Individual Differences in Sentence Processing

by

Melissa L. Troyer

B.S. Cognitive Science
B.S. Psychology
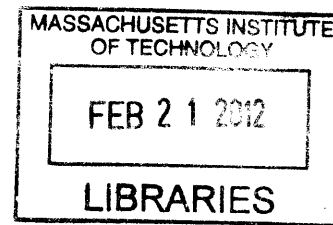B.A. Linguistics, French
Indiana University, 2008

SUBMITTED TO THE DEPARTMENT OF BRAIN AND COGNITIVE SCIENCES IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN COGNITIVE SCIENCE
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEBRUARY 2012

Signature of Author: _____
Department of Brain and Cognitive Sciences
January 15, 2012

Certified by: _____
John Gabrieli
Grover Hermann Professor in Health Sciences and Technology and Cognitive Neuroscience.
Thesis Supervisor

Accepted by: _____
Earl K. Miller
Picower Professor of Neuroscience
Director, Brain and Cognitive Sciences Graduate Program

Individual Differences in Sentence Processing

by

Melissa L. Troyer

ABSTRACT

This thesis aims to elucidate shared mechanisms between retrieval in sentence processing and memory retrieval processes in nonlinguistic domains using an individual differences approach. Prior research in individual differences in sentence processing has provided conflicting evidence as to whether the same memory mechanisms operate in linguistic processing, potentially a quite specialized cognitive domain, and in other, more general areas of cognition (Just & Carpenter, 1992; Caplan & Waters, 1999). This question has been primarily addressed from the point of view of capacity-based theories of working memory (Baddeley, 1986). Under these theories, verbal working memory is either comprised of multiple components including separate components for syntactic and non-syntactic verbal processing, or is dependent on a unitary pool of resources shared across all verbal domains. However, recent memory research has suggested that the capacity-theory architecture may be incorrect. Instead of a three-part memory system composed of focal attention, working memory, and long-term memory, a better model of the memory system may be bipartite, comprising focal attention and long-term memory. In the bipartite theory, working memory is viewed as a set of mechanisms mediating between these two stores, and accurately describes empirical data (McElree, 2006). If the latter hypothesis is correct, then it follows that the bipartite system underlying sentence processing should rely on the same set of working memory mechanisms as in general memory processes. In particular, a number of empirical studies have shown that both general memory and sentence processing are subject to interference from contextually-relevant intervening elements. Such interference is thought to occur at retrieval (as opposed to encoding) both for general memory tasks (e.g., retrieving items from a list) and in sentence processing (e.g., retrieving elements in long-distance syntactic dependencies). However, no systematic attempts have been made to investigate whether this interference results from the same processing limitations. In Study 1, performance on a battery of memory and cognitive tasks is compared to performance on sentence processing tasks. One of the sentence processing tasks correlated with multiple measures likely to rely on general memory mechanisms involved in resolution of retrieval interference. However, low internal reliability of the language tasks in the first study was observed. In Study 2, a series of sentence processing tasks is examined in order to determine which tasks exhibit the highest internal reliability. The results indicate that syntactic complexity manipulations presented in null (isolated) contexts exhibit highest internal reliability and are good candidates for future studies investigating individual differences in sentence processing. Suggestions for future studies investigating shared resources between sentence processing tasks and general memory mechanism are then discussed, informed by the results from these studies.

Acknowledgements

Table of Contents

I. INTRODUCTION

What are the memory mechanisms that underlie human sentence processing? Sentence comprehension is a sequential function and involves understanding words arriving at the ear or eye displaced in time. Understanding a sentence involves keeping track of multiple words, and often multiple syntactic structures, and correctly assigning each word to its appropriate role in the sentence. This involves understanding how words relate to one another within a sentence, how words relate to previous referents from other sentences, and even how words function in multiple ways within a sentence. For instance, in the sentence, *The man that I knew from high school was at the fundraiser*, the noun phrase *the man* functions both as the subject of the main clause (*The man... was at the fundraiser*) and also as the object of the verb of the relative clause (*The man that I knew ___*). Here, at the point of processing the verb *knew*, the comprehender must correctly retrieve the object, *the man*, which occurred several words earlier in the sentence. Similarly, at the point of processing the verb *was*, the noun phrase (NP) *the man*, which occurred many words earlier, must be understood as the subject of the verb *was*. Since sentences contain words that can have multiple dependency relations with other elements that are displaced both temporally and structurally within a sentence, understanding sentences necessarily requires online memory resources in order to correctly map the incoming words onto the appropriate semantic and syntactic structures (Gibson, 1998).

There is reason to believe that the working memory resources involved in sentence processing are at least partially overlapping with those involved in more general memory processes (Just & Carpenter, 1992; Lewis, 1996; c.f. Caplan & Waters, 1999). Lewis and colleagues (2006) urge language researchers to think of language processing as dependent upon other cognitive and memory constraints: "[language processing] is... subject to general

6

processing principles and constraints that govern other domains of memory" (p. 447; see also

Bever, 1970; Lewis, 1996; Marcus, 2006). This thesis tests the hypothesis that sentence that

memory mechanisms operating in sentence processing will be subject to the same constraints as

more general memory mechanisms. In Section I, classic and current models of working memory

are first described, with an attempt to connect them to the sentence processing literature. Next,

several recent studies investigating individual differences in sentence processing are described.

In Sections II-III, two studies are described in which an attempt is made to understand the

relationship between individual differences in comprehending syntactically complex sentences

and in non-linguistic memory processes. Across these studies, performance is asssessed on a

series of tasks tapping linguistic and non-linguistic domains in order to assess co-variation across

individual differences in these areas. Ultimately, the aim of this thesis to elucidate which

cognitive mechanisms are shared between working memory for linguistic processing and

nonlinguistic memory functions.

Models of working memory and sentence processing

Working memory is typically defined as the cognitive system that allows us to store and

retrieve information online. One classic interpretation of findings from working memory

experiments is that it depends on a buffer (or set of buffers) within which information can be

manipulated (Baddeley, 1986, 2007; 2010; Baddeley & Hitch, 1974) by a "central executive", or

the set of functions that manipulates and controls information within these buffers. By this

account, it is assumed that working memory for both general memory and language processing,

as well as working memory in other domains, such as visuo-spatial processing, functions as a

unit distinct both from long-term memory stores and from focal attention. An implicit

consequence of such a system is that working memory for understanding sentences could operate

separately from working memory for other types of verbal material via separate buffers for these different domains[1]. In addition, working memory buffers are thought to have limited capacity, which restricts how much information can be processed at any one point in time. Famously, George Miller's (1956) notion of capacity having limits of "the magical number seven plus or minus two" posited that individuals could hold somewhere between five and nine units of information in working memory.

Some current accounts of working memory have argued that working memory is more constrained than the variable notion of the famous "seven plus or minus two" and that there is a more precise, discrete limited capacity of about four units, or "chunks". Cowan (2010) refers to this as the "magical number four" in short-term memory, citing evidence from memory tasks involving verbal chunks like sentences (Gilchrist, Cowan, & Naveh, Benjamin, 2008), running-span working memory procedures (Cowan, 2001), and memory of words or word-pairs (Chen & Cowan, 2009). In all of these cases, Cowan and colleagues argue that young adults are only able to recall and reproduce between about four (generally between three and five) "chunks" from working memory. Cowan argues that the *number* of chunks that can be recalled is stable regardless of the size of the chunk.

Theories which claim that working memory is based on capacity and buffers have prevailed for many decades, but recently a very different alternative has emerged, in which the whole notion of working memory as an entity distinct from long-term memory and focal attention is called into question. In a recent proposal, McElree (1996; 1998; 2006) and colleagues

---

[1] For instance, Baddeley (1986) has posited at least two separate buffers relying on differing mechanisms: the phonological loop and the visuo-spatial sketchpad. A great deal of research has shown that different pools of resources are used for processing verbal (i.e., sound- and word-based) and visuo-spatial information (Baddeley & Hitch, 1974; Vallar & Shallice, 1990; Hanley, Young, & Pearson, 1991; Jonides, Smith, Koeppe, Awh, Minoshima, & Mintun, 1993; Shah & Miyake, 1996).

(McElree & Dosher, 1989; McElree & Dosher, 1993; McElree, Foraker, & Dyer, 2003; Van Dyke & McElree, 2006, 2011) have argued that a two-part, rather than three-part, memory architecture better accounts for empirical data from the memory and language-processing literatures. In particular, they posit a model in which only one unit of information is available in focal attention while other information remains held in long-term memory, with varying degrees of activation. One of the critical differences between this type of theory and the traditional (i.e., Baddeley-style) theories is that a buffer-based architecture can have a separate linguistic memory store, but the two-part memory account constrains memory operations to be the same for both language and non-linguistic cognition.

McElree (2001) argues that what previous theories have called working memory may be better described by the rapid shunting of information from active states (focal attention) to passive states (long-term memory). Empirical studies support their conclusions. McElree employed an *n*-back task in which participants were asked to respond to whether a target letter had appeared *n* (here, 1 to 3) trials before. Using a speed-accuracy trade-off procedure to estimate the speed with which the cued target was retrieved, McElree found that attention could only be allocated to *one* single memory representation, with speeded retrieval of only the most recently processed item and longer retrieval times for all other items. In another study, McElree and Dosher (1989) tested participants on short lists of words presented sequentially. They observed a fast rate of retrieval when participants were asked to recall the last item in the list, but a slower rate was observed for all other items in the study lists. Such findings converge to suggest differences between the memory resources used for focal attention and the memory search processes used for items not within the scope of focal attention.

Consistent with this model of memory, McElree (2000) argued that in sentence

comprehension, basic binding operations are mediated by a content-addressable memory system, arguing against a serial search process through the contents of working memory during sentence processing (cf. Sternberg, 1966, 1969, 1975). Content-addressable memory structures predict that access to an item in memory can occur *directly* (rather than by using a search process) using cues from the retrieval context. Speed-accuracy trade-off (SAT) paradigms are a useful way of assessing how accuracy varies with retrieval time, allowing retrieval speed to be determined independently of potential differences in activation/memory strength (McElree, 2006). Using a speed-accuracy tradeoff procedure, McElree (2000) showed that although the *accuracy* of the retrieval of words in long-distance dependencies may be affected by the amount of interpolated material, the *speed* at which these items are retrieved is constant across different dependency lengths. That is, at the point of processing a word that his been displaced in a long-distance dependency, such as the object of the verb of a relative clause in object-extracted relative clauses, the amount of intervening material is inconsequential for the speed of retrieval, even though accuracy may vary with the length of the dependency.

McElree (2000; 2006) and others suggest that the mechanism accounting for the accuracy differences is retrieval interference (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Oztekin & McElree, 2007; Van Dyke & McElree, 2011). That is, when more potentially interfering words intervene between points in a long-distance dependency, comprehenders may access the wrong representation before accessing the correct representation due to "…a reduced boost in activation for all items, as a function of the number of distractors that match each retrieval cue" (Lewis et al., 2006, p. 451). This may make it more difficult to use cues in the retrieval context to recover the correct referent; that is, "…an incorrect representation may be retrieved if available cues do not match the desired target as well as they do other items in

memory" (Van Dyke & McElree, 2011, p. 248). However, the *speed* of this retrieval does not vary with the distance between the retrieval site and the item to be retrieved.

The notion of retrieval interference in working memory is not a new one. In basic memory research, Baddeley (1966a) found that phonological interference and, to a lesser extent, semantic interference, could make items in a list more difficult to recall. For instance, the sounds occurring in the letters of the string *BWYKRX* all contain many distinguishing elements from the sounds of the other letters in the string. This type of string is easier to recall than a string like *TCVDBG*, where there is a great deal of phonological overlap among the letters (e.g., each letter contains the sound "ee" (/i/)). However, Baddeley found that strings of words with high semantic overlap, such as *huge-big-long-tall*, were not much more difficult to recall than semantically distinct strings, such as *wet-thin-soft-dark*. Conversely, Baddeley (1966b) found that for long-term memory (as opposed to short-term, or working memory), phonological similarity between competitors was not as relevant for interference; however, items similar in meaning were more likely to interfere. Baddeley found that when participants were asked to remember lists of words over longer intervals of about twenty minutes, lists containing acoustically similar words showed less forgetting than lists containing semantically similar words. Therefore, semantic interference appears to be at play in memory over longer periods whereas phonological interference appears to be relevant for shorter memory intervals. From these studies, Baddeley concluded that mechanisms of encoding are qualitatively different for long-term and short-term memory.

Relevant to the findings described above for basic long-term memory mechanisms, Love & Swinney (1996) stress the importance of non-surface-level (i.e. conceptual, rather than phonologically- or word-based) memory representations in sentence-level working memory. In a cross-modal lexical priming task, they presented participants with sentences containing a lexical

ambiguity in a strongly biased context. At three distinct points in reading a sentence, they tested

whether or not each possible meaning of the word was activated: immediately after the lexical

ambiguity; later in the sentence, occurring much after the main verb; or immediately after the

main verb at a filler-gap construction in which the lexically ambiguous word should be processed

in order to understanding the meaning of the verb and of the sentence. For instance, in the

following sentence, words were inserted at the three points marked with an asterisk:

> (1) The professor insisted that the exam be completed in ink, so Jimmy used the new pen
>
>   * that his mother-in-law recently * purchased * because the multiple colors allowed
>
>   for more creativity.

The authors found that at the first point, both meanings of the lexically ambiguous word *pen*

were activated; at the second point, neither of the meanings was still activated; and at the point

just after the verb, *only* the contextually relevant meaning of the lexically ambiguous word was

re-activated. That is, "…an underlying (deep; non-surface-level) memorial representation of the

sentence is examined during the process of linking an antecedent to a structural position

requiring a referent" (p. 5).

There is further reason to believe that in language processing, as in the long-term

memory and lexical priming studies described above, it is the nature of the meaning (including

both semantic and syntactic features) rather than the nature of the surface-level sounds of a word

that may make it more or less likely to cause retrieval interference with other words in the

sentence. Van Dyke & McElree (2011) argue that syntactic and semantic constraints of a

sentence provide the relevant cues at retrieval sites in sentence processing. When multiple words

in the context match these cues, retrieval interference contributes to difficulty in retrieving the

word and processing the sentence. In particular, they found that intervening (and potentially

interfering) words only contributed to retrieval interference when the syntactic cues matched (i.e., when both the target word and the potentially interfering word were grammatical subjects of the sentence, in their materials). Similarly, Gordon and colleagues (2001, 2004) observed greater difficulty in processing complicated syntax when two noun phrases in a sentence matched in type, such as two personal names or two definite descriptions. This was likely due to interference from the semantic and syntactic overlap between the similar noun phrases.

In another study, Van Dyke and McElree (2006) provide evidence that the locus of such interference effects in sentence processing is at retrieval, and not at encoding. In a dual-task procedure, participants were asked to retain three words in memory while they performed a self-paced reading task. In the reading task, participants read sentences with cleft constructions in which a cleft noun (appearing early in the sentence) functioned as the object of a verb in an embedded clause. Crucially, there were two conditions in the sentence-processing task. In the *interfering* condition, the memory words could also be interpreted as the object of the verb in the embedded clause. In the *non-interfering* condition, this interpretation was not possible. For instance, in the interfering condition, the sentence *It was the boat that the guy who lived by the sea fixed in two sunny days*, the memory list *table-sink-truck* contains words which are also plausible objects of the verb *fixed*. However, in the non-interfering condition, the sentence *It was the boat that the guy who lived by the sea sailed in two sunny days*, the memory list *table-sink-truck* does not contain words which are plausible objects of the verb *sailed*. Van Dyke and McElree observed an increase in reading times at the critical verb (underlined) in the interfering condition compared to the non-interfering condition. They interpreted these findings as resulting from interference from the memory items at the locus of retrieval, rather than at encoding, since the encoding of the correct verb object (and of the memory nouns) was the same in both

13

conditions, and it was only at retrieval that other potential object nouns could interfere.

To summarize, there has been a divide in the literature about how best to construe general working memory, either as a buffer- or capacity-based system as part of a tri-partite structure, or as a mechanism operating within a bi-partite structure of focal attention and a long-term memory store. The best evidence currently suggests that a bi-partite structure supports both general memory mechanisms and also mechanisms of sentence processing. Furthermore, sentence processing appears to be susceptible to interference, which is thought to occur at retrieval. Unlike interference in non-sentence-level verbal information, interference in sentence processing appears to result from semantic- and syntactic-level information.

Individual differences in working memory and sentence processing

Researchers have been interested in addressing the nature of individual differences in high-level language processing for several decades. The first of these studies focused on reading skills. Daneman and Carpenter (1980) found that reading skill appeared to be predicted by working memory capacity (WMC). They measured WMC using a task called *reading span* involving high verbal storage and verbal processing demands[2]. Participants were asked to read a series of sentences aloud while concurrently remembering the last word of each sentence. Daneman and Carpenter observed a strong correlation between scores on the reading span task and several reading comprehension measures, including verbal SAT sores and tasks involving retrieval of factual information and pronominal reference. In general, span tasks tapping both processing and storage capacities of working memory have been shown to be better predictors of comprehension than measures that tap only the storage capacity (Turner & Engle, 1989; Daneman & Merikle, 1996; Engle & Kane, 2004).

---

[2] This type of task is often called a *complex span* task, involving multiple tasks to be performed simultaneously, as opposed to *simple span* tasks, which typically involve only one memory task. Complex span tasks involve a memory task while a concurrent processing task is involved.

King and Just (1991) and Just and Carpenter (1992) suggested that working memory capacity, in addition to predicting overall reading comprehension ability, also predicts high-level sentence processing abilities. King and Just tested participants on a reading span task and had them complete a sentence processing experiment using the self-paced reading paradigm. In particular, they read sentences containing subject-extracted relative clauses (SRCs; syntactically easier) and object-extracted relative clauses (ORCs; syntactically more difficult). Examples are given below with the relative clauses underlined.

Object-extracted relative clause: The reporter that the senator attacked admitted the error.

Subject-extracted relative clause: The reporter that attacked the senator admitted the error.

Object-extracted relative clauses are typically thought to be more difficult because at the point of processing the verb, here *attacked*, the comprehender must retrieve the displaced object of the relative clause, here *the reporter*. Conversely, in subject-extracted relative clauses, the displaced subject of the verb occurs just before it (*The reported that attacked...*) in the canonical order (subject-verb-object (SVO)). Working-memory based theories posit that ORCs are more difficult to process because they require more working memory resources in order to retrieve and integrate the displaced object of the relative clause (Gibson, 1998, 2000[3]).

After grouping participants into high-, medium-, and low-span groups, King and Just observed a main effect of WMC on reading speed in syntactically complex regions of sentences; that is, participants with lower span scores / lower WMC were slower at reading these regions of the sentences. However, there are a few issues with the methodology used in the King and Just study. First, the authors provide no inferential statistics to demonstrate the super-additive

---

[3] See also *surprisal-based theories* (Hale, 2001; Levy, 2008; Gennari & MacDonald, 2008), which posit that ORCs are more difficult because they are less frequent, and therefore increase surprisal at the point of processing the relative clause.

interaction necessary to support their conclusions. Second, the use of a grouping method, rather than a correlational method, could have led to spurious effects due to outliers which may have driven the effects in the high- and low-span groups, in particular. Several researchers have failed to replicate the findings from these initial studies (see Caplan & Waters, 1999, for a review).

Relatedly, in a sentence memory task, Roberts and Gibson (2002) found that participants' memory for two- to five-clause sentences was predicted by a series of short-term storage/verbal working memory tasks. In particular, they computed a composite score for three span tasks that were highly related to the Daneman and Carpenter reading span task and found that this score predicted participants' memory for sentences. In addition, participants completed an n-back task, in which they were presented with short words and asked to respond when the current word matched a word presented n trials before, where n began with 2, and went up to 5, if participants reached criterion performance for each subsequent level. The n-back score was also predictive of participants' performance on the sentence memory task. Roberts and Gibson concluded that general, non-linguistic verbal working memory components support memory for sentence-level linguistic material.

Most individual differences research in sentence processing has used individual differences measures to test capacity-based theories of working memory, with mixed results. There has been surprisingly little work using individual differences to test the clear prediction that differences in general susceptibility to interference in basic memory paradigms may affect sentence processing. However, there have been attempts to experimentally test whether sentence processing might be susceptible to such interference by simultaneously taxing verbal working memory resources that are likely to interference with a sentence while it is being processed. In particular, several studies have used a dual-task paradigm to investigate whether general verbal

working memory resource pools and verbal working memory pools for sentence processing overlap. In an initial study, Gordon and colleagues (2002) had participants perform a concurrent memory task and sentence-processing task. Participants were asked to remember a list of words (either names or descriptions; e.g., *Joel-Craig-Andy* or *poet-cartoonist-voter*). Each word list either matched or did not match the type of nouns used in the sentence (either names or descriptions). At the end of the sentence, participants were required to recall the memory words. They found that for syntactically more complex sentences (as compared to syntactically easier sentences), performance was worse when the type of noun in the memory task matched the type of noun in the sentence, presumably due to increased memory interference when the noun type matched. These findings suggest that the working memory resources used in sentence processing overlap with general memory resources that are used in non-syntactic processes, a result that was replicated by Kemper and Herman (2006). Fedorenko and colleagues (2006) found that such an effect was only present when several (three) words had to be held in memory, but not present when only one memory word was required for recall. Similarly, Wanner and Maratsos (1978) used a sentence-processing task in which sentences were interrupted with a list of memory words. They found that participants were worse at remembering words that had been inserted at more syntactically complex areas of sentences (i.e., in object-extracted relative clauses, compared to simpler subject-extracted relative clauses). These findings all suggest that online sentence processing resources are at least partially overlapping with the memory resources used for non-syntactic verbal processing (e.g., the type of memory required for recalling single words).

Other studies similar in nature to the dual-task studies described above have reported null effects with similar materials in both healthy adults and adults with neurodegenerative disorders

(Waters, Caplan, & Rochon, 1995; Waters & Caplan, 1999). That is, these studies have found no super-additive interactions, meaning that the potentially interfering memory words did not increase the complexity effects found in more difficult syntactic constructions such as object-extracted relative clauses. In addition, DeDe, Caplan, Kemtes, and Waters (2004) used a structural modeling approach to model syntactic processing and language comprehension measures using verbal working memory and age as predictors. While they found that the age of participants had an effect on online sentence processing skill, there was no effect of or interaction with their measures of verbal working memory. However, verbal working memory was a predictor of offline language comprehension, such as performance on comprehension questions.

These studies reporting null results have been interpreted as evidence for separate pools of verbal working memory for language-specific (in particular, online processing-specific) and non-syntactic verbal processing. However, these studies usually required participants to remember a load of digits (typically, at the participants' digit span level, or at digit span minus one), rather than words similar in nature to those in the sentence processing tasks. Compared to the Gordon et al.-type studies, in which potentially interfering words were found to negatively impact sentence processing, what these studies indicate is that only sentence-relevant words held in general verbal working memory may interfere with syntactic processing. Therefore, it is possible that only semantically relevant words are able to potentially interfere with processing a sentence, supporting the results of Van Dyke and McElree (2011).

In the current studies, an attempt is made to resolve some of the unanswered questions from the studies above. In particular, the goal was to disentangle capacity-based theories of working memory from retrieval-interference-based studies of working memory. Using classic

working memory measures from both capacity-based and retrieval based accounts, among other cognitive control measures, the question of which specific memory mechanisms correlate with processing of syntactically complex sentences is addressed. The individual differences approach was chosen because it is a good method for testing whether the same resources underlie linguistic and non-linguistic processing.

Studies 1-2

In an initial study, an extensive array of behavioral tasks was employed, in an attempt to tap into a wide range of cognitive and memory mechanisms. Participants also performed a battery of sentence processing tasks. The ultimate goal of the first study was to test which aspects of memory predicted performance in sentence processing. In a second study, participants completed sentence-processing tasks alone, as the goal was to identify the most reliable sentence processing tasks (including tasks which both had good within-participants reliability and were predictive of performance on other tasks). These results, along with proposed future directions, are then discussed.

II. STUDY 1

A. Tasks

The first study used a large spectrum of tasks designed to measure cognitive control, working memory, and general intelligence in order to compare performance on these tasks with performance on several language tasks. Participants completed these tasks over two days within the span of one week, and completing the entire study took about five hours. The tasks are outlined by type below.

1. *Language Tasks*. Participants performed three separate linguistic tasks. Two of these tasks (Cleft Sentences and Relative Clause Sentences) were self-paced reading tasks, and a third task

(Sentence Ratings) involved reading sentences of varying complexity and judging them on a scale from 1-7 on their "naturalness". In all tasks, participants were given a clear set of instructions on how to perform the task, along with examples, and in all tasks, participants performed comprehension questions for each item in the experiment. All materials from the language tasks are included in Appendices A-C.

Cleft Sentences. Both this task and the Relative Clause Sentences tasks were designed to measure participants' difficulty in processing syntactically complex, versus syntactically easier, sentences. In the Cleft Sentences task, participants read sentences that were presented phrase-by-phrase on a computer screen using the self-paced reading methodology (Wanner & Maratsos, 1978; Gibson, 1998). On each trial, participants were presented with three sentences. The first sentence, which was provided for contextual information, was presented as a whole. At the beginning of a trial, participants saw only dashes covering up all of the words of the sentences. At the press of the space bar, the next sentence, or phrase, in the case of the second and third sentences, was revealed, and the rest of the previously read words were again covered by dashes. This methodology provides a way to measure the time it takes for a participant to read each sentence or phrase.

In this experiment, there were two conditions: object-extracted cleft sentences (syntactically more difficult) and subject-extracted cleft sentences (syntactically easier). The critical regions of interest occurred in the third sentence, which was presented phrase-by-phrase (back-slashes are inserted in the examples below to indicate where the boundaries occurred, and regions of interested are underlined).

        (1) <u>Cleft Sentence Materials</u>

           a.   Object-extracted cleft sentences

A novelist, a cinematographer, and a producer were talking.

Mary: / I heard that / either the novelist / mocked the cinematographer / for wearing a

tie, / or the cinematographer / mocked the novelist / for wearing a tie.

John: / Hmm, / I heard that / it was / the producer / that the novelist mocked / for

wearing a tie.

   b. Subject-extracted cleft sentences

A novelist, a cinematographer, and a producer were talking.

Mary: / I heard that / either the novelist / mocked the cinematographer / for wearing a

tie, / or the cinematographer / mocked the novelist / for wearing a tie.

John: / Hmm, / I heard that / it was / the producer / that mocked the novelist / for

wearing a tie.

This experiment contained 24 such experimental items and 36 filler items of similar syntactic

complexity. A second version of each of the experimental items was created such that the final

clause included the second noun phrase, rather than the first, from the context sentence (e.g.,

cinematographer, rather than *novelist*, in the above example). All subjects saw the same 36 filler

items, but the experimental items were randomized according to a Latin square design such that

participants saw only 12 items from each condition. After reading each item, participants were

asked a true or false comprehension question. The correct answer to half of these questions was

true, and the correct answer was *false* for the remaining questions.fa

   Relative Clause Sentences. In a second language experiment, participants completed

another self-paced reading task involving sentences containing relative clauses. The method

employed was identical to the Cleft Sentences task.

In this experiment, there were two conditions: sentences containing object-extracted

relative clauses (difficult) and sentences containing subject-extracted relative clauses (easier). As in the Cleft Sentences task, participants first saw two context sentences, which were presented as whole sentences. They then saw two more sentences, presented phrase-by-phrase. The critical region of interest occurred in the third sentence, and is underlined in the following example.

(2) <u>Relative Clause Sentence Materials</u>

a. Object-extracted relative clause

At the press conference, a senator and two reporters got into an argument. The senator attacked one of the reporters and then the other reporter attacked the senator.

Mary: / I heard that / the reporter / <u>that the senator attacked</u> / admitted to / making an error.

John: / I am not sure about that. / I heard that / the reporter / that attacked the senator / admitted to / making an error.

b. Subject-extracted relative clause

At the press conference, a senator and two reporters got into an argument.

The senator attacked one of the reporters and then the other reporter attacked the senator.

Mary: / I heard that / the reporter / <u>that attacked the senator</u> / admitted to / making an error.

John: / I am not sure about that. / I heard that / the reporter / that the senator attacked / admitted to / making an error.

This experiment contained 24 such experimental items and 36 filler items of similar length and complexity. All participants saw the same 36 filler items and saw 12 items from each experimental condition, randomized according to a Latin square design. After each item,

participants were presented with a true or false comprehension question. Half of these questions were true, and the other half were false.

Sentence Ratings. This task was designed to get an offline measure of how participants rated sentences of varying syntactic complexity. In this experiment, participants read a series of sentences, which were presented all at once. After reading each sentence, participants were asked to judge it on a scale of 1 to 7 for how natural it sounded. Then they were asked two true or false comprehension questions about the sentence.

There were five sub-experiments within this task. Given that there were many types of sentences, and that all of the sub-experiments were interleaved, no filler sentences were used in this experiment. Each sub-experiment contained materials from two to three conditions. One of these conditions would serve as a baseline for comparison with the other experimental condition. In addition, one of the sub-experiments was in itself a control sub-experiment. Whereas the first four sub-experiments described here manipulated some aspect of syntax which has been claimed to be difficult due to working memory demands, the fifth sub-experiment used agreement and plausibility violations in order to compare breakdowns in sentence processing *not* due to working memory with the tasks manipulating working memory complexity. In addition, the third sub-experiment contained another overall control comparison, including an agreement violation condition in a more complex type of sentence (matched with other materials in that sub-experiment; see the description of Branching vs. Nested Sentential Complements below).

Sentential Complements and Relative Clauses. This sub-experiment manipulated the order of a sentential complement and a relative clause within a sentence. Sentential complements (SCs) embedded within a relative clause (RC) (henceforth SC/RC sentences) are known to be more difficult to understand than relative clauses embedded within sentential complements

(henceforth SC/RC sentences) (Gibson, 1991; Gibson & Thomas, 1998). Gibson & Thomas (1999) propose that this difficulty comes from a greater working memory processing load in the case of the SC/RC sentences. The relevant dependent measures all involve a difference score between the RC/SC (harder) and SC/RC (easier) sentences. Examples are given below.

(3) SC/RC Sentences Materials

a. RC/SC sentences

The intern who the chance that the administrator had lost the medical reports bothered a great deal supervised the nurse.

b. SC/RC sentences

The chance that the administrator who the nurse supervised had lost the medical reports bothered the intern a great deal.

Eighteen such items were generated. Participants saw nine items from each condition, according to a Latin square design.

Branching vs. Nested Relative Clauses. This sub-experiment manipulated whether sentences containing multiple relative clauses were right-branching or doubly-nested. Many studies have observed that nested structures are much more difficult to process than right-branching structures (Bever, 1970; Chomsky & Miller, 1963; Gibson, 1991, 1998; Gibson & Thomas, 1999). This processing difference is thought to be due to working memory resource limitations, such as difficulty in keeping track of multiple incomplete clauses at once in the case of the multiply-nested relative clauses (Gibson & Thomas, 1999). The relevant dependent measures all involve a difference score between the nested relative clause sentences (harder) and the right-branching relative clause sentences (easier). Example sentences follow.

(4) Branching vs. Nested Relative Clauses Materials

a. Nested relative clauses

The pilot who the navigator who the gunner warned alerted broke radio silence.

b. Right-branching relative clauses

The gunner warned the navigator who alerted the pilot who broke radio silence.

Eighteen such items were created. Participants saw nine items from each condition, according to a Latin square design.

Branching vs. Nested Sentential Complements. This sub-experiment was similar to the Branching vs. Nested Relative Clauses sub-experiment, except that it manipulated whether sentential complements were nested or branching within a sentence. Nested sentential complement sentences are thought to be more difficult to process than branching sentential complements for the same reasons as nested vs. branching relative clauses; namely, that in nested structures, it is more difficult to keep track of the multiple incomplete clauses. In addition, an overall "control" condition was included, using branching sentential complements that contained an agreement violation. These sentences were included in order to compare performance on another type of linguistic breakdown not thought to be due to working memory. There were two relevant types of dependent measures: (1) difference scores between the nested sentential complement sentences (harder) and branching sentential complement sentences (easier) and (2) difference scores between the branching sentential complements with agreement violations (violations) and the branching sentential complements (controls). Example sentences follow.

(5) Nested vs. Branching Sentential Complements Materials

a. Nested sentential complements

That whether the seats were gray concerned the engineer annoyed the auto designer.

b. Branching sentential complements

It annoyed the auto designer that it concerned the engineer whether the seats were gray.

   c. Branching sentential complements with agreement violations

   It annoyed the auto designer that it concerned the engineer whether the seats was gray.

Twenty-seven such items were created. Participants saw nine items from each condition, according to a Latin square design.

NP-Island Violations. This sub-experiment compared sentences containing relative clauses with NP-island violations to matched sentences with embedded clauses (initiated by *if*) that were grammatically correct. NP-island violations arise due to a *wh*-noun phrase being moved out of an "island" construction; that is, a construction, such as a relative clause, which blocks the movement of a *wh*-constituent. Some theories posit that *wh*-island violations arise due to working memory constraints similar to the constraints proposed for the difficult or ungrammatical constructions in sub-experiments 1-3, while others propose that island effects stem from the underlying nature of grammar, or syntactic competence (Hofmeister & Sag, 2010; Sprouse, Wagers, & Phillips, in press). The inclusion of this sub-experiment therefore provides a way to simultaneously tease apart these two hypotheses by comparing performance on this task with the non-linguistic working memory tasks and also to examine the extent to which performance on this task correlates with performance on sub-experiments 1-3. The relevant dependent measures involve a difference score between the NP-island violations (ungrammatical) and control sentences. Example sentences are below.

   (6) NP-Island Violations Materials

   a. NP-island violations

The reporter asked who the CEO spread the rumor that the stockholders hated.

b. Control sentences

The reporter asked if the CEO spread the rumor that the stockholders hated the new president.

Eighteen such items were created. Participants saw nine items for each condition, according to a Latin square design.

Control Experiments. This sub-experiment had three conditions, all of which were sentences of similar complexity with either agreement, plausibility, or no violations. The agreement or plausibility violations could come in the first clause, in the second clause, or in both. There were two sets of dependent measures: (1) difference scores between the agreement violations and control sentences, and (2) difference scores between the plausibility violations and control sentences.

(7) Control Experiments Materials

a. Agreement violations

The chef are baking the bread and the waiter are seating the customer.

b. Plausibility violations

The bread is baking the chef and the customer is seating the waiter.

c. Control sentences

The chef is baking the bread and the waiter is seating the customer.

Twenty-seven such items were created. Participants saw nine items from each condition, according to a Latin square design.

2. *Verbal Span tasks.* Participants completed a series of five verbal span tasks, all designed to measure some form of verbal working memory (La Pointe & Engle, 1990; Conway, Kane,

Bunting, Hambrick, Wilhelm, & Engle, 2005; Colom, Rebollo, Abad, & Shih, 2006; Unsworth & Engle, 2006, 2007b; Schmiedeck, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009). Two of these span tasks were *complex span tasks*, and three were *simple span tasks*. As a general rule, complex span tasks require participants to do a processing task while they are simultaneously asked to remember a sequence of words, numbers, or letters. Complex span tasks are more difficult than simple span tasks and have historically been thought to be a better measure of working memory than simple span tasks (Turner & Engle, 1989); however, others have argued that both tasks tap similar processes, including updating processes in working memory along with rehearsal and cognitive control (Colom et al., 2006; Unsworth & Engle, 2007b).

 Reading Span. This complex span task was developed by Daneman & Carpenter (1980) to measure trade-offs between processing and storage limitations in working memory capacity in order to investigate individual differences in reading comprehension (see also Friedman & Miyake 2004, 2005). Here, a computerized version of the task was employed. Participants read a series of sentences aloud from a computer screen. After a sentence was read, the experimenter pressed a button for the next sentence to appear. For each trial on each span size, a participant was asked to read the relevant number of sentences for that span size and try to remember the last word of each sentence. At the end of the trial, participants were asked to recall the last words of the previous set of sentences and to say them out loud in order. Participants began at a span size of two and continued up to a span size of six. For each span size, participants completed five trials at that span size, except for the span size of six, where only three trials were included. Reading span was computed as the highest span size for which the participant got at least three out of five trials completely correct, plus an additional half a point if they got two trials out of

five completely correct at a higher span size. A list of the sentences used in this task can be found in Appendix D.

Operation Span. The second complex span task employed in our studies was developed by Turner and Engle (1989) and, like the reading span task, involves simultaneous memory and processing components. In an adapted, computerized version of their task, participants were required to respond to whether completed arithmetic problems were correct or not (e.g., 3 + 5 = 9?) while remembering a list of words. Participants alternated reading words aloud and performing arithmetic judgments for the length of that span set. As in the Reading Span task, span sets ranged from two to six. Participants were given five trials at each of the span lengths. Span scores were calculated in the same manner as for Reading Span.

Digit Span. This classic task has been used since the 1880s (Galton, 1887; Jacobs, 1887) as a measure of serial recall from short-term memory. In a computerized version of the task, participants were auditorily presented with a sequence of numbers of increasing span lengths. After each trial, they were asked to type back in the numbers that they had heard in the same order they occurred in. Participants completed five trials at each span size. For this task, span sizes from three to eight were tested. Span scores were calculated in the same manner as in the complex span tasks.

Backwards Digit Span. This task was conducted in the same manner as the digit span task. Participants were auditorily presented with a sequence of digits and then were asked to respond by typing in a series of digits. However, they were asked to respond by typing the sequence in reverse order, increasing the difficulty of the task by asking participants to manipulate the contents of working memory. Span sizes ranged from three to eight. Span scores were calculated as in the other span tasks.

Syllable Span. This task is more difficult than a basic span task like Digit Span because participants are required to keep in mind nonsense syllables, rather than familiar numbers. While these nonsense syllables are no longer in phonological length than the digits 1-9 used in the Digit Span and Backwards Digit Span, they have no meaning associated with them and therefore are thought to be a purer measure of short-term phonological memory. The task was a computerized task adapted from the syllable span task used in Fedorenko, Frank, Everett, & Gibson (submitted). The list of syllables used in this version of the task was as follows: *a, i, u, ba, bi, bu, ga, gi, gu, ha, hi, hu, ka, ki, ku, pa, pi, pu, ta, ti,* and *tu.* Span sizes tested ranged from three to eight. Span scores were calculated as in the other span tasks.

3. *Cognitive Control tasks*. Three tasks were used to measure various aspects of executive function, or cognitive control. Two of these tasks were computerized, and an experimenter administered the third. Each of these tasks is thought to measure some aspect of controlling interference (in particular, the Stroop and MSIT tasks), inhibition (in particular, the Go/NoGo and MSIT tasks), or attentional control (all three tasks). The tasks are described in detail below.

MSIT (Multi-Source Interference Task). The MSIT is a relatively new task designed to measure interference from multiple sources (Bush & Shin, 2006). This task reliably activates neural structures known to be involved in attentional control, and is robust within individual participants. In this task, participants are instructed to view three-digit strings composed of the digits 0-3. They are instructed to press a button on a pad (labeled with the digits 1-3) or keyboard corresponding to the digit that does not match the other two digits in the string. In the control condition, two of the digits in the string are "0", and the third digit (i.e., the mismatching digit to which a response should be made) matches the same position as that digit on the keypad. Therefore, in the control condition, participants see only strings like 100, 020, and 003. In the

interference condition, participants see strings in which the two interfering digits are not "0" (i.e., that match other potential response targets) and in which the location of the target digit in the string does *not* match the location of this digit on the keypad. For instance, participants see strings like the following: 131, 212, 232, etc. The relevant dependent measures for this task involve a difference score between the control and interference tasks in reaction times and accuracy.

