention to language input and sets a criteria for the depth of comprehension in a task (cf., comprehension in proof-reading; “good enough parsing” (Ferreira and Patson, 2007, Ferreira et al., 2009). At the same time, the absence of interactions of ST-WM and sentence type indicate that, to the extent that tests of ST-WM measure executive attention, executive attention does not interact in an ongoing way with the skilled comprehension process. This conclusion contradicts that of Novick et al (2010), who presented data from neuroimaging that led to the suggestion that cognitive control mechanisms located in Broca's area “bias . . . comprehension processes when there are strong demands to resolve competition among incompatible characterizations of linguistic stimuli (p. 906).” However, the observations upon which Novick et al base their conclusion are all off-line, and control mechanisms may be biasing response selection well after incremental comprehension processes have taken place. If executive control processes only direct attention towards comprehension -- presumably by activating the components of the functional architecture of the comprehension system -- and do not set levels of activation of particular on-line operations within this functional architecture, this restriction in scope contrasts with results in speech production, where executive control of specific operations has been demonstrated for specific aspects of speech planning and production (Roelofs, 2008). The difference may reflect the automaticity of language comprehension compared to language production. The issue of how control processes affect components of functional architectures in both comprehension and production remains an open question.

Turning to speed of processing, the effect of speed of processing on segments that require revisions points to the success of laboratory tests of speed of processing in capturing individual variation in a basic cognitive capacity. This capacity is thought to result from very basic neurophysiological factors, that applies very broadly. However, speed of processing did not moderate all sentence type effects, indicating that the relation between this basic functional capacity and different aspects of comprehension may differ. This is consistent with the relation of speed of processing to other functions; Faust et al (1999) note that “general slowing is not sufficient to explain the full range of age- related differences in response latencies reported in the literature (p. 780).” We note that tests of speed of processing measure response latencies, which result from the ratio of the amount of information processed and the rate at which it is processed, as well as other factors (e.g., setting response criteria). Exploring the relationship of speed of processing to aspects of comprehension would benefit from studies that separate the determinants of reaction time (Faust et al, 1999).

There was an effect of age that was independent of ST-WM and speed of processing. The effect was found in experiments 1 and 3 but not experiment 2. We suggest that it may have

been mediated by age-related declines in inhibitory abilities. The plausibility of this suggestion depends upon specific analyses of the processes that apply in the three studies reported here and is speculative. In addition, the inhibitory abilities in which age-related decline affects aspects of comprehension would have to be independent of those involved in executive attention that are captured by the ST-WM measures used here. As with the relation of speed of processing to comprehension, the relation of general inhibitory ability to comprehension has been little studied and needs further empirical investigation.

To conclude with a focus on the topic of this Special Issue, the results of the experiments reported here provide consistent evidence against a role for ST-WM in revising structure or meanings during on-line sentence comprehension. ST-WM may play little role in most language comprehension, and instead only be used after comprehension has taken place, such as verbatim repetition or recall while answering questions, or in tasks that require deliberate manipulation of the forms of language, such as writing poetry.

Acknowledgements

This research was supported by NIH Grant AG09661 to David Caplan and Gloria Waters.

References

- Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. 2008; 59:390–412.
- Baddeley, AD.; Hitch, G. Working memory.. In: Bower, GH., editor. *The psychology of learning and motivation: Advances in research and theory*. Vol. 8. Academic Press; New York: 1974. p. 47-89.
- Baddeley AD, Thomson N, Buchanan M. Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*. 1975; 14:575–589.
- Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*. 2013; 68(3):255–278.
- Barrouillet P, Camos V. As Time Goes By: Temporal Constraints in Working Memory, *Current Directions in Psychological Science*. 2012; 21(6):413–419.
- Barrouillet P, Bernardin S, Camos V. Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*. 2004; 133:83–100. [PubMed: 14979753]
- Barrouillet P, Bernardin S, Portrat S, Vergauwe E, Camos V. Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2007; 33:570–585.
- Bates D. Fitting linear mixed models in R. *R news*. 2005; 5(1):27–30.
- Brown GDA, Neath I, Chater N. A temporal ratio model of memory. *Psychological Review*. 2007; 114:539–576. [PubMed: 17638496]
- Camblin CC, Gordon PC, Swaab TY. The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*. 2007; 56(1):103–128. [PubMed: 17218992]
- Caplan D, Dede G, Waters G, Michaud J, Tripodis Y. Effects of age, speed of processing, and working memory on comprehension of sentences with relative clauses. *Psychology and aging*. 2011; 26(2): 439–50. [PubMed: 21480714]
- Caplan D, Waters GS. Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*. 1999:77–126. [PubMed: 11301522]
- Caplan D, Waters GS. Working memory and connectionist models of parsing: A reply to MacDonald and Christiansen. *Psychological Review*. 2002; 109:66–74.