Go/NoGo Task. On this task, participants were shown a series of pictures displayed on a computer monitor (in this version of the task, the pictures were all characters from the children's television show Pokemon; Durston et al., 2002). On most of the trials, participants were instructed to respond to the picture by pressing the space bar. However, participants were instructed *not* to press the space bar when they were shown a particular image (in this version of the task, that of the Pokemon character Meowth; this was called a *no-go* trial). The task therefore measures a participant's ability to disengage from a response, or inhibit the response, that is required for the majority of the trials. The dependent measure was the percent correct for the no-go trials.

Stroop Task. This task was based on the classic Stroop (1935) task, which was designed to measure interference from a highly automatic process (reading) while naming a task-relevant feature (usually, ink color of the word). Participants were shown a series of lists of color words and instructed to name the color of the *ink* the word was printed in (and not simply to read the word). They were shown two lists from each of three conditions. First, subjects always performed the Congruent condition, in which the color words were printed in the matching ink color (for instance, the word "blue" would be printed in blue ink). Then, participants performed the next two conditions, half performing the Incongruent condition second and half performing

the Mixed condition second. In the Incongruent condition, all of the color words mismatched the ink the color was printed in (for instance, the word "blue" might be printed in red ink). In the Mixed condition, half of the words were congruent (word and ink matched) and half were incongruent (word and ink mismatched). The Mixed condition was included due to findings that individuals who appear to have lower working memory perform worse than high-working memory individuals when presented with a mixed version of the Stroop task, but similarly to high-working memory individuals on Stroop tasks when all of the trials match (Unsworth & Engle, 2007a). The dependent measures for Stroop were the difference score on the reaction times to complete the Incongruent vs. Congruent lists and the difference score on the Mixed vs. Congruent lists.

4. *General Intelligence* (Gf) *tasks.* The following two tasks were used as a measure of general fluid intelligence (*Gf*). The first of these tasks, Raven's Advanced Progressive Matrices (RAPM; Raven, 1990, 2000) is a standard, culture-neutral task often used as a measure of general fluid intelligence and has been found to correlate highly with *Gf,* the construct which is thought to reflect an individual's ability to reason and solve complex problems, and is necessary for many cognitive tasks (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). The second task is a recently developed dual *n*-back task that has been found to highly correlate with RAPM, in addition to improving scores on RAPM when individuals undergo weeks of training (Jaeggi et al., 2008; Jaeggi, Etinne, Ozdoba, Perrig, & Nirkko, 2007; Jaeggi, Buschkuehl, Perrig, & Meier, 2010).

RAPM. In this task, participants complete a series of thirty-six matrices. They are shown 3 x 3 grids containing a pattern, in which one of the grid pieces is missing. They are required to pick the correct piece from eight possible choices. As the task continues, the patterns in the matrices become harder to complete. The dependent measure is the number of patterns

completely corrected under a time limit of 40 minutes.

Dual n-*back Task*. This task, developed by Jaeggi and colleagues (2007), employs an *n*-back design while participants concurrently perform two tasks. The task involves responding to stimuli which occurred *n* positions back in an ongoing stream, where *n* can vary depending on the paradigm. In this experiment, *n*s of 1, 2, and 3 were used. On each trial, participants both saw a visual target appear at one of nine positions on a 3 x 3 grid and, at the same time, heard a letter. They were instructed to respond by pressing the "s" button if the spatial cue matched the cue presented *n* trials before and by pressing the "l" key if the letter they heard matched the letter presented *n* trials before. It was also possible for both cues to match, in which case participants were instructed to press both the "s" and "l" keys simultaneously. In general, *n*-back tasks are thought to measure updating processes in working memory. Due to the higher processing demands required for the dual *n*-back task employed here, researchers have hypothesized that it may also be related to general intelligence (Jaeggi et al., 2010) and have argued that it can improve IQ when participants train on it over a period of weeks (Jaeggi et al., 2008). For this task, the dependent measures were both reaction time (for correct responses only) and a *d'* score computed for proportion correct, which was computed as the number of hits (correct responses to valid targets) minus the number of false alarms (incorrect responses to invalid lures).

5. *Spatial Working Memory task*. This task was used as a control working memory task. Unlike the verbal working memory span tasks outlined above, this task had only a spatial component, somewhat similar to the spatial component of the dual *n*-back task. Although it was an adaptive task (not a span task), it is similar in nature to the Corsi block spatial memory test, often used with neuropsychiatric patients, in which participants must tap back spatial sequences on a grid of blocks (Corsi, 1972; Berch, Krikorian, & Huha, 1998).

Corsi Block Task. An adapted, computerized version of the Corsi block task was employed. Participants saw a 4 x 4 grid in which a series of cues would appear. On each trial, participants viewed two distinct series of cues and had to indicate whether the two sequences were the same or different. An adaptive version of this task was used so that as participants correctly identified multiple sequences of the same length, the length of the sequences would increase by one. When participants answered a trial incorrectly, the length of the sequences would decrease by one. There were two dependent measures for this task: first, the total number of trials which the participant answered correctly, and second, the level (i.e., length of sequences) at which they ended the task. Participants all completed sixty trials.

6. *Other tasks.* Participants completed two other tasks involving a verbal component. The first of these, Verbal Fluency, is a task often administered to neuropsychiatric patients as a measure of phonological or semantic fluency or generativity (Lezak, 1995; Rosen & Engle, 1997; Troyer, Moscovitch, & Winocur, 1997; Loonstra, Tarlow, & Sellers, 2001). In addition, it is sometimes considered a test of inhibition, since it requires participants to generate responses from only a certain class of answers (i.e., from some phonological or semantic category) and inhibit other responses (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). Here, we employed only a phonological version of the task. The second task was a simple spelling test involving commonly misspelled words.

Verbal Fluency. The FAS version of the verbal fluency task was administered to participants (Loonstra et al., 2001). Participants were told that they would have one minute to write down as many words as they began beginning with a single letter. Participants first performed the task for the letter *F*, followed by *A* and *S*. Participants were instructed not to use proper nouns or multiple versions of a common word stem (e.g., writing both *fish* and *fished*

would only count as one word). The dependent measures were the total number of eligible words generated for each letter, as well as the overall number of eligible words generated across all of the letters.

Spelling Test. For this spelling test, the author of this thesis recorded a spoken list of 40 commonly misspelled words. Participants listened to these words spoken over headphones and were allowed to replay words if they had any difficulty in understanding them. Participants were given a score sheet and were instructed to write down each word.

B. Methods

1. *Participants.* Sixty participants were recruited from the MIT community, although both students and non-students were invited to participate. Two participants did not complete the entire experiment and were therefore excluded from further analyses. Participants were paid $60 for two days of studies. All participants were native English speakers, meaning that they had been speaking English since at least the age of four. Of these participants, 18 reported they were bilingual, 36 were not bilingual but had studied other languages at school, and four had never studied any language other English. The total number of languages known or studied ranged from one to five, with four participants knowing one language (English), 18 knowing two, 22 knowing three, 13 knowing four, and one participant having knowledge of five languages. Of the 58 participants who completed the experiment, 10 had only a high school degree, 42 had completed a college degree or were in college, 4 had completed Master's degrees, and two did not respond with information about their education. Participants were excluded from analysis if they averaged below 65% accuracy on comprehension questions for the two sentence-processing tasks, Cleft Sentences and Relative Clause Sentences.

2. *Materials.* Materials for each of the fifteen tasks are described in the preceding section. Except

for the language tasks, participants all saw the same materials for each of the tasks. For the language tasks, participants all saw the same items, but saw different versions of some items such that only one condition of each item was presented to each participant. The materials for the language tasks, the reading span task, and the spelling test are given in Appendices A-E.

3. *Procedure*. All participants completed the tasks in the same order. On the first day, they completed the Cleft Sentences, Digit Span, Operation Span, Dual *n*-back, Corsi Block, RAPM, Verbal Fluency, and MSIT tasks. Completing these tasks took between two and three hours. On the second day, participants completed the Relative Clause Sentences, Backwards Digit Span, Go/NoGo, Syllable Span, Sentence Ratings, Reading Span, Stroop, and Spelling Test tasks, along with the a questionnaire collecting language and demographic information. The second session also took between two and three hours two complete. Detailed descriptions of each task are given in the preceding section.

C. Results & Discussion by Task

Region:	Hmm,	I heard that	it was	the producer	that the novelist mocked / that mocked the novelist	for wearing a tie.
Object	404.27 (142.50)	412.02 (130.12)	409.26 (134.75)	537.62 (297.27)	1444.07 (994.22)	1161.68 (847.70)
Subject	415.07 (161.30)	418.48 (161.45)	419.33 (125.54)	548.65 (332.28)	1129.83 (859.02)	1019.03 (709.01)

Table 1: Reading times (in ms) by region for the Cleft Sentences task from Study 1.

Results for each are first described along with specific predictions for each task. Next, reliability and correlations within the language tasks are discussed. Finally, correlations of the non-language task measures and language task measures are given.

Results from each task

1. *Cleft Sentences*. Reading times were calculated for each region of the final sentence. Reading times that were more than three standard deviations from the average of each region were discarded and excluded from analysis. The primary interest was in the difference between reading times for the embedded clause in the object and subject cleft conditions. In addition, the difference between the question accuracies for the object and subject cleft was analyzed.

Reading times along with standard deviation by region are given in Table 1 for the final sentence of the cleft items (John's utterance). Of particular interest, the average reading time for the object-extracted embedded clause regions was 1444.07 ms while the average reading time for the subject-extracted embedded clause regions was only 1129.83 ms, a difference of 314.24 ms. Using a linear mixed-effects model with subject and item as random effects and the condition (object or subject cleft), question accuracy, and length of the region as fixed effect predictors, condition was found to be a significant predictor of the critical region reading time, $t(1322) = 7.945$, $p < .0001$. The length of the region in characters was also a significant predictor, $t(1322) = 4.641$, $p < .0001$.

Region	Mary:	I heard that	the reporter	that the senator attacked / that attacked the senator	admit-ted to	making an error.
Object	484.27 (310.52)	460.25 (251.18)	509.83 (454.96)	1556.51 (1330.88)	862.80 (720.57)	1091.20 (821.73)
Subject	479.81 (280.38)	453.36 (208.71)	506.40 (239.95)	1280.51 (993.44)	829.66 (654.52)	1000.35 (720.99)

Table 2: Reading times (in ms) by region for the Relative Clause Sentences experiment from Study 1.

Question accuracies followed the same pattern. For object cleft sentences, mean accuracy was only 77.45% (SD = 41.81%), whereas for subject cleft sentences, mean accuracy was 91.67% (SD = 27.66%), a difference of 14.22%. Accuracy for filler sentences was 88.79% (SD =

31.56%). A linear mixed-effects model with subject and item as random effects and condition as a fixed effect indicated that the difference between object and subject cleft sentences was significant, $t(1296) = 7.706$, p < .0001.

The results from the Cleft Sentences task thus confirmed our prediction that object cleft sentences would be more difficult to process, and lead to lower question accuracies, than the subject cleft sentences.

2. *Relative Clause Sentences.* Reading times were calculated for each region of the third sentence. Reading times that were more than three standard deviations from the average of each region were discarded and excluded from analysis. The primary interest was the difference between reading times for the relative clause in the object- and subject-extracted relative clause conditions. In addition, the difference between the question accuracies for the object- and subject-extracted relative clause sentences was analyzed.

Reading times along with standard deviations are given in Table 2 for the third sentence of the relative clause items (Mary's utterance). Of particular interest, the average reading time for the object-extracted relative clause regions was 1556.51 ms whereas the average reading time for the subject-extracted relative clause regions was only 1280.51 ms, for a difference of 276 ms. Using a linear mixed-effects model with subject and item as random effects and the condition (object or subject RC), question accuracy, and region length as fixed effect predictors, condition was a significant predictor of RC region reading time, $t(1262) = 5.916$, p < .0001. The length of the region ($t(1272) = 3.62$, p < .001) and whether or not the question was correctly answered ($t(1272) = 2.218$, p < .05) were also significant predictors.

Question accuracies followed the same pattern. For object-extracted RC sentences, mean accuracy was only 82.02% (SD = 41.81%), whereas accuracy was 84.23% (SD = 27.66%) for

subject-extracted RC sentences, for a difference of 2.21%. However, this trend was not significant, $t(1292) = 1.14$, p >.05. Mean accuracy for filler sentences was 87.60% (SD = 31.56%).

The results from the reading times of the Relative Clause Sentence task confirmed our prediction that object-extracted RCs would be more difficult to process than subject-extracted RCs. However, our prediction that this difficulty would also be reflected in the comprehension question accuracies was not supported by this study.

3. *Sentence Ratings*. For each of the sub-experiments in this task, four measures were obtained: the reading time for the whole sentence, the rating given by the participant for how "natural" the sentence sounded, and the accuracies for two comprehension questions. Although the primary measure of interest was the sentence ratings, the results from each of these measures are described below.

Whole-sentence reading times are presented in Table 3, along with standard deviations, by sub-experiment and condition. For the Sentential Clauses and Relative Clauses sub-

Sub-experiment	Condition	Whole-sentence RT (in ms)
Sentential Complements	SC/RC	10736.98 (7828.76)
and Relative Clauses	RC/SC	11506.07 (8635.23)
Branching vs. Nested	Nested RCs	10494.67 (7160.62)
Relative Clauses	Branching RCs	8006.84 (6022.62)
Branching vs. Nested	Nested SCs	7801.39 (4930.04)
Sentential Complements	Branching SCs	7319.47 (6249.97)
	Branching SCs w/ Violations	6727.48 (4856.66)
NP-Island Violations	NP-Island Violations	9075.91 (5749.56)
	Control Sentences	8042.14 (6001.63)
Agreement and	Agreement Violations	4951.82 (3416.64)
Plausibility Violations	Plausibility Violations	5840.36 (7828.76)
	Control Sentences	4911.78 (4060.00)

Table 3. Whole sentence reading times from the Sentence Ratings task from Study 1.

experiment, we predicted that the RC/SC order would be more difficult than the SC/RC order,

and therefore would result in longer reading times. The mean whole-sentence reading times for these two conditions were similar (RC/SC mean: 11506.07 ms; SC/RC mean: 10736.98 ms) and a linear mixed-effects model with subject and item as random effects and reading time and sentence length (in characters) as fixed effects indicated that this difference approached significance, $t(948) = 1.930$, $p < .06$. Sentence length also approached significance as a predictor, $t(948) = 1.922$, $p < .06$. Similar mixed-effects models were used to analyze all of the whole-sentence reading time data.

For the Branching vs. Nested Relative Clauses sub-experiment, mean reading time was 19494.67 ms for the nested sentences but only 8006.84 ms for the branching sentences. As predicted, this effect of sentence difficulty was significant, $t(948) = 7.013$, $p < .0001$. Sentence length was also a significant predictor, $t(948) = 2.583$, $p < .01$.

There were two comparisons of interest within the Branching vs. Nested Sentential Complements sub-experiment: nested vs. branching sentences and branching sentences vs. branching sentences with agreement violations. Each comparison was made with reference to the third condition. The mean sentence reading times were 7801.39 ms for the nested condition, 7319.47 ms for the branching condition, and 6727.48 ms for the branching/violations condition. The predicted difficulty was observed for the branching vs. nested conditions, such that the nested conditions took longer to read, $t(960) = 2.364$, $p < .05$. Sentence length was also a significant predictor of reading time in this analysis, $t(960) = 2.540$, $p < .05$. The predicted difference was also observed for the branching vs. branching/violation conditions, $t(963) = 2.021$, $p < .05$. Sentence length was again a significant predictor, $t(963) = 3.170$, $p < .01$.

For the NP-Island Violations sub-experiment, mean reading time was 9075.91 ms for the NP-Island Violation condition and only 8042.14 ms for the Control condition. As predicted, NP-

island violations took longer to read, $t(951) = 3.108$, p < .01. No other fixed effects were significant predictors of reading time.

Finally, for the Agreement and Plausibility Violations sub-experiment, mean reading times were 4951.82 ms for the agreement violations, 5840.36 for the plausibility violations, and 4911.78 ms for the control condition. As predicted, the plausibility violations took longer to read than the controls, $t(957) = 4.321$, p < .0001. Sentence length was also a significant predictor, $t(957) = 4.409$, p< .0001. However, the agreement violations did not take a significantly longer amount of time to read than the controls, $t(962) = .2.99$, p >.05. Sentence length was again a significant predictor, $t(962) = 4.176$, p < .0001.

To summarize the whole-sentence reading time findings, nearly all of the predicted main effects were observed, with the exception of the agreement violations. The agreement violations in this sub-experiment were likely the easiest of the "violation" or complexity manipulations. It is likely that participants were able to read these sentences quite easily, and the absence of a

Sub-experiment	Condition	Z-scored rating
Sentential Complements	RC/SC	-0.828
and Relative Clauses	SC/RC	0.248
Branching vs. Nested	Nested RCs	-0.940
Relative Clauses	Branching RCs	0.276
Branching vs. Nested	Nested SCs	-0.351
Sentential Complements	Branching SCs	0.447
	Branching SCs w/ Violations	-0.461
NP-Island Violations	NP-Island Violations	-0.673
	Control Sentences	0.975
Agreement and Plausibility	Agreement Violations	0.142
Violations	Plausibility Violations	-0.088
	Control Sentences	1.289

Table 4. Z-scored ratings from the Sentence Ratings Task from Study 1.

main effect between the reading time for these sentences and the corresponding controls is not

worrisome.

Each participant's ratings were z-scored so that they could be compared in as similar a space as possible to the other participants' ratings. The z-scored ratings for each of the sub-experiments are presented in Table 4. Over all the data, z-scored ratings ranged from -3.477 to 2.728.

The average ratings for the Sentential Complements and Relative Clauses task were -0.828 for the RC/SC (more difficult) condition and 0.248 for the SC/RC (easier) condition. A linear mixed-effects model with subject and item as random effects and condition and sentence length as fixed effects revealed that this difference was significant, $t(948) = 24.342$, p < .0001. Similar mixed-effects models were used for all of the z-score analyses. Sentence length was not a significant predictor for any of the ratings measures.

For the Branching vs. Nested Relative Clauses sub-experiment, average ratings were -0.940 for the nested RCs and 0.276 for the branching RCs. As predicted, this effect was significant, $t(948) = 28.329$, p < .0001.

For the Branching vs. Nested Sentential Complements, the ratings were -0.351 for the nested SCs, 0.447 for the branching SCs, and -0.461 for the branching SCs with violations. As predicted, branching SCs were rated higher than both nested SCs ($t(972) = 15.863$, p < .0001) and branching SCs with agreement violations ($t(972) = 21.989$, p < .0001).

The average ratings for the NP-Island Violation sub-experiment were -0.673 for the NP-island violations and 0.975 for the control sentences. As predicted, the NP-island violations were rated worse than the control sentences, $t(972) = 26.975$, p < .0001.

In the Agreement and Plausibility Violations sub-experiment, the ratings were 0.142 for the agreement violations, -0.088 for the plausibility violations, and 1.289 for the control

sentences. As predicted, the control sentences were rated higher than both the agreement (t(972) = 26.882, p < .0001) and plausibility violations (t(972) = 30.519, p < .0001).

Unlike the whole-sentence reading times, complexity effects were observed in every planned comparison for the ratings. These z-scored ratings were the primary measure of interest in the Sentence Ratings study and it is this measure that will be tested against measures from the other tasks.

Mean accuracies to the first question by condition are presented in Table 5. For the Sentential Complements and Relative Clauses sub-experiment, mean accuracy for the RC/SC (harder) condition was 74.42%, and mean accuracy for the SC/RC (easier) condition was 86.32%. A linear mixed-effects model with subject and item as random effects and condition and sentence length as fixed effects revealed that this difference in question accuracies was significant, t(948) = 4.963, p < .0001. Similar mixed-effects models were used to analyze all of the question accuracy scores.

Sub-experiment	Condition	Question 1 Accuracy
Sentential Complements and Relative Clauses	RC/SC	74.42% (43.68%)
	SC/RC	86.32% (34.40%)
Branching vs. Nested Relative Clauses	Nested RCs	71.49% (45.19%)
	Branching RCs	84.94% (35.81%)
Branching vs. Nested Sentential Complements	Nested SCs	88.13% (32.38%)
	Branching SCs	87.71% (32.87%)
	Branching SCs w/ Violations	88.41% (32.05%)
NP-Island Violations	NP-Island Violations	60.88% (48.85%)
	Control Sentences	73.57% (44.14%)
Agreement and Plausibility Violations	Agreement Violations	91.53% (27.87%)
	Plausibility Violations	92.48% (26.39%)
	Control Sentences	94.77% (22.29%)

Table 5. Question 1 Accuracies from the Sentence Ratings Task from Study 1.

For the Branching vs. Nested Relative Clauses sub-experiment, question accuracies were 71.49% for the nested RCs and 84.94% for the branching RCs. This difference was significant, $t(948) = 5.321$, p < .0001.

For the Branching vs. Nested Sentential Complements sub-experiment, question accuracies were 88.13% for the nested SCs, 87.71% for the branching SCs, and 88.41% for the branching SCs with agreement violations. The differences between the branching and nested conditions, and between the branching and branching with violations conditions, were not significant, both ps >.05.

For the NP-Island Violations sub-experiment, question accuracies were 60.88% for the NP-island violation condition and 73.57% for the control sentences. This difference was significant, $t(951) = 3.267$, p < .002.

Finally, for the Agreement and Plausibility Violations sub-experiment, question accuracies were very high across the board: 91.53% for agreement violations, 92.48% for plausibility violations, and 94.77% for the control sentences. The difference between the

Sub-experiment	Condition	Question 2 Accuracy
Sentential Complements	RC/SC	69.98% (45.88%)
and Relative Clauses	SC/RC	83.79% (36.89%)
Branching vs. Nested	Nested RCs	70.00% (45.87%)
Relative Clauses	Branching RCs	80.75% (39.47%)
Branching vs. Nested	Nested SCs	88.96% (31.37&
Sentential Complements	Branching SCs	87.92% (32.63%)
	Branching SCs w/ Violations	85.30% (35.45%)
NP-Island Violations	NP-Island Violations	66.74% (47.17%)
	Control Sentences	76.96% (42.16%)
Agreement and Plausibility	Agreement Violations	89.67% (30.47%)
Violations	Plausibility Violations	88.10% (32.41%)
	Control Sentences	92.68% (27.08%)

Table 6. Question 2 Accuracies from the Sentence Ratings Task from Study 1.

agreement violations and the control sentences reached significance, $t(962) = 2.020$, p < .05. The difference between the plausibility violations and the control sentences did not reach significance, $t(957) = 1.452$, p >.05.

To summarize the results from the first comprehension question, nearly all comparisons showed an effect. The lack of a difference between the RC/SC (harder) and SC/RC (easier) sentences may be attributable to the fact that even the "easier" versions of these materials were long and rather complex, with multiple embedded clauses, one of which was an object-extracted relative clause. As for the lack of an effect between the plausibility violations and control sentences, it is possible that the implausible words in these violation conditions made them *more* memorable. Unlike the sentences in the SC/RC task, the plausibility violation and control sentences were much simpler.

The answers to the second comprehension question were analyzed in the same way, and mean accuracies by condition are presented in Table 6. For the Sentential Complements and Relative Clauses sub-experiment, question accuracies were 69.98% for the RC/SC (harder) condition and 83.79% for the SC/RC (easier) condition. As predicted, this difference was significant, $t(948) = 5.688$, p < .0001.

For the Branching vs. Nested Relative Clauses sub-experiment, accuracies were 70.00% for the nested RCs and 80.75% for the branching RCs. This difference was also significant, $t(948) = 4.326$, p < .0001.

For the Branching vs. Nested Sentential Complements sub-experiment, mean accuracies were 88.96% for the nested SCs, 87.92% for the branching SCs, and 85.30% for the branching SCs with agreement violations. Neither the difference between nested and branching SCs nor the difference between the branching SCs and branching SCs with agreement violations was

significant, both ps >.05.

For the NP Island Violations sub-experiment, mean accuracies were 66.74% for the NP-island violations and 76.96% for the control sentences. As predicted, control sentence questions were answered significantly better than NP-island violation question, $t(951) = 3.359$, p < .001.

For the Agreement and Plausibility Violations sub-experiment, mean accuracies were 89.67% for agreement violations, 88.10% for plausibility violations, and 92.68% for control sentences. There was no significant difference between the agreement violations and controls sentences (p > .05), but there was a significant difference between the plausibility violations and control sentence, $t(957) = 2.430$, p < .05.

Given that more time had elapsed between reading the sentence and answering the question for question 2, as opposed to question 1, it seemed plausible that performance would be worse overall for the second question. On average, mean accuracy for all sub-experiments for question 1 was 82.93% (SD = 37.62%), and mean accuracy for all sub-experiments for question 2 was 81.78% (SD = 38.60%). This difference approached significance, $t(11462) = 1.66$, p < .10. Given that the second question might be slightly harder, in general, it seemed that this question would be most likely to reveal difficulty differences between the more difficult conditions and their corresponding control sentences. However, the only instances in which a significant difference appeared in the second question, but not the first, were for the plausibility violations. In addition, for one comparison, the agreement violations, the difference only appeared in the first comprehension question. The agreement and plausibility violation materials were undoubtedly the simplest materials in the study, so it is possible that there is a subtle tendency for each of the violations to be more difficult than the control sentences, but that this difference is too small to appear consistently.

4. *Reading Span*. For the reading span task, a span score was determined for each participant. "Span length" refers to the number of sentences for which participants were required to remember the last word for a given trial. The span score was calculated as the highest span length for which a participant was able to correctly recall all of the final words on three out of a possible five trials. An additional half a point was awarded if they were able to completely correctly recall two out of five lists on trials from any higher span length. For this task, span scores ranged from 1.5 (floor performance) to 5.5 across subjects, with a mean span score of 3.40 (SD = 0.92).

5. *Operation Span*. For the operation span task, span score was determined in the same way as for the reading span task. The range of scores was from 3 to 6 (which was ceiling performance). The mean span score for operation span was 5.18 (SD = 0.77).

6. *Digit Span*. For the digit span task, span score was determined in the same way as for the reading and operation span tasks, although the possible range of scores was different (3 to 8, here, compared with 2 to 6 for the complex span tasks). The range of scores across participants was from 4.5 to 8 (ceiling), with a mean of 7.43 (SD = 0.87).

7. *Backwards Digit Span*. For the backwards digit span task, span score was determined in the same way as in the other span tasks, and possible scores ranged from 3 to 8. The range of scores across participants was 4.5 to 8, as in the forward digit span task, and the mean span score was 6.90 (SD = 1.05).

8. *Syllable Span*. For the syllable span task, span score was determined in the same way as in the other span tasks, and possible scores again ranged from 3 to 8. The range of scores across participants was lower than in the other two simple span measures: 2.5 to 6. The mean span score was 3.69 (SD = 1.00).

9. *MSIT*. There were two measures of interest in the MSIT task: the difference in accuracies and reaction times (for the accurate trials) between the interference trials and the no-interference trials. Mean accuracy was 97.37% (SD = 0.16%) for the no-interference trials and 90.21% (SD = 0.30%) for the interference trials, for a mean difference score of 7.16%. As predicted, a linear mixed-effects model with condition as a fixed effect and subject as a random effect revealed that accuracy was worse for the interference trials, $t(20140) = 23.14$, $p < .0001$.

For accurate trials only, mean reaction time for the no-interference trials was 537.17 ms (SD = 173.62), and mean reaction time for the interference trials was 843.76 ms (SD = 211.38). As predicted, a linear mixed-effects model with condition (no-interference or interference) as a fixed effect and subject as a random effect revealed that responses were slower for the interference trials, $t(18889) = 124.10$, $p < .0001$.

List type	List 1 (s)	List 2 (s)	Two-list Average (s)
Congruent	21.10 (4.73)	19.16 (4.68)	20.13 (0.62)
Incongruent	25.35 (6.31)	25.85 (5.98)	25.60 (0.80)
Mixed	25.81 (5.51)	24.35 (5.80)	25.08 (0.74)

Table 7. Mean reading times for Congruent, Incongruent, and Mixed lists from the Stroop task in Study 1.

10. *Go/NoGo*. For the Go/NoGo task, the relevant dependent measure was accuracy on the "NoGo" trials (i.e., trials on which participants were not supposed to respond). Mean accuracy was 76.19% (SD = 18.25%).

11. *Stroop*. For the Stroop task, the time taken to read each of six lists, two for each condition, was recorded: Congruent, Incongruent, and Mixed color words. These times are reported in Table 7. The average of each of the two lists for each condition was computed, and two

48

difference scores were obtained: the difference between the incongruent and congruent conditions, and the difference between the mixed condition and the congruent condition. The average difference between the incongruent and congruent conditions was 5.47 s, and the average difference between the mixed and congruent conditions was 4.95 s. Linear mixed-effects models with condition (congruent, incongruent, or mixed) as a fixed effect and subject as a random effect revealed that as predicted, there were significant differences between both the congruent and mixed conditions ($t(108) = 7.872$, p < .0001) and also the congruent and mixed conditions ($t(118) = 8.214$, p < .0001). There was no significant difference between the incongruent and mixed conditions, p > .05.

12. *RAPM*. For the RAPM task, the dependent measure of interest was the total number of items answered correctly within the allotted forty minutes for the task. Participants ranged from 18 to 36 (ceiling) answers correct on this task. The average score was 27.17 (SD = 4.55).

Block / N	N = 1	N = 2	N = 3
Block 1	.686 (.354)	.638 (.299)	.421 (.309)
Block 2	.883 (.246)	.700 (.231)	.440 (.259)
Block 3	.897 (.239)	.735 (.276)	.428 (.234)
Average	.822 (.299)	.691 (.272)	.430 (.268)

Table 8. Mean accuracy (d') by condition (N=1, N=2, and N=3) and block for the Dual *n*-back task from Study 1.

13. *Dual* n-*back*. For this task, the relevant dependent measures were accuracy (d'; hits – false alarms) and reaction time for targets that were correctly responded to. Accuracy by block and condition (*n*-level) are given in Table 8. For *n*=1, mean accuracy was highest, at .822. For *n*=2, accuracy was lower, at .691. And finally, for *n*=3, mean accuracy was lowest, at .430. Linear mixed-effects models with condition ($n = 1$, 2, or 3) as a fixed effect and subject as a random

effect revealed that there were significant differences for all pair-wise comparisons. Accuracy for

n=1 was significantly higher than accuracy for both n=2 (t(359) = 5.161, p < .0001) and n=3

(t(359) = 14.77, p < .0001). Accuracy for n=2 was also higher than accuracy for n=3, t(360) =

11.34, p < .0001.

Mean reaction time for accurate trials are given in Table 9. Mean reaction times were

1399 ms for n=1, 1642 ms for n=2, and 1759 ms for n=3. Linear mixed-effects models with

condition as a fixed effect and subject as a random effect revealed that there were significant

differences for all pair-wise comparisons, such that reaction time was faster for n=1 than for n=2

(t(355) = 10.33, p < .0001) and for n=3 (t(358) = 11.74, p < .0004) and faster for n=2 than for

n=3 (t(357) = 3.689, p < .0004).

14. *Corsi Block*. There were two relevant dependent measures from the Corsi Block task, each a

different type of "span" measure. The first of these was the maximum span level achieved over

the course of the task ("High Span"). Since the task was adaptive, increasing the level of

Block Number / N	N = 1	N = 2	N = 3
Block 1	1479 (315)	1697 (294)	1797 (377)
Block 2	1368 (314)	1619 (326)	1776 (337)
Block 3	1349 (322)	1611 (319)	1705 (400)
Average	1399 (321)	1642 (314)	1759 (372)

Table 9. Mean reaction time (in ms) for accurate trials by condition (N = 1, 2, or 3) and block for the Dual n-back task from Study 1.

difficulty as participants performed better on the task and decreasing the difficulty once

participants failed to perform correctly on multiple trials at a given difficulty level, the maximum

span level (i.e., the maximum number of spatial blocks they were able to correctly respond to)

was a reasonable measure for a participants spatial working memory. The second dependent

measure was the span level attained at the end of the task, which could be lower than their

absolute maximum ("End Span"). As all participants were given the same number of trials, they each had the opportunity to practice the same amount throughout the experiment, making performance on the final trial a reasonable metric for estimating spatial working memory span. The range for the High Span measure was 7 to 17, and the mean was 10.56 (SD = 2.17). The range for the End Span measure was 4 to 17, and the mean was 8.04 (SD = 2.69).

15. *Verbal Fluency*. There were four relevant dependent measures for the verbal fluency task. First, there were three measures each taken from a different letter from the "F", "A", and "S" tasks. Finally, a composite measure taking the mean performance for (i.e., mean number of words produced to) each of the three letters was computed. For the "F" task, which was performed first, the range was 6 to 24 words produced, with a mean number of words produced at 15.41 (SD = 3.90). Performance for the "A" task ranged from 6 to 23 words, and the mean was 13.30 words (SD = 3.68). Performance for the "S" task ranged from 6 to 25 words, and the mean was 16.83 words (SD = 4.37). Average performance across the three tasks was 15.18 (SD = 3.54).

16. *Spelling test*. The dependent measure for the spelling test was simply the number of words (out of a possible 40) correct. Participants could take as much time as they wanted for this task and were able to listen to words as many times as they wished, at their leisure. The maximum number of words spelled correctly was 39, and the minimum was 16. Mean performance was 28.81 words (SD = 5.53).

D. Results from correlations within language measures

In order to make sense of correlations between linguistic measures and non-linguistic measures, it was necessary to ascertain a) that the linguistic measures exhibited reasonable split-half reliability (that is, that a given measure was reliable within each participant, measured by

estimating how performance on a random sample of half of the measures from the task correlated with the remaining samples) and b) that there were reasonable correlations between the two self-paced reading language tasks and within several of the measures from the Sentence Ratings task.

1. *Split-half reliability of the Cleft Sentences task.* In order to estimate the split-half reliability on each of the sentence processing tasks, reading times (RTs) from the 24 critical regions (the embedded clause region, i.e., region 6 of John's utterance) were isolated for each participant. Twelve of these RTs were from the harder object cleft condition, and twelve were from the easier subject cleft condition. For each condition, six RTs were randomly sampled, and a difference score was computed from the mean RT of the object cleft condition – the mean RT of the subject cleft condition. The same process was performed on the remaining six RTs from each condition, resulting in two difference scores for each participant. Then, a Pearson's r coefficient was computed for each participant. This process was repeated 100 times using a bootstrapping procedure. Unfortunately, the result from this procedure was estimated to be $r = -.273$, indicating that there was little (if any) reliability within a participant for this task.

2. *Split-half reliability of the Relative Clause Sentences task.* The same bootstrapping procedure as the one described above for the Cleft Sentences task was employed to estimate the split-half reliability of the Relative Clause Sentences task. The result was $r = .085$. While this value indicates at least a positive correlation within the Relative Clause Sentences task, the measure is very low, and indicates that there was little if any reliability within a participant for this task, as well as the Cleft Sentences task.

3. *Correlation between the sentence processing tasks.* Given that the split-half reliability estimates were so low for both the Clefts Sentences and Relative Clause Sentences tasks, it is difficult to interpret correlations between either of these tasks and other, non-linguistic tasks. In

addition, the correlation between the two tasks was low. The difference score from the reading times during the embedded/relative clause for the objects (harder) minus the subjects (easier) conditions of each of these tasks was obtained, and the Pearson's correlation for the Clefts Sentences task and the Relative Clause Sentences task for these measures was $r = 0.168$, p $> .05$. Because of this low correlation and the low split-half reliabilities described above, it will be difficult to interpret correlations between these linguistic measures and other non-linguistic measures. However, it is possible that due to the limited size of the samples used in the split-half correlations estimates, these are simply too noisy to be of use (for each estimate, only six difference scores were able to be computed per participant whereas in correlations between each of the sentence processing tasks and other measures, all twelve difference scores factor into the mean difference score). Therefore, correlations between these linguistic measures and the non-linguistic tasks were still computed and are discussed below.

R	Agr. Viol.	Plaus. Viol.	NP Isl. Viol.	SC Agr. Viol.	Nested SC	Nested RC	SC/RC
Agr. Viol.	1.000	-0.192	**0.424**	**0.579**	-0.185	-0.006	-0.023
Plaus. Viol.	---	1.000	-0.193	**-0.297**	0.038	0.227	-0.116
NP Isl. Viol.	---	---	1.000	**0.530**	0.182	**0.258**	-0.012
SC Agr. Viol.	---	---	---	1.000	**0.280**	0.036	0.045
Nested SC	---	---	---	---	1.000	0.196	0.120
Nested RC	---	---	---	---	---	1.000	**0.346**
SC/RC	---	---	---	---	---	---	1.000

p-values	Simple Agr.	Simple Plaus.	NP Island Viol.	SC Agr.	Nested SC	Nested RC	SC/RC
Simple Agr.	---	0.164	**0.001**	**0.000**	0.181	0.963	0.867
Simple Plaus.	---	---	0.161	**0.029**	0.788	0.227	0.404
NP Isl. Viol.	---	---	---	**0.000**	0.189	**0.060**	0.932
SC Agr.	---	---	---	---	**0.040**	0.795	0.747
Nested SC	---	---	---	---	---	0.156	0.389
Nested RC	---	---	---	---	---	---	**0.010**
SC/RC	---	---	---	---	---	---	---

Table 10. Ratings correlations from the Sentence Ratings task from Study 1 (sub-experiments testing working memory manipulations are highlighted, and significant correlations are presented in bold).

4. *Correlations within the Sentence Ratings task.* The primary measures of interest from the Sentence Ratings task were the naturalness ratings (on a scale from 1 to 7) that participants gave to each sentence. As described in the results section for the Sentence Ratings task, difference scores consisting of the z-scored rating for an easier or more well-formed control condition minus the z-scored rating for each experimental condition were obtained for all sub-experiments. The correlations (Pearson's r) for these difference scores for each sub-experiment are presented in Table 10, along with the corresponding p-values. Correlations that reached or closely approached significance are presented in bold. Of particular interest was whether correlations within the sub-experiments testing a working memory manipulation (Sentential Complements and Relative Clauses, Branching vs. Nested RCs, and Branching vs. Nested SCs) would be higher than correlations between these sub-experiments and the other sub-experiments that were not designed to test working memory manipulations (Agreement Violations, Plausibility Violations, NP-Island Violations, and SC Agreement Violations). As predicted, some correlations were observed between the three working memory manipulations. Branching vs. Nested RCs and Branching vs. Nested SCs exhibited a correlation of $r = 0.196$ that nearly approached significance, $p < .160$. Branching vs. Nested RCs and Sentential Complements and Relative Clauses exhibited a correlation of $r = 0.346$, $p < .05$, and Branching vs. Nested SCs and Sentential Complements and Relative Clauses exhibited a non-significant correlation of $r = 0.120$, $p > .05$. Overall, correlations between these sub-experiments and the non-working memory sub-experiments were somewhat lower, ranging from -0.185 to 0.280. A Welch two-sample t-test indicated that the difference between the within-working memory and between-working memory

and non-working memory tasks approached significance, $t(3.951) = -2.429$, p < .08.

E. Results from correlations between all tasks

All relevant correlation coefficients (r values) are listed in Tables 11(a-c). Correlations are emphasized when they reach or approach significance (see tables for a key). Groups of correlations from each type of non-linguistic task are discussed below. There is some redundancy as each type of task is discussed in its own section as well as when relevant to the discussion of other groups of tasks. Correlations with the language tasks are also discussed.

1. *Verbal Span tasks*. As predicted, all five of the span tasks correlated with one another to some extent, all ps < .05. Furthermore, the simple span and complex span composite measures were very highly correlated, $r = .620$, p < .001. The two complex span measures (Reading Span and Operation Span) were both highly correlated with RAPM ($r = .319$, p < .05 for Reading Span; $r = .464$, p < .001 for Operation Span). RAPM is thought to measure complex reasoning ability

Rs	R-Span	RAPM	Spel	Str1	Str2	VF-F	VF-A	VF-S	VF-Avg	Clefts Diff	RCs Diff	Clefts QDiff	RCs QDiff	RCSC SCRC	Bran./ Nest RC	Ctrl-Plaus	Ctrl-Agr
RSpan[†]	-	**0.319**	0.224	-0.171	**-0.297**	0.167	*0.480*	*0.257*	**0.333**	0.026	*-0.255*	-0.073	-0.187	-0.171	-0.077	-0.076	0.080
RAPM	-	-	**0.281**	-0.120	-0.115	0.125	*0.266*	*0.256*	*0.243*	-0.159	-0.149	**-0.357**	-0.042	0.104	-0.064	-0.123	0.148
Spel	-	-	-	-0.213	*-0.248*	0.200	**0.297**	*0.265*	**0.285**	**-0.334**	-0.083	-0.271	0.193	0.190	0.026	-0.079	0.214
Str1	-	-	-	-	*0.815*	0.038	*-0.133*	-0.011	-0.037	0.049	-0.148	0.105	0.035	0.089	0.111	0.022	-0.135
Str2	-	-	-	-	-	-0.056	*-0.245*	-0.054	-0.128	-0.049	-0.023	0.060	0.035	0.073	0.076	-0.015	-0.150
VF-F	-	-	-	-	-	-	*0.659*	*0.714*	*0.888*	-0.072	0.090	-0.007	0.141	0.085	0.112	-0.069	0.132
VF-A	-	-	-	-	-	-	-	*0.683*	*0.869*	-0.027	0.012	0.011	0.070	-0.062	0.110	-0.007	**0.285**
VF-S	-	-	-	-	-	-	-	-	*0.909*	-0.142	0.136	-0.117	0.019	0.040	0.109	-0.004	0.086
VF-Avg	-	-	-	-	-	-	-	-	-	-0.094	0.093	-0.047	0.084	0.026	0.124	-0.030	0.182
Clefts Diff	-	-	-	-	-	-	-	-	-	-	0.168	-0.028	0.091	-0.219	-0.054	0.093	**-0.330**
RCs Diff	-	-	-	-	-	-	-	-	-	-	-	-0.139	**0.269**	-0.139	*0.226*	0.129	-0.137
Clefts QDiff	-	-	-	-	-	-	-	-	-	-	-	-	-0.123	-0.188	0.193	0.056	-0.192
RCs Qdiff	-	-	-	-	-	-	-	-	-	-	-	-	-	0.092	0.107	-0.028	-0.137
RCSC/SCRC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**0.346**	-0.116	-0.023
Branch/NestRC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-0.168	-0.007
Ctrl-Plaus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-0.192

Table 11a. Correlations between all tasks from Study 1, part 1. (*p < .001*, *p < 01*, **p < .05**, *p < .10*)

[1†] RSpan = Reading Span; RAPM = # correct on Raven's Advanced Progressive Matrices; Spel = # correct on the Spelling Test; Str1 = Differences between RT for the Stroop Incongruent – Congruent lists; Str2 = Differences between RT for the Stroop Mixed – Congruent lists; VF-F = # of responses for Verbal Fluency "F"; VF-A = # of responses for Verbal Fluency "A"; VF-S = # of responses for Verbal Fluency "S"; VF-Avg = Average # of responses for Verbal Fluency F, A, and S lists; Clefts Diff = RT difference of Object – Subject Cleft versions at the critical region; RCs Diff = RT difference of Object – Subject RC versions at the critical region; Clefts QDiff = accuracy difference between Subject – Object Cleft questions; RCs QDiff = accuracy difference between Subject – Object RC questions; RCSC/SCRC = rating score difference of SC/RC – RC/SC sentences; Branching/NestRCs = rating score difference of branching – nested RCs; Ctrl-Plaus = rating score difference of control – plausibility violation sentences; Ctrl-Agr = rating score difference of control – agreement violation sentences.

Rs	NPIsl†	Br/Nest SC	Br/Viol SC	DSF	Syl	BDS	OS	MSIT	Go/NoGo	1Bk	2Bk	3Bk	Corsi End	Corsi High	Comp Span	Sim Span	All Span
RSpan	0.075	-0.137	0.003	0.336	0.357	0.464	0.370	-0.144	-0.146	0.115	0.167	0.223	0.029	-0.041	0.856	0.466	0.691
RAPM	0.162	-0.131	-0.134	0.372	0.532	0.423	0.464	-0.008	-0.226	0.424	0.438	0.567	0.230	0.315	0.465	0.542	0.570
Spel	0.051	-0.076	0.065	0.310	0.214	0.438	0.344	-0.020	0.152	0.300	0.240	0.255	0.108	-0.009	0.337	0.357	0.407
Str1	-0.231	-0.001	-0.116	-0.254	-0.086	-0.313	-0.152	0.046	-0.058	0.024	-0.028	-0.231	0.101	-0.000	-0.196	-0.256	-0.268
Str2	-0.314	0.012	-0.178	-0.283	-0.198	-0.280	-0.127	0.074	-0.069	-0.077	-0.058	-0.132	0.026	0.0316	-0.264	-0.263	-0.311
VF-F	0.087	-0.078	-0.078	0.294	0.220	0.142	0.092	-0.016	-0.147	0.268	0.194	0.249	0.110	-0.025	0.159	0.1979	0.2236
VF-A	0.375	-0.084	0.111	0.344	0.387	0.243	0.227	0.067	-0.084	0.246	0.274	0.433	0.005	-0.120	0.438	0.342	0.440
VF-S	0.054	-0.118	-0.162	0.316	0.183	0.260	0.219	0.038	-0.198	0.309	0.234	0.315	0.036	-0.088	0.289	0.336	0.352
VF-Avg	0.184	-0.106	-0.057	0.357	0.286	0.243	0.202	0.033	-0.165	0.308	0.262	0.371	0.057	-0.087	0.329	0.329	0.379
Clefts Diff RCs	-0.034	0.128	-0.070	-0.166	-0.098	-0.236	0.070	-0.201	-0.117	-0.151	-0.299	-0.174	0.024	0.097	0.056	-0.220	-0.128
Diff Clefts	0.082	0.189	-0.037	-0.128	-0.043	-0.166	-0.132	-0.012	-0.069	-0.199	-0.235	0.031	-0.066	-0.065	-0.239	-0.177	-0.218
QDiff RCs	-0.060	0.107	-0.091	-0.072	-0.135	-0.205	-0.357	0.360	-0.065	-0.173	-0.061	-0.099	-0.046	-0.016	-0.246	-0.206	-0.235
QDiff RCSC	0.009	0.233	0.016	0.234	0.017	-0.116	0.082	-0.098	0.254	0.275	-0.074	0.040	-0.114	-0.061	-0.075	0.015	-0.019
SCRC	-0.012	0.120	0.045	0.062	0.148	0.088	-0.003	0.153	0.161	-0.116	-0.016	0.0362	0.105	0.086	-0.113	0.135	0.046
Br./Nest RC	0.258	0.195	0.036	0.092	0.147	-0.126	-0.030	0.240	0.179	-0.303	-0.054	0.143	0.118	0.083	-0.066	-0.047	-0.036
Ctrl-Plaus	-0.194	0.037	-0.298	-0.253	-0.303	-0.139	-0.332	0.082	-0.112	-0.177	-0.088	-0.170	-0.283	-0.306	-0.234	-0.331	-0.314
Ctrl-Agr	0.424	-0.185	0.579	0.190	0.347	0.197	-0.065	-0.136	-0.003	0.266	0.299	0.222	0.105	-0.020	0.015	0.307	0.222

Table 11b. Correlations between all tasks from Study 1, part 2. (*p < .001*, *p < .01*, **p < .05**, *p < .10*)

---

† NPIsl = rating score difference of control – NP-island violation sentences; Br/NestSC = rating score difference of branching – nested SC sentences; Br/ViolSC = rating score difference of branching SC – branching SC with violations sentences; DSF = forward digit span; Syl = syllable span; BDS = backwards digit span; OS = operation span; MSIT = difference score for accuracy of control – interference conditions; Go/NoGo = accuracy on the "NoGo"trials of the Go/NoGo task; 1Bk = accuracy (d') on the 1-back conditions of the dual n-back task; 2Bk = accuracy (d') on the 2-back conditions of the dual n-back task; 3Bk = accuracy (d') on the 3-back conditions of the dual n-back task; CorsiEnd = final "span" score on the Corsi Block task; CorsiHigh = highest "span" score on the Corsi Block task; CompSpan = complex span average of both complex verbal span tasks; SimSpan = simple span average of all three simple verbal span tasks; AllSpan = composite score averaging all five verbal span tasks.

Rs	NPIsl	Br/Ne st SC	Br/Vio 1 SC	DSF	Syl	BDS	OS	MSIT	Go/ NoGo	1Bk	2Bk	3Bk	Corsi End	Corsi High	Comp Span	Sim Span	All Span
NPIsl	-	0.182	*0.530*	0.167	*0.389*	0.001	0.062	0.072	0.094	-0.100	0.024	*0.232*	0.006	-0.028	0.083	0.213	0.185
Br/Ne st SC	-	-	**0.280**	-0.015	-0.035	-0.080	-0.198	-0.013	**0.281**	-0.120	*-0.273*	-0.096	0.015	0.050	-0.200	-0.033	-0.112
Br/Vio 1 SC	-	-	-	0.057	*0.269*	0.029	-0.110	-0.039	*0.255*	0.000	-0.053	-0.076	0.041	-0.048	-0.059	0.171	0.089
DSF	-	-	-	-	**0.301**	*0.595*	*0.528*	-0.001	-0.024	*0.506*	*0.427*	0.267	-0.064	-0.123	*0.512*	*0.781*	*0.758*
Syl	-	-	-	-	-	**0.285**	*0.368*	0.119	-0.110	*0.274*	0.189	*0.449*	0.115	0.128	*0.441*	*0.660*	*0.630*
BDS	-	-	-	-	-	-	*0.477*	**-0.303**	-0.046	*0.423*	*0.433*	0.271	0.001	-0.033	*0.568*	*0.782*	*0.794*
OS	-	-	-	-	-	-	-	-0.209	0.036	**0.325**	0.215	**0.314**	0.124	0.081	*0.797*	*0.569*	*0.730*
MSIT	-	-	-	-	-	-	-	-	-0.107	**-0.303**	-0.059	0.130	-0.213	*-0.235*	-0.210	-0.099	-0.157
Go/No Go	-	-	-	-	-	-	-	-	-	-0.046	*-0.257*	**-0.299**	-0.192	*-0.275*	-0.075	-0.077	-0.085
1Bk	-	-	-	-	-	-	-	-	-	-	**0.432**	*0.241*	-0.020	0.002	*0.264*	*0.518*	*0.469*
2Bk	-	-	-	-	-	-	-	-	-	-	-	**0.433**	*0.238*	*0.241*	0.226	*0.395*	*0.380*
3Bk	-	-	-	-	-	-	-	-	-	-	-	-	*0.227*	*0.253*	0.320	*0.396*	*0.415*
Corsi End	-	-	-	-	-	-	-	-	-	-	-	-	-	*0.817*	0.088	0.002	0.048
Corsi High	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.019	-0.011	0.001
Comp Span	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*0.620*	*0.856*
Sim Span	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*0.930*

Table 11c. Correlations between all tasks from Study 1, part 3. (*$p \leq .001$*, *$p < .01$*, **$p < .05$**, *$p < .10$*)

58

and fluid intelligence (Gf), and given the high correlation between RAPM and Operation Span, it is perhaps not surprising that Operation Span also correlated highly with two measures from the dual n-back task, accuracy (d') on the 1-back and 3-back tasks. The dual n-back task has previously been shown to correlate highly with, and improve scores on, the RAPM task (Jaeggi et al., 2008, 2010). The reading span task was also highly correlated with an executive function measure, namely the difference between the Mixed and Congruent conditions of the Stroop task ($r = -.297$, p < .05). This measure from the Stroop task should function as an index not only of inhibition (i.e., repressing the urge to read a word rather than name the color of the ink) but also of fluid switching between easier trials (where the color of the ink matches the word) and the more difficult, incongruent trials. Such processes may be at play in the reading span task, which likely requires suppression of task-irrelevant words (i.e., words other than the last word in each sentence that is read aloud) but activation of task-relevant words (i.e., the last word of each sentence). In addition, the reading span task was correlated with several measures from the verbal fluency task (the highest correlation, with number of words produced for the letter "A", reaching a correlation of $r = .480$, p < .001). Finally, the operation span task was also highly correlated with the spelling task, $r = .344$, p < .05.

The simple span tasks were correlated with many of the same measures as the complex span tasks, though to varying degrees. Syllable span, thought to be a good measure of pure phonological working memory, was correlated quite strongly with the two measures thought to be the best indexes for general fluid intelligence (Gf): RAPM ($r = .532$, p < .001) and the Dual 3-back task ($r = .449$, p < .001). Digit span, a task which is easier than the syllable span task, was quite strongly correlated with the easier Dual 1-back task, $r = .506$, p < .001, and correlated less strongly with the Dual 2-back task, $r = .427$, p < .01. Interestingly, the backwards digit span task,

which is more difficult than forwards digit span but easier than syllable span, was correlated with all three n-back task levels: Dual 1-back (r = .423, p < .01), Dual 2-back (r = .433, p < .01), and Dual 3-back (r = .0271, p < .05). In addition, digit span was also correlated with verbal fluency (all measures; ps < .05), RAPM (p < .01), spelling (p < .05), and the Mixed – Congruent difference on the Stroop task (p < .05). Syllable span was also correlated with verbal fluency (reaching the highest correlation for the letter "A", p < .01). Backwards digit span was also correlated with spelling (p < .001), RAPM (p < .01), and several executive function measures, including MSIT (p < .05) and both Stroop measures (both ps < .05).

2. *Cognitive control tasks.* The three cognitive control tasks unfortunately did not exhibit high correlations amongst themselves (all ps > .10), with the exception of the two Stroop measures exhibiting a very high correlation, as expected, r = .815, p < .001.

However, Stroop (Congruent – Incongruent) exhibited a high correlation with the backwards digit span and the combined span (from all span measures) scores, both ps < .05. Presumably, the backwards span task, more than the other span tasks which do not require mentally rearranging digits or other symbols, taps the ability to inhibit interference from certain digits while accessing digits occurring in a backwards order. The second Stroop measure (Congruent – Mixed) showed many correlations with other tasks, including reading span, NP-island violations, digit span, backwards digit span, and the composite span measure comprising all of the span tasks. This Stroop measure taps the ability to switch between inhibiting a prepotent response (i.e., reading the name of the word, and not the ink color) and responding to an easier component of the task, when both the name and ink color of the word match. This task switching skill may tap resources that are more general than the inhibition skills required for the Stroop Incongruent task.

The MSIT task was correlated with only a few tasks, including the backwards digit span task ($r = -.303$, p < .05), the dual 1-back task ($r = -.303$, p < .05), and interestingly, the Clefts question accuracy difference score ($r = .360$, p < .01). The dual 1-back task may require inhibitory skills similar to those necessary for the backwards digit span task and the dual 1-back task (the accuracy measure, d', also incorporates false alarms, which may measure such inhibitory skills). As for the Clefts question accuracy difference, it may be the case that interference resulting from the more difficult (object-extracted) syntactic constructions overlaps in some way with the type of more general interference experienced during the MSIT task. However, this was the only cognitive control task correlating with a language measure, so these results should be interpreted precautionarily.

The Go/NoGo task also correlated with one of the n-back measures (dual 3-back), $r = -.299$, p < .05, as well as the Branching vs. Nested SCs ratings measure, $r = .281$, p < .05. The dual 3-back task requires a great deal of cognitive effort and inhibition of multiple sounds (letters) and visual cues at a time, so inhibiting a prepotent response to a lure and inhibiting a response to the "NoGo" stimuli may require similar inhibition mechanisms. The relationship between the Go/NoGo task and the Branching vs. Nested SCs ratings measure is more difficult to interpret, particularly because similar ratings measures (e.g., Branching vs. Nested RCs and the RC/SC order manipulations) were not correlated with any of the cognitive control tasks.

3. *General intelligence tasks.* The RAPM task is one of the best tasks for estimating *Gf*, or general fluid intelligence. Unsurprisingly, it therefore correlated with a host of other tasks from several categories. Of the span tasks measures, it correlated very highly with syllable span ($r = .532$, p < .001), operation span ($r = .464$, p < .001), the composite complex span score ($r = .465$, p < .001); the composite simple span score ($r = .542$, p < .001), and the composite all span score (r

= .570, p < .001); it also correlated slightly less strongly with digit span (r = .372, p < .01) and

backwards digit span (r = .423, p < .01), and more moderately with reading span (r = .319,

p < .05). These correlations between Raven's and verbal span tasks are similar to those that have

been observed in other studies (Kane, Hambrick, Tuholski, Wilhelm, Payen, & Engle, 2004;

Kane, Conway, Miura, & Colflesh, 2007; Shamosh, DeYoung, Green, Reis, Johnson, Conway,

Engle, Braver, & Gray, 2008). In addition, RAPM correlated to varying degrees with all of the

measures from the dual n-back task, including dual 3-back (r = .567, p < .001), dual 2-back (r = .

438, p < .01), and dual 1-back (r = .424, p < .01), similar to previous findings investigating the

relationship between these two tasks (Jaeggi et al., 2010). RAPM correlated more moderately

with several other measures, as well, including the spelling task (r = .281, p < .05) and the high

span measure from the Corsi Block task (r = .315, p < .05). Most surprisingly, RAPM was one of

the few tasks to correlate with a question accuracy difference score from one of the sentence

processing tasks, the Cleft Sentences task, r = -.357, p < .01. That is, the better a score on the

RAPM, the smaller the difference between question accuracies for harder (object-extracted) and

easier (subject-extracted) cleft sentence questions.

Some, but not all, of the dual n-back measures were highly inter-correlated. Dual 1-back

correlated highly with the dual 2-back task (r = .432, p < .01), but the correlation with the more

complex dual 3-back task did not reach significance (p < .10). Dual 2-back and dual 3-back were

also highly correlated, (r = .432, p < .01). This pattern suggests that, as anticipated, the n-back

tasks got progressively more difficult and perhaps even recruited different mechanisms in

performance on the more difficult dual 3-back task compared to the easier dual 1-back task.

The dual n-back measures also correlated highly with several other working memory and

cognitive control tasks. The dual 1-back task, the simplest task, correlated very highly with

forward digit span ($r = .506$, p < .001), composite simple span ($r = .518$, p < .001), and the composite measure from all span tasks ($r = .469$, p < .001). The correlations between the 1-back task and other span tasks were slightly smaller but still highly significant, including backwards digit span ($r = .423$, p < .01) and operation span ($r = .325$, p < .05). Other nonlinguistic tasks correlating with the dual 1-back were RAPM ($r = .424$, p < .01), spelling ($r = .300$, p < .05), verbal fluency "S" ($r = .309$, p < .05), verbal fluency average ($r = .308$, p < .05), and MSIT ($r = -.303$, p < .05). The negative correlation with MSIT is to be expected, as a larger difference score between congruent and incongruent trials is more likely to be associated with lower scores on taxing tasks such as the dual n-back tasks. Interestingly, the dual 1-back task was also correlated with the branching vs. nested RC ratings ($r = -.303$, p < .05). It is plausible that keeping track of multiple cues in the 1-back task requires some of the processes necessary to keep track of doubly-embedded structures that involve keeping track of multiple long-distance dependencies at once in the nested RC sentences. However, given that the 1-back task did not correlate highly with any other similar linguistic ratings measure, this finding should be interpreted with caution.

The dual 2-back task also correlated to various degrees with many of the span measures, including forward digit span ($r = .427$, p < .01), backwards digit span ($r = .433$, p < .01), composite simple span ($r = .395$, p < .01), and the composite measure of all the span tasks ($r = .380$, p < .01). In addition, the 2-back task correlated highly with the RAPM ($r = .438$, p < .01), supporting previous findings (Jaeggi et al., 2010). Interestingly, the dual 2-back task was one of the few tasks to correlate highly with an online sentence processing measure, the Clefts Sentences extraction effect ($r = -.299$, p < .05). The negative correlation would be expected, as a larger extraction effect would indicate more difficulty on related tasks. The dual 2-back task requires that individuals remember a visual cue and spoken letter while similar

material intervenes for one trial, and both this task and the Clefts Sentences task may require similar resistance to interference from intervening material. Although the dual 2-back task did not correlate highly with the similar Relative Clause Sentences task, the correlation between the two sentence processing tasks was also low, indicating that internal reliability in the sentence processing tasks may be an issue in finding all correlations with nonlinguistic tasks.

Finally, the dual 3-back task was quite highly correlated with RAPM ($r = .567$, p < .001), as observed in prior studies (Jaeggi et al., 2008, 2010). In addition, dual 3-back was correlated to varying degrees with many verbal span task measures, including syllable span ($r = .449$, p < .001), composite simple span ($r = .396$, p < .01), the composite measure for all of the span scores ($r = .415$, p < .01), backwards digit span ($r = .271$, p < .05), and operation span ($r = .314$, p < .05). In addition, dual 3-back was correlated with several of the verbal fluency measures, including the verbal fluency average ($r = .371$, p < .01), verbal fluency "A" ($r = .433$, p < .01), and verbal fluency "S" ($r = .315$, p < .05). Finally, the dual 3-back task was moderately correlated with the Go/NoGo task ($r = -.299$, p < .05), perhaps due to the fact that the ability to inhibit a prepotent response on the Go/NoGo task became crucial on the very taxing dual 3-back task.

4. *Spatial working memory task.* Unsurprisingly, the two Corsi block measures, Corsi high span and Corsi end span, correlated quite highly with each other, $r = .817$, p < .001. The Corsi high span measure was correlated with the RAPM task, $r = .315$, p < .05 (and the correlation between the Corsi end span measure and RAPM approached significance, with p < .10). This correlation is unsurprising given that while thought to measure domain-general fluid intelligence, the RAPM is a spatial reasoning task. Surprisingly, the Plausibility Violations ratings scores correlated significantly with both the Corsi high span measure ($r = -.306$, p < .05) and the Corsi end span

measure ($r = -.283$, p < .05). This correlation was not predicted, and is difficult to interpret. One possible explanation is that the fluid visuo-spatial skills tapped in the Corsi block task were useful in interpreting plausibility in the Plausibility Violations ratings task, as the items for the plausibility materials were in large part composed of highly imageable nouns and verbs (see Appendix C). It is therefore possible that fluidity in imagery related to general fluid visuo-spatial working memory skills contributes to making plausibility judgments. This correlation, should it be corroborated in future studies, is quite interesting, though not relevant for the current primary research question.

5. *Other tasks*. As described in preceding sections, several measures from the verbal fluency task correlated with many of the span scores. In particular, the verbal fluency score for the "A" task was most frequently highly correlated with other measures, including Reading Span, Syllable Span, the Dual 3-back task, the spelling task, the digit span task, and all composite span measures, as well as with the other letters from the FAS Verbal Fluency task (all ps < .05). The "F" and "S" tasks were more moderately correlated with a subset of these tasks, including digit span, Dual 1- and 3-back tasks, the spelling task, and composite span measures (all ps < .05).

The spelling task was correlated with many of the verbal span measures, some of the general intelligence measures, and several other verbal and linguistic tasks. These findings converge to suggest that spelling may involve mechanisms that are relevant to both tasks tapping general fluid intelligence and also verbal, and even syntactic, skills. In particular, the span tasks that correlated with the spelling task were backwards digit span ($r = .438$, p < .001), composite simple span ($r = .407$, p < .01), the composite measure for all span scores ($r = .407$, p < .01), digit span ($r = .310$, p < .05), operation span ($r = .344$, p < .01), and composite complex span ($r = .337$, p < .05). In addition, spelling correlated with RAPM ($r = .281$, p < .05) and the dual 1-

back task ($r = .300$, p < .05) and with two verbal fluency measures, verbal fluency average ($r = .285$, p < .05) and verbal fluency "A" ($r = .297$, p < .05). Finally, spelling was also correlated with both the sentence processing extraction effect from the Cleft Sentences task ($r = -.334$, p < .05) and with the question accuracy difference for the Cleft Sentences task ($r = -.271$, p < .05). These negative scores mean that the better an individual's spelling skills, the smaller the difference between processing and answering questions about the more difficult (object-extracted) versus easier (subject-extracted) cleft sentences. This finding is consistent with the hypothesis that good spellers have more stable representations of words (Perfetti & Lesgold 1977; Perfetti, 1985, 2007), which may be important for sentence comprehension and perhaps particularly for keeping track of words in complex syntactic dependencies (as in the object-extracted clefts and RCs materials in the current studies; Perfetti & Lesgold, 1977; Van Dyke, Johns, & Kukona, 2010). This hypothesis is compatible with retrieval interference explanations of sentence complexity (in, e.g., the standard object-/subject-extraction effect), and is further explored in the General Discussion.

F. General Discussion

Study 1 was a first pass toward investigating a wide array of tasks tapping memory and cognitive mechanisms, such as susceptibility to interference, and to compare performance on these tasks with performance on an array of linguistic tasks that manipulated various kinds of syntactic complexity. In particular, the aim was to test which aspects of memory predicted performance in sentence processing. A second goal was to determine which of the sentence processing tasks were internally reliable and would therefore be optimal for further individual difference studies of language processing.

In terms of addressing the latter goal, the two sentence processing tasks (Cleft Sentences

and RC Sentences) that were chosen for this experiment did not exhibit high split-half reliability. These materials manipulated the syntax of either a cleft or relative clause construction in more difficult (object-extracted) or easier (subject-extracted) constructions in supportive contexts (following a set-up in which the referents in the relevant syntactic constructions were described before the syntactic manipulation). In previous sentence processing studies, Fedorenko and Gibson (personal communication) observed that performance on such materials in supportive contexts exhibited a quantitatively (and perhaps qualitatively) different type of extraction effect, (i.e., the difference between the more difficult (object-) and simpler (subject-extracted) versions). First, the extraction effect was larger using materials in supportive contexts, and second, items presented in supportive contexts were thought to exhibit higher split-half reliability than items presented in null contexts (i.e., in single sentences). However, smaller sample sizes and less rigorous tests of split-half reliability led to these findings. The current study does not replicate the latter finding. Therefore, Study 2 set out to explicitly test whether performance on sentence processing tasks involving syntactic manipulations presented in null contexts may be more internally reliable than performance on similar materials presented in supportive contexts.

Even though the sentence processing materials used in this study did not exhibit high internal reliability, the reader should recognize that it might be difficult to accurately estimate this split-half reliability due to the limited number of items used in the study. That is, since each condition of each sentence processing experiment only had twelve items, correlations were being computed over difference scores composed of an average of six items from the syntactically difficult condition minus the average of six items from the syntactically easier condition. Given that the two sentence processing tasks that were chosen have exhibited replicable extraction effects in many sentence processing studies (Wanner & Maratsos, 1978; Gibson, 1998, 2000;

Gordon et al., 2001, 2002, 2004; Fedorenko et al., 2006, 2007), the correlations between these tasks and other memory and cognitive tasks provide relevant data toward answering the primary research question: which memory and cognitive mechanisms predict sentence processing skill?

A chief aim of this work is to disentangle capacity-based theories of working memory and retrieval-interference-based theories of working memory. Current research in psycholinguistics has shown that retrieval interference is a crucial phenomenon for sentence processing, particularly in syntactically complex constructions (McElree, 2000, 2006; Lewis & Vasishth, 2005; Lewis et al., 2006; Van Dyke & McElree, 2006, 2011). Therefore, it seems crucial to conceptualize language processing as operating under the same constraints as general memory processes, which are subject to certain limitations and properties such as susceptibility to interference in retrieval. Below, the profile of the correlations between the general cognitive and memory tasks and each language task is discussed in detail, with particular attention given to correlations between sentence processing tasks and other tasks.

Neither dependent measure (RTs and question accuracies) from the Cleft Sentences task was highly correlated with many other measures. The clefts RT difference score was correlated with three other tasks: the spelling task, the agreement violations sub-experiment from the Sentence Ratings task, and the dual 2-back measure from the Dual n-back task. At least one, and possibly two, of these correlations supports the hypothesis that variability in sentence processing is mediated by susceptibility to retrieval interference.

The first of these is the correlation between the cleft RT extraction effect and the 2-back task. The dual n-back task is perhaps the most complicated task included in the battery of nonlinguistic tasks. It requires attending to two streams of information at once (a simultaneous presentation of a spatial cue in a grid displayed on a computer screen and a stream of auditory

cues), holding multiple streams of information in memory at once, and correctly recognizing one (and at times, two) target cues that match the memory items at the appropriate sequential location. In addition, it requires inhibiting responses to intervening cues that may interfere with the correct target, either because the cues are similar in content (e.g., a similar-sounding auditory letter or a spatial cue in a similar location) or because they are in an incorrect, but nearby, location in the sequence. In particular, in the case of the dual 2-back and dual 3-back measures, one or two intervening cues must be encoded in memory as a potential target but must *not* be used for comparison in judging whether the current cue is a correct target for a match. A similar phenomenon must occur in language processing, particularly at the point of processing displaced words in long-distance dependencies. For instance, in one of the object cleft sentences from the current study, *...I heard that it was the producer that the novelist mocked for wearing a tie*, at the point of processing the verb *mocked*, the noun phrase *the producer* must be understood as the object of *mocked*. This noun phrase was encountered multiple words earlier in the sentence. It has already been understood as the object of the verb *was* and the head of the embedded clause *that the novelist mocked* and must now also be understood in a third role, the object of the verb *mocked*, while it is no longer available for direct access in the comprehender's current visible section of text. This process therefore is directly comparable to a similar situation in the 2-back task, where the participant must recognize that a current target matches a target seen or heard previously (two targets back) in a continuous stream of stimuli. Should such a correlation between similar sentence processing tasks and *n*-back-type tasks be replicated in future work, this finding can be interpreted as evidence for a more general mechanism in sentence processing that likely taps into an individual's ability to inhibit irrelevant, interfering materials and to correctly access a relevant target encountered recently in a continuous stream of information.

The second correlation which may support the hypothesis that sentence processing relies on domain-general abilities of retrieval mechanisms to inhibit interference from competing information is not as straightforward. The spelling task was included in the current study for two reasons. First, there was a general interest in testing this language-related ability against many other memory and cognitive mechanisms. Second, there is some evidence that the quality of lexical representations, including the quality of both phonetic/phonological representations and a representation of the orthography, is directly relevant to retrieving words in sentence processing. A great deal of research has suggested that neighborhood density effects, or effects relating to how many phonological competitors a word has, strongly influence the ability to recognize an isolated word in normal spoken word recognition and especially in spoken word recognition in degraded conditions, such as in noise (Pisoni, Nusbaum, Luce, & Slowiaczek, 1985; Sommers, Kirk, & Pisoni, 1997; Luce & Pisoni, 1998). Part of the difficulty in accessing words in highly dense phonological neighborhoods is thought to result from transient activation followed by subsequent inhibition of competitors. Relatedly, if a word's phonological form and orthography are not as firmly fixed in an individual's lexical representation of that word, it logically follows that such a word may be even more susceptible to interference from competing representations. Work by Perfetti and colleagues (Perfetti & Lesgold, 1977; Perfetti & Hart, 2002; Perfetti, 1985, 2007) has suggested that individuals with poorer lexical representations perform worse on tests of reading comprehension. Recently, some work investigating individual differences in sentence comprehension has also suggested that poor lexical quality may lead to poorer comprehension. In an individual differences study using many memory and non-syntactic verbal measures in an attempt to predict performance on sentence comprehension, Van Dyke and colleagues found that poor receptive vocabulary, as measured by a standard vocabulary test, interacted with syntactic

complexity such that participants who had poorer vocabulary skills performed worse in answering questions about sentences containing highly complex syntactic constructions (Van Dyke et al., 2010). In particular, the sentence materials used in this study manipulated syntactic constructions thought to index difficulty in resolving retrieval interference from competing material in the sentences (materials were taken from Van Dyke et al. 2006). The finding that poorer lexical representations, as indexed by poor receptive vocabulary, related to difficulty in understanding more syntactically complicated sentences in this study was interpreted in the following way. Poorer lexical representations are less specific and may lead to a noisier retrieval cue-target match, which increases interference (cf. Crowder, 1976). Poor receptive vocabulary is thought to be related to poorer overall lexical representations, including partially overlapping representations of orthography, phonology, and meaning, and perhaps less tightly linked representations of the three (Seidenberg & McClelland, 1989). As spelling skills provide one index of such lexical representations, it follows that if lexical representation quality is relevant for retrieval in complicated syntactic constructions, such as long-distance dependencies, the spelling test included in the current set of studies should correlate with performance on the sentence processing tasks. This interpretation should be taken with caution, as the relationship between syntactic processing and spelling ability is likely mediated by phonological representations and is, therefore, not straightforward. In future research, this relationship should be further explored by testing spelling skills more thoroughly and investigating their connection with several sentence-processing tasks. More research is need to fully test this hypothesis.

The third task to correlate with the cleft RTs, the agreement violations sub-experiment from the Sentence Ratings task, is not directly predicted by any hypothesis relating to working memory capacity or retrieval interference. However, performance on the sentence-processing

task is measured somewhat implicitly via reaction times whereas performance on the ratings task involves a conscious decision on the part of the participant. The two tasks were negatively correlated, indicating that a larger difference in processing the easier vs. more difficult sentences in the clefts task was proportionate to a smaller difference score in overtly judging the agreement violations as being less grammatical than control sentences. One interpretation of this finding is that participants who had an easier time on the complex syntactic structures were more sensitive to any errors in agreement whereas participants who experience more difficulty on the sentence-processing task were less sensitive to (or arguably more tolerant of) grammatical errors in the agreement violations task. As these results were not predicted by any known hypothesis, they are interpreted with skepticism. Future studies could include a variety of grammatical violations similar to the agreement violations and test whether the relationship with complex sentence processing skills replicates.

The second dependent measure from the Cleft Sentences task, the comprehension question accuracy difference score, correlated highly with more measures than did the difference scores from reading times, including correlations with the Raven's matrices, operation span, the MSIT, and the spelling test. Before discussing these correlations, there is a key difference to point out between the comprehension question measure and the reaction times measure, namely that reaction times index an online processing skill whereas comprehension question accuracy reflects offline processing. Caplan & Waters (1999), among others (see Fodor, 1983; Forster, 1979; Frazier, 1990), distinguish between these two types of processes in sentence comprehension in "...the distinction between the extraction of meaning from a linguistic signal, which [they] call 'interpretive processing,' and in the use of that meaning to accomplish other tasks, such as storing information in long-term semantic memory, reasoning, planning actions,

and other functions, which [they] call 'post-interpretive processing.' By 'interpretive processing,' [they] refer to the processes of recognizing words and appreciating their meanings and syntactic features; constructing syntactic and prosodic representations; and assigning thematic roles, focus, and other aspects of propositional discourse-level semantics... [T]he processes involved in interpretive process [may be] distinct from those involved in other verbally mediated functions" (Caplan & Waters, 1999, p. 78). Given the distinction that has been drawn between "online" measures and "offline" measures, the tasks predicting scores on the comprehension question measure should not be interpreted as predicting working memory skills in sentence processing but rather as predicting other sentence comprehension skills.[4]

As post-interpretive language processing is thought to involve many complex cognitive mechanisms, including reasoning skills, it is not surprising that post-interpretive question accuracies measure correlated highly with the Raven's progressive matrices (RAPM) measure. RAPM is thought to be one of the best tasks for estimating general fluid intelligence (Gf), which correlates with many high-level cognitive abilities. Another task correlating with the clefts question accuracy scores was operation span, a complex span task known to correlate highly with RAPM (a finding replicated in the current study, as well). MSIT, a task known for its neuro-cognitive reliability in activating the brain network involved in attention and measuring interference effects (Bush & Shin, 2006) was also highly correlated with the clefts question accuracy score. This finding may be more reflective of an individual's ability to concentrate on comprehension questions, rather than being a sign of resistance to interference from inappropriate responses. Finally, the correlation between spelling and cleft comprehension

---

[4] NB: It is interpretive processing, rather than post-interpretive processing, that many researchers have claimed to be modular with respect to other cognitive systems (see Caplan & Waters, 1999). That post-interpretive processing relies on domain-general processes is more widely accepted.

questions may be mediated by the correlations between each task and general intelligence. It is also possible that the relationship between spelling skills and more stable lexical representations allows individuals to encode a more accurate representation of the sentence they have just read, leading to better performance on the comprehension questions.

The online reading time measure from the Relative Clause Sentences task was correlated with only one other measure, the relative clause comprehension question difference score from the same task. This correlation is difficult to interpret, as there was no main effect of condition on comprehension question accuracies (although the mean accuracy of each condition trended in the predicted direction; i.e., accuracy was numerically worse for object-extracted RC questions than for subject-extracted RC questions). This difference score for comprehension questions was not correlated with any other task.

The Sentence Ratings task in the current study exhibited correlations within the working-memory sub-experiments that were moderate at best compared to previous findings from the same lab (E. Gibson, N. Twarog, and E. Fedorenko, personal communication). Previous studies had indicated that there were higher correlations among the Sentential Complements / Relative Clauses, Nested vs. Branching Sentential Complements, and Nested vs. Branching Relative Clauses sub-experiments between $r = .219$ and $r = .440$, all ps < .05. The current findings of correlations ranging from $r = .120$ to $r = .346$, with only one correlation reaching significance, were less promising. Apart from the within-task correlations among other sub-experiments from the Sentence Ratings task, ratings scores from this task correlated with a variety of other tasks, including the dual 1- and 2-back measure, several span measures, spatial working memory measures, verbal fluency measures, Stroop, the Go/NoGo accuracy measure, and (in one case) the clefts RT difference score measure. Each of these correlations is considered below.

The SCRC / RCSC ratings difference score was correlated with one measure within the same task, the Branching vs. Nested RCs sub-experiment. This correlation was predicted by the hypothesis that similar working memory resources should be manipulated in both sub-experiments.

In addition to the SCRC / RCSC ratings measure, the Branching vs. Nested RCs sub-experiment was also correlated with the dual 1-back task. It was the only Sentence Ratings sub-experiment to correlate with this measure. The correlation was negative, as should be expected given that a *smaller* difference in naturalness ratings between the easier vs. more difficult RCs might correlate with *higher* accuracy scores on measures tapping related cognitive mechanisms.. If this correlation results from retrieval interference difficulty in both tasks, then the correlation could be interpreted in the following way. The dual 1-back task requires keeping in mind two simultaneously-presented cues from one trial to the next and inhibiting other irrelevant cues that have been processed at other points during the experiment. This requires the participant to ignore items that potentially interfere due to proactive interference buildup (Keppel & Underwood, 1962). However, the 1-back task was the easiest version administered of the dual n-back task, and if this theory were true, the other n-back measures should have correlated highly with the Branching vs. Nested RCs sub-experiment as well. Given the lack of further correlations, it is difficult to interpret the relationship between the 1-back task and the Branching / Nested RCs task.

The Plausibility Violations sub-experiment correlated with many measures from other tasks, namely, the two spatial working memory measures (the "end span" and "high span" measures from the Corsi block task) and most of the span measures: syllable span, operation span, the composite simple span measure, and the composite measure incorporating all of the

span tasks. These correlations were not predicted by any of our hypotheses about working memory and language but seem to indicate that the acceptability of plausibility violations is mediatedß by a general component of working memory, particular given that both spatial working memory and verbal working memory measures predicted performance on the plausibility judgments task. Though this set of findings was not predicted, it is possible that some aspect of working memory, such as the ability to keep separate entities distinct from one another in short-term memory, is related to keeping elements separate in making plausibility judgments. That is, rating two sentences like *The chef is baking the cookies* and *The cookies are baking the chef* more similarly (leading to a lower difference score) could indicate an aptitude for easily imagining the implausible scenario, perhaps indicating a facility for inhibiting the plausible interpretation in the case of the implausible items[5] (reversing the subject and object in the implausible items always led to a plausible interpretation). If such inhibition / interference mechanisms are shared across the span tasks, the correlations between the plausibility judgments and the span tasks would be expected. Further research is required to support this hypothesis.