- Caplan D, Waters G. Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin & Review*. 2013; 20(2):243–68. [PubMed: 23319178]
- Caplan D, Waters G, Howard D. Slave Systems in Verbal Short Term Memory. *Aphasiology*. 2012; 26:279–316.
- Caplan D, Waters GS. Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin Review*. 2013; 20(2):243–68. [PubMed: 23319178]
- Christianson K, Williams CC, Zacks RT, Ferreira F. Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*. 2006; 42:205–238. [PubMed: 17203135]
- Clark HH. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*. 1973; 12:335–359.
- Cohen, J.; Cohen, P.; West, SG.; Aiken, LS. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd ed.. Erlbaum; Mahwah, NJ: 2003.
- Cohen J, MacWhinney B, Flatt M, Provost J. PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods*. 1993; 25(2):257–271.
- Cowan N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*. 2000; 24:87–185. [PubMed: 11515286]
- Cowan N, Nugent LD, Elliott EM, Geer T. Is there a temporal basis of the word length effect? A response to Service (1998). *Quarterly Journal of Experimental Psychology*. 2000; 53A:647–660. [PubMed: 10994222]
- Crowder RG. The demise of short term memory. *Acta Psychologica*. 1982; 50:291–323. [PubMed: 7124433]
- Cummings I, Felser C. The role of working memory in the processing of reflexives. *Language and Cognitive Processes*. 2013; 28(1-2):188–219.
- Daneman M, Carpenter PA. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*. 1980; 19:450–466.
- Daneman M, Merickle PM. Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*. 1996; 3:422–433. [PubMed: 24213976]
- DeDe G, Caplan D, Kemtes K, Waters G, et al. The relationship between age, verbal working memory, and language comprehension. *Psychology and Aging*. 2004; 19:601–616. [PubMed: 15584786]
- Earles JL, Connor LT, Frieske D. Age differences in inhibition: Possible causes and consequences. *Aging, Neuropsychology, and Cognition*. 1997; 4:45–57.
- Engle, RW.; Kane, MJ. Executive attention, working memory capacity, and a two-factor theory of cognitive control. In: Ross, B., editor. *The Psychology of Learning and Motivation*. Vol. 44. Elsevier; New York: 2004. p. 145-199.
- Ericsson KA, Kintsch W. Long-term working memory. *Psychological Review*. 1995; 102(2):211. [PubMed: 7740089]
- Evans WS, Caplan D, Waters G. Effects of concurrent arithmetical and syntactic complexity on self-paced reaction times and eye fixations. *Psychonomic Bulletin & Review*. 2011; 18(6):1203–11. [PubMed: 21956381]
- Farmer, T.; Fine, A.; Jaeger, T. The 33rd annual meeting of the cognitive science society (cogsci11). Boston, MA: 2011. Implicit context-specific learning leads to rapid shifts in syntactic expectations.; p. 2055-2060.
- Ferreira, F.; Engelhardt, PE.; Jones, MW. Good enough language processing: A satisficing approach.. In: Taatgen, N.; Rijn, H.; Nerbonne, J.; Schomaker, L., editors. *Proceedings of the 31st Annual conference of the Cognitive Science Society*. Cognitive Science Society; Austin, TX: 2009.
- Ferreira, F.; Henderson, JM. Syntactic reanalysis, thematic processing, and sentence comprehension.. In: Fodor, JD.; Ferreira, F., editors. *Reanalysis in sentence processing (73-100)*. Kluwer; Boston, MA: 1998.
- Ferreira F, Patson N. The good enough approach to language comprehension. *Language and Linguistics Compass*. 2007; 1:71–83.