The agreement violations sub-experiment was also correlated with several measures from other tasks, namely the verbal fluency "A" measure, the clefts difference score, syllable span, the dual 2-back measure, and the composite simple span score. Given the range of tasks the agreement violations ratings correlate with, the interpretation is not clear. It is possible that

---

[5] This interpretation is related to work by Ferreira and colleagues (Ferreira, 2003; Christiansen, Luke, & Ferreira, 2010), finding that when occurring in less common grammatical structures (e.g., as in passive, vs. active, sentences), implausible sentences may be interpreted as plausible variants with a more common grammatical structure. For instance, the sentence *The dog was bitten by the man* might be interpreted as a man biting a dog. Ferreira and colleagues attribute these grammatically incorrect but semantically more plausible interpretations as resulting from "good-enough" representations in language processing. In "good-enough" representations, semantic roles for referents in a sentence may be assigned before syntax is fully parsed, leading to interpretations not dictated by the syntax (cf. Kim & Osterhout, 2005, Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006).

individuals who have less tolerance for simple agreement violations are less able to mentally repair these violations, which may correlate with general intelligence. However, although the Raven's matrices, the best measure of general intelligence included in the current study, is highly correlated with many of these tasks (verbal fluency, the span tasks, and the dual 2-back task), it does not correlate with the agreement violations sub-experiment itself, which is difficult to interpret.

The NP-island violations sub-experiment was correlated with three measures from other tasks, two of which were also correlated with the agreement violations ratings. These were verbal fluency "A", syllable span, and a Stroop measure (mixed – congruent lists). Interestingly, the NP-islands ratings measure was correlated positively with the syllable span measure, meaning that the larger the difference score between ratings for the grammatical control sentences and the NP-island violation sentences, the better participants performed on the syllable span task. This could indicate that participants with a weaker grasp on grammar may not have performed as highly on the syllable span task and were not as adversely affected by the NP-island violations (or perhaps that the violations were not noticed, to an extent). Similarly, there was a positive correlation between the NP-islands ratings and performance on the verbal fluency "A" measure, indicating that participants who were more attuned to the NP-island violations (and therefore exhibited larger difference scores on this task) performed better on the verbal fluency task, providing more acceptable answers. However, the correlation between the NP-islands ratings and the Stroop effect was negative, indicating that a smaller interference effect on Stroop materials (presumably resulting from a stronger ability to inhibit the prepotent response of reading the color word aloud) led to a larger effect on the NP-island ratings, indicating that participants who were more attuned to the NP-island effect also had more difficulty on the Stroop task. This

pattern of results is consistent with an interpretation whereby higher difference scores on the NP-island violations sub-experiment are due to difficulty in processing the grammatical error, perhaps due to variability in more general cognitive mechanisms, given the correlations across three quite distinct cognitive tasks.

Finally, one of the comparisons within the Branching vs. Nested SCs task (nested SCs – branching SCs) exhibited a correlation with one other task, the Go/NoGo task. This was the only Sentence Ratings measure to correlate with Go/NoGo. This correlation was not directly predicted by any of the hypotheses regarding working memory in language but is consistent with the interpretation that the inhibition to a prepotent response required for the Go/NoGo task is involved in inhibiting interference that arises in the more difficult nested SCs. However, since the correlation was positive, individuals who were more sensitive to the difference between nested and branching SCs (and who presumably thought the nested SCs were more difficult) performed better / more accurately on the Go/NoGo task, making this interpretation difficult. The branching SCs with violations / branching SCs difference score was not correlated with measures from any other task.

To summarize, Study 1 attempted to discern which of a variety of measures from tasks tapping cognitive and memory mechanisms would predict performance on sentence processing tasks. In addition, it attempted to estimate the internal reliability of the sentence processing measures, which are known to exhibit a great deal of noise. Although high correlations between sentence processing and two non-linguistic tasks (namely, the 2-back measure from the dual *n*-back task and the spelling measure) were observed, it is difficult to interpret these results fully without consistently higher internal reliability in the sentence processing measures. As a result, Study 2 was conducted to systematically test multiples types of sentence processing materials in

an attempt to find the most internally reliable measures.

III. STUDY 2

The second study focused on testing several sentence processing tasks to discern which would exhibit the best internal reliability, measured by the same split-half reliability metric used for the sentence processing tasks in Study 1. First, the same Cleft Sentences and Relative Clause Sentence tasks from Study 1 were examined using a larger sample size. These tasks were presented using exactly the same materials, presented in supportive contexts. Next, the same critical sentences were presented in null contexts, so that the split-half reliability between the two types of tasks could be compared directly. Finally, some of the materials from the Sentence Ratings task from Study 1 were adapted to a self-paced reading procedure in order to determine whether the working memory manipulations employed in those materials might also exhibit high split-half reliability and be useful in comparisons with nonlinguistic tasks in future studies.

A. Tasks

All of the experiments in Study 2 were run using Amazon.com's Mechanical Turk (AMT) online crowd-sourcing service. AMT provides an easy way to collect data from a large number of participants in a short period of time (Munro, 2010) and has been replicated in a number of previous psycholinguistic studies (Gibson, 2011, personal communication).

EXPERIMENT 1: Clefts and RCs in Supportive Contexts. The materials used in Experiment 1 were identical to those used in the Cleft Sentences and RC Sentences tasks in Study 1 (see example materials (1-2) and Appendices A-B).

EXPERIMENT 2: Clefts and RCs in Null Contexts. The materials used in Experiment 2 were similar to those used in Experiment 1, but presented in null, rather than supportive, contexts. Examples of these sentences are presented below, and a full listing of all of the

experimental materials for Experiment 2 are listed in Appendix F (Clefts in Null Contexts) and G (RCs in Null Contexts). These materials were also presented phrase-by-phrase, and phrase boundaries are indicated below by forward slashes.

(3) Clefts in Null Contexts Materials

c. Object-extracted cleft

It was / the producer / that the novelist mocked / for wearing a tie.

d. Subject-extracted cleft

It was / the producer / that mocked the novelist / for wearing a tie.

(4) RCs in Null Contexts Materials

a. Object-extracted relative clause

The reporter / that the senator attacked / admitted to / making an error.

b. Subject-extracted relative clause

The reporter / that attacked the senator / admitted to / making an error.

Materials from both of these tasks were interleaved in Experiment 2. As in the previous sentence processing tasks in Study 1, the items were randomized by participant according to a Latin-square design such that each participant only saw one version of each item, for a total of 48 items (12 object-extracted clefts, 12 subject-extracted clefts, 12 object-extracted RCs, and 12 subject-extracted RCs). As in all of the other sentence processing experiments, participants completed a comprehension question after reading each sentence phrase-by-phrase.

EXPERIMENT 3: Nested RCs, SCs and RCs, and Controls. Experiment 3 used materials adapted for a self-paced reading paradigm from the Sentence Ratings task from Study 1. Materials from the Nested vs. Branching RCs and SC/RC tasks were the experimental materials, and plausibility violations were included as a control. Eighteen items from each sub-

experiment were included, randomized according to a Latin-square design such that each participant only saw one condition from each item. Examples are given below and the full listing of experimental materials is given in Appendix H.

(4) Nested vs. Branching Relative Clauses Materials

a. Nested RCs

The pilot who / the navigator who / the gunner / warned / alerted / broke radio silence.

b. Branching RCs

The gunner / warned / the navigator who / alerted / the pilot who / broke radio silence.

(5) Sentential Complements and Relative Clauses Materials

a. RC / SC

The intern who / the chance that / the administrator / had lost / the medical reports / bothered / a great deal / supervised / the nurse / on Monday.

b. SC / RC

The chance that / the administrator who / the nurse / supervised / had lost / the medical reports / bothered / the intern / a great deal / on Monday.

(6) Plausibility Violation Materials

a. Plausibility Violation

The chef / is baking / the bread / and the customer / is seating / the waiter.

b. Control

The chef / is baking / the bread / and the waiter / is seating / the customer.

Materials from all three tasks were interleaved in Experiment 3. Items were randomized across participants according to a Latin-square design such that each participant only saw one version of each item, for a total of 54 items. Comprehension questions appeared after each sentence.

EXPERIMENT 4: Nested RCs, Celfts and RCs in Null Contexts, and Controls.

Experiment 4 took the three most promising working memory manipulations from the sentence

processing materials (Nested vs. Branching RCs, Clefts in Null Contexts, and Relative Clauses in

Null Contexts), combined with a control manipulation (Plausibility Violations). One further task

was included as a pilot for another study and is not discussed here. The three working memory

manipulations and control task included 24 items each, and the fifth task also contained an

additional 24 items. However, only a subset of 60 (half) of these items was shown to each

participant so that the task would not be overly lengthy and taxing for the online participants.

The additional 6 items from each of the working memory manipulations and the control

manipulation are given in Appendix I. Comprehension questions appeared after each sentence.

B. Methods

1. *Participants*. All data was collected using a pool of participants from Amazon.com's

Mechanical Turk who were residents of the United States and were native speakers of English.

The number of experiments varied for each experiment: 240 participants (Experiment 1), 85

participants (Experiment 2), 82 participants (Experiment 3), and 80 participants (Experiment 4).

Participants were excluded if they scored less than 65% correct on a combination of all

comprehension questions, leaving 183 participants for Experiment 1, 73 participants for

Experiment 2, 80 participants for Experiment 3, and 80 participants for Experiment 4. Regardless

of whether they were included in experimental analyses, all participants were paid for their time

as long as they completed at least 75% of the questions correctly at a rate of $1.50 for

Experiment 1, $0.60 for Experiment 2, $1.00 for Experiment 3, and $0.60 for Experiment 4.

2. *Materials*. Materials for each of the four experiments are described in the preceding section.

For all of these tasks, participants saw the same items, but saw different versions of some items

such that only one condition of each item was presented to each participant. The full set of materials for each of these experiments is given in Appendices A-B and F-I.

3. *Procedure.* For each of the experiments, all materials were interleaved and presented to participants in a randomized order. The sentence materials were presented to participants using Flash self-paced reading software designed by Hal Tily to be used in Amazon.com Mechanical Turk psycholinguistic experiments. Participants were first given general information about the experiment, at which point they were able to choose whether or not to participate. If they chose to complete the study, they were then directed to the self-paced reading program, where they were given additional instructions about the experiment. They were told that they would be reading a series of sentences, or, in the case of Experiment 1, sets of sentences, with phrase-by-phrase presentation, and that they would then be asked to answer questions about the sentences. Participants were told how many sentences or sets of sentences they would be reading, and as they advanced through the experiment, a progress bar at the top of the screen indicated the relative number of sentences left in the experiment.

C. Results & Discussion by Experiment

1. *Experiment 1: Clefts and RCs in Supportive Contexts.* The primary goal of this experiment was to further test the split-half analyses that had been performed in Study 1 in order to

Region:	Hmm,	I heard that	it was	the producer	that the novelist mocked / that mocked the novelist	for wearing a tie.
Object	479.88 (387.85)	473.71 (319.67)	482.78 (250.00)	1086.69 (1144.79)	2519.31 (2358.33)	1207.49 (1297.13)
Subject	443.05 (255.86)	451.06 (331.77)	475.73 (235.42)	1088.16 (1125.09)	1965.46 (1867.81)	1010.48 (1048.75)

Table 12: Reading times (in ms) by region for the Clefts in Supportive Contexts task from Study 2, Experiment 1.

determine the internal reliability of the Clefts and Relative Clauses in Supportive Contexts tasks. An advantage of re-testing these tasks on Amazon.com's Mechanical Turk was that a much larger participant pool could be tapped, and data from over 200 subjects could be collected relatively quickly. First, descriptive statistics and predicted main effects from these tasks are discussed, followed by a description of the split-half reliability of the two tasks and their relationship to one another.

For the Clefts in Supportive Contexts sub-experiment, reading times were calculated for each region of the final sentence. Reading times that were more than three standard deviations from the mean reading time for each region were discarded and excluded from analysis. The primary region of interest was the embedded clause region. As in Study 1, this region was preceded by two full sentences and the beginning of the "John" utterance, such that it was the sixth region in the third sentence, e.g. "that the novelist mocked" (object cleft) versus "that mocked the novelist" (subject cleft). Reading times along with standard deviation by region are given in Table 12. Of particular interest, the mean reading time for the object-extracted embedded clause regions was 2519.31 ms, while the mean reading time for the subject-extracted embedded clause regions was only 1965.46 ms, a difference of 553.85 ms. Using a linear mixed-effects model with subject and item as random effects and the condition (object or subject cleft), question accuracy, and length of the region as fixed effect predictors, condition was found to be a significant predictor of the critical region reading time, $t(4233) = 13.733$, p < .0001. Question accuracy ($t(4233) = 3.503$, p <.001) and region length ($t(4233) = 3.694$, p < .001) were also significant predictors.

Question accuracies for the experimental conditions followed the same pattern. As

predicted, for object cleft sentences, mean accuracy was worse, at 78.64% (SD = 7.86%), than

for subject cleft sentences, with a mean accuracy of 91.62% (SD = 8.45%), for a difference of

12.98%. A linear mixed-effects model with subject and item as random effects and condition as a

fixed effect indicated that the difference between object and subject cleft sentences was

significant, $t(4233) = 13.19$, $p < .0001$.

The results from the Clefts in Supportive Contexts task thus replicated the results from

the Cleft Sentences task from Study 1 in that object cleft sentences were more difficult to process

and led to lower question accuracies than subject cleft sentences.

For the RCs in Supportive Contexts sub-experiment, reading times were calculated for

regions of the third sentence. Reading times that were more than three standard deviations from

the mean of each region were discarded and excluded from analysis. In particular, the primary

interest was the difference between reading times for the relative clause in the object- and

subject-extracted relative clause conditions. In addition, the difference in question accuracies for

the object- and subject-extracted relative clause conditions was analyzed.

Reading times along with standard deviations are provided in Table 13 for the third

sentence (Mary's utterance). Of particular interest, the mean reading time for the object-extracted

relative clause regions was 2901.78 ms whereas the mean reading time for the subject-extracted

relative clause regions was only 2493.63 ms, a difference of 408.15 ms. Using a linear mixed-

Region	Mary:	I heard that	the reporter	that the senator attacked / that attacked the senator	admitted to	making an error.
Object	657.10 (612.22)	519.30 (312.81)	867.60 (1232.94)	2901.78 (2677.87)	1168.56 (1313.82)	1696.47 (1914.14)
Subject	637.07 (670.76)	518.58 (388.42)	842.32 (1022.99)	2493.63 (2427.12)	1091.74 (1231.36)	1559.04 (1654.74)

Table 13: Reading times (in ms) by region for the RCs in Supportive Contexts sub-experiment from Experiment 1 of Study 2.

effects model with subject and item as random effects and the condition (object- or subject-extracted RC), question accuracy, and region length as fixed effect predictors, condition was a significant predictor of RC region reading time, $t(4255) = 8.159$, p < .0001. The length of the region ($t(4255) = 5.578$, p < .0001) and question accuracy ($t(4255) = 4.381$, p < .0001) were also significant predictors.

Question accuracies did not follow the same pattern for the RCs in Supportive Contexts sub-experiment, with both conditions having similar accuracies. For object-extracted relative clauses, the mean accuracy was 83.01% (SD = 12.61%), which was not statistically different from the mean accuracy for subject-extracted relative clauses, which had a mean accuracy of 82.56% (SD = 13.53%) ($t(4255) = 0.20$, p > .05).

The results from the reading times of the RCs in Supportive Contexts sub-experiment replicated the results from Study 1 in the finding that object-extracted RCs were more difficult to process than subject-extracted RCs. In addition, the prediction that the processing difficulty would also be reflected in the comprehension question accuracies was not supported in this study, a further replication of Study 1. It is not clear why this null effect persisted in both studies, particularly given than many prior studies have found a complexity effect in object- vs. subject-extracted RC reading times and question accuracies (e.g. Gordon et al., 2002, 2004; Fedorenko et al., 2006). One difference between the two RC studies described thus far and the studies cited above is that previous studies presented the critical sentences in null, not supportive, contexts. It is possible that the RC complexity effect is not reflected in the studies described here because the supportive context, including the follow-up sentence, makes the overall complexity of the two conditions similar. In particular, each set of sentences contains both an object- and a subject-

extracted relative clause. Although the critical analysis in the studies focused on the third sentence, where the first relative clause occurs, the second sentence contains a second relative clause. In the case of the easier condition (where the first relative clause is a subject-extracted RC), the following sentence contains a more difficult object-extracted relative clause, meaning that the complexity effect that exhibits itself in the self-paced reading times of the third sentence may no longer lead to an overall complexity effect by the time the comprehension question must be answered (that is, after the fourth sentence, containing the more difficult type of relative clause, has been read). See the results for Experiment 2 in this study for a comparison with comprehension question accuracies in relative clauses appearing in null contexts.

Split-half reliability for both the Clefts and RCs in Supportive Contexts sub-experiments was calculated the same way as in Study 1 for the Cleft Sentences and Relative Clause Sentences sub-experiments. That is, a bootstrapping method was employed to randomly split each dataset in half such that half of the object- and half of the subject-extracted items were extracted for each participant, difference scores were calculated, and these difference scores were correlated with the difference calculated from the remaining items for each participant. This procedure was repeated many items (here, the number of iterations was 100) so that a true split-half correlation could be estimated by taking the mean of the r values obtained from each iteration.

For the Clefts in Supportive Contexts sub-experiment, the split-half reliability was much higher than in Study 1, estimated to be $r = .276$. Split-half reliability was also moderately higher for the RCs in Supportive Contexts sub-experiment, estimated to be $r = .140$. Encouragingly, the correlation between the two sub-experiments was also much higher in Study 2 than in Study 1, with the correlation between the Clefts and RCs sub-experiments reaching $r = .167$, $p < .05$.

The split-half correlations for each sub-experiment in Experiment 1 of Study 2 were

therefore higher than those observed in Study 1, which is an unsurprising finding given that the number of participants was nearly quadrupled. Most promisingly, a correlation between the two sub-experiments was observed. Given that each sub-experiment (notably the Clefts sub-experiment) has reasonably high within-experiment split-half reliability, this correlation can be taken as meaningful. Experiment 2 sought to uncover whether such high reliability and between-sub-experiment correlations would be observed in syntactic manipulations occurring in null, not supportive, contexts.

2. *Experiment 2: Clefts and RCs in Null Contexts*. The primary objective of Experiment 2 was to test whether internal reliability (and inter-sub-experiment correlations) in the object/subject cleft and RC manipulations would be higher when presented in null contexts. First, descriptive statistics and main effects from the experiment are presented, followed by a discussion of the split-half reliability and correlation between the two tasks.

For the Clefts in Null Contexts sub-experiment, reading times were calculated for each region. Reading times that were more than three standard deviations from the mean of each region were discarded and excluded from analysis. In particular, the primary interest was in the difference between reading times for the embedded clause (the third region) in the object and

Region:	It was	the producer	that the novelist mocked / that mocked the novelist	for wearing a tie.
Object	884.12 (952.47)	1171.68 (1082.01)	2977.47 (2599.85)	2315.43 (2161.09)
Subject	896.62 (1016.40)	1185.57 (1125.00)	2519.44 (2204.09)	2143.98 (1962.78)

Table 14: Reading times (in ms) by region for the Clefts in Null Contexts task from Study 2, Experiment 2.

subject cleft conditions. In addition, the difference score between the question accuracies for the object and subject cleft sentences was analyzed.

Reading times along with standard deviations by region are given in Table 14 for the entire sentence. Of particular interest, the mean reading time for the object-extracted embedded clause regions was 2977.47 ms, while the mean reading time for subject-extracted embedded clause regions was only 2519.44, a difference of 458.03 ms. A linear mixed-effects model with subject and item as random effects and the condition (object or subject cleft), question accuracy, and length of the region as fixed effect predictors revealed that condition was a significant predictor of the critical region reading time, $t(1677) = 6.829$, p < .0001. Region length was also a significant predictor, $t(1677) = 5.578$, p < .0001.

Question accuracies followed the same pattern. For object cleft sentences, mean accuracy was only 88.13% (SD = 11.28%) whereas accuracy was 92.24% (SD = 9.45%) for subject cleft sentences, for a difference of 4.11%. A linear mixed-effects model with subject and item as random effects and condition as a fixed effect indicated that the difference between object and subject cleft sentences was significant, $t(1677) = 2.89$, p < .005).

The results from the Clefts in Null Contexts sub-experiment were quite similar to the previous studies presenting materials in supportive contexts. Object clefts both were more

Region	The reporter	that the senator attacked / that attacked the senator	admitted to	making an error.
Object	1237.42 (1400.23)	3096.39 (2753.50)	1980.99 (2444.14)	1647.71 (1632.19)
Subject	1186.08 (1219.56)	2635.91 (2299.44)	1777.55 (1683.71)	1555.59 (1579.02)

Table 15: Reading times (in ms) by region for the RCs in Null Contexts sub-experiment from Experiment 2 of Study 2.

difficult to process and led to more difficulty in comprehension questions compared to subject clefts.

For the RCs in Null Contexts sub-experiment, reading times were calculated for each region. Reading times that were more than three standard deviations from the mean of each region were discarded and excluded from analysis. In particular, the primary interest was in the difference between reading times for the relative clause (the second region) in the object- and subject- extracted relative clause conditions. In addition, the difference score between the question accuracies for the object and subject cleft sentences was analyzed.

Reading times along with standard deviations are given in Table 15 for the entire sentence. Of particular interest, the mean reading time for the object-extracted relative clause regions was 3096.39 ms, while the mean reading time for the subject-extracted relative clause regions was only 2635.91 ms, a difference of 460.48 ms. A linear mixed-effects model with subject and item as random effects and condition (object or subject RC), question accuracy, and length of the region as fixed effects revealed that condition was a significant predictor, $t(1704) =$ 7.077, p < .0001. Length of the region in characters was also a significant predictor, $t(1704) =$ 7.422, p < .0001, as was question accuracy, $t(1704) = 2.944$, p < .005.

Question accuracies followed the same pattern. For object-extracted relative clause sentences, mean accuracy was only 90.18% (SD = 8.93%) whereas accuracy was 95.32% (7.86%) for subject-extracted relative clause sentences. A linear mixed-effects model with subject and item as random effects and condition as a fixed effect indicated that the difference between object and subject relative clause sentences was significant, $t(1704) = 4.40$, p < .0001).

The finding that object relative clause sentences were more difficult to process than subject-extracted relative clause sentences replicated findings from Study 1 and Study 2,

Experiment 1 with relative clause sentences in supportive contexts. However, in opposition to the earlier studies, Study 2 from Experiment 2 indicated that object relative clause sentences led to more difficulty in comprehension questions than subject relative clause sentences, as well, when the relative clauses were presented in null contexts. Given that the overall complexity of both conditions was greatly reduced from the earlier studies, this is not a surprising finding (see the discussion of the relative clause sub-experiment from Study 2, Experiment 1).

Split-half reliability for both the Clefts in Null Context and RCs in Null Context sub-experiments was calculated the same way as in Experiment 1 for the Clefts and RCs in Supportive Contexts sub-experiments. The same bootstrapping method was employed to randomly split each dataset in half such that half of the items from each condition were extracted for each participant. Then, difference scores were calculated and correlated with the difference scores calculated from the remaining half of the items for each participant. This procedure was repeated many times (here, the number of iterations was 100) so that a true split-half correlation could be estimated by taking the mean of the r values.

For the Clefts in Null Contexts sub-experiment, the split-half reliability was slightly higher than in Experiment 1, $r = .306$. Split-half reliability was also higher than in Experiment 1 for the RCs in Null Contexts sub-experiment, $r = .404$. The correlation between the two sub-experiments was also moderately higher than in Experiment 1, $r = .256$, $p < .05$. This finding is particularly encouraging given that the split-half reliability for each sub-experiment was relatively high.

The split-half correlations in Experiment 2 were therefore higher than the split-half correlations in both Study 2, Experiment 1 and Study 1 (in the Cleft Sentences and Relative Clause Sentences tasks). Of particular note, even with a smaller sample size ($n = 73$ in

Experiment 2, compared to $n = 183$ in Experiment 1), the split-half reliability was higher when cleft and relative clause sentences were presented in null, rather than supportive, contexts. One plausible explanation for this finding is that when such materials are presented in supportive contexts, item variability may be greater due to difficulty in integrating different referents into the context. For instance, it may be easier to build up a coherent representation for more common or familiar words whereas less familiar words may not be as easy to maintain in memory. If this is the cause, then it is plausible that less familiar words cause more difficulty in more difficult syntactic structures (i.e., syntactic structures such as object-extracted relative clauses and clefts that consume more working memory resources) in supportive contexts when a boost from context is not as useful (cf. work by Hofmeister (2010) showing that more descriptive, detailed noun phrases reduce sentence processing complexity effects). Furthermore, it is possible that there is greater variability across individual subjects in the case of sentences presented in supportive contexts, as individuals may be able to use the context to greater or lesser degrees. Future studies could investigate these hypotheses directly by predicting performance of individual participants based on knowledge of the specific words used in individual items and by looking at individual differences in syntactic complexity effects in materials presented in null vs. supportive contexts. The account described herein would predict that more individual variability would exist in materials presented in supportive contexts compared to null contexts. However, as the current study aimed to establish the most reliable sentence-processing measures (where the variability occurs across, not within, subjects), the final two experiments of Study 2 focus on sentence materials presented in null contexts. Experiment 3 tests additional working memory manipulations in sentences in null contexts, and Experiments 4 directly compares the sub-experiments from Experiment 2 with the most promising tasks from Experiment 3.

3. *Experiment 3: Nested RCs, SCs and RCs, and Controls*. This experiment focused on testing

additional working memory manipulations in sentences adapted from the Sentence Ratings task

from Study 1. The sub-experiments included were the nested vs. branching RC manipulation and

the sentential complement / relative clause order manipulation. In addition, the plausibility

violation sub-experiment was included as a control. First, descriptive statistics of reading times

for each of the three sub-experiments are discussed, with special emphasis on regions of interest.

Then, split-half reliabilities of the reading times for the critical regions are provided, followed by

a discussion of the relationship of the sub-experiments to one another.

For the Nested vs. Branching Relative Clauses sub-experiment, reading times were

calculated for each region. Reading times that were more than three standard deviations from the

mean reading time of each region were discarded and excluded from analysis. As this was the

first attempt at analyzing phrase-by-phrase reading times in nested and branching sentences

containing multiple relative clauses, two potential critical regions were tested. The difficulty in

processing nested sentences is thought to be due to difficulty in integrating arguments of the verb

at the point of processing each of the verbs. In particular, the primary site of this difficulty

should be observed at the final verb, when all previous referents need to be integrated into the

Region	The pilot who / The gunner	the navigator who / warned	the gunner / the navigator who	warned / alerted	alerted / the pilot who	broke radio silence.
Nested	1595.81 (1827.98)	2221.20 (2350.52)	1644.41 (1781.69)	1769.31 (1962.72)	1340.04 (1450.66)	2251.47 (2430.54)
Branching	1391.35 (1679.80)	1036.45 (1067.40)	1808.85 (1953.43)	1254.25 (1247.33)	1777.78 (2092.00)	1818.20 (2089.11)

Table 16: Reading times (in ms) by region for the Branching vs. Nested RCs sub-experiment
from Experiment 3 of Study 2.

appropriate argument positions in order for the sentence to be correctly understood. Therefore,

both the final region (including the final verb as well as an object of the verb, and occurring in region six in both conditions) as well as a combination of all three verb regions (regions four to six in the nested relative clause conditions, and regions two, four, and six in the branching relative clauses) were tested. Difference scores for both measures were computed.

Reading times along with standard deviations by region for the Nested vs. Branching RCs sub-experiment are given in Table 16. For the nested (more difficult) sentences, the total reading times of the three verb regions averaged 5360.81 ms whereas the total reading times of the three verb regions averaged only 4108.90 ms for the branching (easier) sentences, a difference of 1251.91 ms. A linear mixed-effects model with condition (nested or branching), region length, and question accuracy as fixed effects and subject and item as random effects revealed that this difference was significant, $t(3767) = 9.213$, p < .0001.

For the nested sentences, reading time for the final region (i.e., the final verb) was 2251.47 ms, compared to only 1818.20 ms for the branching sentences, a difference of 433.27 ms. This difference was also analyzed with a linear mixed-effects model with condition, region length, and question accuracy as fixed effects and subject and item as random effects, revealing that the difference was significant, $t(1247) = 8.598$, p < .0001. In this analysis, question accuracy was also a significant predictor of reading time, $t(1247) = 2.263$, p < .05.

Question accuracies followed the same pattern, with performance on questions following nested RC sentences leading to lower accuracies than on questions following branching RC sentences. Mean accuracy on nested RC sentences was 67.02% (SD = 47.02%) compared to 85.51% (SD = 35.20%) on branching RC sentences, a difference of 18.49%. A linear mixed-effects model with condition as a fixed effect and subject and item as random effects revealed that this difference was significant, $t(1247) = 5.792$, p < .0001.

For the Sentential Complements / Relative Clauses sub-experiment, reading times were calculated for each region. Reading times that were more than three standard deviations from the mean reading time of each region were discarded and excluded from analysis. As this was the first attempt at analyzing self-paced reading data from these materials, two different potential critical regions were tested. First, all verbal regions as well as the adverbial modifying the main verb of the sentential complement were analyzed (regions four, six, seven, and eight for the RC/SC (more difficult) sentences, and regions four, five, seven, and nine for the SC/RC (easier) sentences). Second, the region containing the verb taking the head of the SC as its subject, plus the adverbial following it, were analyzed (regions six and seven for the RC/SC sentences, and regions seven and nine for the SC/RC sentences). This verb is where the bulk of the processing difficulty should arise, as the head of the relative clause must also be understood as the object of this verb after processing multiple complicated syntactic embeddings (the SC and the RC). Difference scores for both measures were computed.

Reading times along with standard deviations by region for the Sentential Complements / Relative Clauses sub-experiment are given in Table 17. For Relative Clause / Sentential

Region	The intern who / The chance that	the chance that / the administrator who	the administrator / the nurse	had lost / supervised	the medical reports / had lost
RC/SC	1770.30 (2105.58)	2126.76 (2322.82)	1573.80 (1776.95)	1577.98 (1933.65)	1402.27 (1861.97)
SC/RC	1869.23 (2285.07)	1895.86 (2186.01)	1540.25 (1640.52)	1568.49 (1633.12)	1519.12 (1758.72)

	bothered / the medical reports	a great deal / bothered	supervised / the intern	the nurse / a great deal	on Monday.
RC/SC	1337.88 (1666.47)	1271.58 (1662.12)	1217.95 (1459.205)	1224.83 (1502.30)	1371.78 (1773.36)
SC/RC	1337.95 (1606.69)	1253.38 (1438.36)	1405.45 (1863.15)	1194.73 (1722.92)	1246.66 (1763.04)

Table 17: Reading times (in ms) by region for the Sentential Complement / Relative Clause sub-experiment from Experiment 3 of Study 2.

Complement (more difficult) sentences, the total reading times of the three verb regions + the final adverbial averaged 5405.39 ms, and the total reading times of the three verb regions + the adverbial modifying the SC verb averaged 5535.72 ms for the Sentential Complement / Relative Clause (easier) sentences, a difference of 130.33 ms. This difference was not in the direction predicted by the hypothesis that SC/RC sentences should be easier to process than RC/SC sentences. A linear mixed-effects model with condition (RC/SC or SC/RC), region length, and question accuracy as fixed effects and subject and item as random effects revealed that this difference was not significant, $t(5039) = 0.322$, p >.05.

The second analysis of only the verb taking the head of the SC as its subject and the following adverbial yielded a different result. For the RC/SC sentences, the mean reading time for the SC verb + adverbial was 2609.46 ms, and the mean reading time was 2448.110 ms for the SC/RC regions. This difference was in the predicted direction, and the difference nearly reached significance, $t(2541) = 1.940$, p < .06.

As for the comprehension questions, mean accuracy was 78.55% (SD = 41.05%) for RC/SC (more difficult) sentences and 85.52% (SD = 35.19%) for SC/RC (easier) sentences, a difference of 6.97%. This difference was statistically significant by a linear mixed-effects model with condition as a fixed effect and subject and item as random effects, $t(1280) = .3.741$, p < . 001.

For the Plausibility Violations sub-experiment, reading times were calculated for each region. Reading times that were more than three standard deviations from the mean reading time of each region were discarded and excluded from further analysis. As this was the first attempt at

analyzing phrase-by-phrase reading times in plausibility violation materials, two critical regions were analyzed. The first of these included only the final region (region six) containing the object of the verb of the second clause. The second of these included the last two regions (regions five and six), including the verb and object of the second clause. Both of these measures were taken because it is likely that the plausibility violation might be perceptible by the main verb of the implausible clause, in the plausibility violations condition; however, it is also possible that the violation might not be fully processed until the object of the verb is realized, and so the last two regions as a whole were also analyzed.

Reading times along with standard deviations by region for the Plausibility Violations sub-experiment are given in Table 18. For the sentences with plausibility violations, the mean reading time of the last region alone (the object of the second verb) was 1215.31 ms whereas the mean reading time for the last region in the control condition was only 994.40 ms, a difference of 219.91 ms. A linear mixed-model with the condition (plausibility violation or control), length of the region, and question accuracy as fixed effects and subject and item as random effects

Region	The chef	is baking	the bread	and the waiter / and the customer	is seating	the waiter.
Violation	1168.35 (1536.82)	1024.20 (1122.35)	987.06 (1058.22)	1131.01 (1197.95)	1099.60 (1292.05)	1214.31 (1324.05)
Control	1168.57 (1420.87)	1054.27 (1157.28)	976.16 (909.53)	1176.89 (1261.99)	949.29 (1109.86)	994.40 (1027.23)

Table 18: Reading times (in ms) by region for the Plausibility Violations sub-experiment from Experiment 3 of Study 2.

revealed that this difference was significant, $t(1263) = 2.787$, p < .01. In this analysis, region length was also a significant predictor, $t(1263) = 4.573$, p < .0001.

In the second analysis, the total reading time of regions five and six (the verb and object of the second clause) was 2313.91 ms for the violation condition and only 1943.69 ms for the control condition, a difference of 370.22 ms. A linear mixed-model with condition, length of the region, and question accuracy as fixed effects and subject and item as random effects revealed that this difference was also significant, $t(2494) = 4.506$, p < .0001. Region length was also a significant predictor in this analysis, $t(2494) = 3.689$, p < .001.

Question accuracies followed the same numeric pattern, with performance on questions following plausibility violation sentences leading to lower accuracies than on questions following control sentences. Mean accuracy on plausibility violation sentences was 97.06% (SD = 16.90%) compared to 98.12% (SD = 13.59%) on control sentences, a difference of 1.06%. However, a linear mixed-effects model with condition as a fixed effect and subject and item as random effects revealed that this difference was not significant, $t(1238) = .000$, p > .05. The lack of a difference in question accuracy on the plausibility violation materials is likely due to the fact that the materials are much less syntactically complex than all of the other materials in the sentence processing sub-experiments, and it is likely not much more taxing to keep track of a short (five-word) clause containing a verb with an argument structure that is implausible.

Split-half reliability for each dependent measure for each sub-experiment was calculated the same way as in the previous clefts and RCs sub-experiments. The same bootstrapping method was employed to randomly split the relevant measures from each dataset in half such that half of the items from each condition were extracted for each participant. Then, difference scores were calculated and correlated with the difference scores calculated from the remaining half of the items for each participant. This procedure was repeated many times (here, the number of iterations was 100) so that a true split-half correlation could be estimated by taking the mean of

the *r* values. The six estimates of split-half reliability (two for each sub-experiment) are described below, followed by a summary of the correlations between the sub-experiments (as estimated by each of the two dependent measures from each sub-experiment).

For the Nested vs. Branching RCs sub-experiment, the split-half reliability for the final verb region was estimated to be $r = .222$. The split-half reliability for all three verb regions was much higher, estimated to be $r = .459$. This was the highest split-half reliability observed in any of the sub-experiments described herein. For the Sentential Complements / Relative Clauses sub-experiment, the split-half reliability for the SC head verb and adverbial was estimated to be $r = .$ 132. The split-half reliability for all three verb regions plus the adverbial was slightly higher, estimated to be $r = .216$. Split-half reliability was overall quite low for the plausibility violations. For the final region alone, the split-half reliability was estimated to be $r = .052$, and for the final two regions, the split-half reliability was estimated to be $r = .040$. This incredibly low split-half reliability may reflect the fact that the plausibility items hinged on strikingly obvious implausible events, but that these events may have varied in the magnitude of the implausibility. In future

Region	SC/RC: SC Head Vb. + Adv.	SC/RC: All Verbs + Adv.	Plaus.: Final Region	Plaus.: Final Two Regions
Nested/Branch. RCs: Final Verb	*.425*	*.377*	*.392*	*.462*
Nested/Branch. RCs: All Verbs	.285	.120	.152	*.203*
SC/RC: SC Head Vb. + Adv.	---	---	-.148	-.154
SC/RC: All Verbs + Adv.	---	---	-.123	-.177

Table 19: Correlations between measures from the three sub-experiments from Study 2, Experiment 3. (*p < .001*, *p < .01*, p < .05, *p < .10*)

studies, it may be beneficial to norm such items for plausibility of the semantic fit between the verb and its arguments in the implausible versions of the items.

Correlations between the sub-experiments are presented in Table 19. Of particular note, measures between the two working memory tasks were mostly highly correlated, with the highest correlation of these being between the SC head verb and adverbial measure from the SC/RC task and the three-verb measure from the Nested/Branching RCs task, $r = .425$, p < .001. However, measures between the plausibility violations and the Nested/Branching RCs sub-experiment also reached significance, with the correlation between the final region in the Plausibility Violations sub-experiment and the final verb region from the Nested/Branching RCs sub-experiment reaching $r = .462$, p < .001. No measures from the Plausibility Violations and SC/RC sub-experiments were significantly correlated, all ps > .05. It should be taken under consideration that the split-half correlations were very high for the Nested/Branching RCs sub-experiment, middling for the SC/RC sub-experiment, and very low for the Plausibility Violations sub-experiment. As a result, correlations between the Plausibility Violations sub-experiment, in particular, with the other sub-experiments are difficult to interpret. However, the observation that the highest correlations overall were observed between the two working memory manipulations in the SC/RC and Nested/Branching RCs sub-experiments is encouraging and in line with predictions. In addition, the very high split-half reliability observed in the Nested/Branching RCs task makes it an ideal task to use in individual differences studies investigating working memory in sentence processing. Therefore, this sub-experiment was included, along with the plausibility violations condition as a control, in Experiment 4, which directly compared these two sub-experiments with the Clefts and RCs sub-experiments from Experiment 2.

4. *Experiment 4: Nested RCs, Clefts and RCs in Null Contexts, and Controls.* The primary objective of Experiment 4 was to directly test the correlation between the Nested vs. Branching RCs sub-experiment from Experiment 3 and the Clefts and RCs in Null Contexts sub-

experiments from Experiment 2. First, descriptive statistics and main effects from the experiment are presented, followed by a discussion of the correlations between the sub-experiments.

For the Clefts in Null Contexts sub-experiment, reading times were calculated for each region. Reading times that were more than three standard deviations from the mean of each region were discarded and excluded from further analysis. In particular, the primary interest was in the difference between reading times for the embedded clause (the third region) in the object and subject cleft conditions. In addition, the difference score between the question accuracies for the object and subject cleft sentences was analyzed.

Reading times along with standard deviations by region are given in Table 20 for the entire sentence. Of particular interest, the mean reading time for the object-extracted embedded clause regions was 2396.23 ms, while the mean reading time for subject-extracted embedded clause regions was only 1852.08 ms, a difference of 544.15 ms. A linear mixed-effects model with subject and item as random effects and the condition (object or subject cleft), question

Region:	It was	the producer	that the novelist mocked / that mocked the novelist	for wearing a tie.
Object	690.14 (579.83)	889.52 (762.20)	2396.23 (2042.09)	1967.93 (1514.36)
Subject	616.60 (408.63)	872.33 (1058.85)	1852.08 (1524.04)	1652.09 (1195.31)

Table 20: Reading times (in ms) by region for the Clefts in Null Contexts task from Study 2, Experiment 4.

accuracy, and length of the region as fixed effect predictors and subject and item as random effects revealed that condition was a significant predictor of the critical region reading time, $t(869) = 6.034$, p < .0001. Region length in characters was also a significant predictor, $t(869) =$

4.585, p < .0001.

Question accuracies followed the same pattern. For object cleft sentences, mean accuracy was only 77.85% (SD = 21.50%) whereas accuracy was 85.96% (SD = 15.17%) for subject cleft sentences, a difference of 8.11%. A linear mixed-effects model with subject and item as random effects and condition as a fixed effect indicated that the difference between object and subject cleft sentences was significant, $t(869) = 3.01$, p < .005.

The results from the Clefts in Null Contexts sub-experiment replicated the basic findings from Experiment 2. Object clefts both were more difficult to process and led to more difficulty in comprehension questions compared to subject clefts.

For the RCs in Null Contexts sub-experiment, reading times were calculated for each region. Reading times that were more than three standard deviations from the mean of each region were discarded and excluded from further analysis. In particular, the primary interest was in the difference between reading times for the relative clause (the second region) in the object

Region	The reporter	that the senator attacked / that attacked the senator	admitted to	making an error.
Object	889.02 (912.14)	2361.55 (1860.54)	1355.28 (1085.40)	1375.43 (1355.91)
Subject	801.40 (668.59)	1975.05 (1482.79)	1280.81 (980.95)	1297.51 (1113.49)

Table 21: Reading times (in ms) by region for the RCs in Null Contexts sub-experiment from Experiment 4 of Study 2.

and subject relative clause conditions. In addition, the difference score between the question accuracies for the object and subject relative clause sentences was analyzed.

Reading times along with standard deviations by region are given in Table 21 for the

entire sentence. Of particular interest, the mean reading time for the object-extracted relative clause regions was 1375.43 ms, while the mean reading time for subject-extracted relative clause regions was only 1297.51 ms, a difference of 77.92 ms. A linear mixed-effects model with subject and item as random effects and the condition (object or subject relative clause), question accuracy, and length of the region as fixed effect predictors and subject and item as random effects revealed that condition was a significant predictor of the critical region reading time, $t(876) = 4.872$, $p < .0001$. Region length in characters was also a significant predictor, $t(876) = 6.105$, $p < .0001$.

Question accuracies did not follow the same pattern. For object relative clause sentences, mean accuracy was 86.84% (SD = 87.94%) and accuracy was 87.94% (SD = 17.13%) for subject relative clause sentences, a difference of 1.1%. A linear mixed-effects model with subject and item as random effects and condition as a fixed effect indicated that the difference between object and subject relative clause sentences was not significant, $t(876) = .450$, $p > .05$.

Region	The pilot who / The gunner	the navigator who / warned	the gunner / the navigator who	warned / alerted	alerted / the pilot who	broke radio silence.
Nested	1182.78 (1182.58)	2175.47 (2238.22)	1598.12 (1385.88)	1681.41 (1742.01)	1211.70 (1353.10)	2175.49 (2535.64)
Branching	988.85 (958.47)	967.85 (1131.78)	1607.64 (1860.83)	1034.28 (1058.87)	1371.64 (1400.20)	1617.14 (1619.81)

Table 22: Reading times (in ms) by region for the Branching vs. Nested RCs sub-experiment from Experiment 4 of Study 2.

The RCs in Null Contexts sub-experiment replicated the primary finding from Experiment 2, with object-extracted relative clauses being more difficult to process than subject-extracted relative clauses. However, the finding that this difficulty was also reflected in accuracy

on the comprehension questions was not replicated. Given that the effect was small in Experiment 2, it is perhaps not surprising that it was not replicated in the second study. Crucially, the large extraction effect in the self-paced reading portion of the experiment exists in each experiment.

For the Nested vs. Branching RCs sub-experiment, reading times were calculated for each region. Reading times that were more than three standard deviations from the mean of each region were discarded and excluded from further analysis. In particular, the primary interest was in the difference between reading times the three verb regions in each condition (regions four to six in the nested relative clause conditions, and regions two, four, and six in the branching relative clauses). In addition, the difference score between the question accuracies for the branching and nested RCs was analyzed.

Reading times along with standard deviations by region are given in Table 22 for the entire sentence. Of particular interest, the mean reading time for the nested RC verb regions was 5068.60 ms, while the mean reading time for the branching RC verb regions was only 3619.27 ms, a difference of 1449.33 ms. A linear mixed-effects model with subject and item as random effects and the condition (nested or branching RCs), question accuracy, and length of the region as fixed effect predictors and subject and item as random effects revealed that condition was a significant predictor of the critical region reading time, $t(2661) = 9.771$, $p < .0001$. Region length in characters was also a significant predictor, $t(2661) = 13.515$, $p < .0001$, as was question accuracy, $t(2661) = 3.209$, $p < .01$.

Question accuracies followed the same pattern. For nested RC sentences, mean accuracy was 71.49% (SD = 19.78%) and accuracy was 86.40% (SD = 15.57%) for branching RC sentences, a difference of 14.91%. A linear mixed-effects model with subject and item as random

effects and condition as a fixed effect indicated that the difference between nested and branching RC sentences was significant, $t(893) = 6.193$, $p < .0001$.

The Nested vs. Branching RCs sub-experiment therefore replicated findings from Experiment 2. Nested sentences were more difficult as revealed both by sentence reading times for the verb regions of the sentences and also by question accuracies.

For the Plausibility Violations sub-experiment, reading times were calculated for each region. Reading times that were more than three standard deviations from the mean of each region were discarded and excluded from further analysis. In particular, the primary interest was in the difference between reading times for the last two regions in each condition (the verb and object of the second clause, where the plausibility violations occurred). In addition, the difference score between the question accuracies for the plausibility violations and control sentences was analyzed.

Reading times along with standard deviations by region are given in Table 23 for the entire sentence. Of particular interest, the mean reading time for the plausibility violations final

Region	The chef	is baking	the bread	and the waiter / and the customer	is seating	the waiter.
Violation	778.79 (605.66)	841.61 (896.17)	800.91 (584.61)	968.34 (761.27)	787.86 (662.51)	1130.45 (1068.47)
Control	735.52 (472.52)	810.27 (754.79)	842.00 (867.51)	929.44 (754.50)	728.73 (794.64)	895.73 (965.08)

Table 23: Reading times (in ms) by region for the Plausibility Violations sub-experiment from Experiment 4 of Study 2.

regions was 1918.31 ms, while the mean reading time for the control final regions was only 1624.46 ms, a difference of 293.85 ms. A linear mixed-effects model with subject and item as random effects and the condition (plausibility violations or controls), question accuracy, and

length of the region as fixed effect predictors and subject and item as random effects revealed that condition was a significant predictor of the critical region reading time, $t(1760) = 3.863$, p < .001. Region length was also a significant predictor, $t(1760) = 2.495$, p < .05.

Question accuracies numerically followed the same pattern. For plausibility violations, mean accuracy was 95.61% (SD = 8.76%) and accuracy was 97.37% (SD = 7.23%) for control sentences, a difference of 1.76%. A linear mixed-effects model with subject and item as random effects and condition as a fixed effect indicated that the difference between plausibility violation and control sentences did not reach significance, $t(892) = 1.69$, p > .05.

The plausibility violations sub-experiment therefore replicated the results from Experiment 3. Implausible items were more difficult to process as estimated by reading times but did not lead to serious impairment on comprehension question accuracies.

Finally, the relationship between the three experiments was analyzed, summarized in Table 24. None of the correlations reached significance, though some of the correlations nearly approached significance. Of particular interest were the correlations between the three working

Region	RCs Diff.	Nest/Branch RC Diff.	Plaus. Viol. Diff.
Clefts Diff.	.164	.072	.152
RCs Diff.	---	.159	-.126
Nest/Branch RC Diff.	---	---	-.051

Table 24: Correlations between measures from the four sub-experiments from Study 2, Experiment 4. ($p < .001$, $p < .01$, p < .05, $p < .10$)

memory manipulations (the clefts, RCs, and nested/branching RCs sub-experiments). The finding from Experiment 2 that the Clefts and RCs in Null Contexts were correlated was not replicated, though this correlation was close to approaching significance, $r = .164$, p < .16. The correlation between the RCs in Null Contexts and Nested vs. Branching RCs sub-experiments was similar in magnitude, $r = .159$, p < .18. Interestingly, the Clefts in Null Contexts sub-

experiment was more highly correlated with the Plausibility Violations sub-experiment than with the Nested vs. Branching RCs sub-experiment, $r = .152$, $p < .20$. No other sub-experiments were correlated, all $ps > .25$.

Unfortunately, Experiment 4 did not replicate the promising finding from Experiment 2 that the Clefts and RCs in Null Contexts tasks were correlated. However, it is possible that the lower correlations observed in this experiment are in part due to the fact that participants received on average only six items from each condition. In a future study, it would be useful to replicate Experiment 4 using all of the materials such that each participant sees twelve items from each condition in order to better estimate performance on each task.

D. General Discussion

The second study aimed to test a number of sentence processing tasks with the goal of finding tasks with the best split-half reliability for use in future studies involving individuals differences in sentence processing. To this end, the study was successful in finding that working memory manipulations, such as object versus subject cleft and relative clause constructions, were more internally reliable when presented in isolated, or null, contexts, rather than in the type of supportive contexts that were used in Study 1. In addition, a related goal of the study was to find additional sentence processing materials with working memory manipulations that also exhibited high internal reliability. To do so, an additional set of materials was tested and a third task with multiple nested versus branching relative clauses also exhibited high internal reliability.

A second goal of the study was to establish that multiple sets of materials with different working memory manipulations were highly correlated with one another. To this end, a high correlation was found between cleft and relative clause sentence manipulations presented in null contexts, and more marginal correlations were observed between these tasks and a third task, the

nested vs. branching relative clauses task.

In the future, these three tasks should be used in conjunction with a variety of other memory and cognitive tasks, informed by the results from the battery of tasks used in Study 1.

IV. Conclusions and Future Directions

This thesis investigated two questions related to individual differences in sentence processing. The first of these, addressed in Study 1, was directed at a large theoretical question, namely, what cognitive and memory mechanisms are shared between sentence processing and more general cognitive and memory processes? A primary focus of this question was investigating to what extent retrieval interference mechanisms are shared between general memory function and working memory in sentence processing. The second question, addressed in Study 2, was more methodological in nature and focused on determining which types of sentence processing tasks would lead to high internal reliability and might therefore be prime candidates for investigating individual differences in sentence processing in future studies.

In Study 1, performance on a large battery of memory and cognitive tasks was compared to performance on a battery of language tasks, including two online sentence-processing tasks and a task primarily involving offline measures, the Sentence Ratings task. Unfortunately, internal reliability as measured by a split-half reliability metric was low on both of the sentence-processing tasks, making it difficult to conclusively interpret correlations with other tasks as meaningful. As such, this study should be taken as a first pass towards investigating individual differences in these sentence-processing tasks and other cognitive and memory measures. However, given that it may be difficult to estimate such reliability, the correlations which did occur between the sentence-processing tasks (the Cleft Sentences task, in particular) and other cognitive and memory tasks are still considered in detail in this thesis and provide a starting

point for future individual differences studies.

In particular, the correlations between the Cleft Sentences reading time measure and two of the memory and cognitive function tasks, namely the 2-back measure from the Dual *n*-back task and the spelling test measure provide new insights into the shared mechanisms between sentence processing and other cognitive functions. This is the first study to show that a direct, online measure of working memory difficulty in sentence processing correlates with either of these tasks.

Dual *n*-back tasks are thought to be a good measure of general working memory ability due to the fact that the nature of dual-task paradigms may prevent the use of rehearsal strategies, leading to a purer measure of working memory than other types of tasks; furthermore, the dual *n*-back task employed in Study 1 has also been shown correlate highly with and improve performance on tasks measuring general fluid intelligence (Oberauer, Lange, & Engle, 2004; Jaeggi et al., 2008, 2010). In general, any *n*-back task requires participants to continuously update items that are active in memory and to keep track of the temporal order of each element so that it may be correctly understood as being *n* items back in the stream of cues when it matches a current target. In addition, the dual *n*-back task used in the Study 1 required participants to keep track of two such streams at once, taxing the system and likely preventing participants from relying solely on rehearsal strategies. While phonological rehearsal strategies may be useful in accessing sentence representations that are already built up (i.e., what Caplan and Waters (1999) term *post-interpretive* processing), many researchers agree that phonological rehearsal strategies are not as useful in online sentence processing (i.e., *interpretive* processing) (Gathercole & Baddeley, 1994). Therefore, it is possible that the use of the dual *n*-back task in Study 1 was able to tap what Baddeley-style theories would term the central executive, involved

in the coordination of information within working memory. It is therefore possible that the mechanisms involved in general memory retrieval and coordination, tapped by the dual *n*-back task, are (at least partially) the same as the mechanisms used in sentence processing.

The correlation between the Cleft Sentences reading time and the spelling task from Study 1 suggests that the specificity and well-formedness of lexical items may also impact high-level sentence processing mechanisms. This interpretation supports theories such as the Lexical Quality Hypothesis (Perfetti & Hart, 2002) and Hofmeister's (2008; 2010) theory that representational complexity affects memory retrieval in language comprehension.

The Lexical Quality Hypothesis suggests that lexical representations are an important source of individual differences in many tasks tapping reading comprehension, meaning that factors involved in individual word reading (such as spelling ability, among other things) are also important for more high-level text comprehension (Perfetti, 1985, 1992). Perfetti and colleagues have found that good spellers also have more stable representations of words, in general, including "…well-specified and partly redundant representations of form (orthography and phonology) and flexible representations of meaning, allowing for rapid and reliable meaning retrieval" (Perfetti, 2007, p. 357). It is possible that variability in lexical quality has direct consequences for retrieval of words in sentence processing, particularly across complicated syntactic constructions such as long-distance dependencies.

New work by Hofmeister (2008, 2010) suggests that there is reason to believe that representational complexity, or specificity, of noun phrases is important for sentence processing. Hofmeister showed that more detailed noun phrases were more difficult to encode but easier to access than simpler noun phrases. In his studies, this level of detail, or specificity, was indexed by linguistic complexity such as number of words used to describe the noun phrase or the general

specificity of the word. For instance, in the first study, more descriptors were added to a noun phrase to increase the complexity of the noun phrase. For example, the noun phrase *a communist* is less complex than the noun phrase *an alleged communist*, which is in turn less complex than the noun phrase *an alleged Venezuelan communist*. In the second study, more complex noun phrases always used a more specific descriptor (such as *which soldier*) whereas the less complex noun phrase was also a more generic descriptor (*which person*). In both cases, reading times for regions of a sentence in which noun phrase had to be retrieved were increased for the less specific noun phrases, indicating that having a more complex, and therefore more specific, representation of the noun phrase encoded in memory facilitated retrieval.

Taken together, the findings from the work of both Perfetti and from Hofmeister converge to suggest that the quality of the memory representation, influenced both by a) persistent individual differences in lexical quality and b) contextual, instance-specific representational features (such as the descriptive complexity) of referents, or even variability in long-term knowledge of particular referents, are important factors that may influence sentence processing complexity. As a result, future studies should take these findings into consideration in investigating individual differences in sentence processing, including measures that test both individual differences in lexical quality and in the ability to form more specific, richer representations.

In addition, recent work by Van Dyke and colleagues (2010) suggests that individual variability in the quality of representations may be particularly important for susceptibility to interference from other items, supporting retrieval interference theories about sentence processing complexity (Lewis & Vasishth, 2005; Lewis, et al., 2006; Oztekin & McElree, 2007; Van Dyke & McElree, 2011). In particular, given that poorer lexical representations are less

specific, they lead to a noisier match between the target and retrieval cues in the context, which makes them more likely to be susceptible to interference from other potential cue-target matches. That is, even though each potentially interfering word in a sentence does not have a higher independent match with retrieval cues, the overall likelihood of interference does, because the likelihood of retrieving the correct target is lowered due to noise. As a result, future studies should investigate the relationship between more general interference in memory paradigms and interference in sentence processing.

In Study 2, a number of sentence processing tasks were investigated with the aim of elucidating the most internally consistent tasks, including variations of the two sentence processing tasks investigated in Study 1. Contrary to prior beliefs, materials presented in null (i.e., isolated) contexts were found to be more internally reliable. One plausible reason for this finding is that less variability in processing materials occurs when there is minimal information in the context (i.e., when less information is known about each referent). More variability between items might occur once these items are processed more fully, as in the supportive context presentations. Furthermore, Study 2 showed that a third manipulation, the nested vs. branching relative clauses task, was also highly internally reliable and correlated to some extent with other sentence processing manipulations. Future studies investigating individual differences in sentence processing should make use of these findings and consider them for inclusion in a sentence processing battery.f

In general, future studies investigating individual differences in sentence processing should focus in part on the mechanisms and representations underlying retrieval interference. Classic tasks from the memory literature including list memory (focusing on variability in release from proactive interference), false memories, and memory for whole sentences should be

included, along with tasks tapping variability in vocabulary and spelling skills. These tasks, along with their potential relevance for understanding sentence processing, are described next.

Simple list memory tasks, in which participants are required to memorize a short list of materials, often followed by a period in which participants perform an interpolated task (to prevent direct rehearsal of the memory items) before retrieval is tested, tap various properties of general memory skills. In particular, tasks created by Gardiner and colleagues measured a property of memory they termed release from proactive interference (Gardiner, Craik, & Birtwistle, 1972). In this type of task, short lists of materials from a common category are presented to a participant. These materials must subsequently be recalled, typically after a short period of time on the order of seconds. Over a series of such trials, proactive interference may build up, meaning that interference from lists of words from the same category may arise, causing more difficulty on memory for lists occurring later in the sequence. However, when the category of words changes, *even at the stage of retrieval*, interference lessens, and words are much easier to recall. This phenomenon is known as release from proactive interference. Interestingly, Gardiner and colleagues found that even when the general category of the words was left the same, but a *subcategory* pertinent to the most recently studied list was provided to participants just before retrieval, participants still exhibited release from proactive interference and exhibited higher recall than on identical trials in which no subcategory change was mentioned to the participants. These findings suggest that the cues at retrieval are important for memory of items, and when the cues become more specific, recent items are remembered more easily, with less retrieval interference from previously seen items. It is likely that such mechanisms are also at play in sentence processing, where the syntactic and semantic cues at retrieval are typically specific enough to eliminate interference from many other items in the

sentence. The relationship between release from proactive interference and sentence processing has not been systematically explored by comparing individual differences in the Gardiner-type tasks and in sentence processing tasks.

It is plausible that individual variability in keeping memory representations distinct (which is likely connected to variability in lexical quality and to the preciseness, or distinctiveness, of a particular representation) is also related to individual variability in susceptibility to false memories. Using a paradigm similar to the basic list memory paradigm described above, Roediger and McDermott (1995) examined false recall rates for words similar to words presented in memory lists.[6] For instance, in a list containing words like *sour, candy*, and *sugar*, the related word *sweet* is often falsely recalled. Such studies have found high rates of false recall and false recognition up to 40-55% of the time, and individual differences in such false memory predict real-life susceptibility to falsely recalled memories (Geraerts, Lindsay, Merckelbach, Jelicic, Raymaekers, Arnold, & Schooler, 2009). One theory used to explain the prevalence of such false memories is that activation spreads to words related to the words in memory lists, including false lures such as *sweet*, in the above example (Collins & Loftus, 1975). It is plausible that imperviousness to false memories might rely on resistance to interference from other activated words, which is plausibly related to resistance to interference in sentence processing. Investigating the relationship between false memories and sentence processing may shed light on such interference.

Related to the false memory of lures in DRM paradigms, false memory for particular lexical items in sentence memory has also been extensively studied. Potter and Lombardi (1990) tested the claim that immediate recall for whole sentences involves a reconstruction from

---

[6] This paradigm was modeled after work by Deese (1959), and is often referred to as the "DRM" (Deese-Roediger-McDermott) paradigm.

114

conceptual memory, and not from a concrete phonological representation. They found that when

presented with synonyms of a word occurring in a sentence as lures (either before or after the

sentence), participants often recalled the sentences with these lure words intruding. This task is

related to the false memory tasks described above in that it requires keeping particular words

separate from other related words, here, synonyms. It is therefore likely that this task may also

predict performance in sentence processing.

Finally, future studies investigating individual differences in sentence processing should

also include measures of spelling and vocabulary, as such measures are thought to affect lexical

quality (Perfetti & Hart, 2002), which may affect susceptibility to interference in sentence

processing. Vocabulary has previously been found to correlate with a sentence processing

measure (Van Dyke et al., 2010), and in Study 1, spelling was also found to correlate with

sentence processing in the Cleft Sentences task.

In summary, the current set of studies achieved two goals. First, Study 1 found

correlations between an online sentence processing measure and two nonlinguistic measures,

namely the dual 2-back task and the spelling test. These measures provide some support for

retrieval interference theories of sentence processing complexity, suggesting that resources and

mechanisms are shared between a) retrieval processes in sentence processing and in other, more

general memory mechanisms (as indicated by the correlation with the dual 2-back task) and, to

an extent, b) access to lexical representations (as indicated by the correlation with the spelling

task). In addition, Study 2 methodically tested several sentence processing measures, finding that

sentence materials in null contexts exhibit higher reliability than materials presented in

supportive contexts. These findings suggest that future studies of individual differences in

sentence processing should systematically investigate general retrieval mechanisms and compare

them to sentence processing tasks such as those investigated here.

References

Baddeley, A.D. (1966a). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology, 18*: 362-365.

Baddeley, A.D. (1966b). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology, 18*: 302-309.

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

Baddeley, A.D. (2007). *Working Memory, Thought and Action*. Oxford: Oxford University Press.

Baddeley, A.D. (2010). Working memory. *Current Biology, 20*(4): R136-R140.

Baddeley, A.D. & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

Berch, D.B., Krikorian, R., & Huha, E.M. (1998). The Corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition, 38*: 317-338.

Bever, T.G. (1970). The cognitive basis for linguistic structures. In Hayes, J.R. (Ed.), *Cognition and the Development of Language* (pp. 279-362). New York: Wiley.

Bush, G. & Shin, L.M. (2006). The Multi-Source Interference Task: an fMRI task that reliably activates the cingulo-frontal-parietal cognition/attention network. *Nature Protocols, 1*(1): 308-313.

Caplan, D. & Waters, G.S. (1999). Verbal working memory and sentence comprehension. *Brain and Behavioral Sciences, 22*: 77-126.

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

Chen, Z. & Cowan, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6): 1235-1249.

Christianson, K., Luke, S.G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(2): 538-544.

Colom, R, Rebollo, I., Abad, F.J., & Shih, P.C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory & Cognition, 34*(1): 158-171.

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5): 769-786.

Corsi, P.M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34*(2): 819B. University Microfilms No. AAI05-77717.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*: 87-185.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*: 51-57.

Crowder, R.G. (1976). *Principles of Learning and Memory*. Hillsdale, NJ: Erlbaum.

Daneman, M. & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4): 450-466.

Daneman, M. & Merikle, P.M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4): 422-433.

DeDe, G., Caplan, D., Kemtes, K., & Waters, G. (2004). The relationship between age, verbal working memory, and language comprehension. *Psychology and Aging, 19*(4): 601-616.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*: 17-22.

Durston, S., Thomas, K.M., Yang, Y., Uluğ, A.M., Zimmerman, R.D., & Casey, B.J. (2002). A neural basis for the development of inhibitory control. *Developmental Science, 5*(4): F9-F16.

Engle, R.W. & Kane, M.J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *The psychology of learning and motivation, 44*, 145-199.

Fedorenko, E., Frank, M., Everett, D., & Gibson, E. (Submitted). Short-term memory abilities in an isolated Amazonian tribe.

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

Fedorenko, E., Gibson, E., & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language, 56*: 246-269.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*: 164-203.

Fodor, J.A. (1983). *The modularity of mind*. MIT Press.

Forster, K.I. (1979). Levels of processing and the structure of the language processor. In W.E. Cooper & E.C.T. Walker (Eds.), *Sentence processing*. Erlbaum.

Frazier, L. (1990). Exploring the architecture of the language-processing system. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. MIT Press/Bradford Books.

Friedman, N.P. & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language, 51*: 136-158.

Friedman, N.P. & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods, Instruments, & Computers, 37*: 581-590.

Galton, F. (1887). Notes on prehension in idiots. *Mind, 12*: 79-82.

Gardiner, J.M., Craik, F.A., & Birtwistle, J. (1972). Retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior, 11*: 778-783.

Gathercole, S.E. & Baddeley, A.D. (1994). *Working Memory and Language*. Hillsdale: Lawrence Erlbaum.

Gennari, S.P. & MacDonald, M.C. (2008). Semantic indeterminacy in relative clauses. *Journal of Memory and Language, 58*: 161-187.

Geraerts, E., Lindsay, D.S., Merckelbach, H., Jelicic, M., Raymaekers, L., Arnold, M.M., & Schooler, J.W. (2009). Cognitive mechanisms underlying recovered-memory experiences of childhood sexual abuse. *Psychological Science, 20*(1): 92-98.

Gibson, E. (1991). A computational theory of human linguistic processing: Memory limitations and processing breakdown. PhD Dissertation, Carnegie Mellon University, Pittsburgh, PA.

Gibson, E. (1998). Linguisic complexity: locality of syntactic dependencies. *Cognition, 68*: 1-76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95-126), Cambridge, MA: MIT Press.

Gibson, E. & Thomas, J. (1998). *The complexity of English nested structures*. Unpublished manuscript, Massachusetts Institute of Technology, Cambridge, MA.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes, 14*(3): 225-248.

Gilchrist, A.L., Cowan, N., & Naveh-Benjamin, M. (2008). Working memory capacity for spoken sentences decreases with adult aging: Recall of fewer, but not smaller chunks in older adults. *Memory, 16*(7): 773-787.

Gordon, P.C, Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6): 1411-1423.

Gordon, P.C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*: 97-114.

Gordon, P.C., Hendrick, R., Levine, W.H. (2002). Memory-load interference in syntactic processing. *Psychological Science, 13*(5): 425-430.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL, 2*: 159-166.

Hanley, J.R., Young, A., & Pearson, N.A. (1991). Impairment of the visuo-spatial scratchpad. *Quarterly Journal of Experimental Psychology, 43A*: 101-125.

Hofmeister, P. (2008). Representational complexity and memory retrieval in language processing. PhD Dissertation, Carnegie Mellon University, Pittsburgh, PA.

Hofmeister, P. (2010). Representational complexity and memory retrieval in language processing. *Language and Cognitive Processes, 26*(3): 376-405.

Hofmeister, P. & Sag, I.A. (2010). Cognitive constraints and island effects. *Language, 86*(2): 366-415.

Jacobs, J. (1887). Experiments in "prehension." *Mind, 12*: 75-79.

Jaeggi, S.M., Buschkuehl, M., Jonides, J., & Perrig, W.J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Science, 105*(19): 6829-6833.

Jaeggi, S.M., Buschkuehl, M., Perrig, W.J., & Meier, B. (2010). The concurrent validity of the *N*-back task as a working memory measure. *Memory, 18*(4): 394-412.

Jaeggi, S.M., Etienne, A., Ozdoba, C., Perrig, W.J., & Nirkko, A.C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive, Affective, & Behavioral Neuroscience, 7*(2), 75-89.

Jonides, J., Smith, E.E., Koeppe, R.A., Awh, E., Minoshima, S., & Mintun, M.A. (1993). Spatial working memory in humans as revealed by PET. *Nature, 363*, 623-625.

Just, M.A. & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1): 122-149.

Kane, M.J., Conway, A.R.A., Miura, T.K., & Colflesh, G.J.H. (2007). Working memory, attention control, and the *N*-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3): 615-622.

Kane, M.J., Hambrick, D.Z., Tuholski, S.W., Wilhelm, O., Payne, T.W., & Engle, R.W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2): 189-217.

Keppel, G. & Underwood, B.J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1*(3): 153-161.

Kemper, S. & Herman, R.E. (2006). Age differences in memory-load interference effects in syntactic processing. *Journal of Gerontology B Psychological Sciences, 61*(6): P327-P332.

Kim, A. & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*: 205-225.

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

Kuperberg, G.R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P.J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes, 21*(5): 489-530.

La Pointe, L.B. & Engle, R.W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(6): 1118-1133.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*: 1126-1177.

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

Lewis, R.L. & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*: 375-419.

Lewis, R.L., Vasishth, S., & Van Dyke, J.A. (2006). Computational principles of working memory in sentence comprehension. *TRENDS in Cognitive Science, 10*(10): 447-454.

Lezak, M. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Loonstra, A.S., Tarlow, A.R., & Sellers, A.H. (2001). COWAT metanorms across age, education, and gender. *Applied Neuropsychology, 8*(3): 161-166.

Love, T. & Swinney, D. (1996). Coreference processing and levels of analysis in object-relative constructions; Demonstration of antecedent reactivation with the cross-modal priming paradigm. *Journal of Psycholinguistic Research, 25*(1): 5-24.

Luce, P.A. & Pisoni, D.B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear and Hearing, 19*(1): 1-36.

Marcus, G. (2006). Cognitive architecture and descent with modification. *Cognition, 101*: 443-465.

McCabe, D.P., Roediger, H.L., McDaniel, M.A., Balota, D.A., & Hambrick, D.Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology, 24*(2): 222-243.

McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time-course analysis. *Memory & Cognition, 24*(2): 173-187.

McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language, 38*: 225-252.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*(2): 111-123.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(3): 817-835.

McElree, B. (2006). Accessing recent events. In B.H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46). San Diego: Academic Press.

McElree, B. & Dosher, B.A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General, 118*(4): 346-373.

McElree & Dosher, B.A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General, 122*(3): 291-315.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*: 67-91.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2): 81-97.

Munro, R. (2010). Crowdsourced translation for emergency response and beyond. *NSF Workshop on crowdsourcing and translation*, University of Maryland.

Oberauer, K., Lange, E., & Engle, R.W. (2004). Working memory capacity and resistance to interference. *Journal of Memory and Language, 51*(1): 80-96.

Öztekin, I. & McElree, B. (2007). Proactive interference slows recognition by eliminating fast assessments of familiarty. *Journal of Memory and Language, 57*: 126-149.

Perfetti, C. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. (1992). The representational problem in reading acquisition. In P.B. Gough, L.C. Ehri, & R. Treiman (Eds.), *Reading Acquisition*, pp. 145-174. Hillsdale, NJ: Erlbaum.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4): 357-383.

Perfetti, C.A. & Hart, L. (2002). The lexical quality hypothesis. In L. Vehoeven, C. Elbrow, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam/Philadelphia: John Benjamins.

Perfetti, C. & Lesgold, A.M. (1977). Discourse comprehension and sources of individual differences. In M.A. Just & P.A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 141-183). Hillsdale, NJ: Erlbaum.

Pisoni, D.B., Nusbaum, H.C., Luce, P.A., & Slowiaczek, L.M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication, 4*(1-3): 75-95.

Raven, J.C. (1990). *Advanced Progressive Matrices: Sets I, II*. Oxford University Press: Oxford.

Raven, J.C. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41*: 1-48.

Roberts, R. & Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research, 31*(6), 573-598.

Roediger, H.L. & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4): 803-814.

Rosen, V. & Engle, R.W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126*: 211-227.

Schmiedeck, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 1089-1096.

Seidenberg, M.S. & McClelland, J.L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Shah & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 129*(1): 4-27.

Shamosh, N.A., DeYoung, C.G., Green, A.E., Reis, D.L., Johnson, M.R., Conway, A.R.A., Engle, R.W., Braver, T.S., & Gray, J.R. (2008). Individual differences in delay discounting. *Psychological Science, 19*(9), 904-911.

Sommers, M.S., Kirk, K.I., & Pisoni, D.B. (1997). Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear & Hearing, 18*(2): 89-99.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*(3736): 652-654.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), *Attention and performance II* (pp. 276-315). Amsterdam: North Holland.

Sternberg, S. (1975). Memory-scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27*: 1-32.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*: 643-662.

Troyer, A.K., Moscovitch, M., & Winocur, G. (1997). Cluster and switching as two components of verbal fluency: Evidence from young and older healthy adults. *Neuropscyhology, 11*(1): 138-146.

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

Unsworth, N. & Engle, R.W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language, 54*: 68-80.

Unsworth, N. & Engle, R.W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1): 104-132.

Unsworth, N. & Engle, R.W. (2007b). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*(6): 1038-1066.

Vallar, G. & Shallice, T. (Eds.) (1990). *Neuropsychological impairments of short-term memory*. New York: Cambridge University Press.

Van Dyke, J.A., Johns, C.L., & Kukona, A. (2010). Individual differences in sentence comprehension: A retrieval interference approach. Talk presented at the 23rd Annual CUNY Conference on Human Sentence Processing.

Van Dyke, J.A. & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language, 55*: 157-166.

Van Dyke, J.A. & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language, 65*: 247-263.

Wanner, E. & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119-161). Cambridge, MA: MIT Press.

APPENDIX A.

The following set of 24 experimental items and 36 filler items was used in the Cleft Sentences experiment in Study 1 and in Experiment 1 of Study 2. For each experimental item, the two context sentences are provided first, followed by the four different types of sentences (object clefts and sentences clefts with two different noun phrases from the context sentences in the relative clause region. The filler items are presented with all three sentences, and were exactly the same for all participants.

Experimental items

1. A novelist, a cinematographer, and a producer were talking.
Mary: I heard that either the novelist mocked the cinematographer for wearing a tie, or the cinematographer mocked the novelist for wearing a tie.
Object cleft, version 1: John: Hmm, I heard that it was the producer that the novelist mocked for wearing a tie.
Object cleft, version 2: John: Hmm, I heard that it was the producer that the cinematographer mocked for wearing a tie.
Subject cleft, version 1: John: Hmm, I heard that it was the producer that mocked the novelist for wearing a tie.
Subject cleft, version 2: John: Hmm, I heard that it was the producer that mocked the cinematographer for wearing a tie.

2. A mime, a guitarist, and a juggler were having fun.
Mary: I heard that either the mime entertained the guitarist with a new joke, or the guitarist entertained the mime with a new joke.
Object cleft, version 1: John: Hmm, I heard that it was the juggler that the mime entertained with a new joke.
Object cleft, version 2: John: Hmm, I heard that it was the juggler that the guitarist entertained with a new joke.
Subject cleft, version 1: John: Hmm, I heard that it was the juggler that entertained the mime with a new joke.
Subject cleft, version 2: John: Hmm, I heard that it was the juggler that entertained the guitarist with a new joke.

3. A duke, an earl, and a queen were dining.
Mary: I heard that either the duke ignored the earl for most of the night, or the earl ignored the duke for most of the night.
Object cleft, version 1: John: Hmm, I heard that it was the queen that the duke ignored for most of the night.
Object cleft, version 2: John: Hmm, I heard that it was the queen that the earl ignored for most of the night.
Subject cleft, version 1: John: Hmm, I heard that it was the queen that ignored the duke for most of the night.
Subject cleft, version 2: John: Hmm, I heard that it was the queen that ignored the earl for most of the night.

4. A director, an actor, and a screenwriter were filming a movie.
Mary: I heard that either the director irritated the actor by losing the camera battery, or the actor irritated the director by losing the camera battery.
Object cleft, version 1: John: Hmm, I heard that it was the screenwriter that the director irritated by losing the camera battery.
Object cleft, version 2: John: Hmm, I heard that it was the screenwriter that the actor irritated by losing the camera battery.
Subject cleft, version 1: John: Hmm, I heard that it was the screenwriter that irritated the director by losing the camera battery.
Subject cleft, version 2: John: Hmm, I heard that it was the screenwriter that irritated the actor by losing the camera battery.

5. A trainer, a coach, and an agent were meeting.
Mary: I heard that either the trainer harassed the agent about the team's performance, or the agent harassed the trainer about the team's performance.
Object cleft, version 1: John: Hmm, I heard that it was the coach that the trainer harassed about the team's performance.
Object cleft, version 2: John: Hmm, I heard that it was the coach that the agent harassed about the team's performance.
Subject cleft, version 1: John: Hmm, I heard that it was the coach that harassed the trainer about the team's performance.
Subject cleft, version 2: John: Hmm, I heard that it was the coach that harassed the agent about the team's performance.

6. A referee, a linebacker, and a quarterback were tending to an injured player.
Mary: I heard that either the referee assisted the quarterback in getting the player on his feet, or the quarterback assisted the referee in getting the player on his feet.
Object cleft, version 1: John: Hmm, I heard that it was the linebacker that the referee assisted in getting the player on his feet.
Object cleft, version 2: John: Hmm, I heard that it was the linebacker that the quarterback assisted in getting the player on his feet.
Subject cleft, version 1: John: Hmm, I heard that it was the linebacker that assisted the referee in getting the player on his feet.
Subject cleft, version 2: John: Hmm, I heard that it was the linebacker that assisted the quarterback in getting the player on his feet.

7. A cook, a pastry-chef, and a manager were making a dinner.
Mary: I heard that either the cook deceived the manager about paying for ingredients, or the manager deceived the cook about paying for ingredients.
Object cleft, version 1: John: Hmm, I heard that it was the pastry-chef that the cook deceived about paying for ingredients.
Object cleft, version 2: John: Hmm, I heard that it was the pastry-chef that the manager deceived about paying for ingredients.
Subject cleft, version 1: John: Hmm, I heard that it was the pastry-chef that deceived the cook about paying for ingredients.

Subject cleft, version 2: John: Hmm, I heard that it was the pastry-chef that deceived the manager about paying for ingredients.

8. A baker, a brewer, and a butcher were exchanging recipes.
Mary: I heard that either the baker praised the butcher for an old recipe from Italy, or the butcher praised the baker for an old recipe from Italy.
Object cleft, version 1: John: Hmm, I heard that it was the brewer that the baker praised for an old recipe from Italy.
Object cleft, version 2: John: Hmm, I heard that it was the brewer that the butcher praised for an old recipe from Italy.
Subject cleft, version 1: John: Hmm, I heard that it was the brewer that praised the baker for an old recipe from Italy.
Subject cleft, version 2: John: Hmm, I heard that it was the brewer that praised the butcher for an old recipe from Italy.

9. A surgeon, an oncologist, and an orderly were discussing a patient's case.
Mary: I heard that either the oncologist misled the surgeon about the location of the X-rays, or the surgeon misled the oncologist about the location of the X-rays.
Object cleft, version 1: John: Hmm, I heard that it was the orderly that the oncologist misled about the location of the X-rays.
Object cleft, version 2: John: Hmm, I heard that it was the orderly that the surgeon misled about the location of the X-rays.
Subject cleft, version 1: John: Hmm, I heard that it was the orderly that misled the oncologist about the location of the X-rays.
Subject cleft, version 2: John: Hmm, I heard that it was the orderly that misled the surgeon about the location of the X-rays.

10. A photographer, a filmmaker, and a correspondent were attending a press conference.
Mary: I heard that either the filmmaker shoved the photographer in order to talk to the mayor, or the photographer shoved the filmmaker in order to talk to the mayor.
Object cleft, version 1: John: Hmm, I heard that it was the correspondent that the filmmaker shoved in order to talk to the mayor.
Object cleft, version 2: John: Hmm, I heard that it was the correspondent that the photographer shoved in order to talk to the mayor.
Subject cleft, version 1: John: Hmm, I heard that it was the correspondent that shoved the filmmaker in order to talk to the mayor.
Subject cleft, version 2: John: Hmm, I heard that it was the correspondent that shoved the photographer in order to talk to the mayor.