- Folstein MF, Folstein SE, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12(3):189–198. [PubMed: 1202204]
- Frazier L, Rayner K. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*. 1982; 14:178–210.
- Gordon PC, Grosz BJ, Gilliom L. Pronouns, Names, and the Centering of Attention in Discourse. *Cognitive Science*. 1993; 17(3):311–347.
- Gordon P, Hendrick R. The Representation and Processing of Discourse in Coreference. *Cognitive Science*. 1998; 22(4):389–424.
- Gordon PC, Searce KA. Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory & Cognition*. 1995; 23(3):313–23. [PubMed: 7791600]
- Grosz B, Joshi A, Weinstein S. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*. 1995; 2(21):203–225.
- Hasher, L.; Lustig, C.; Zacks, RT. Inhibitory mechanisms and the control of attention.. In: Conway, A.; Jarrold, C.; Kane, M.; Miyake, A.; Towse, J., editors. *Variation in working memory*. Oxford University Press; New York: 2007. p. 227-249.A.
- Henson RNA. Short-term memory for serial order: The start-end model. *Cognitive Psychology*. 1998; 36:73–137. [PubMed: 9721198]
- Inhoff AW, Weger UW. Memory for word location during reading: Eye movements to previously read words are spatially selective but not precise. *Memory & Cognition*. 2005; 33:447–461. [PubMed: 16156180]
- Jaeger TF. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*. 2008; 59:434–446. [PubMed: 19884961]
- Jones, DM. Objects, streams and threads of auditory attention.. In: Baddeley, AD.; Weiskrantz, L., editors. *Attention: Selection, awareness and control*. Oxford University Press, Clarendon Press; Oxford: 1993. p. 87-104.
- Jones DM, Macken WJ. Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1993; 19:369–381.
- Jones DM, Macken WJ, Murray AC. Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory & Cognition*. 1993; 21:318–328. [PubMed: 8316094]
- Jones D, Hughes RW, Macken WJ. Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory and Language*. 2006; 54:265–281.
- Jones DM, Hughes RW, Macken WJ. The phonological store abandoned. *Quarterly Journal of Experimental Psychology*. 2007; 60:505–511.
- Joshi, A.; Prasad, R.; Miltsakaki, E. Anaphora resolution: Centering theory approach.. In: Brown, K., editor. *Encyclopedia of language and linguistics*. Vol. 1. Elsevier; Oxford: 2006. p. 223-230.
- Just MA, Carpenter PA. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*. 1992; 99(1):122–149. [PubMed: 1546114]
- Kamp, H.; Reyle, U. *Introduction to model theoretic semantics of natural language, formal logic, and discourse representation theory*. Kluwer Academic Publishers; Dordrecht: 1993. From discourse to logic..
- Kemper S, Crow A, Kemtes KA. Eye-fixation patterns of high- and low-span young and older adults: Down the garden path and back again. *Psychology and Aging*. 2004; 19:157–170. [PubMed: 15065939]
- Kemtes KA, Kemper S. Younger and older adults' on-line processing of syntactic ambiguities. *Psychology and Aging*. 1997; 12:362–371. [PubMed: 9189996]
- Kennison SM, Gordon PC. Comprehending referential expressions during reading: Evidence from eye tracking. *Discourse Processes*. 1997; 24(2-3):229–252.
- King J, Just MA. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*. 1991; 30:580–602.

- Lashley, KS. The problem of serial order in behavior.. In: Jeffress, LA., editor. Cerebral mechanisms in behavior. Wiley; New York: 1951.
- Lawrence, M. a; Klein, RM. Isolating exogenous and endogenous modes of temporal attention. *Journal of Experimental Psychology. General*. 2013; 142(2):560–72. [PubMed: 22746699]
- Levy R, Fedorenko E, Gibson E. The syntactic complexity of Russian relative clauses. *Journal of Memory and Language*. 2013; 69(4):461–495. doi:10.1016/j.jml.2012.10.005. [PubMed: 24711687]
- Lewandowsky S, Oberauer K. The word-length effect provides no evidence for decay in short-term memory. *Psychonomic Bulletin and Review*. 2008; 15(5):875–888. [PubMed: 18926980]
- Lewis RL. Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*. 1996; 25(1):93–115. [PubMed: 8789368]
- Lombardi L, Potter MC. The regeneration of syntax in short term memory. *Journal of Memory and Language*. 1992; 31:713–733.
- Long D, Prat C. Individual differences in syntactic ambiguity resolution: Readers vary in their use of plausibility information. *Memory & Cognition*. 2008; 36(2):375–391. [PubMed: 18426067]
- MacDonald MC, Just MA, Carpenter PA. Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*. 1992; 24:56–98. [PubMed: 1537232]
- MacDonald MC, Christiansen MH. Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review*. 2002; 109:35–54. [PubMed: 11863041]
- Macken WJ, Jones DM. Functional characteristics of the inner voice and the inner ear: Single or double agency? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1995; 21:436–448.
- Marslen Wilson W, Tyler L. Memory and Levels of Processing in a Psycholinguistic Context. *Journal of Experimental Psychology: Human Learning and Memory*. 1976; 2:112–119.
- Martin RC, He T. Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language*. 2004; 89:76–82. [PubMed: 15010239]
- Meseguer E, Carreiras M, Clifton C. Overt reanalysis strategies during the reading of mild garden path sentences. *Memory & Cognition*. 2002; 30:551–561. [PubMed: 12184556]
- Miller, GA.; Chomsky, N. Finitary models of language users.. In: Duncan Luce, R.; Bush, Robert R.; Galanter, Eugene, editors. *Handbook of mathematical psychology*. Vol. 2. Wiley; New York, NY: 1963. p. 419-491.
- Mitchell DC, Shen X, Green MJ, Hodgson T. Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the selective reanalysis hypothesis. *Journal of Memory and Language*. 2008; 59:266–293.
- Murray WS, Kennedy A. Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye movement analyses. *Quarterly Journal of Experimental Psychology*. 1988; 40A:693–718. [PubMed: 3212210]
- Neath I. Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin and Review*. 2000; 7:403–423. [PubMed: 11082850]
- Novick JM, Trueswell JC, Thompsen-Schill SL. Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*. 2010; 4(10):906–924.
- Oberauer K, Lewandowsky S. Modelling working memory: A computational implementation of the time-based resource-sharing theory. *Psychonomic Bulletin and Review*. 2011; 18:10–45. [PubMed: 21327362]
- Page MPA, Norris D. The primacy model: A new model of immediate serial recall. *Psychological Review*. 1998; 105:761–781. [PubMed: 9830378]
- Page, M.; Henson, RN. Models of short-term memory: modelling immediate serial recall of verbal material.. In: Andrade, J., editor. *Working Memory in Perspective*. Psychology Press Ltd; Hove, UK: 2001. p. 177-198.
- Payne BR, Grison S, Gao X, Christianson K, Morrow DG, Stine-Morrow EAL. Aging and individual differences in binding during sentence understanding: Evidence from temporary and global syntactic attachment ambiguities. *Cognition*. 2014; 130(2):157–173. [PubMed: 24291806]