11. A postdoc, an RA, and a graduate student were contemplating who should apply for a new research position.
Mary: I heard that either the RA recommended the postdoc as a good applicant for the job, or the postdoc recommended the RA as a good applicant for the job.
Object cleft, version 1: John: Hmm, I heard that it was the graduate student that the RA recommended as a good applicant for the job.
Object cleft, version 2: John: Hmm, I heard that it was the graduate student that the postdoc

recommended as a good applicant for the job.
Subject cleft, version 1: John: Hmm, I heard that it was the graduate student that recommended the RA as a good applicant for the job.
Subject cleft, version 2: John: Hmm, I heard that it was the graduate student that recommended the postdoc as a good applicant for the job.

12. A biologist, a chemist, and a psychologist were reading a paper.
Mary: I heard that either the chemist enlightened the biologist about the meaning of the paper, or the biologist enlightened the chemist about the meaning of the paper.
Object cleft, version 1: John: Hmm, I heard that it was the psychologist that the chemist enlightened about the meaning of the paper.
Object cleft, version 2: John: Hmm, I heard that it was the psychologist that the biologist enlightened about the meaning of the paper.
Subject cleft, version 1: John: Hmm, I heard that it was the psychologist that enlightened the chemist about the meaning of the paper.
Subject cleft, version 2: John: Hmm, I heard that it was the psychologist that enlightened the biologist about the meaning of the paper.

13. A cheerleader, an acrobat, and a dancer were fighting for gym space.
Mary: I heard that either the acrobat disliked the dancer for wanting too much space, or the dancer disliked the acrobat for wanting too much space.
Object cleft, version 1: John: Hmm, I heard that it was the cheerleader that the acrobat disliked for wanting too much space.
Object cleft, version 2: John: Hmm, I heard that it was the cheerleader that the dancer disliked for wanting too much space.
Subject cleft, version 1: John: Hmm, I heard that it was the cheerleader that disliked the acrobat for wanting too much space.
Subject cleft, version 2: John: Hmm, I heard that it was the cheerleader that disliked the dancer for wanting too much space.

14. A maid, a bellhop, and a janitor were getting ready for guests.
Mary: I heard that either the bellhop distracted the janitor while they were trying to work, or the janitor distracted the bellhop while they were trying to work.
Object cleft, version 1: John: Hmm, I heard that it was the maid that the bellhop distracted while they were trying to work.
Object cleft, version 2: John: Hmm, I heard that it was the maid that the janitor distracted while they were trying to work.
Subject cleft, version 1: John: Hmm, I heard that it was the maid that distracted the bellhop while they were trying to work.
Subject cleft, version 2: John: Hmm, I heard that it was the maid that distracted the janitor while they were trying to work.

15. A jogger, a biker, and a skater were arguing about an accident.
Mary: I heard that either the biker pushed the skater for causing the accident, or the skater pushed the biker for causing the accident.
Object cleft, version 1: John: Hmm, I heard that it was the jogger that the biker pushed for

causing the accident.
Object cleft, version 2: John: Hmm, I heard that it was the jogger that the skater pushed for causing the accident.
Subject cleft, version 1: John: Hmm, I heard that it was the jogger that pushed the biker for causing the accident.
Subject cleft, version 2: John: Hmm, I heard that it was the jogger that pushed the skater for causing the accident.

16. A machinist, a programmer, and a technician were outlining a project.
Mary: I heard that either the programmer liked the technician for being easy to work with, or the technician liked the programmer for being easy to work with.
Object cleft, version 1: John: Hmm, I heard that it was the machinist that the programmer liked for being easy to work with.
Object cleft, version 2: John: Hmm, I heard that it was the machinist that the technician liked for being easy to work with.
Subject cleft, version 1: John: Hmm, I heard that it was the machinist that liked the programmer for being easy to work with.
Subject cleft, version 2: John: Hmm, I heard that it was the machinist that liked the technician for being easy to work with.

17. A radiologist, a cardiologist, and a consultant were deciding whether to run a test.
Mary: I heard that either the consultant persuaded the radiologist to run the test, or the radiologist persuaded the consultant to run the test.
Object cleft, version 1: John: Hmm, I heard that it was the cardiologist that the consultant persuaded to run the test.
Object cleft, version 2: John: Hmm, I heard that it was the cardiologist that the radiologist persuaded to run the test.
Subject cleft, version 1: John: Hmm, I heard that it was the cardiologist that persuaded the consultant to run the test.
Subject cleft, version 2: John: Hmm, I heard that it was the cardiologist that persuaded the radiologist to run the test.

18. A physicist, an astronomer, and a mathematician were designing an experiment.
Mary: I heard that either the mathematician directed the physicist about the execution of the experiment, or the physicist directed the mathematician about the execution of the experiment.
Object cleft, version 1: John: Hmm, I heard that it was the astronomer that the mathematician directed about the execution of the experiment.
Object cleft, version 2: John: Hmm, I heard that it was the astronomer that the physicist directed about the execution of the experiment.
Subject cleft, version 1: John: Hmm, I heard that it was the astronomer that directed the mathematician about the execution of the experiment.
Subject cleft, version 2: John: Hmm, I heard that it was the astronomer that directed the physicist about the execution of the experiment.

19. A plumber, an electrician, and a painter were working on a new house.
Mary: I heard that either the painter annoyed the plumber during long hours on the job, or the

plumber annoyed the painter during long hours on the job.
Object cleft, version 1: John: Hmm, I heard that it was the electrician that the painter annoyed during long hours on the job.
Object cleft, version 2: John: Hmm, I heard that it was the electrician that the plumber annoyed during long hours on the job.
Subject cleft, version 1: John: Hmm, I heard that it was the electrician that annoyed the painter during long hours on the job.
Subject cleft, version 2: John: Hmm, I heard that it was the electrician that annoyed the plumber during long hours on the job.

20. An illustrator, a biographer, and a publisher were analyzing a book's success.
Mary: I heard that either the publisher alarmed the illustrator about the number of copies sold, or the illustrator alarmed the publisher about the number of copies sold.
Object cleft, version 1: John: Hmm, I heard that it was the biographer that the publisher alarmed about the number of copies sold.
Object cleft, version 2: John: Hmm, I heard that it was the biographer that the illustrator alarmed about the number of copies sold.
Subject cleft, version 1: John: Hmm, I heard that it was the biographer that alarmed the publisher about the number of copies sold.
Subject cleft, version 2: John: Hmm, I heard that it was the biographer that alarmed the illustrator about the number of copies sold.

21. A valet, a bartender, and a waiter were telling jokes outside a restaurant.
Mary: I heard that either the waiter provoked the bartender with a dirty joke, or the bartender provoked the waiter with a dirty joke.
Object cleft, version 1: John: Hmm, I heard that it was the valet that the waiter provoked with a dirty joke.
Object cleft, version 2: John: Hmm, I heard that it was the valet that the bartender provoked with a dirty joke.
Subject cleft, version 1: John: Hmm, I heard that it was the valet that provoked the waiter with a dirty joke.
Subject cleft, version 2: John: Hmm, I heard that it was the valet that provoked the bartender with a dirty joke.

22. A welder, a woodworker, and a blacksmith were preparing for a job.
Mary: I heard that either the blacksmith trained the woodworker to build a gate, or the woodworker trained the blacksmith to build a gate.
Object cleft, version 1: John: Hmm, I heard that it was the welder that the blacksmith trained to build a gate.
Object cleft, version 2: John: Hmm, I heard that it was the welder that the woodworker trained to build a gate.
Subject cleft, version 1: John: Hmm, I heard that it was the welder that trained the blacksmith to build a gate.
Subject cleft, version 2: John: Hmm, I heard that it was the welder that trained the woodworker to build a gate.

23. An actress, a playwright, and a choreographer were lamenting their play's failure.
Mary: I heard that either the choreographer ridiculed the playwright for having predicted a huge success, or the playwright ridiculed the choreographer for having predicted a huge success.
Object cleft, version 1: John: Hmm, I heard that it was the actress that the choreographer ridiculed for having predicted a huge success.
Object cleft, version 2: John: Hmm, I heard that it was the actress that the playwright ridiculed for having predicted a huge success.
Subject cleft, version 1: John: Hmm, I heard that it was the actress that ridiculed the choreographer for having predicted a huge success.
Subject cleft, version 2: John: Hmm, I heard that it was the actress that ridiculed the playwright for having predicted a huge success.

24. A psychic, a hypnotist, and a snake-charmer were showing off their talents.
Mary: I heard that either the snake-charmer tricked the hypnotist by using a spell, or the hypnotist tricked the snake-charmer by using a spell.
Object cleft, version 1: John: Hmm, I heard that it was the psychic that the snake-charmer tricked by using a spell.
Object cleft, version 2: John: Hmm, I heard that it was the psychic that the hypnotist tricked by using a spell.
Subject cleft, version 1: John: Hmm, I heard that it was the psychic that tricked the snake-charmer by using a spell.
Subject cleft, version 2: John: Hmm, I heard that it was the psychic that tricked the hypnotist by using a spell.
Subject cleft, version 1: John: Hmm, I heard that the historian and the librarian preferred the Southern states.

Filler items

1. The historian, the librarian and the lecturer scrutinized a map together.
Mary: I heard that the historian and the librarian said they favored the East coast, while the lecturer said that he preferred the Southern states.
John: Hmm, I heard that the historian and the librarian preferred the Southern states.

2. The tightrope walker, the clown and the acrobats bowed to the audience.
Mary: I heard that the clown opened the circus with an introductory act, and then the tightrope walker and the acrobats ended the show with an amazing finale.
John: That's right. I also heard that the tightrope walker did not want to be in the circus when he was little.

3. The chauffeur, the beauty queen and the actor sat in the limousine together.
Mary: I heard that the beauty queen and the chauffeur exchanged private looks, while the actor talked on the phone.
John: That's right. I heard that the beauty queen and the chauffeur are planning to elope together.

4. The cartoonist, the filmmaker and the producer expressed interest in producing a movie together.

Mary: I heard that the cartoonist and the producer wanted to make a movie for young children; however, the filmmaker wanted to make an R-rated movie.
John: That's right. But I heard that every movie that the filmmaker produces is a total failure on opening weekend.

5. The chemist, the physicist and the biologist conducted several experiments.
Mary: I heard that the chemist and the physicist came up with a new hypothesis, while the biologist offered an alternative account.
John: Yes, that's right. I also heard that the chemist and the biologist won a Nobel prize for some important discoveries.

6. On a snowy day Karen, Daniel and Collin sculpted a snowman together.
Mary: I heard that Karen and Daniel were enthusiastic about making the snowman, but Collin did not want to be engaged in the activity because he hates snow.
John: You're wrong. Collin didn't want to make the snowman because he's jealous of Karen and Daniel's blooming relationship.

7. The bachelor, the babysitter and the governess wanted to do something fun together.
Mary: I heard that the bachelor and the babysitter wanted to go skating while the governess wanted to attend a concert.
John: That's right. I heard that the babysitter is one of the bachelor's favorite nieces.

8. The vegetarian, the waiter and the cook perused a menu together.
Mary: I heard that the vegetarian stated that he did not eat meat, and then the waiter and the cook recommended vegetarian dishes to him.
John: Yes, that's true. I heard that the vegetarian is a difficult customer and also notorious for being a very picky eater.

9. The culprit, the lawyer and the judge convened in court.
Mary: I heard that the lawyer defended the culprit's innocence; however, the judge was not persuaded by the argument.
John: That's right. I heard that the culprit hired a top-notch lawyer and paid five million dollars.

10. The teenager scheduled an appointment with the dermatologist and the dietician.
Mary: I heard that the teenager was really depressed about her acne problem, and the dermatologist and the dietician proposed several remedies and solutions.
John: I think you're wrong. The teenager refuses to allow her acne problem to lower her self-esteem.

11. Jeannie, Esther and David discovered a club that they have never been to.
Mary: I heard that Jeannie and Esther entered the club while David lingered outside because he doesn't like dancing.
John: You're wrong. David didn't enter the club because he saw his ex-girlfriend enter the club with another guy.

12. The queen, the princess and the prince prepared for the grand arrival of the king.

Mary: I heard that the queen applied make-up while the princess and the prince checked up on the festive meal.
John: Yeah, and I heard that the queen spent all day primping for the arrival of the king.

13. The mayor, the president and the orator gathered together for a meeting.
Mary: I heard that the mayor and the president wanted to improve their public speaking skills; and so, the orator was giving them advice.
John: You're right. The mayor's and the president's speeches improved significantly after meeting with the orator.

14. The environmentalist, the zoologist and the philanthropist worked on a plan for saving the endangered African elephants.
Mary: I heard that after the environmentalist and the zoologist described the worsening situation, the philanthropist wrote a check for 23 million dollars.
John: That's right. I heard that the environmentalist and the zoologist have been working together for many years.

15. The surfer, the life-guard and the professional swimmer challenged each other to a swim race.
Mary: I heard that the surfer and the life-guard tied for first place while the professional swimmer came last because of a cramp.
John: Hmm, I think that's wrong. I heard that the swimmer came last because she experienced a cramp in the middle of the race.

16. The prodigy, the inventor and the researcher received a grant for their project.
Mary: I heard that the prodigy and the inventor wanted to spend the money on their ongoing research; however, the researcher wanted to start a new project.
John: Yeah, I heard that the prodigy and the inventor spent several hours trying to negotiate a compromise with the researcher.

17. At the ice rink Stephanie, Derek and Tom skated together.
Mary: I heard that Stephanie and Derek fell several times, while Tom only stumbled once.
John: You're right. In fact, I heard that Tom used to be a pro ice-skater and skated in the Olympics twice.

18. The bouncer, the waitress and the DJ conversed during their break.
Mary: I heard that the bouncer playfully teased the waitress, and then the DJ warned the bouncer about his inappropriate behavior.
John: Hmm, that's weird, I heard that the bouncer is asexual and does not much care for anyone.

19. Ann, Tim and Steve had a frank conversation.
Mary: I heard that Ann confided that she was in love with Tim, while Tim and Steve confessed that they were both in love with Ann.
John: You're right. I heard that Tim and Steve got into a fight after Ann confessed her feelings for Tim.

20. The model, the actress and the singer signed a contract.
Mary: I heard that the model and the actress were reluctant to sign, but the singer signed without a moment's hesitation.
John: That's right. I heard that the model and the actress are wanted by several competing agencies.

21. Todd, Eva and Gary rode a roller-coaster together.
Mary: I heard that afterwards, Todd got sick while Eva and Gary observed with disgust.
John: You right. I heard that Todd got sick because he has motion sickness and is afraid of heights.

22. The mom, the dad and the daughter threw snowballs at each other.
Mary: I heard that the mom got hit in the face pretty hard, and the daughter and the dad ended the snowball fight.
John: That's correct. I heard the mom was angry because she got hurt and refuses to participate again.

23. The librarian, the volunteer and the freshman carried the books to the table.
Mary: I heard that the librarian and the volunteer said that only 10 books can be checked out at a time; however, the freshman begged to take out 20.
John: Yes, you're right. I heard that the freshman has a presentation on the negative impacts of global warming coming up in her seminar class.

24. The bully, the cheerleader and the jock complained about the negative stereotypes associated with them.
Mary: I heard that the bully and the jock claimed that they are intelligent, while the cheerleader stated that she is friendly.
John: That's true. In fact, I heard that the bully got an A on a hard test last week.

25. The spendthrift, the unemployed and the millionaire compared their spending habits.
Mary: I heard that the spendthrift and the unemployed spent less money on buying luxury items, while the millionaire spent more money on buying unnecessary things.
John: Yeah, that's right. I heard that the millionaire is a self-made man and has worked hard ever since he was little.

26. The grandmother, the grandfather and the grandchild attended the city fair.
Mary: I heard that the grandmother and the grandchild took pictures with the clown, and then the grandfather bought them cotton candy.
John: That's true. I heard that it was the grandmother's and the grandfather's first time at the fair.

27. The stalker, the performer and the cop arrived at an agreement.
Mary: I heard that the stalker promised that he would not harass the performer, and then the performer and the cop would drop the past charges.
John: That's right. I also heard that the stalker was really obsessed with the performer and knew every single detail about the performer's life.

28. Christine, Annie and Sarah appeared annoyed and frustrated.
Mary: I heard that Christine and Annie were engaged in a heated argument, while Sarah tried to mediate.
John: Yeah, I heard that Christine stole Annie's boyfriend for the second time this year.

29. The interviewer, the interpreter and the pro-golfer met for an interview.
Mary: I heard that the pro-golfer and the interpreter greeted each other with a kiss, while the interviewer looked away in embarrassment.
John: That's true. The pro-golfer said he would formally ask the interpreter out if he won the championship.

30. Kathy, John and Sally arrived at the meeting an hour late.
Mary: I heard that Kathy and John apologized for being tardy, but Sally refused to apologize.
John: Yeah, I heard that it was the second time that Kathy and John were late because they slept in.

31. The drummer, the guitarist and the pianist led praise worship on Sunday.
Mary: I heard that the drummer and the guitarist cried with joy, and then the pianist felt blessed.
John: Yeah, I heard that the drummer and the guitarist have been going to church ever since they were young.

32. Esther, David and Sandy played a video game together at Sandy's house.
Mary: I heard that Esther and David damaged the controllers, and Sandy became furious.
John: That's right. I heard that Esther and David rarely show respect for other people's property.

33. Chelsea, Rachel and Kevin took an exam last week.
Mary: I heard that Chelsea received a high grade on the exam, while Rachel and Kevin got caught for cheating.
John: That's true, and I heard that as punishment, Rachel and Kevin got suspended from school.

34. Kendra, Brad and Meg were together in the library.
Mary: I heard that Kendra studied for her upcoming math exam, while Brad and Meg made paper airplanes.
John: That's right. I heard that Brad and Meg were being obnoxious and quite loud in the library.

35. Sally, Daniel and Laura chatted together over coffee.
Mary: I heard that Sally spilled coffee on Daniel, and then Laura helped clean the coffee spill with a napkin.
John: I think you're right. I heard that Daniel was upset that Sally spilled coffee on his new trousers.

36. Brian, Tiffany and Molly embarrassed themselves at the conference.
Mary: I heard that Brian couldn't answer the questions well, while Tiffany and Molly stuttered during their presentation.
John: Yeah, that's right, and I heard that Tiffany and Molly went to a bar immediately after the conference.

APPENDIX B.

The following 24 experimental items were used in the Relative Clause Sentences experiment from Study 1 and in Experiment 1 of Study 2. For each experimental item, the four versions of each item is given (object and subject relative clause, with two versions each). The 36 filler items used in this experiment are also given.

Experimental items

1. Subject relative clause, version 1: At the press-conference, a senator and two reporters got into an argument.
The senator attacked one of the reporters and then the other reporter attacked the senator.
Mary: I heard that the reporter that attacked the senator admitted to making an error.
John: I am not sure about that. I heard that the reporter that the senator attacked admitted to making an error.

Subject relative clause, version 2: At the press-conference, a reporter and two senators got into an argument.
The reporter attacked one of the senators and then the other senator attacked the reporter.
Mary: I heard that the senator that attacked the reporter admitted to making an error.
John: I am not sure about that. I heard that the senator that the reporter attacked admitted to making an error.

Object relative clause, version 1: At the press-conference, a senator and two reporters got into an argument.
The senator attacked one of the reporters and then the other reporter attacked the senator.
Mary: I heard that the reporter that the senator attacked admitted to making an error.
John: I am not sure about that. I heard that the reporter that attacked the senator admitted to making an error.

Object relative clause, version 2: At the press-conference, a reporter and two senators got into an argument.
The reporter attacked one of the senators and then the other senator attacked the reporter.
Mary: I heard that the senator that the reporter attacked admitted to making an error.
John: I am not sure about that. I heard that the senator that attacked the reporter admitted to making an error.

2. Subject relative clause, version 1: During the interview, a newscaster and two musicians had a brief discussion.
The newscaster insulted one of the musicians and then the other musician insulted the newscaster.
Mary: I heard that the musician that insulted the newscaster left the building after the interview.
John: I am not sure about that. I heard that the musician that the newscaster insulted left the building after the interview.

Subject relative clause, version 2: During the interview, a musician and two newscasters had a

brief discussion.
The musician insulted one of the newscasters and then the other newscaster insulted the musician.
Mary: I heard that the newscaster that insulted the musician left the building after the interview.
John: I am not sure about that. I heard that the newscaster that the musician insulted left the building after the interview.

Object relative clause, version 1: During the interview, a newscaster and two musicians had a brief discussion.
The newscaster insulted one of the musicians and then the other musician insulted the newscaster.
Mary: I heard that the musician that the newscaster insulted left the building after the interview.
John: I am not sure about that. I heard that the musician that insulted the newscaster left the building after the interview.

Object relative clause, version 2: During the interview, a musician and two newscasters had a brief discussion.
The musician insulted one of the newscasters and then the other newscaster insulted the musician.
Mary: I heard that the newscaster that the musician insulted left the building after the interview.
John: I am not sure about that. I heard that the newscaster that insulted the musician left the building after the interview.

3. Subject relative clause, version 1: After the lecture, a scientist and two interns went over the problem set
The scientist confused one of the interns and then the other intern confused the scientist.
Mary: I heard that the intern that confused the scientist worked at a famous lab at Harvard University.
John: I am not sure about that. I heard that the intern that the scientist confused worked at a famous lab at Harvard University.

Subject relative clause, version 2: After the lecture, an intern and two scientists went over the problem set
The intern confused one of the scientists and then the other scientist confused the intern.
Mary: I heard that the scientist that confused the intern worked at a famous lab at Harvard University.
John: I am not sure about that. I heard that the scientist that the intern confused worked at a famous lab at Harvard University.

Object relative clause, version 1: After the lecture, a scientist and two interns went over the problem set
The scientist confused one of the interns and then the other intern confused the scientist.
Mary: I heard that the intern that the scientist confused worked at a famous lab at Harvard University.
John: I am not sure about that. I heard that the intern that confused the scientist worked at a famous lab at Harvard University.

Object relative clause, version 2: After the lecture, an intern and two scientists went over the problem set
The intern confused one of the scientists and then the other scientist confused the intern.
Mary: I heard that the scientist that the intern confused worked at a famous lab at Harvard University.
John: I am not sure about that. I heard that the scientist that confused the intern worked at a famous lab at Harvard University.

4. Subject relative clause, version 1: After the incident, an officer and two detectives talked about possible suspects.
The officer approached one of the detectives and then the other detective approached the officer.
Mary: I heard that the detective that approached the officer had a good record in solving similar cases.
John: I am not sure about that. I heard that the detective that the officer approached had a good record in solving similar cases.

Subject relative clause, version 2: After the incident, a detective and two officers talked about possible suspects.
The detective approached one of the officers and then the other officer approached the detective.
Mary: I heard that the officer that approached the detective had a good record in solving similar cases.
John: I am not sure about that. I heard that the officer that the detective approached had a good record in solving similar cases.

Object relative clause, version 1: After the incident, an officer and two detectives talked about possible suspects.
The officer approached one of the detectives and then the other detective approached the officer.
Mary: I heard that the detective that the officer approached had a good record in solving similar cases.
John: I am not sure about that. I heard that the detective that approached the officer had a good record in solving similar cases.

Object relative clause, version 2: After the incident, a detective and two officers talked about possible suspects.
The detective approached one of the officers and then the other officer approached the detective.
Mary: I heard that the officer that the detective approached had a good record in solving similar cases.
John: I am not sure about that. I heard that the officer that approached the detective had a good record in solving similar cases.

5. Subject relative clause, version 1: Over the weekend, a dentist and two pediatricians talked about various medications.
The dentist called one of the pediatricians and then the other pediatrician called the dentist.
Mary: I heard that the pediatrician that called the dentist left a message about the recommended dosage.

John: I am not sure about that. I heard that the pediatrician that the dentist called left a message about the recommended dosage.

Subject relative clause, version 2: Over the weekend, a pediatrician and two dentists talked about various medications.
The pediatrician called one of the dentists and then the other dentist called the pediatrician.
Mary: I heard that the dentist that called the pediatrician left a message about the recommended dosage.
John: I am not sure about that. I heard that the dentist that the pediatrician called left a message about the recommended dosage.

Object relative clause, version 1: Over the weekend, a dentist and two pediatricians talked about various medications.
The dentist called one of the pediatricians and then the other pediatrician called the dentist.
Mary: I heard that the pediatrician that the dentist called left a message about the recommended dosage.
John: I am not sure about that. I heard that the pediatrician that called the dentist left a message about the recommended dosage.

Object relative clause, version 2: Over the weekend, a pediatrician and two dentists talked about various medications.
The pediatrician called one of the dentists and then the other dentist called the pediatrician.
Mary: I heard that the dentist that the pediatrician called left a message about the recommended dosage.
John: I am not sure about that. I heard that the dentist that called the pediatrician left a message about the recommended dosage.

6. Subject relative clause, version 1: Before the surgery, a physician and two neurologists had a long discussion.
The physician helped one of the neurologists and then the other neurologist helped the physician .
Mary: I heard that the neurologist that helped the physician worked at MGH for the last ten years.
John: I am not sure about that. I heard that the neurologist that the physician helped worked at MGH for the last ten years.

Subject relative clause, version 2: Before the surgery, a neurologist and two physicians had a long discussion.
The neurologist helped one of the physicians and then the other physician helped the neurologist.
Mary: I heard that the physician that helped the neurologist worked at MGH for the last ten years.
John: I am not sure about that. I heard that the physician that the neurologist helped worked at MGH for the last ten years.

Object relative clause, version 1: Before the surgery, a physician and two neurologists had a long discussion.

The physician helped one of the neurologists and then the other neurologist helped the physician .
Mary: I heard that the neurologist that the physician helped worked at MGH for the last ten years.
John: I am not sure about that. I heard that the neurologist that helped the physician worked at MGH for the last ten years.

Object relative clause, version 2: Before the surgery, a neurologist and two physicians had a long discussion.
The neurologist helped one of the physicians and then the other physician helped the neurologist.
Mary: I heard that the physician that the neurologist helped worked at MGH for the last ten years.
John: I am not sure about that. I heard that the physician that helped the neurologist worked at MGH for the last ten years.

7. Subject relative clause, version 1: During the event, an author and two critics argued about the point of a book.
The author annoyed one of the critics and then the other critic annoyed the author.
Mary: I heard that the critic that annoyed the author had strong opinions about many things.
John: I am not sure about that. I heard that the critic that the author annoyed had strong opinions about many things.

Subject relative clause, version 2: During the event, a critic and two authors argued about the point of a book.
The critic annoyed one of the authors and then the other author annoyed the critic.
Mary: I heard that the author that annoyed the critic had strong opinions about many things.
John: I am not sure about that. I heard that the author that the critic annoyed had strong opinions about many things.

Object relative clause, version 1: During the event, an author and two critics argued about the point of a book.
The author annoyed one of the critics and then the other critic annoyed the author.
Mary: I heard that the critic that the author annoyed had strong opinions about many things.
John: I am not sure about that. I heard that the critic that annoyed the author had strong opinions about many things.

Object relative clause, version 2: During the event, a critic and two authors argued about the point of a book.
The critic annoyed one of the authors and then the other author annoyed the critic.
Mary: I heard that the author that the critic annoyed had strong opinions about many things.
John: I am not sure about that. I heard that the author that annoyed the critic had strong opinions about many things.

8. Subject relative clause, version 1: Before the class, a teacher and two students gathered to discuss the homework.
The teacher greeted one of the students and then the other student greeted the teacher.

Mary: I heard that the student that greeted the teacher gave an interview to the school newspaper recently.
John: I am not sure about that. I heard that the student that the teacher greeted gave an interview to the school newspaper recently.

Subject relative clause, version 2: Before the class, a student and two teachers gathered to discuss the homework.
The student greeted one of the teachers and then the other teacher greeted the student.
Mary: I heard that the teacher that greeted the student gave an interview to the school newspaper recently.
John: I am not sure about that. I heard that the teacher that the student greeted gave an interview to the school newspaper recently.

Object relative clause, version 1: Before the class, a teacher and two students gathered to discuss the homework.
The teacher greeted one of the students and then the other student greeted the teacher.
Mary: I heard that the student that the teacher greeted gave an interview to the school newspaper recently.
John: I am not sure about that. I heard that the student that greeted the teacher gave an interview to the school newspaper recently.

Object relative clause, version 2: Before the class, a student and two teachers gathered to discuss the homework.
The student greeted one of the teachers and then the other teacher greeted the student.
Mary: I heard that the teacher that the student greeted gave an interview to the school newspaper recently.
John: I am not sure about that. I heard that the teacher that greeted the student gave an interview to the school newspaper recently.

9. Subject relative clause, version 1: At the debate, a congressman and two governors argued about the national election.
The congressman antagonized one of the governors and then the other governor antagonized the congressman.
Mary: I heard that the governor that antagonized the congressman apologized for being too aggressive.
John: I am not sure about that. I heard that the governor that the congressman antagonized apologized for being too aggressive.

Subject relative clause, version 2: At the debate, a governor and two congressmen argued about the national election.
The governor antagonized one of the congressmen and then the other congressman antagonized the governor.
Mary: I heard that the congressman that antagonized the governor apologized for being too aggressive.
John: I am not sure about that. I heard that the congressman that the governor antagonized apologized for being too aggressive.

Object relative clause, version 1: At the debate, a congressman and two governors argued about the national election.
The congressman antagonized one of the governors and then the other governor antagonized the congressman.
Mary: I heard that the governor that the congressman antagonized apologized for being too aggressive.
John: I am not sure about that. I heard that the governor that antagonized the congressman apologized for being too aggressive.

Object relative clause, version 2: At the debate, a governor and two congressmen argued about the national election.
The governor antagonized one of the congressmen and then the other congressman antagonized the governor.
Mary: I heard that the congressman that the governor antagonized apologized for being too aggressive.
John: I am not sure about that. I heard that the congressman that antagonized the governor apologized for being too aggressive.

10. Subject relative clause, version 1: Earlier this month, an interpreter and two ambassadors planned a trip.
The interpreter contacted one of the ambassadors and then the other ambassador contacted the interpreter.
Mary: I heard that the ambassador that contacted the interpreter lived in Africa for many years.
John: I am not sure about that. I heard that the ambassador that the interpreter contacted lived in Africa for many years.

Subject relative clause, version 2: Earlier this month, an ambassador and two interpreters planned a trip.
The ambassador contacted one of the interpreters and then the other interpreter contacted the ambassador.
Mary: I heard that the interpreter that contacted the ambassador lived in Africa for many years.
John: I am not sure about that. I heard that the interpreter that the ambassador contacted lived in Africa for many years.

Object relative clause, version 1: Earlier this month, an interpreter and two ambassadors planned a trip.
The interpreter contacted one of the ambassadors and then the other ambassador contacted the interpreter.
Mary: I heard that the ambassador that the interpreter contacted lived in Africa for many years.
John: I am not sure about that. I heard that the ambassador that contacted the interpreter lived in Africa for many years.

Object relative clause, version 2: Earlier this month, an ambassador and two interpreters planned a trip.
The ambassador contacted one of the interpreters and then the other interpreter contacted the

ambassador.

Mary: I heard that the interpreter that the ambassador contacted lived in Africa for many years.

John: I am not sure about that. I heard that the interpreter that contacted the ambassador lived in Africa for many years.

11. Subject relative clause, version 1: After the meeting, an administrator and two managers examined the accounting books.

The administrator questioned one of the managers and then the other manager questioned the administrator.

Mary: I heard that the manager that questioned the administrator had a problem with the company.

John: I am not sure about that. I heard that the manager that the administrator questioned had a problem with the company.

Subject relative clause, version 2: After the meeting, a manager and two administrators examined the accounting books.

The manager questioned one of the administrators and then the other administrator questioned the manager.

Mary: I heard that the administrator that questioned the manager had a problem with the company.

John: I am not sure about that. I heard that the administrator that the manager questioned had a problem with the company.

Object relative clause, version 1: After the meeting, an administrator and two managers examined the accounting books.

The administrator questioned one of the managers and then the other manager questioned the administrator.

Mary: I heard that the manager that the administrator questioned had a problem with the company.

John: I am not sure about that. I heard that the manager that questioned the administrator had a problem with the company.

Object relative clause, version 2: After the meeting, a manager and two administrators examined the accounting books.

The manager questioned one of the administrators and then the other administrator questioned the manager.

Mary: I heard that the administrator that the manager questioned had a problem with the company.

John: I am not sure about that. I heard that the administrator that questioned the manager had a problem with the company.

12. Subject relative clause, version 1: During the lecture, an anthropologist and two historians discussed the article.

The anthropologist challenged one of the historians and then the other historian challenged the anthropologist.

Mary: I heard that the historian that challenged the anthropologist published a book on the same

topic.

John: I am not sure about that. I heard that the historian that the anthropologist challenged published a book on the same topic.

Subject relative clause, version 2: During the lecture, a historian and two anthropologists discussed the article.
The historian challenged one of the anthropologists and then the other anthropologist challenged the historian.
Mary: I heard that the anthropologist that challenged the historian published a book on the same topic.
John: I am not sure about that. I heard that the anthropologist that the historian challenged published a book on the same topic.

Object relative clause, version 1: During the lecture, an anthropologist and two historians discussed the article.
The anthropologist challenged one of the historians and then the other historian challenged the anthropologist.
Mary: I heard that the historian that the anthropologist challenged published a book on the same topic.
John: I am not sure about that. I heard that the historian that challenged the anthropologist published a book on the same topic.

Object relative clause, version 2: During the lecture, a historian and two anthropologists discussed the article.
The historian challenged one of the anthropologists and then the other anthropologist challenged the historian.
Mary: I heard that the anthropologist that the historian challenged published a book on the same topic.
John: I am not sure about that. I heard that the anthropologist that challenged the historian published a book on the same topic.

13. Subject relative clause, version 1: At the laboratory, a statistician and two engineers talked about the project.
The statistician impressed one of the engineers and then the other engineer impressed the statistician.
Mary: I heard that the engineer that impressed the statistician received a reward at a recent conference.
John: I am not sure about that. I heard that the engineer that the statistician impressed received a reward at a recent conference.

Subject relative clause, version 2: At the laboratory, an engineer and two statisticians talked about the project.
The engineer impressed one of the statisticians and then the other statistician impressed the engineer.
Mary: I heard that the statistician that impressed the engineer received a reward at a recent conference.

John: I am not sure about that. I heard that the statistician that the engineer impressed received a reward at a recent conference.

Object relative clause, version 1: At the laboratory, a statistician and two engineers talked about the project.
The statistician impressed one of the engineers and then the other engineer impressed the statistician.
Mary: I heard that the engineer that the statistician impressed received a reward at a recent conference.
John: I am not sure about that. I heard that the engineer that impressed the statistician received a reward at a recent conference.

Object relative clause, version 2: At the laboratory, an engineer and two statisticians talked about the project.
The engineer impressed one of the statisticians and then the other statistician impressed the engineer.
Mary: I heard that the statistician that the engineer impressed received a reward at a recent conference.
John: I am not sure about that. I heard that the statistician that impressed the engineer received a reward at a recent conference.

14. Subject relative clause, version 1: A week ago, a writer and two artists met at the museum.
The writer upset one of the artists and then the other artist upset the writer.
Mary: I heard that the artist that upset the writer left the museum in a bad mood.
John: I am not sure about that. I heard that the artist that the writer upset left the museum in a bad mood.

Subject relative clause, version 2: A week ago, an artist and two writers met at the museum.
The artist upset one of the writers and then the other writer upset the artist.
Mary: I heard that the writer that upset the artist left the museum in a bad mood.
John: I am not sure about that. I heard that the writer that the artist upset left the museum in a bad mood.

Object relative clause, version 1: A week ago, a writer and two artists met at the museum.
The writer upset one of the artists and then the other artist upset the writer.
Mary: I heard that the artist that the writer upset left the museum in a bad mood.
John: I am not sure about that. I heard that the artist that upset the writer left the museum in a bad mood.

Object relative clause, version 2: A week ago, an artist and two writers met at the museum.
The artist upset one of the writers and then the other writer upset the artist.
Mary: I heard that the writer that the artist upset left the museum in a bad mood.
John: I am not sure about that. I heard that the writer that upset the artist left the museum in a bad mood.

15. Subject relative clause, version 1: At the gym, a gymnast and two wrestlers were training for

an upcoming meet.
The gymnast observed one of the wrestlers and then the other wrestler observed the gymnast.
Mary: I heard that the wrestler that observed the gymnast attended the college on a scholarship.
John: I am not sure about that. I heard that the wrestler that the gymnast observed attended the college on a scholarship.

Subject relative clause, version 2: At the gym, a wrestler and two gymnasts were training for an upcoming meet.
The wrestler observed one of the gymnasts and then the other gymnast observed the wrestler.
Mary: I heard that the gymnast that observed the wrestler attended the college on a scholarship.
John: I am not sure about that. I heard that the gymnast that the wrestler observed attended the college on a scholarship.

Object relative clause, version 1: At the gym, a gymnast and two wrestlers were training for an upcoming meet.
The gymnast observed one of the wrestlers and then the other wrestler observed the gymnast.
Mary: I heard that the wrestler that the gymnast observed attended the college on a scholarship.
John: I am not sure about that. I heard that the wrestler that observed the gymnast attended the college on a scholarship.

Object relative clause, version 2: At the gym, a wrestler and two gymnasts were training for an upcoming meet.
The wrestler observed one of the gymnasts and then the other gymnast observed the wrestler.
Mary: I heard that the gymnast that the wrestler observed attended the college on a scholarship.
John: I am not sure about that. I heard that the gymnast that observed the wrestler attended the college on a scholarship.

16. Subject relative clause, version 1: Before the reunion, a chef and two caterers talked about food.
The chef frustrated one of the caterers and then the other caterer frustrated the chef.
Mary: I heard that the caterer that frustrated the chef was famous for his mushroom soup recipe.
John: I am not sure about that. I heard that the caterer that the chef frustrated was famous for his mushroom soup recipe.

Subject relative clause, version 2: Before the reunion, a caterer and two chefs talked about food.
The caterer frustrated one of the chefs and then the other chef frustrated the caterer.
Mary: I heard that the chef that frustrated the caterer was famous for his mushroom soup recipe.
John: I am not sure about that. I heard that the chef that the caterer frustrated was famous for his mushroom soup recipe.

Object relative clause, version 1: Before the reunion, a chef and two caterers talked about food.
The chef frustrated one of the caterers and then the other caterer frustrated the chef.
Mary: I heard that the caterer that the chef frustrated was famous for his mushroom soup recipe.
John: I am not sure about that. I heard that the caterer that frustrated the chef was famous for his mushroom soup recipe.

Object relative clause, version 2: Before the reunion, a caterer and two chefs talked about food. The caterer frustrated one of the chefs and then the other chef frustrated the caterer.
Mary: I heard that the chef that the caterer frustrated was famous for his mushroom soup recipe.
John: I am not sure about that. I heard that the chef that frustrated the caterer was famous for his mushroom soup recipe.

17. Subject relative clause, version 1: After the shoot, an editor and two models looked over the images.
The editor complimented one of the models and then the other model complimented the editor.
Mary: I heard that the model that complimented the editor worked for a leading fashion magazine.
John: I am not sure about that. I heard that the model that the editor complimented worked for a leading fashion magazine.

Subject relative clause, version 2: After the shoot, a model and two editors looked over the images.
The model complimented one of the editors and then the other editor complimented the model.
Mary: I heard that the editor that complimented the model worked for a leading fashion magazine.
John: I am not sure about that. I heard that the editor that the model complimented worked for a leading fashion magazine.

Object relative clause, version 1: After the shoot, an editor and two models looked over the images.
The editor complimented one of the models and then the other model complimented the editor.
Mary: I heard that the model that the editor complimented worked for a leading fashion magazine.
John: I am not sure about that. I heard that the model that complimented the editor worked for a leading fashion magazine.