- Pearlmutter N, MacDonald M. Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*. 1995; 34:521–542.
- Potter MC, Lombardi L. Regeneration in the short-term recall of sentences. *Journal of Memory and Learning*. 1990; 29:633–654.
- Preacher KJ, Rucker DD, MacCallum RC, Nicewander WA. Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychological Methods*. 2005; 10(2):178–92. [PubMed: 15998176]
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2008. URL <http://www.R-project.org>
- Ratcliff R, Thapar A, McKoon G. Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin and Review*. 2006; 13:626–635. [PubMed: 17201362]
- Ratcliff R, Thapar A, McKoon G. Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*. 2010; 60:127–157. [PubMed: 19962693]
- Roelofs A. Attention to spoken word planning: Chronometric and neuroimaging evidence. *Language and Linguistics Compass*. 2008; 2:389–405.
- Rohde, D. Linger: A flexible platform for language processing experiments. MIT; 2003. URL <http://tedlab.mit.edu/~Linger/>, version 2.94.
- Royall, R. Statistical evidence: A likelihood paradigm. Chapman and Hall; New York, NY: 1997.
- Salame P, Baddeley AD. Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society*. 1986; 24:263–265.
- Salthouse TA. The processing-speed theory of adult age differences in cognition. *Psychological Review*. 1996; 3:403–428. [PubMed: 8759042]
- Service E. The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology*. 1998; 51A:283–304.
- Sturt P. The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*. 2003; 48:542–562.
- Sturt P, Crocker M. Monotonic syntactic processing: A cross-linguistic study of attachment. *Language and Cognitive Processes*. 1996; 11(5):449–494.
- Swets B, Desmet T, Hambrick DZ, Ferreira F. The role of working memory in syntactic ambiguity resolution: a psychometric approach. *Journal of Experimental Psychology*. 2007; 136(1):64–81. [PubMed: 17324085]
- Traxler MJ, Williams RS, Blozis SA, Morris RK. Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*. 2005; 53:204–224.
- Verhaeghen P, Basak C. Aging and switching of the focus of attention in working memory: Results from a modified n-back task. *Quarterly Journal of Experimental Psychology (A)*. 2005; 58:134–154.
- Waters G, Caplan D. The capacity theory of sentence comprehension: A reply to Just and Carpenter (1992). *Psychological Review*. 1996a; 103:761–772. [PubMed: 8888653]
- Waters GS, Caplan D. The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*. 1996b; 49A:51–74. [PubMed: 8920099]
- Waters GS, Caplan D. The reliability and stability of operation span measures. *Behavioral Research Methods, Instruments & Computers*. 2003; 35(4):550–564.
- Weger UW, Inhoff AW. Long-range regressions to previously read words are guided by spatial and verbal memory. *Memory & Cognition*. 2007; 35:1293–1306. [PubMed: 18035628]
- von der Malsburg T, Vasishth S. What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*. 2011; 65(2):109–127.

Figure 1.

Confirmatory factor analysis path coefficients: structural model of relationships between short-term working memory, speed of processing, and age. SOP = speed of processing, ST-WM= short-term working memory, LC = letter comparison, PC = pattern comparison, Alpha Item = alphabet span summed item score, Rdg Item = reading span summed item score. $p < .05$ for all path coefficients. $\chi^2 = 7.77$, $df = 7$, $p = 0.354$; SRMR = 0.0213; RMSEA = 0.023; RMSEA Upper 90% Limit = 0.092; TLI = 0.997; CFI = 0.998.

Figure 2.

Plots of interaction effects between subject-level and sentence-level factors on critical “P6” region reading times in experiment 1 (as presented in table 2). Speed of processing and age interactions with sentence type are significant. SOP = speed of processing, ST-WM= short-term working memory. Changes in subject-level factors reflect a ± 1 standard deviation change of these continuous variables from the mean.

Figure 3.

Plots of interaction effects between subject-level and sentence-level factors on critical sentence 2 reading times in experiment 2 (as presented in table 4). No interactions with sentence type are significant. SOP = speed of processing, ST-WM= short-term working memory, Sub-Name= proper noun in subject position, Ob-Name = proper noun in object position, Sub-Pro= pronoun in subject position, Ob-Pro = pronoun in object position. Changes in subject-level factors reflect a ± 1 standard deviation change of these continuous variables from the mean.

Figure 4.

Plots of interaction effects between subject-level and sentence-level factors on critical “Verb 2” reading times in experiment 3 (as presented in table 6). Speed of processing and age interactions with sentence type are significant. SOP = speed of processing, ST-WM= short-term working memory. Changes in subject-level factors reflect a ± 1 standard deviation change of these continuous variables from the mean.

Table 1

Results, experiment 1

Mean discourse region self-paced reading time, comprehension accuracy, and standard errors, separately by condition.		
Condition:	Continuation	Shift
<i>Discourse region reading time (ms)</i>		
P1	1646 (22)	1638 (22)
P2	1143 (13)	1160 (13)
P3	1495 (16)	1508 (16)
P4	1308 (16)	1300 (16)
P5	1137 (11)	1136 (11)
P6	1255 (13)	1372 (16)
P7	1211 (14)	1239 (13)
P8	1085 (13)	1091 (14)
Comprehension accuracy (% correct)		
	91.8 (0.6)	92.1 (0.6)

Note: Standard errors in parenthesis. Discourse region P6 contained the condition contrast in bold (Continuation vs. Shift), while all other regions were identical across conditions. Comprehension accuracy for verification questions presented following each trial.

Table 2

Fixed effect parameter estimates and standard errors for mixed effects models of reading time on critical disambiguating region “P6” and question accuracy in experiment 1 (Pronoun Revision).

	Reading Times (ms) Estimate (SE)	Accuracy (log odds) Estimate (SE)
<i>Main-effects only models</i>		
Intercept	369.09 (132.48) ***	3.10 (.25) ***
Condition	126.00 (19.27) ***	0.00 (.12)
ST-WM	-137.05 (33.01) ***	0.17 (.11)
SOP	-122.39 (40.08) **	0.27 (.13) *
Age	62.54 (34.40)	0.19 (.11)
Length	30.50 (2.53) ***	-
Freq	38.78 (25.05)	-
<i>ST-WM by Condition interaction models</i>		
Intercept	361.19 (132.26) **	3.09 (0.25) ***
Condition	126.94 (19.21) ***	0.02 (0.12)
ST-WM	-120.38 (33.34) ***	0.23 (0.14).
SOP	-109.86 (37.79) **	0.20 (0.12)
Age	67.65 (32.43) *	0.16 (0.10)
Length	30.45 (2.52) ***	-
Freq	38.96 (25.03)	-
Condition × ST-WM	1.31 (20.89)	0.09 (0.13)
“Bits of Evidence”	-2.88	-2.34
<i>SOP by Condition interaction models</i>		
Intercept	362.74 (131.96) **	3.10 (0.25) ***
Condition	123.93 (19.24) ***	-0.01 (0.12)
SOP	-129.58 (37.79) ***	0.31 (0.16) *
ST-WM	-101.80 (30.55) ***	0.18 (0.11).
Age	74.44 (32.79) *	0.18 (0.11)
Length	30.51 (2.52) ***	-
Freq	39.29 (24.95)	-
Condition × SOP	-66.87 (22.40) **	-0.13 (0.14)
“Bits of Evidence”	9.58	-2.02
<i>Age by Condition interaction models</i>		
Intercept	367.07 (131.82) **	3.11 (0.25) ***
Condition	125.03 (18.87) ***	-0.01 (0.12)
Age	85.38 (34.02) *	0.18 (0.13)

	Reading Times (ms) Estimate (SE)	Accuracy (log odds) Estimate (SE)
ST-WM	-109.05 (32.13) ***	0.17 (0.11)
SOP	-120.05 (39.17) **	0.29 (0.13) *
Length	30.50 (2.51) ***	-
Freq	39.05 (24.91)	-
Condition × Age	72.43 (18.85) ***	0.07 (0.12)
“Bits of Evidence”	17.44	-2.54

Note: Condition = Shift vs. Continuation (reference level). ST-WM= working memory, SOP = speed of processing, Length = phrase length in characters, Freq = region average log frequency. Random effects for main-effects only models were (in R lme4 syntax) “(Condition + 1 | Subject) + (1 | Item)” and for interaction models were “(Individual Difference Factor * Condition + 1 | Subject) + (1 | Item).”

*
p < .05

**
p < .01

p < .001.

Table 3

Experiment 2 mean sentence reading time, comprehension accuracy, and standard errors, separately by condition.

	Reading time (<i>ms</i>)		Comprehension accuracy (% <i>correct</i>)	
	Sentence 1	Sentence 2	Sentence 3	
Condition (Subject position/ Object position)				
A. Name/ Name	3738 (58)	3389 (42)	2230 (28)	93.2 (0.6)
D. Name/ Pronoun	3669 (55)	3297 (38)	2200 (27)	92.6 (0.6)
C. Pronoun/ Name	3640 (55)	3236 (39)	2232 (27)	92.4 (0.6)
B. Pronoun/ Pronoun	3682 (60)	3137 (40)	2229 (28)	92.7 (0.6)

Note: Standard errors in parenthesis. Repeated name vs. pronoun in Subject and Object position manipulation present in sentence 2 (in bold). Sentences 1 and 3 were identical across conditions. Comprehension accuracy for verification questions presented following each trial.