Object relative clause, version 2: After the shoot, a model and two editors looked over the images.
The model complimented one of the editors and then the other editor complimented the model.
Mary: I heard that the editor that the model complimented worked for a leading fashion magazine.
John: I am not sure about that. I heard that the editor that complimented the model worked for a leading fashion magazine.

18. Subject relative clause, version 1: At the playground, a boy and two girls played in the sandbox.
The boy teased one of the girls and then the other girl teased the boy.
Mary: I heard that the girl that teased the boy was very spoiled and difficult.
John: I am not sure about that. I heard that the girl that the boy teased was very spoiled and difficult.

Subject relative clause, version 2: At the playground, a girl and two boys played in the sandbox.

The girl teased one of the boys and then the other boy teased the girl.
Mary: I heard that the boy that teased the girl was very spoiled and difficult.
John: I am not sure about that. I heard that the boy that the girl teased was very spoiled and difficult.

Object relative clause, version 1: At the playground, a boy and two girls played in the sandbox.
The boy teased one of the girls and then the other girl teased the boy.
Mary: I heard that the girl that the boy teased was very spoiled and difficult.
John: I am not sure about that. I heard that the girl that teased the boy was very spoiled and difficult.

Object relative clause, version 2: At the playground, a girl and two boys played in the sandbox.
The girl teased one of the boys and then the other boy teased the girl.
Mary: I heard that the boy that the girl teased was very spoiled and difficult.
John: I am not sure about that. I heard that the boy that teased the girl was very spoiled and difficult.

19. Subject relative clause, version 1: A week ago, a planner and two decorators began the preparations for the event.
The planner telephoned one of the decorators and then the other decorator telephoned the planner.
Mary: I heard that the decorator that telephoned the planner intended to leave the company.
John: I am not sure about that. I heard that the decorator that the planner telephoned intended to leave the company.

Subject relative clause, version 2: A week ago, a decorator and two planners began the preparations for the event.
The decorator telephoned one of the planners and then the other planner telephoned the decorator.
Mary: I heard that the planner that telephoned the decorator intended to leave the company.
John: I am not sure about that. I heard that the planner that the decorator telephoned intended to leave the company.

Object relative clause, version 1: A week ago, a planner and two decorators began the preparations for the event.
The planner telephoned one of the decorators and then the other decorator telephoned the planner.
Mary: I heard that the decorator that the planner telephoned intended to leave the company.
John: I am not sure about that. I heard that the decorator that telephoned the planner intended to leave the company.

Object relative clause, version 2: A week ago, a decorator and two planners began the preparations for the event.
The decorator telephoned one of the planners and then the other planner telephoned the decorator.
Mary: I heard that the planner that the decorator telephoned intended to leave the company.

John: I am not sure about that. I heard that the planner that telephoned the decorator intended to leave the company.

20. Subject relative clause, version 1: In the yard, a cat and two dogs got in a fight.
The cat bit one of the dogs and then the other dog bit the cat.
Mary: I heard that the dog that bit the cat was taken to a veterinarian.
John: I am not sure about that. I heard that the dog that the cat bit was taken to a veterinarian.

Subject relative clause, version 2: In the yard, a dog and two cats got in a fight.
The dog bit one of the cats and then the other cat bit the dog.
Mary: I heard that the cat that bit the dog was taken to a veterinarian.
John: I am not sure about that. I heard that the cat that the dog bit was taken to a veterinarian.

Object relative clause, version 1: In the yard, a cat and two dogs got in a fight.
The cat bit one of the dogs and then the other dog bit the cat.
Mary: I heard that the dog that the cat bit was taken to a veterinarian.
John: I am not sure about that. I heard that the dog that bit the cat was taken to a veterinarian.

Object relative clause, version 2: In the yard, a dog and two cats got in a fight.
The dog bit one of the cats and then the other cat bit the dog.
Mary: I heard that the cat that the dog bit was taken to a veterinarian.
John: I am not sure about that. I heard that the cat that bit the dog was taken to a veterinarian.

21. Subject relative clause, version 1: At the family reunion, a brother and two sisters greeted each other.
The brother kissed one of the sisters and then the other sister kissed the brother.
Mary: I heard that the sister that kissed the brother spent two years in the Peace Corps in South Africa.
John: I am not sure about that. I heard that the sister that the brother kissed spent two years in the Peace Corps in South Africa.

Subject relative clause, version 2: At the family reunion, a sister and two brothers greeted each other.
The sister kissed one of the brothers and then the other brother kissed the sister.
Mary: I heard that the brother that kissed the sister spent two years in the Peace Corps in South Africa.
John: I am not sure about that. I heard that the brother that the sister kissed spent two years in the Peace Corps in South Africa.

Object relative clause, version 1: At the family reunion, a brother and two sisters greeted each other.
The brother kissed one of the sisters and then the other sister kissed the brother.
Mary: I heard that the sister that the brother kissed spent two years in the Peace Corps in South Africa.
John: I am not sure about that. I heard that the sister that kissed the brother spent two years in the Peace Corps in South Africa.

Object relative clause, version 2: At the family reunion, a sister and two brothers greeted each other.
The sister kissed one of the brothers and then the other brother kissed the sister.
Mary: I heard that the brother that the sister kissed spent two years in the Peace Corps in South Africa.
John: I am not sure about that. I heard that the brother that kissed the sister spent two years in the Peace Corps in South Africa.

22. Subject relative clause, version 1: In the field, a rancher and two farmhands rounded up the animals.
The rancher yelled to one of the farmhands and then the other farmhand yelled to the rancher.
Mary: I heard that the farmhand that yelled to the rancher fell off a horse last month and broke his leg.
John: I am not sure about that. I heard that the farmhand that the rancher yelled to fell off a horse last month and broke his leg.

Subject relative clause, version 2: In the field, a farmhand and two ranchers rounded up the animals.
The farmhand yelled to one of the ranchers and then the other rancher yelled to the farmhand.
Mary: I heard that the rancher that yelled to the farmhand fell off a horse last month and broke his leg.
John: I am not sure about that. I heard that the rancher that the farmhand yelled to fell off a horse last month and broke his leg.

Object relative clause, version 1: In the field, a rancher and two farmhands rounded up the animals.
The rancher yelled to one of the farmhands and then the other farmhand yelled to the rancher.
Mary: I heard that the farmhand that the rancher yelled to fell off a horse last month and broke his leg.
John: I am not sure about that. I heard that the farmhand that yelled to the rancher fell off a horse last month and broke his leg.

Object relative clause, version 2: In the field, a farmhand and two ranchers rounded up the animals.
The farmhand yelled to one of the ranchers and then the other rancher yelled to the farmhand.
Mary: I heard that the rancher that the farmhand yelled to fell off a horse last month and broke his leg.
John: I am not sure about that. I heard that the rancher that yelled to the farmhand fell off a horse last month and broke his leg.

23. Subject relative clause, version 1: During the military parade, a sergeant and two colonels met after many years.
The sergeant embraced one of the colonels and then the other colonel embraced the sergeant.
Mary: I heard that the colonel that embraced the sergeant had received many medals and honors.
John: I am not sure about that. I heard that the colonel that the sergeant embraced had received

many medals and honors.

Subject relative clause, version 2: During the military parade, a colonel and two sergeants met after many years.
The colonel embraced one of the sergeants and then the other sergeant embraced the colonel.
Mary: I heard that the sergeant that embraced the colonel had received many medals and honors.
John: I am not sure about that. I heard that the sergeant that the colonel embraced had received many medals and honors.

Object relative clause, version 1: During the military parade, a sergeant and two colonels met after many years.
The sergeant embraced one of the colonels and then the other colonel embraced the sergeant.
Mary: I heard that the colonel that the sergeant embraced had received many medals and honors.
John: I am not sure about that. I heard that the colonel that embraced the sergeant had received many medals and honors.

Object relative clause, version 2: During the military parade, a colonel and two sergeants met after many years.
The colonel embraced one of the sergeants and then the other sergeant embraced the colonel.
Mary: I heard that the sergeant that the colonel embraced had received many medals and honors.
John: I am not sure about that. I heard that the sergeant that embraced the colonel had received many medals and honors.

24. Subject relative clause, version 1: At the hospital, a therapist and two nurses were remembering their escapades in high school.
The therapist embarrassed one of the nurses and then the other nurse embarrassed the therapist.
Mary: I heard that the nurse that embarrassed the therapist used to have an alcohol problem.
John: I am not sure about that. I heard that the nurse that the therapist embarrassed used to have an alcohol problem.

Subject relative clause, version 2: At the hospital, a nurse and two therapists were remembering their escapades in high school.
The nurse embarrassed one of the therapists and then the other therapist embarrassed the nurse.
Mary: I heard that the therapist that embarrassed the nurse used to have an alcohol problem.
John: I am not sure about that. I heard that the therapist that the nurse embarrassed used to have an alcohol problem.

Object relative clause, version 1: At the hospital, a therapist and two nurses were remembering their escapades in high school.
The therapist embarrassed one of the nurses and then the other nurse embarrassed the therapist.
Mary: I heard that the nurse that the therapist embarrassed used to have an alcohol problem.
John: I am not sure about that. I heard that the nurse that embarrassed the therapist used to have an alcohol problem.

Object relative clause, version 2: At the hospital, a nurse and two therapists were remembering their escapades in high school.

The nurse embarrassed one of the therapists and then the other therapist embarrassed the nurse.
Mary: I heard that the therapist that the nurse embarrassed used to have an alcohol problem.
John: I am not sure about that. I heard that the therapist that embarrassed the nurse used to have an alcohol problem.

Fillers
1. During the photo shoot the model, the photographer and the hairdresser discussed possible hairstyles.
The model liked her hair straight; however, the photographer and the hairdresser wanted her hair to be curly.
Mary: I heard that the model is stubborn and hard to work with.
John: Yeah, that's right, and I also heard that she gets paid ten thousand dollars for every photo shoot.

2. During the summer the lifeguard, the swimmer and the surfer became good friends.
The lifeguard began dating the swimmer and then the surfer envied the lifeguard.
Mary: I heard that the swimmer secretly liked the surfer.
John: Yeah, that's right, a few weeks later she broke up with the lifeguard and started dating the surfer.

3. In the movie theater Renee, Joanna and Renee's brother watched a movie together.
Renee's brother threw popcorn at Joanna and then Renee hit the boy for misbehaving.
Mary: I heard that Renee's brother was kicked out of the movie theater.
John: I'm not sure about that, I heard that Renee made her brother leave.

4. During the concert the singer, the dancer and the violinist performed together.
The violinist accompanied the singer and then the dancer joined in with a beautiful piece.
Mary: I heard that the singer is actually dating the dancer.
John: I think you're right, I heard they have been going out for two months.

5. At the basketball game a cheerleader, an injured player and the cheerleader's father sat together.
The father disliked the injured player because he was dating his daughter.
Mary: I heard that the cheerleader's father dislikes all of his daughter's boyfriends.
John: That's true, and he really hated her ex-boyfriend Jonathan.

6. At the grocery store the cashier, the manager and the customer had a dispute.
The customer complained about the service to the manager while the cashier defended himself.
Mary: I heard that the customer pushed the cashier.
John: Yeah, that's right, and I also heard that the cashier fell and broke his arm.

7. At the park the nanny, the dog walker and the pickpocket were sitting in the same area.
The dog walker punched the pickpocket after the pickpocket stole the nanny's wallet.
Mary: I heard that the pickpocket was finally arrested five days later.
John: Yeah, that's right, and I also heard that he has stolen over ten thousand dollars.

8. Since the snowboarding competition the coach, the snowboarder and the sports psychologist met weekly.
The sports psychologist gave a lecture and then the coach showed a video of the snowboarder's race.
Mary: I heard that the snowboarder hopes to win the gold medal at the Olympics this year.
John: That's right, and I heard he won a bronze medal at the Olympics last year.

9. After graduation the teacher, the student and the student's mother talked.
The student introduced his mother, and then his mother and the teacher shook hands.
Mary: I heard that the teacher asked the mother on a date.
John: Yeah, that's right, and I also heard that now they are engaged.

10. During the ski lesson the ski instructor, George and George's sister skied down the bunny slope together.
The ski instructor and the sister finished the course quickly; however, George fell down twice before the finish line.
Mary: I heard that it was George's first time skiing ever.
John: That's not true, I heard he had tried to ski once before but broke his arm.

11. At the circus the clown, the elephant and the acrobat were performing an act together.
The clown rode the elephant while the acrobat swung from a rope.
Mary: I heard that the acrobat used to be an Olympic gymnast.
John: That's right, and I heard that he used to be a street performer when he was young.

12. At the zoo Tim, John and Mike walked around together.
John stopped to buy some ice-cream while Mike and Tim looked at the lions.
Mary: I heard that Tim loves to go to the zoo.
John: I think you're right, I heard that he has a season pass.

13. Before prom Laurie, Ashley and Ashley's boyfriend were getting ready together.
Ashley's boyfriend got dressed while Ashley and Laurie did their make-up.
Mary: I heard that Laurie's dress cost over a thousand dollars.
John: Hmm, that's weird, I heard that it cost seven hundred dollars.

14. During the play the actress, the singer and the dog had a scene together.
The actress held the dog while the singer serenaded the actress.
Mary: I heard that the actress has become very popular.
John: Yeah, that's right, and I also heard that all of her performances are sold out for the next two months.

15. On the yacht the supermodel, the businessman and the businessman's wife drank red wine together.
The businessman flirted with the supermodel while the wife rolled her eyes.
Mary: I heard that the businessman dated five different models in the last month.
John: Yeah, that's right, and I also heard that his wife would divorce him if he wasn't so rich.

16. While the song was playing, the producer, the singer and the drummer listened carefully. The producer decided to change the lyrics after the drummer complained about the singer's performance.
Mary: I heard that the singer is no longer talking to the drummer.
John: Yeah, that's right, and I also heard that the band might break up.

17. Before the party the actress, the model and the rock star held a pre-party together. The actress talked to the rock star and then the model danced with the rock star.
Mary: I heard that the actress is now dating the rock star.
John: Yeah, that's right, and I also heard that the model just divorced her husband.

18. During the board meeting the CEO, the strategist and the trustee argued. The trustee and the strategist agreed upon a solution but the CEO preferred his own idea.
Mary: I heard that the trustee donated ten million dollars to the company.
John: That's not true, I heard that he only donated two million dollars.

19. Before the team dinner the goalie, the midfielder and the fullback made dessert together. The goalie and the fullback baked the cake and then the midfielder decorated the cake with icing.
Mary: I heard that the goalie wanted to make the cake for the soccer coach's birthday.
John: Yeah, that's right, the team really loves their coach.

20. After the test the teacher, the student and the student's mother had a conference. The teacher and the mother talked about the test while the student daydreamed.
Mary: I heard that the student is failing every single class this semester.
John: Yeah, that's right, and I also heard that his mother is thinking of hiring a tutor.

21. After the wedding the groom, the bride and the bridesmaid celebrated at the reception. The bride and groom danced while the bridesmaid looked on.
Mary: I heard that the groom cheated on the bride with the bridesmaid.
John: That's not true, the bride cheated on the groom with her ex-boyfriend.

22. Between classes the principal, the hall monitor and the janitor patrolled the hallways. The hall monitor asked the janitor to clean up after a food fight and then informed the principal of the incident.
Mary: I heard that it took the janitor two hours to clean up the mess.
John: That's true, and I also heard that the principal could not identify the students involved in the food fight.

23. During recess Emily, Jack and Dan played soccer together. Dan pushed Emily after Emily insulted Jack.
Mary: I heard that after she was pushed, Emily fell down and broke her nose.
John: Yeah, that's right, and I also heard that Dan was suspended for a week.

24. At the art gallery the customer, the artist and the gallery owner looked at a painting. The gallery owner asked the artist for more of his paintings after the customer bought six paintings.

Mary: I heard that the artist sells his paintings for over five thousand dollars each.
John: I think you're right, I heard that they are very expensive.

25. At the dance the chaperone and two students were having an argument.
The chaperone was yelling at the students because the students had been dancing inappropriately.
Mary: I heard that the chaperone yelled at most of the students at the dance.
John: I think you're right, she is very strict.

26. After the ballet performance Laura, her sister and her mother returned home and saw that they had been robbed.
The mother called the police because Laura and her sister noticed the broken window.
Mary: I heard that Laura's sister was missing her brand-new laptop.
John: Yeah, that's right, and I also heard that Laura was missing her pearl necklace.

27. At the aquarium the shark, the seal and the animal feeder were in the same tank together.
The shark attacked the seal while the animal feeder tried to distract it with shark food.
Mary: I heard that the shark bit the animal feeder on the leg.
John: I think you're right, I heard that the animal feeder was suing the aquarium.

28. On the train the fugitive, the policeman and the schoolgirl were in the same train car.
The policeman tackled the fugitive and then the schoolgirl screamed.
Mary: I heard that the policeman arrested the fugitive.
John: Hmm, that's weird, I heard that the fugitive managed to escape.

29. On the farm the farmer, the banker and the mayor had a meeting.
The farmer talked to the banker while the mayor surveyed the farm.
Mary: I heard that the mayor wants to buy the farm and turn it into a community park.
John: Yeah, that's right, and I also heard that the mayor wants to start construction this month.

30. During the filming the producer, the writer and the actress watched the opening scene.
The writer disliked the actress but the producer totally adored her.
Mary: I heard that the actress is being paid twenty million dollars to star in the movie.
John: I think you're right, she is really famous.

31. During the field trip a student and two teachers argued.
The student wanted to buy chocolate ice cream; however, the two teachers told him to wait until lunch.
Mary: I heard that the student really annoyed one of the teachers.
John: Yeah, I think so. I heard that the other teacher was more patient.
32. At orientation the university president, the chief of campus police and the dean spoke to the incoming freshmen.
The chief's speech was sobering but the other two speeches were more uplifting.
Mary: I heard that the university president is retiring this year.
John: Yeah, that's right, and I also heard that they have already found a replacement.

33. In the kitchen Sarah's parents, Sarah's sister and Sarah were eating dinner.

Sarah's sister talked to her mother while her father read the newspaper.
Mary I heard that her father was fired from his job last week.
John: Yeah, that's right, and I also heard that it was because he stole money from the company.

34. In the emergency room the physician assistant, the surgeon and the patient discussed several treatment options.
The physician assistant and the surgeon recommended the surgery; however, the patient remained skeptical.
Mary: I heard that the physician assistant and the surgeon have recently ended their romantic relationship.
John: You're right. The physician assistant has been dating another doctor since they broke up.

35. In the hotel lobby the concierge, the ballerina and the flutist were engaged in an argument.
The ballerina and the flutist stated they made room reservations; however, the concierge claimed that they did not.
Mary: I heard that the concierge was fired for being rude to the ballerina and the flutist.
John: You're right. The ballerina and the flutist complained to the manager about the horrible guest service.

36. In the seminar room the professor, the neurologist and the scientist gave a presentation on neurodegenerative diseases.
While presenting, the neurologist fainted and then the professor and the scientist called the ambulance.
Mary: I heard that the neurologist fainted because she has stage fright.
John: I don't think so. She often faints because she has low blood pressure.

APPENDIX C.

The experimental items for the Sentence Ratings experiment in Study 1 are given below, grouped by sub-experiment. Each condition for each item is shown. No extra fillers were included in this experiment.

Sub-experiment 1: Sentential Complements and Relative Clauses

1. SC/RC: The chance that the administrator who the nurse supervised had lost the medical reports bothered the intern a great deal.
RC/SC: The intern who the chance that the administrator had lost the medical reports bothered a great deal supervised the nurse.

2. SC/RC: The discovery that the babysitter who the parents were fond of had spanked the child disturbed the neighbors to some degree.
RC/SC: The neighbors who the discovery that the babysitter had spanked the child disturbed to some degree were fond of the parents.

3. SC/RC: The news that the defense lawyer who the district attorney influenced would defend the murderer affected the celebrity case quite strongly.
RC/SC: The celebrity case that the news that the defense lawyer would defend the murderer affected quite strongly influenced the district attorney.

4. SC/RC: The suggestion that the preacher who the voters adored will give a moving speech might unnerve the politician a bit.
RC/SC: The politician who the suggestion that the preacher will give a moving speech might unnerve a bit adored the voters.

5. SC/RC: The likelihood that the graduate student who the professor respected had plagiarized the article upset the secretary quite a lot.
RC/SC: The secretary who the likelihood that the graduate student had plagiarized the article upset quite a lot respected the professor.

6. SC/RC: The disclosure that the union members who the news stories criticized had agreed upon a contract encouraged the executive tremendously.
RC/SC: The executive who the disclosure that the union members had agreed upon a contract encouraged tremendously criticized the news stories.

7. SC/RC: The possibility that the actor who the director yelled at would try to reorganize the scene maddened the producers beyond words.
RC/SC: The producers who the possibility that the actor would try to reorganize the scene maddened beyond words yelled at the director.

8. SC/RC: The message that the janitor who the doorman chatted with forgot to empty the trash can infuriated the supervisor all night long.
RC/SC: The supervisor who the message that the janitor forgot to empty the trash can infuriated

all night long chatted with the doorman.

9. SC/RC: The notion that the bus-driver who the policeman screamed at might hit the pedestrian scared the passenger out of his wits.
RC/SC: The passenger who the notion that the bus-driver might hit the pedestrian scared out of his wits screamed at the policeman.

10. SC/RC: The knowledge that the athlete who the doctor helped had scored the winning goal inspired the little boy to great achievements.
RC/SC: The little boy who the knowledge that the athlete had scored the winning goal inspired to great achievements helped the doctor.

11. SC/RC: The report that the fans who the hooligans ridiculed cheered for the home team excited the star player a huge amount.
RC/SC: The star player who the report that the fans cheered for the home team excited a huge amount ridiculed the hooligans.

12. SC/RC: The myth that the guard who the fair maiden married had killed the dragon impressed the knight immensely.
RC/SC: The knight who the myth that the guard had killed the dragon impressed immensely married the fair maiden.

13. SC/RC: The inference that the mindless lackey who the hero fought with would bungle the job exasperated the evil genius tremendously.
RC/SC: The evil genius who the inference that the mindless lackey would bungle the job exasperated tremendously fought with the hero.

14. SC/RC: The accusation that the cheerleader who the quarterback dated had snubbed the nerds aggravated the class president a great deal.
RC/SC: The class president who the accusation that the cheerleader had snubbed the nerds aggravated a great deal dated the quarterback.

15. SC/RC: The implication that the dictator who the diplomat insulted had closed the embassy embarrassed the prime minister in parliament.
RC/SC: The prime minister who the implication that the dictator had closed the embassy embarrassed in parliament insulted the diplomat.

16. SC/RC: The threat that the beggar boy who the genie served would speak the magic words terrified the king immensely.
RC/SC: The king who the threat that the beggar boy would speak the magic words terrified immensely served the genie.

17. SC/RC: The hunch that the serial killer who the waitress trusted might hide the body frightened the FBI agent into action.
RC/SC: The FBI agent who the hunch that the serial killer might hide the body frightened into action trusted the waitress.

18. SC/RC: The confession that the housewife who the whole town gossiped about bought cosmetics distracted the Avon lady from her work.
RC/SC: The Avon lady who the confession that the housewife bought cosmetics distracted from her work gossiped about the whole town.

Sub-experiment 2: Branching vs. Nested Relative Clauses

1. Nested: The pilot who the navigator who the gunner warned alerted broke radio silence.
Branching: The gunner warned the navigator who alerted the pilot who broke radio silence.

2. Nested: The beatnik who the writer who everyone seemed to have met fascinated wore women's clothing.
Branching: Everyone seemed to have met the writer who fascinated the beatnik who wore women's clothing.

3. Nested: The tourist who the rug-merchant who the local children teased mystified took dozens of pictures.
Branching: The local children teased the rug-merchant who mystified the tourist who took dozens of pictures.

4. Nested: The terrorist who the spy who the assassin shot at provoked planted a bomb.
Branching: The assassin shot at the spy who provoked the terrorist who planted a bomb.

5. Nested: The small town girl who the mysterious visitor who strangers never talk to intrigued eats live frogs.
Branching: Strangers never talk to the mysterious visitor who intrigued the small town girl who eats live frogs.

6. Nested: The drummer who the singer who groupies love disgusted spat on the crowd.
Branching: Groupies love the singer who disgusted the drummer who spat on the crowd.

7. Nested: The medic who the victim who the doctor yelled to startled dropped the bandage.
Branching: The doctor yelled to the victim who startled the medic who dropped the bandage.

8. Nested: The frightened child who the old woman who the rescue worker looked for comforted survived the crash.
Branching: The rescue worker looked for the old woman who comforted the frightened child who survived the crash.

9. Nested: The leading man who the comedian who the starlet adored entertained forgot the crucial line.
Branching: The starlet adored the comedian who entertained the leading man who forgot the crucial line.

10. Nested: The violinist who the conductor who the corporate sponsors flattered disappointed

skipped the rehearsal.
Branching: The corporate sponsors flattered the conductor who disappointed the violinist who skipped the rehearsal.

11. Nested: The student who the teaching assistant who the professor met with tormented disrupted the class.
Branching: The professor met with the teaching assistant who tormented the student who disrupted the class.

12. Nested: The suspected mobster who the ambitious attorney who the media attacked investigated had been to jail.
Branching: The media attacked the ambitious attorney who investigated the suspected mobster who had been to jail.

13. Nested: The burglar who the paranoid neighbor who the returning couple startled alarmed dropped the TV.
Branching: The returning couple startled the paranoid neighbor who alarmed the burglar who dropped the TV.

14. Nested: The teenager who the bossy contractor who the architect hired irritated ruined the lawn.
Branching: The architect hired the bossy contractor who irritated the teenager who ruined the lawn.

15. Nested: The child who the divorcing parents who the psychologist talked to worried lost a lot of sleep.
Branching: The psychologist talked to the divorcing parents who worried the child who lost a lot of sleep.

16. Nested: The settlers who the tribe who the Sun-god protected frightened were not taken seriously.
Branching: The Sun-god protected the tribe who frightened the settlers who were not taken seriously.

17. Nested: The parishioners who the actress who the preacher was sleeping with scandalized were kept quiet.
Branching: The preacher was sleeping with the actress who scandalized the parishioners who were kept quiet.

18. Nested: The invasion that the children who the aliens kidnapped knew about fell to pieces.
Branching: The aliens kidnapped the children who knew about the invasion that fell to pieces.

Sub-experiment 3: Branching vs. Nesting Sentential Complements

1. Nested: That whether the seats were gray concerned the engineer annoyed the auto designer.
Branching-Control: It annoyed the auto designer that it concerned the engineer whether the seats

were gray.
Branching-Violation: It annoyed the auto designer that it concerned the engineer whether the seats was gray.

2. Nested: Whether that the law passed frustrated the senator distracted the aide.
Branching-Control: It distracted the aide whether it frustrated the senator that the law passed.
Branching-Violation: It distracted the aide whether it frustrating the senator that the law passed.

3. Nested: That whether the painting was genuine nagged at the authenticator surprised the curator.
Branching-Control: It surprised the curator that it nagged at the authenticator whether the painting was genuine.
Branching-Violation: It surprised the curator that it nagging at the authenticator whether the painting were genuine.

4. Nested: Whether that the evidence was thrown out upset the attorney worried the defendant.
Branching-Control: It worried the defendant whether it upset the attorney that the evidence was thrown out.
Branching-Violation: It worry the defendant whether it upset the attorney that the evidence was thrown out.

5. Nested: That whether the child was safe concerned the kidnapper comforted the negotiator.
Branching-Control: It comforted the negotiator that it concerned the kidnapper whether the child was safe.
Branching-Violation: It comforted the negotiator that it concerning the kidnapper whether the child were safe.

6. Nested: Whether that the temple had fallen frightened the worshippers weighed on the storm goddess.
Branching-Control: It weighed on the storm goddess whether it frightened the worshippers that the temple had fallen.
Branching-Violation: It weighed on the storm goddess whether it frightened the worshippers that the temple falling.

7. Nested: That whether the recruits were prepared mattered to the sergeant pleased the general.
Branching-Control: It pleased the general that it mattered to the sergeant whether the recruits were prepared.
Branching-Violation: It pleased the general that it mattering to the sergeant whether the recruits was prepared.

8. Nested: Whether that the customer asked for fries bothered the waiter worried the manager.
Branching-Control: It worried the manager whether it bothered the waiter that the customer asked for fries.
Branching-Violation: It worry the manager whether it bothered the waiter that the customer asking for fries.

9. Nested: That whether the affair was revealed to the public scared the candidate amused the reporter.
Branching-Control: It amused the reporter that it scared the candidate whether the affair was revealed to the public.
Branching-Violation: It amuse the reporter that it scaring the candidate whether the affair was revealed to the public.

10. Nested: Whether that the tour guide got lost annoyed the tourists frustrated the travel agent.
Branching-Control: It frustrated the travel agent whether it annoyed the tourists that the tour guide got lost.
Branching-Violation: It frustrated the travel agent whether it annoying the tourists that the tour guide got lost.

11. Nested: That whether the theorem was provable perplexed the student stunned the instructor.
Branching-Control: It stunned the instructor that it perplexed the student whether the theorem was provable.
Branching-Violation: It stun the instructor that it perplexing the student whether the theorem was provable.

12. Nested: Whether that the rock band changed styles would alienate the fans frightened the band manager.
Branching-Control: It frightened the band manager whether it would alienate the fans that the rock band changed styles.
Branching-Violation: It frightened the band manager whether it would alienate the fans that the rock band changing styles.

13. Nested: That whether the diplomat agreed to the terms worried the ambassador bothered the president.
Branching-Control: It bothered the president that it worried the ambassador whether the diplomat agreed to the terms.
Branching-Violation: It bother the president that it worried the ambassador whether the diplomat agreeing to the terms.

14. Nested: Whether that the tumor had shrunk relieved the patient weighed on the doctor.
Branching-Control: It weighed on the doctor whether it relieved the patient that the tumor had shrunk.
Branching-Violation: It weighing on the doctor whether it relieving the patient that the tumor had shrunk.

15. Nested: That whether the king was safe concerned the guards comforted the queen.
Branching-Control: It comforted the queen that it concerned the guards whether the king was safe.
Branching-Violation: It comforted the queen that it concerning the guards whether the king was safe.

16. Nested: Whether that the news story was a fake would outrage the readers terrified the editor.

Branching-Control: It terrified the editor whether it would outrage the readers that the news story was a fake.
Branching-Violation: It terrify the editor whether it would outrage the readers that the news story was a fake.

17. Nested: That whether the play would work concerned the coach discouraged the quarterback.
Branching-Control: It discouraged the quarterback that it concerned the coach whether the play would work.
Branching-Violation: It discourage the quarterback that it concerned the coach whether the play would work.

18. Nested: Whether that the engine was failing would upset the crew worried the pilot.
Branching-Control: It worried the pilot whether it would upset the crew that the engine was failing.
Branching-Violation: It worry the pilot whether it would upset the crew that the engine are falling.

19. Nested: That whether the principal was getting married excited the students entertained the teacher's aide.
Branching-Control: It entertained the teacher's aide that it excited the students whether the principal was getting married.
Branching-Violation: It entertain the teacher's aide that it excited the students whether the principal was getting married.
20. Nested: Whether that the police car was dented would offend the sheriff perplexed the police officer.
Branching-Control: It perplexed the police officer whether it would offend the sheriff that the police car was dented.
Branching-Violation: It perplexed the police officer whether it would offending the sheriff that the police car was dented.

21. Nested: That whether the groupies got in a fist fight infuriated the rock star impressed the manager.
Branching-Control: It impressed the manager that it infuriated the rock star whether the groupies got in a fist fight.
Branching-Violation: It impressed the manager that it infuriated the rock star whether the groupies getting in a fist fight.

22. Nested: Whether that the teacher let class out early would disturb the principal puzzled the freshman.
Branching-Control: It puzzled the freshman whether it would disturb the principal that the teacher let class out early.
Branching-Violation: It puzzle the freshman whether it would disturbing the principal that the teacher let class out early.

23. Nested: That whether the table was properly carved bored the carpenter saddened the apprentice.

Branching-Control: It saddened the apprentice that it bored the carpenter whether the table was properly carved.
Branching-Violation: It saddened the apprentice that it boring the carpenter whether the table are properly carved.

24. Nested: Whether that the summer book club was starting early inspired the children fascinated the librarian.
Branching-Control: It fascinated the librarian whether it inspired the children that the summer book club was starting early.
Branching-Violation: It fascinating the librarian whether it inspired the children that the summer book club are starting early.

25. Nested: That whether the ethics seminar would be cancelled distressed the employees relieved the business owner.
Branching-Control: It relieved the business owner that it distressed the employees whether the ethics seminar would be cancelled.
Branching-Violation: It relieve the business owner that it distressed the employees whether the ethics seminar would be cancelled.

26. Nested: Whether that the embezzlement had taken place humiliated the company haunted the CEO.
Branching-Control: It haunted the CEO whether it humiliated the company that the embezzlement had taken place.
Branching-Violation: It haunted the CEO whether it humiliating the company that the embezzlement had taken place.

27. Nested: That whether the toddler would behave interested the parents delighted the nanny.
Branching-Control: It delighted the nanny that it interested the parents whether the toddler would behave.
Branching-Violation: It delighted the nanny that it interested the parents whether the toddler would behaving.

Sub-experiment 4: NP-Island Violations

1. NP-island violation: The reporter asked who the CEO spread the rumor that the stock holders hated.
Control: The reporter asked if the CEO spread the rumor that the stock holders hated the new president.

2. NP-island violation: The secretary checked what the administrator heard the complaint that the office needed.
Control: The secretary checked if the administrator heard the complaint that the office needed a copier.

3. NP-island violation: The customer wondered what the clerk was aware of the possibility that the store had run out of.

Control: The customer wondered if the clerk was aware of the possibility that the store had run out of razors.

4. NP-island violation: The mother asked what her husband was informed of the accusation that their son was taking.
Control: The mother asked if her husband was informed of the accusation that their son was taking drugs.

5. NP-island violation: The doctor wondered what the patient had seen the report that his prescription was.
Control: The doctor wondered if the patient had seen the report that his prescription was a placebo.

6. NP-island violation: The governor checked who the investigator knew about the allegation that the political group bribed.
Control: The governor checked if the investigator knew about the allegation that the political group bribed the senator.

7. NP-island violation: The assistant wondered what the magician sensed the chance that the audience saw.
Control: The assistant wondered if the magician sensed the chance that the audience saw the trap door..
8. NP-island violation: The negotiator asked who the police had considered the threat that the terrorist would assassinate.
Control: The negotiator asked if the police had considered the threat that the terrorist would assassinate the delegate.

9. NP-island violation: The attorney asked who the witness was aware of the claim that the suspect had threatened.
Control: The attorney asked if the witness was aware of the claim that the suspect had threatened the victim.

10. NP-island violation: The villagers wondered what the children knew the story that the fortune teller practiced.
Control: The villagers wondered if the children knew the story that the fortune teller practiced witchcraft.

11. NP-island violation: The editor checked who the cartoonist had heard the news that his cartoon upset.
Control: The editor checked if the cartoonist had heard the news that his cartoon upset some readers.

12. NP-island violation: The cheerleader asked who her teammate started the rumor that the quarterback was dating.
Control: The cheerleader asked if her teammate started the rumor that the quarterback was dating the athlete.

13. NP-island violation: The sea goddess wondered what her subjects sensed the possibility that the storm would destroy.
Control: The sea goddess wondered if her subjects sensed the possibility that the storm would destroy the temple.

14. NP-island violation: The mayor asked what the farmers agreed with the claim that locusts decimated.
Control: The mayor asked if the farmers agreed with the claim that locusts decimated the crops.

15. NP-island violation: The patron checked what the painter believed the prediction that his mentor would paint.
Control: The patron checked if the painter believed the prediction that his mentor would paint a self-portrait.

16. NP-island violation: The tour guide asked who the tourists heard the myth that the king murdered.
Control: The tour guide asked if the tourists heard the myth that the king murdered his predecessor.

17. NP-island violation: The detective wondered who the suspect was aware of the report that her husband found.
Control: The detective wondered if the suspect was aware of the report that her husband found the murder weapon.

18. NP-island violation: The writer checked what his editor made the implication that the manuscript required.
Control: The writer checked if his editor made the implication that the manuscript required rewriting.

Sub-experiment 5: Control experiments

1. Agreement violation: The chef are baking the bread and the waiter are seating the customer.
Plausibility violation: The bread is baking the chef and the customer is seating the waiter.
Control: The chef is baking the bread and the waiter is seating the customer.

2. Agreement violation: The cowboys is roping the bulls and the clowns is juggling the batons.
Plausibility violation: The bulls are roping the cowboys and the batons are juggling the clowns.
Control: The cowboys are roping the bulls and the clowns are juggling the batons.

3. Agreement violation: The farmers is milking the cows and the shepherds is corralling the sheep.
Plausibility violation: The cows are milking the farmers and the sheep are corralling the shepherds.
Control: The farmers are milking the cows and the shepherds are corralling the sheep.

4. Agreement violation: The conductor are leading the orchestra and the violinists is playing the instruments.
Plausibility violation: The orchestra is leading the conductor and the instruments are playing the violinists.
Control: The conductor is leading the orchestra and the violinists are playing the instruments.

5. Agreement violation: The students is learning the rules and the teacher are grading the tests.
Plausibility violation: The rules are learning the students and the tests are grading the teacher.
Control: The students are learning the rules and the teacher is grading the tests.

6. Agreement violation: The boy scouts is gathering the wood and the ranger are lighting the fire.
Plausibility violation: The wood is gathering the boy scouts and the fire is lighting the ranger.
Control: The boy scouts are gathering the wood and the ranger is lighting the fire.

7. Agreement violation: The dog are chewing the bone and the cat are chasing the yarn.
Plausibility violation: The bone is chewing the dog and the yarn is chasing the cat.
Control: The dog is chewing the bone and the cat is chasing the yarn.

8. Agreement violation: The gardener are watering the plant and the man are painting the house.
Plausibility violation: The plant is watering the gardener and the house is painting the man.
Control: The gardener is watering the plant and the man is painting the house.

9. Agreement violation: The artists is sketching the landscape and the potters is shaping the clay.
Plausibility violation: The landscape is sketching the artists and the clay is shaping the potters.
Control: The artists are sketching the landscape and the potters are shaping the clay.

10. Agreement violation: The bankers is reading the newspaper and the accountants are calculating the sums.
Plausibility violation: The newspaper is reading the bankers and the accountants are calculating the sums.
Control: The bankers are reading the newspaper and the accountants are calculating the sums.

11. Agreement violation: The playwright are writing the scene and the actors are reciting the lines.
Plausibility violation: The scene is writing the playwright and the actors are reciting the lines.
Control: The playwright is writing the scene and the actors are reciting the lines.

12. Agreement violation: The bees is stinging the children and the wasp is building the nest.
Plausibility violation: The children are stinging the bees and the wasp is building the nest.
Control: The bees are stinging the children and the wasp is building the nest.

13. Agreement violation: The mother are feeding the baby and the father is mowing the grass.
Plausibility violation: The baby is feeding the mother and the father is mowing the grass.
Control: The mother is feeding the baby and the father is mowing the grass.

14. Agreement violation: The cyclist are riding the bicycle and the kayaker is paddling the boat.

Plausibility violation: The bicycle is riding the cyclist and the kayaker is paddling the boat.
Control: The cyclist is riding the bicycle and the kayaker is paddling the boat.

15. Agreement violation: The judges is banging the gavels and the lawyers are unzipping the briefcases.
Plausibility violation: The gavels are banging the judges and the lawyers are unzipping the briefcases.
Control: The judges are banging the gavels and the lawyers are unzipping the briefcases.

16. Agreement violation: The cleaner are washing the windows and the electricians are repairing the televisions.
Plausibility violation: The windows are washing the cleaner and the electricians are repairing the televisions.
Control: The cleaner is washing the windows and the electricians are repairing the televisions.