Table 4

Fixed effect parameter estimates and standard errors for mixed effects models of reading time on critical sentence 2 and question accuracy in experiment 2 (Repeated Name Penalty).

	Reading Times (ms) Estimate (SE)	Accuracy (log odds) Estimate (SE)
<i>Main-effects only models</i>		
Intercept	889.68 (254.40) ***	3.13 (.32) ***
Subject	130.28 (27.77) ***	0.09 (.10)
Object	67.76 (27.90) *	0.07 (.10)
ST-WM	-292.42 (86.69) ***	0.44 (.08) ***
SOP	-351.75 (105.48) ***	0.12 (.10)
Age	82.00 (90.46)	0.32 (.08) ***
Length	41.40 (4.96) ***	-
<i>Random effects</i>		(1/Subject)+(1/Item)
<i>ST-WM by Condition interaction models</i>		
Intercept	945.04 (285.82) ***	3.55 (0.56) ***
Subject	91.79 (86.07)	-0.17 (0.32)
Object	26.80 (84.98)	-0.20 (0.32)
ST-WM	-240.47 (181.11)	0.58 (0.58)
SOP	-297.63 (96.11) **	0.12 (0.10)
Age	105.43 (83.64)	0.33 (0.08) ***
Length	41.43 (4.95) ***	-
Subject × Object	27.96 (54.44)	0.17 (0.21)
Subject × ST-WM	-38.45 (99.32)	0.02 (0.36)
Object × ST-WM	-22.83 (96.82)	-0.05 (0.37)
Subject × Object × ST-WM	-8.90 (61.78)	-0.04 (0.23)
<i>Random effects</i>		(1/Subject)+(1/Item)
“Bits of Evidence”	-3.96	-7.08
<i>SOP by Condition interaction models</i>		
Intercept	915.67 (285.44) **	3.52 (0.55) ***
Subject	107.76 (86.51)	-0.17 (0.32)
Object	37.30 (85.06)	-0.19 (0.32)
SOP	-500.30 (188.31) **	0.25 (0.60)
ST-WM	-327.14 (76.11) ***	0.44 (0.08) ***
Age	104.49 (82.17)	0.33 (0.08) ***
Length	41.38 (4.97) ***	-
Subject × Object	15.13 (55.34)	0.17 (0.20)
Subject × SOP	168.98 (104.93)	0.01 (0.38)

	Reading Times (ms) Estimate (SE)	Accuracy (log odds) Estimate (SE)
Object × SOP	167.80 (101.70)	-0.08 (0.38)
Subject × Object × SOP	-121.19 (67.33)	-0.01 (0.24)
<i>Random effects</i>		<i>(1/Subject)+(1/Item)</i>
“Bits of Evidence”	-4.00	-7.69
<i>Age by Condition interaction models</i>		
Intercept	945.02 (284.72) ***	3.51 (0.55) ***
Subject	92.28 (85.12)	-0.17 (0.32)
Object	29.85 (85.04)	-0.18 (0.32)
Age	179.61 (163.42)	0.01 (0.51)
ST-WM	-326.69 (89.93) ***	0.44 (0.08) ***
SOP	-383.55 (109.46) ***	0.12 (0.10)
Length	41.40 (4.96) ***	-
Subject × Object	25.72 (53.73)	0.17 (0.20)
Subject × Age	-72.21 (85.06)	0.17 (0.32)
Object × Age	-114.35 (85.27)	0.20 (0.32)
Subject × Object × Age	65.04 (53.89)	-0.11 (0.20)
<i>Random effects</i>	<i>(1/Subject)+(1/Item)</i>	<i>(1/Subject)+(1/Item)</i>
“Bits of Evidence”	-4.74	-8.03

Note: Subject and Object = Name vs. Pronoun (reference level) in their respective positions. ST-WM= working memory, SOP = speed of processing, Length = sentence length in characters, Random effects = final random effects structures (as written in lmer syntax) for models where initial “maximal” structures did not converge on a unique model specification. Otherwise, random effects for main-effects only models were (in R lme4 syntax) “(Condition + 1 | Subject) + (1 | Item)”, and for interaction models were “(Individual Difference Factor * Subject * Object + 1 | Subject) + (1 | Item).”

*
p<.05

**
p < .01

p < .001.

Table 5

Results: Experiment 3

Mean residual self-paced word reading time, comprehension accuracy, and standard errors, separately by condition.		
Condition:	Late Closure	Early Closure
<i>Residual word reading time (ms)</i>		
Comp	–43(2.2)	–45(2.0)
DN1	–44(1.8)	–47(1.6)
N1	–48(2.5)	–58(2.2)
V1	–39(2.2)	–37(2.1)
DN2	–27(1.7)	–24(1.8)
N2	–10(2.8)	–12(2.7)
DN3	9(2.2)	NA
N3	–35(2.5)	NA
V2	–26(2.4)	14(2.7)
End	166(5.9)	192(6.3)
<i>Comprehension accuracy (% correct)</i>		
	81 (.01)	86 (.01)

Note: Standard errors in parenthesis. Word “V2” point of condition contrast (in bold). Reading times presented as residuals with presentation order regressed by subject to correct for the single order used, which contained a higher proportion of early closure sentences near beginning of list. Comprehension accuracy for verification questions presented following one third of trials.

Table 6

Fixed effect parameter estimates and standard errors for mixed effects models of reading time on disambiguating word “Verb 2” and question accuracy in experiment 3 (Early Closure/ Late Closure).

	V2 Reading Times (ms) Estimate (SE)	Accuracy (log odds) Estimate (SE)
<i>Main-effects only models</i>		
Intercept	500.77 (12.04) ***	2.33 (.48) ***
Presentation Order	-1.06 (.11) ***	-
Condition	39.39 (8.94) ***	0.35 (.39)
ST-WM	-31.74 (14.06) *	0.08 (.10)
SOP	-43.03 (17.14) *	-0.10 (.12)
Age	59.93 (14.71) ***	0.66 (.10) ***
<i>ST-WM by Condition interaction models</i>		
Intercept	497.94 (11.53) ***	2.33 (0.48) ***
Presentation Order	-0.93 (0.10) ***	-
Condition	40.46 (9.15) ***	0.50 (0.39)
ST-WM	-41.15 (11.74) ***	0.08 (0.10)
SOP	-32.18 (12.32) **	-0.10 (0.11)
Age	50.45 (10.83) ***	0.67 (0.11) ***
Condition × ST-WM	-7.19 (5.78)	0.21 (0.16)
“Bits of Evidence”	-0.724	-1.26
<i>SOP by Condition interaction models</i>		
Intercept	496.43 (11.91) ***	2.33 (0.48) ***
Presentation Order	-1.00 (0.09) ***	-
Condition	37.99 (8.91) ***	0.39 (0.38)
ST-WM	-28.07 (10.61) **	0.67 (0.10) ***
SOP	-39.01 (14.27) **	0.08 (0.10)
Age	56.44 (12.08) ***	-0.18 (0.13)
Condition × SOP	-17.79 (6.40) **	0.17 (0.17)
“Bits of Evidence”	7.58	-1.77
<i>Age by Condition interaction models</i>		
Intercept	494.92 (11.63) ***	2.34 (0.48) ***
Presentation Order	-0.91 (0.09) ***	-
Condition	40.85 (9.15) ***	0.39 (0.38)
ST-WM	-24.96 (10.42) *	0.69 (0.09) ***
SOP	-32.78 (12.74) *	-0.11 (0.11)

	V2 Reading Times (ms) Estimate (SE)	Accuracy (log odds) Estimate (SE)
Age	64.59 (12.80) ***	0.07 (0.11)
Condition × Age	20.46 (5.29) ***	−0.09 (0.13)
“Bits of Evidence”	−2.88	−2.34

Note: Condition = Early Closure vs. Late Closure (reference level). ST-WM= working memory, SOP = speed of processing. Presentation Order = centered item presentation order. Random effects for main-effects only models were (in R lme4 syntax) “(Condition + 1 | Subject) + (Condition + 1 | Item)” and for interaction models were “(Individual Difference Factor * Condition + 1 | Subject) + (Individual Difference Factor * Condition + 1 | Item).”

*
p < .05

**
p < .01

p < .001.