17. Agreement violation: The cook are grilling the meat and the guests are eating the hamburgers.
Plausibility violation: The meat is grilling the cook and the guests are eating the hamburgers.
Control: The cook is grilling the meat and the guests are eating the hamburgers.

18. Agreement violation: The dentists is drilling the teeth and the nurse is giving the injection.
Plausibility violation: The teeth is drilling the dentists and the nurse is giving the injection.
Control: The dentists are drilling the teeth and the nurse is giving the injection.

19. Agreement violation: The pilot is flying the plane and the flight attendant are pushing the cart.
Plausibility violation: The pilot is flying the plane and the cart is pushing the flight attendant.
Control: The pilot is flying the plane and the flight attendant is pushing the cart.

20. Agreement violation: The chickens are pecking the ground and the pigs is sniffing the mud.
Plausibility violation: The chickens are pecking the ground and the mud is sniffing the pigs.
Control: The chickens are pecking the ground and the pigs are sniffing the mud.

21. Agreement violation: The technicians are solving the problems and the engineers is assessing the situation.
Plausibility violation: The technicians are solving the technicians and the situation is assessing the engineers.
Control: The technicians are solving the problems and the engineers are assessing the situation.

22. Agreement violation: The orphan is shining the shoes and the conmen is fooling the audience.
Plausibility violation: The orphan is shining the shoes and the audience is fooling the conmen.
Control: The orphan is shining the shoes and the conmen are fooling the audience.

23. Agreement violation: The journalists are reporting the story and the editor are correcting the article.

Plausibility violation: The journalists are reporting the story and the article is correcting the editor.
Control: The journalists are reporting the story and the editor is correcting the article.

24. Agreement violation: The carpenters are sanding the tables and the plumber are unclogging the pipes.
Plausibility violation: The carpenters are sanding the tables and the pipes are unclogging the plumber.
Control: The carpenters are sanding the tables and the plumber is unclogging the pipes.

25. Agreement violation: The robber is stealing the money and the vandal are breaking the window.
Plausibility violation: The robber is stealing the money and the window is breaking the vandal.
Control: The robber is stealing the money and the vandal is breaking the window.

26. Agreement violation: The explorer is sailing the ocean and the navigator are interpreting the maps.
Plausibility violation: The explorer is sailing the ocean and the maps are interpreting the navigator.
Control: The explorer is sailing the ocean and the navigator is interpreting the maps.

27. Agreement violation: The earl is lifting the teacups and the duchesses is admiring the saucers.
Plausibility violation: The earl is lifting the teacups and the saucers are admiring the duchesses.
Control: The earl is lifting the teacups and the duchesses are admiring the saucers.

APPENDIX D.

The following sentences were used in the Reading Span task. Sentences were presented in sets of 2-6, and grouped below accordingly. The words to be remembered are underlined.

Sets of 2

The weather was very unpredictable that summer so no one made plans too far in <u>advance</u>.
The entire town arrived to see the appearance of the controversial political <u>candidate</u>.

According to the results from the survey, Robert Redford is the most liked Hollywood <u>star</u>.
The lieutenant sat beside the man with the walkie-talkie and stared at the muddy <u>ground</u>.

Jane's relatives had decided that her gentleman friend was not one of high <u>status</u>.
After passing all the exams, the class celebrated for an entire week without <u>resting</u>.

In a flash of fatigue and fantasy, he saw a fat Indian sitting beside a <u>campfire</u>.
When at last his eyes opened, there was no gleam of triumph, no shade of <u>anger</u>.

I was so surprised at this unaccountable apparition, that I was speechless for a <u>while</u>.
The words of human love have been used by the saints to describe their vision of <u>God</u>.

Sets of 3

It was shortly after this that an unusual pressure of business called me into <u>town</u>.
I imagine that you have a shrewd suspicion of the object of my early <u>visit</u>.
I turned my memories over at random like pictures in a photograph <u>album</u>.

Filled with these dreary forebodings, I fearfully opened the heavy wooden <u>door</u>.
The taxi turned up Michigan Avenue, where they had a clear view of the <u>lake</u>.
I'm not certain what went wrong but I think it was my cruel and bad <u>temper</u>.

He pursued this theme, still pretending to seek for information to quiet his own <u>doubts</u>.
In order to postpone the business trip, he canceled his engagements for the <u>week</u>.
The courses are designed as much for professional engineers as for amateur <u>enthusiasts</u>.

The brilliant trial attorney dazzled the jury with his astute knowledge of the <u>case</u>.
I found the keynote speaker incredibly boring, inarticulate and not well <u>read</u>.
The incorrigible child was punished brutally for his lack of respect for his <u>elders</u>.

In a moment of complete spontaneity, she developed a thesis for her <u>paper</u>.
Circumstantial evidence indicated that there was a conspiracy to eliminate <u>him</u>.
Her mother nagged incessantly about her lack of concern for the welfare of the <u>children</u>.

Sets of 4

At the conclusion of the musicians' performance, the enthusiastic crowd <u>applauded</u>.

Without any hesitation, he plunged into the difficult mathematics assignment <u>blindly</u>.
The devastating effects of the flood were not fully realized until months <u>later</u>.
When I got to the big tobacco field I saw that it had not suffered <u>much</u>.

The products of digital electronics will play an important role in your <u>future</u>.
The old lady talked to her new neighbor on her weekly walks from <u>church</u>.
One problem with this explanation is that there appears to be no defense against <u>cheating</u>.
To determine the effects of the medication, the doctor hospitalized his <u>patient</u>.

Sometimes the scapegoat is an outsider who has been taken into the <u>community</u>.
The entire construction crew decided to lengthen their work day in order to have <u>lunch</u>.
Without tension there could be no balance either in nature or in mechanical <u>design</u>.
Two or three substantial pieces of wood smoldered on the hearth, for the night was <u>cold</u>.

A small oil lamp burned on the floor and two men crouched against the wall, watching <u>me</u>.
I wish there existed someone to whom I could say that I felt very <u>sorry</u>.
Here, as elsewhere, the empirical patterns are important and abundantly <u>documented</u>.
The rain and howling wind kept beating against the rattling window <u>panes</u>.

They attended the theater habitually except for circumstances beyond their <u>control</u>.
He covered his heart with both hands to keep anyone from hearing the noise it <u>made</u>.
I will not shock my readers with a description of the cool-blooded butchery that <u>followed</u>.
The smokers were asked to refrain from their habit until the end of the <u>production</u>.

Sets of 5

Despite the unusually cold weather, the campers continued their canoe <u>trip</u>.
I should not be able to make any one understand how exciting it all <u>was</u>.
There was still more than an hour before breakfast, and the house was silent and <u>asleep</u>.
I imagined that he had been thinking things over while the secretary was with <u>us</u>.
The young business executive was determined to develop his housing projects within the <u>year</u>.

The stories all deal with a middle-aged protagonist who attempts to withdraw from <u>society</u>.
The lumbermen worked long hours in order to obtain the necessary amount of <u>wood</u>.
In comparison to his earlier works, the musician had developed a unique enthralling <u>style</u>.
The boisterous laughter of the children was disturbing to the aged in the <u>building</u>.
The sound of an approaching train woke him, and he started to his <u>feet</u>.

These splendid melancholy eyes were turned upon me from the mirror with a haughty <u>stare</u>.
The intervals of silence grew progressively longer; the delays became very <u>maddening</u>.
He sometimes considered suicide but the thought was too oppressive to remain in his <u>mind</u>.
He had patronized her when she was a schoolgirl and teased her when she was a <u>student</u>.
Slicing it out carefully with his knife, he folded it without creasing the <u>face</u>.

He tolerated another intrusion and thought himself a paragon of patience for doing <u>so</u>.
He leant on the parapet of the bridge and the two policemen watched him from a <u>distance</u>.

He laughed sarcastically and looked as if he could have poisoned me for my <u>errors</u>.
On the desk where she wrote her letters was a clutter of objects coated in <u>dust</u>.
With shocked amazement and appalled fascination Marion looked at the <u>pictures</u>.

And now that a man had died some unimaginably different state of affairs must come to <u>be</u>.
The girl hesitated for a moment to taste the onions because her husband hated the <u>smell</u>.
All students that passed the test were exempt from any further seminars that <u>semester</u>.
The reader may suppose that I had other motives, besides the desire to escape the <u>law</u>.
What would come after this day would be inconceivably different, would be real <u>life</u>.

Sets of 6

It was your belief in the significance of my suffering that kept me <u>going</u>.
To do so in directions that are adaptive for mankind would be a realistic <u>objective</u>.
His imagination had so abstracted him that his name was called twice before he <u>answered</u>.
Sometimes I get so tired of trying to convince him that I love him and shall for <u>ever</u>.
It is possible, of course, that life did not arise on the earth at <u>all</u>.
He stood there at the edge of the crowd while they were singing, and he looked <u>bitter</u>.

John became annoyed with Karen's bad habits of biting her nails and chewing <u>gum</u>.
When in trouble, children naturally hope for a miraculous intervention by a <u>superhuman</u>.
After all he had not gone far, and some of his walking had been <u>circular</u>.
The poor lady was thoroughly persuaded that she was not long to survive this <u>vision</u>.
The announcement of it would resound throughout the world, penetrate to the remotest <u>land</u>.
He listened carefully because he had the weird impression that he knew the <u>voices</u>.

There are days when the city where I live wakes in the morning with a strange <u>look</u>.
He had an odd elongated skull which sat on his shoulders like a pear on a <u>dish</u>.
Due to his gross inadequacies, his position as director was terminated <u>abruptly</u>.
We boys wanted to warn them, but we backed down when it came to the <u>pinch</u>.
He stuffed his denim jacket into his pants and fastened the stiff, new snaps <u>securely</u>.
The basic characteristic of the heroes in the preceding stories is their <u>sensitivity</u>.

APPENDIX E.

The following list of forty commonly misspelled words was given to participants for the spelling test. The words were randomized, but the randomization was kept constant for each participant.

absence
beneficial
cemetery
collectible
conscientious
definitely
discipline
embarrass
exceed
excellent
exhilarate
fascinating
guarantee
height
hierarchy
interference
jealousy
jewelry
leisure
license
maintenance
minuscule
mischievous
neighbor
occurrence
parallel
pastime
personnel
playwright
privilege
questionnaire
receipt
rhythm
rough
schedule
separate (read as an adjective)
sergeant
supersede
undoubtedly
vacuum

APPENDIX F.

The following set of 24 experimental items comprises the clefts materials used in Study 1, Experiment 2. These materials were all presented in null contexts, and no filler materials were used in this experiment.

1. Object cleft, version 1: It was the producer that the novelist mocked for wearing a tie.
Object cleft, version 2: It was the novelist that the producer mocked for wearing a tie.
Subject cleft, version 1: It was the producer that mocked the novelist for wearing a tie.
Subject cleft, version 2: It was the novelist that mocked the producer for wearing a tie.

2. Object cleft, version 1: It was the juggler that the mime entertained with a new joke.
Object cleft, version 2: It was the mime that the juggler entertained with a new joke.
Subject cleft, version 1: It was the juggler that entertained the mime with a new joke.
Subject cleft, version 2: It was the mime that entertained the juggler with a new joke.

3. Object cleft, version 1: It was the queen that the duke ignored for most of the night.
Object cleft, version 2: It was the duke that the queen ignored for most of the night.
Subject cleft, version 1: It was the queen that ignored the duke for most of the night.
Subject cleft, version 2: It was the duke that ignored the queen for most of the night.

4. Object cleft, version 1: It was the screenwriter that the director irritated by losing the camera battery.
Object cleft, version 2: It was the director that the screenwriter irritated by losing the camera battery.
Subject cleft, version 1: It was the screenwriter that irritated the director by losing the camera battery.
Subject cleft, version 2: It was the director that irritated the screenwriter by losing the camera battery.

5. Object cleft, version 1: It was the coach that the trainer harassed about the team's performance.
Object cleft, version 2: It was the trainer that the coach harassed about the team's performance.
Subject cleft, version 1: It was the coach that harassed the trainer about the team's performance.
Subject cleft, version 2: It was the trainer that harassed the coach about the team's performance.

6. Object cleft, version 1: It was the linebacker that the referee assisted in getting the player on his feet.
Object cleft, version 2: It was the referee that the linebacker assisted in getting the player on his feet.
Subject cleft, version 1: It was the linebacker that assisted the referee in getting the player on his feet.
Subject cleft, version 2: It was the referee that assisted the linebacker in getting the player on his feet.

7. Object cleft, version 1: It was the pastry-chef that the cook deceived about paying for ingredients.

Object cleft, version 2: It was the cook that the pastry-chef deceived about paying for ingredients.
Subject cleft, version 1: It was the pastry-chef that deceived the cook about paying for ingredients.
Subject cleft, version 2: It was the cook that deceived the pastry-chef about paying for ingredients.

8. Object cleft, version 1: It was the brewer that the baker praised for an old recipe from Italy.
Object cleft, version 2: It was the baker that the brewer praised for an old recipe from Italy.
Subject cleft, version 1: It was the brewer that praised the baker for an old recipe from Italy.
Subject cleft, version 2: It was the baker that praised the brewer for an old recipe from Italy.

9. Object cleft, version 1: It was the orderly that the oncologist misled about the location of the X-rays.
Object cleft, version 2: It was the oncologist that the orderly misled about the location of the X-rays.
Subject cleft, version 1: It was the orderly that misled the oncologist about the location of the X-rays.
Subject cleft, version 2: It was the oncologist that misled the orderly about the location of the X-rays.

10. Object cleft, version 1: It was the correspondent that the filmmaker shoved in order to talk to the mayor.
Object cleft, version 2: It was the filmmaker that the correspondent shoved in order to talk to the mayor.
Subject cleft, version 1: It was the correspondent that shoved the filmmaker in order to talk to the mayor.
Subject cleft, version 2: It was the filmmaker that shoved the correspondent in order to talk to the mayor.

11. Object cleft, version 1: It was the graduate student that the RA recommended as a good applicant for the job.
Object cleft, version 2: It was the RA that the graduate student recommended as a good applicant for the job.
Subject cleft, version 1: It was the graduate student that recommended the RA as a good applicant for the job.
12. Subject cleft, version 2: It was the RA that recommended the graduate student as a good applicant for the job.
Object cleft, version 1: It was the psychologist that the chemist enlightened about the meaning of the paper.
Object cleft, version 2: It was the chemist that the psychologist enlightened about the meaning of the paper.
Subject cleft, version 1: It was the psychologist that enlightened the chemist about the meaning of the paper.
Subject cleft, version 2: It was the chemist that enlightened the psychologist about the meaning of the paper.

13. Object cleft, version 1: It was the cheerleader that the acrobat disliked for wanting too much space.
Object cleft, version 2: It was the acrobat that the cheerleader disliked for wanting too much space.
Subject cleft, version 1: It was the cheerleader that disliked the acrobat for wanting too much space.
Subject cleft, version 2: It was the acrobat that disliked the cheerleader for wanting too much space.

14. Object cleft, version 1: It was the maid that the bellhop distracted while they were trying to work.
Object cleft, version 2: It was the bellhop that the maid distracted while they were trying to work.
Subject cleft, version 1: It was the maid that distracted the bellhop while they were trying to work.
Subject cleft, version 2: It was the bellhop that distracted the maid while they were trying to work.

15. Object cleft, version 1: It was the jogger that the biker pushed for causing the accident.
Object cleft, version 2: It was the biker that the jogger pushed for causing the accident.
Subject cleft, version 1: It was the jogger that pushed the biker for causing the accident.
Subject cleft, version 2: It was the biker that pushed the jogger for causing the accident.

16. Object cleft, version 1: It was the machinist that the programmer liked for being easy to work with.
Object cleft, version 2: It was the programmer that the machinist liked for being easy to work with.
Subject cleft, version 1: It was the machinist that liked the programmer for being easy to work with.
Subject cleft, version 2: It was the programmer that liked the machinist for being easy to work with.

17. Object cleft, version 1: It was the cardiologist that the consultant persuaded to run the test.
Object cleft, version 2: It was the consultant that the cardiologist persuaded to run the test.
Subject cleft, version 1: It was the cardiologist that persuaded the consultant to run the test.
Subject cleft, version 2: It was the consultant that persuaded the cardiologist to run the test.

18. Object cleft, version 1: It was the astronomer that the mathematician directed about the execution of the experiment.
Object cleft, version 2: It was the mathematician that the astronomer directed about the execution of the experiment.
Subject cleft, version 1: It was the astronomer that directed the mathematician about the execution of the experiment.
Subject cleft, version 2: It was the mathematician that directed the astronomer about the

execution of the experiment.

19. Object cleft, version 1: It was the electrician that the painter annoyed during long hours on the job.
Object cleft, version 2: It was the painter that the electrician annoyed during long hours on the job.
Subject cleft, version 1: It was the electrician that annoyed the painter during long hours on the job.
Subject cleft, version 2: It was the painter that annoyed the electrician during long hours on the job.

20. Object cleft, version 1: It was the biographer that the publisher alarmed about the number of copies sold.
Object cleft, version 2: It was the publisher that the biographer alarmed about the number of copies sold.
Subject cleft, version 1: It was the biographer that alarmed the publisher about the number of copies sold.
Subject cleft, version 2: It was the publisher that alarmed the biographer about the number of copies sold.

21. Object cleft, version 1: It was the valet that the waiter provoked with a dirty joke.
Object cleft, version 2: It was the waiter that the valet provoked with a dirty joke.
Subject cleft, version 1: It was the valet that provoked the waiter with a dirty joke.
Subject cleft, version 2: It was the waiter that provoked the valet with a dirty joke.

22. Object cleft, version 1: It was the welder that the blacksmith trained to build a gate.
Object cleft, version 2: It was the blacksmith that the welder trained to build a gate.
Subject cleft, version 1: It was the welder that trained the blacksmith to build a gate.
Subject cleft, version 2: It was the blacksmith that trained the welder to build a gate.
23. Object cleft, version 1: It was the actress that the choreographer ridiculed for having predicted a huge success.
Object cleft, version 2: It was the choreographer that the actress ridiculed for having predicted a huge success.
Subject cleft, version 1: It was the actress that ridiculed the choreographer for having predicted a huge success.
Subject cleft, version 2: It was the choreographer that ridiculed the actress for having predicted a huge success.

24. Object cleft, version 1: It was the psychic that the snake-charmer tricked by using a spell.
Object cleft, version 2: It was the snake-charmer that the psychic tricked by using a spell.
Subject cleft, version 1: It was the psychic that tricked the snake-charmer by using a spell.
Subject cleft, version 2: It was the snake-charmer that tricked the psychic by using a spell.

APPENDIX G.

The following set of 24 experimental items comprises the RCs materials used in Study 1, Experiment 2. These materials were all presented in null contexts, and no filler materials were used in this experiment.

1. Object RC, version 1: The reporter that the senator attacked admitted to making an error.
Object RC, version 2: The senator that the reporter attacked admitted to making an error.
Subject RC, version 1: The reporter that attacked the senator admitted to making an error.
Subject RC, version 2: The senator that attacked the reporter admitted to making an error.

2. Object RC, version 1: The musician that the newscaster insulted left the building after the interview.
Object RC, version 2: The newscaster that the musician insulted left the building after the interview.
Subject RC, version 1: The musician that insulted the newscaster left the building after the interview.
Subject RC, version 2: The newscaster that insulted the musician left the building after the interview.

3. Object RC, version 1: The intern that the scientist confused worked at a famous lab at Harvard University.
Object RC, version 2: The scientist that the intern confused worked at a famous lab at Harvard University.
Subject RC, version 1: The intern that confused the scientist worked at a famous lab at Harvard University.
Subject RC, version 2: The scientist that confused the intern worked at a famous lab at Harvard University.

4. Object RC, version 1: The detective that the officer approached had a good record in solving similar cases.
Object RC, version 2: The officer that the detective approached had a good record in solving similar cases.
Subject RC, version 1: The detective that approached the officer had a good record in solving similar cases.
Subject RC, version 2: The officer that approached the detective had a good record in solving similar cases.

5. Object RC, version 1: The pediatrician that the dentist called left a message about the recommended dosage.
Object RC, version 2: The dentist that the pediatrician called left a message about the recommended dosage.
Subject RC, version 1: The pediatrician that called the dentist left a message about the recommended dosage.
Subject RC, version 2: The dentist that called the pediatrician left a message about the recommended dosage.

6. Object RC, version 1: The neurologist that the physician helped worked at MGH for the last ten years.
Object RC, version 2: The physician that the neurologist helped worked at MGH for the last ten years.
Subject RC, version 1: The neurologist that helped the physician worked at MGH for the last ten years.
Subject RC, version 2: The physician that helped the neurologist worked at MGH for the last ten years.

7. Object RC, version 1: The critic that the author annoyed had strong opinions about many things.
Object RC, version 2: The author that the critic annoyed had strong opinions about many things.
Subject RC, version 1: The critic that annoyed the author had strong opinions about many things.
Subject RC, version 2: The author that annoyed the critic had strong opinions about many things.

8. Object RC, version 1: The student that the teacher greeted gave an interview to the school newspaper recently.
Object RC, version 2: The teacher that the student greeted gave an interview to the school newspaper recently.
Subject RC, version 1: The student that greeted the teacher gave an interview to the school newspaper recently.
Subject RC, version 2: The teacher that greeted the student gave an interview to the school newspaper recently.

9. Object RC, version 1: The governor that the congressman antagonized apologized for being too aggressive.
Object RC, version 2: The congressman that the governor antagonized apologized for being too aggressive.
Subject RC, version 1: The governor that antagonized the congressman apologized for being too aggressive.
Subject RC, version 2: The congressman that antagonized the governor apologized for being too aggressive.

10. Object RC, version 1: The ambassador that the interpreter contacted lived in Africa for many years.
Object RC, version 2: The interpreter that the ambassador contacted lived in Africa for many years.
Subject RC, version 1: The ambassador that contacted the interpreter lived in Africa for many years.
Subject RC, version 2: The interpreter that contacted the ambassador lived in Africa for many years.

11. Object RC, version 1: The manager that the administrator questioned had a problem with the company.
Object RC, version 2: The administrator that the manager questioned had a problem with the

company.
Subject RC, version 1: The manager that questioned the administrator had a problem with the company.
Subject RC, version 2: The administrator that questioned the manager had a problem with the company.

12. Object RC, version 1: The historian that the anthropologist challenged published a book on the same topic.
Object RC, version 2: The anthropologist that the historian challenged published a book on the same topic.
Subject RC, version 1: The historian that challenged the anthropologist published a book on the same topic.
Subject RC, version 2: The anthropologist that challenged the historian published a book on the same topic.

13. Object RC, version 1: The engineer that the statistician impressed received a reward at a recent conference.
Object RC, version 2: The statistician that the engineer impressed received a reward at a recent conference.
Subject RC, version 1: The engineer that impressed the statistician received a reward at a recent conference.
Subject RC, version 2: The statistician that impressed the engineer received a reward at a recent conference.

14. Object RC, version 1: The artist that the writer upset left the museum in a bad mood.
Object RC, version 2: The writer that the artist upset left the museum in a bad mood.
Subject RC, version 1: The artist that upset the writer left the museum in a bad mood.
Subject RC, version 2: The writer that upset the artist left the museum in a bad mood.

15. Object RC, version 1: The wrestler that the gymnast observed attended the college on a scholarship.
Object RC, version 2: The gymnast that the wrestler observed attended the college on a scholarship.
Subject RC, version 1: The wrestler that observed the gymnast attended the college on a scholarship.
Subject RC, version 2: The gymnast that observed the wrestler attended the college on a scholarship.

16. Object RC, version 1: The caterer that the chef frustrated was famous for his mushroom soup recipe.
Object RC, version 2: The chef that the caterer frustrated was famous for his mushroom soup recipe.
Subject RC, version 1: The caterer that frustrated the chef was famous for his mushroom soup recipe.

Subject RC, version 2: The chef that frustrated the caterer was famous for his mushroom soup recipe.

17. Object RC, version 1: The model that the editor complimented worked for a leading fashion magazine.
Object RC, version 2: The editor that the model complimented worked for a leading fashion magazine.
Subject RC, version 1: The model that complimented the editor worked for a leading fashion magazine.
Subject RC, version 2: The editor that complimented the model worked for a leading fashion magazine.

18. Object RC, version 1: The girl that the boy teased was very spoiled and difficult.
Object RC, version 2: The boy that the girl teased was very spoiled and difficult.
Subject RC, version 1: The girl that teased the boy was very spoiled and difficult.
Subject RC, version 2: The boy that teased the girl was very spoiled and difficult.

19. Object RC, version 1: The decorator that the planner telephoned intended to leave the company.
Object RC, version 2: The planner that the decorator telephoned intended to leave the company.
Subject RC, version 1: The decorator that telephoned the planner intended to leave the company.
Subject RC, version 2: The planner that telephoned the decorator intended to leave the company.

20. Object RC, version 1: The dog that the cat bit was taken to a veterinarian.
Object RC, version 2: The cat that the dog bit was taken to a veterinarian.
Subject RC, version 1: The dog that bit the cat was taken to a veterinarian.
Subject RC, version 2: The cat that bit the dog was taken to a veterinarian.

21. Object RC, version 1: The sister that the brother kissed spent two years in the Peace Corps in South Africa.
Object RC, version 2: The brother that the sister kissed spent two years in the Peace Corps in South Africa.
Subject RC, version 1: The sister that kissed the brother spent two years in the Peace Corps in South Africa.
Subject RC, version 2: The brother that kissed the sister spent two years in the Peace Corps in South Africa.

22. Object RC, version 1: The farmhand that the rancher yelled to fell off a horse last month and broke his leg.
Object RC, version 2: The rancher that the farmhand yelled to fell off a horse last month and broke his leg.
Subject RC, version 1: The farmhand that yelled to the rancher fell off a horse last month and broke his leg.
Subject RC, version 2: The rancher that yelled to the farmhand fell off a horse last month and broke his leg.

23. Object RC, version 1: The colonel that the sergeant embraced had received many medals and honors.
Object RC, version 2: The sergeant that the colonel embraced had received many medals and honors.
Subject RC, version 1: The colonel that embraced the sergeant had received many medals and honors.
Subject RC, version 2: The sergeant that embraced the colonel had received many medals and honors.

24. Object RC, version 1: The nurse that the therapist embarrassed used to have an alcohol problem.
Object RC, version 2: The therapist that the nurse embarrassed used to have an alcohol problem.
Subject RC, version 1: The nurse that embarrassed the therapist used to have an alcohol problem.
Subject RC, version 2: The therapist that embarrassed the nurse used to have an alcohol problem.

APPENDIX H.

The experimental items for Experiment 3 from Study 2 are given below, grouped by sub-experiment. Each condition for each item is shown. No extra fillers were included in this experiment.

Sub-experiment 1: Nested vs. Branching Relative Clauses

1. Nested: The pilot who the navigator who the gunner warned alerted broke radio silence.
Branching: The gunner warned the navigator who alerted the pilot who broke radio silence.

2. Nested: The beatnik who the writer who everyone seemed to have met fascinated wore women's clothing.
Branching: Everyone seemed to have met the writer who fascinated the beatnik who wore women's clothing.

3. Nested: The tourist who the rug-merchant who the local children teased mystified took dozens of pictures.
Branching: The local children teased the rug-merchant who mystified the tourist who took dozens of pictures.

4. Nested: The terrorist who the spy who the assassin shot at provoked planted a bomb.
Branching: The assassin shot at the spy who provoked the terrorist who planted a bomb.

5. Nested: The small town girl who the mysterious visitor who strangers never talk to intrigued eats live frogs.
Branching: Strangers never talk to the mysterious visitor who intrigued the small town girl who eats live frogs.

6. Nested: The drummer who the singer who groupies love disgusted spat on the crowd.
Branching: Groupies love the singer who disgusted the drummer who spat on the crowd.

7. Nested: The medic who the victim who the doctor yelled to startled dropped the bandage.
Branching: The doctor yelled to the victim who startled the medic who dropped the bandage.

8. Nested: The frightened child who the old woman who the rescue worker looked for comforted survived the crash.
Branching: The rescue worker looked for the old woman who comforted the frightened child who survived the crash.

9. Nested: The leading man who the comedian who the starlet adored entertained forgot the crucial line.
Branching: The starlet adored the comedian who entertained the leading man who forgot the crucial line.

10. Nested: The violinist who the conductor who the corporate sponsors flattered disappointed

skipped the rehearsal.
Branching: The corporate sponsors flattered the conductor who disappointed the violinist who skipped the rehearsal.

11. Nested: The student who the teaching assistant who the professor met with tormented disrupted the class.
Branching: The professor met with the teaching assistant who tormented the student who disrupted the class.

12. Nested: The suspected mobster who the ambitious attorney who the media attacked investigated had been to jail.
Branching: The media attacked the ambitious attorney who investigated the suspected mobster who had been to jail.

13. Nested: The burglar who the paranoid neighbor who the returning couple startled alarmed dropped the TV.
Branching: The returning couple startled the paranoid neighbor who alarmed the burglar who dropped the TV.

14. Nested: The teenager who the bossy contractor who the architect hired irritated ruined the lawn.
Branching: The architect hired the bossy contractor who irritated the teenager who ruined the lawn.

15. Nested: The child who the divorcing parents who the psychologist talked to worried lost a lot of sleep.
Branching: The psychologist talked to the divorcing parents who worried the child who lost a lot of sleep.

16. Nested: The settlers who the tribe who the Sun-god protected frightened were not taken seriously.
Branching: The Sun-god protected the tribe who frightened the settlers who were not taken seriously.

17. Nested: The parishioners who the actress who the preacher was sleeping with scandalized were kept quiet.
Branching: The preacher was sleeping with the actress who scandalized the parishioners who were kept quiet.

18. Nested: The invasion that the children who the aliens kidnapped knew about fell to pieces.

Branching: The aliens kidnapped the children who knew about the invasion that fell to pieces.

Sub-experiment 2: Sentential Complements and Relative Clauses

1. SC/RC: The chance that the administrator who the nurse supervised had lost the medical

reports bothered the intern a great deal on Monday.
RC/SC: The intern who the chance that the administrator had lost the medical reports bothered a great deal supervised the nurse on Monday.

2. SC/RC: The discovery that the babysitter who the parents were fond of had spanked the child disturbed the neighbors to some degree for a long time.
RC/SC: The neighbors who the discovery that the babysitter had spanked the child disturbed to some degree were fond of the parents for a long time.

3. SC/RC: The news that the defense lawyer who the district attorney influenced would defend the murderer affected the celebrity case quite strongly last week.
RC/SC: The celebrity case that the news that the defense lawyer would defend the murderer affected quite strongly influenced the district attorney last week.

4. SC/RC: The suggestion that the preacher who the voters adored will give a moving speech might unnerve the politician a bit at the rally.
RC/SC: The politician who the suggestion that the preacher will give a moving speech might unnerve a bit adored the voters at the rally.

5. SC/RC: The likelihood that the graduate student who the professor respected had plagiarized the article upset the secretary quite a lot all year long.
RC/SC: The secretary who the likelihood that the graduate student had plagiarized the article upset quite a lot respected the professor all year long.

6. SC/RC: The disclosure that the union members who the news stories criticized had agreed upon a contract encouraged the executive tremendously after the meeting.
RC/SC: The executive who the disclosure that the union members had agreed upon a contract encouraged tremendously criticized the news stories after the meeting.

7. SC/RC: The possibility that the actor who the director yelled at would try to reorganize the scene maddened the producers beyond words this morning.
RC/SC: The producers who the possibility that the actor would try to reorganize the scene maddened beyond words yelled at the director this morning.

8. SC/RC: The message that the janitor who the doorman chatted with forgot to empty the trash can infuriated the supervisor all night long on Saturday.
RC/SC: The supervisor who the message that the janitor forgot to empty the trash can infuriated all night long chatted with the doorman on Saturday.

9. SC/RC: The notion that the bus-driver who the policeman screamed at might hit the pedestrian scared the passenger out of his wits at the school.
RC/SC: The passenger who the notion that the bus-driver might hit the pedestrian scared out of his wits screamed at the policeman at the school.

10. SC/RC: The knowledge that the athlete who the doctor helped had scored the winning goal

inspired the little boy to great achievements this year.
RC/SC: The little boy who the knowledge that the athlete had scored the winning goal inspired to great achievements helped the doctor this year.

11. SC/RC: The report that the fans who the hooligans ridiculed cheered for the home team excited the star player a huge amount last night.
RC/SC: The star player who the report that the fans cheered for the home team excited a huge amount ridiculed the hooligans last night.

12. SC/RC: The myth that the guard who the fair maiden married had killed the dragon impressed the knight immensely at the castle.
RC/SC: The knight who the myth that the guard had killed the dragon impressed immensely married the fair maiden at the castle.

13. SC/RC: The inference that the mindless lackey who the hero fought with would bungle the job exasperated the evil genius tremendously this week.
RC/SC: The evil genius who the inference that the mindless lackey would bungle the job exasperated tremendously fought with the hero this week.

14. SC/RC: The accusation that the cheerleader who the quarterback dated had snubbed the nerds aggravated the class president a great deal last year.
RC/SC: The class president who the accusation that the cheerleader had snubbed the nerds aggravated a great deal dated the quarterback last year.

15. SC/RC: The implication that the dictator who the diplomat insulted had closed the embassy embarrassed the prime minister in parliament during the last session.
RC/SC: The prime minister who the implication that the dictator had closed the embassy embarrassed in parliament insulted the diplomat during the last session.

16. SC/RC: The threat that the beggar boy who the genie served would speak the magic words terrified the king immensely once upon a time.
RC/SC: The king who the threat that the beggar boy would speak the magic words terrified immensely served the genie once upon a time.

17. SC/RC: The hunch that the serial killer who the waitress trusted might hide the body frightened the FBI agent into action yesterday morning.
RC/SC: The FBI agent who the hunch that the serial killer might hide the body frightened into action trusted the waitress yesterday morning.

18. SC/RC: The confession that the housewife who the whole town gossiped about bought cosmetics distracted the Avon lady from her work for a while.
RC/SC: The Avon lady who the confession that the housewife bought cosmetics distracted from her work gossiped about the whole town for a while.

Sub-experiment 3: Plausibility Violations

1. Control: The chef is baking the bread and the waiter is seating the customer.
Plaus. Violation: The chef is baking the bread and the customer is seating the waiter.

2. Control: The cowboys are roping the bulls and the clowns are juggling the batons.
Plaus. Violation: The cowboys are roping the bulls and the batons are juggling the clowns.

3. Control: The farmers are milking the cows and the shepherds are corralling the sheep.
Plaus. Violation: The farmers are milking the cows and the sheep are corralling the shepherds.

4. Control: The conductor is leading the orchestra and the violinists are playing the instruments.
Plaus. Violation: The conductor is leading the orchestra and the instruments are playing the violinists.

5. Control: The students are learning the rules and the teacher is grading the tests .
Plaus. Violation: The students are learning the rules and the tests are grading the teacher.

6. Control: The gardener is watering the plant and the man is painting the house.
Plaus. Violation: The gardener is watering the plant and the house is painting the man.

7. Control: The artists are sketching the landscape and the potters are shaping the clay.
Plaus. Violation: The artists are sketching the landscape and the clay is shaping the potters.

8. Control: The playwright is writing the scene and the actors are reciting the lines.
Plaus. Violation: The playwright is writing the scene and the lines are reciting the actors.

9. Control: The bees are stinging the children and the wasp is building the nest.
Plaus. Violation: The bees are stinging the children and the nest is building the wasp.

10. Control: The judges are banging the gavels and the lawyers are unzipping the briefcases.
Plaus. Violation: The judges are banging the gavels and the briefcases are unzipping the lawyers.

11. Control: The cook is grilling the meat and the guests are eating the hamburgers.
Plaus. Violation: The cook is grilling the meat and the hamburgers are eating the guests.
12. Control: The dentists are drilling the teeth and the nurse is giving the injection.
Plaus. Violation: The dentists are drilling the teeth and the injection is giving the nurse.

13. Control: The pilot is flying the plane and the flight attendant is pushing the cart.
Plaus. Violation: The pilot is flying the plane and the cart is pushing the flight attendant.

14. Control: The orphan is shining the shoes and the conmen are fooling the audience.
Plaus. Violation: The orphan is shining the shoes and the audience is fooling the conmen.

15. Control: The journalists are reporting the story and the editor is correcting the article.
Plaus. Violation: The journalists are reporting the story and the article is correcting the editor.

16. Control: The carpenters are sanding the tables and the plumber is unclogging the pipes.

Plaus. Violation: The carpenters are sanding the tables and the pipes are unclogging the plumber.

17. Control: The explorer is sailing the ocean and the navigator is interpreting the maps.
Plaus. Violation: The explorer is sailing the ocean and the maps is interpreting the navigator.

18. Control: The earl is lifting the teacups and the duchesses are admiring the saucers.
Plaus. Violation: The earl is lifting the teacups and the saucers are admiring the duchesses.

APPENDIX I.

The additional experimental items for Experiment 4 from Study 2 are given below, grouped by sub-experiment. Each condition for each item is shown. No fillers were included in this experiment.

Sub-experiment 1: Nested vs. Branching Relative Clauses

19. Nested: The lullaby that the famous country singer who the record label had signed with sang was written several years ago.
Branching: The record label had signed with the famous country singer who sang the lullaby that was written several years ago.

20. Nested: The game that the child who the lawnmower startled was playing lasted for hours.
Branching: The lawnmower startled the child who was playing the game that lasted for hours.

21. Nested: The picture that the artist who the school had expelled was copying was printed.
Branching: The school had expelled the artist who was copying the picture that was printed.

22. Nested: The crime that the gangster who the newspaper had profiled had planned was quickly solved.
Branching: The newspaper had profiled the gangster who had planned the crime that was quickly solved.

23. Nested: The apartment that the maid who the service sent over was cleaning was well-decorated.
Branching: The service sent over the maid who was cleaning the apartment that was well-decorated.

24. Nested: The shirt that the seamstress who the immigration officer had investigated was mending needed to be dry-cleaned.
Branching: The immigration officer had investigated the seamstress who was mending the shirt that needed to be dry-cleaned.

Sub-experiment 2: Clefts in Null Contexts

19. Object cleft, version 1: It was the electrician that the painter annoyed during long hours on the job.
Object cleft, version 2: It was the painter that the electrician annoyed during long hours on the job.
Subject cleft, version 1: It was the electrician that annoyed the painter during long hours on the job.
Subject cleft, version 2: It was the painter that annoyed the electrician during long hours on the job.

20. Object cleft, version 1: It was the biographer that the publisher alarmed about the number of

copies sold.

Object cleft, version 2: It was the publisher that the biographer alarmed about the number of copies sold.

Subject cleft, version 1: It was the biographer that alarmed the publisher about the number of copies sold.

Subject cleft, version 2: It was the publisher that alarmed the biographer about the number of copies sold.

21. Object cleft, version 1: It was the valet that the waiter provoked with a dirty joke.
Object cleft, version 2: It was the waiter that the valet provoked with a dirty joke.
Subject cleft, version 1: It was the valet that provoked the waiter with a dirty joke.
Subject cleft, version 2: It was the waiter that provoked the valet with a dirty joke.

22. Object cleft, version 1: It was the welder that the blacksmith trained to build a gate.
Object cleft, version 2: It was the blacksmith that the welder trained to build a gate.
Subject cleft, version 1: It was the welder that trained the blacksmith to build a gate.
Subject cleft, version 2: It was the blacksmith that trained the welder to build a gate.

23. Object cleft, version 1: It was the actress that the choreographer ridiculed for having predicted a huge success.
Object cleft, version 2: It was the choreographer that the actress ridiculed for having predicted a huge success.
Subject cleft, version 1: It was the actress that ridiculed the choreographer for having predicted a huge success.
Subject cleft, version 2: It was the choreographer that ridiculed the actress for having predicted a huge success.

24. Object cleft, version 1: It was the psychic that the snake-charmer tricked by using a spell.
Object cleft, version 2: It was the snake-charmer that the psychic tricked by using a spell.
Subject cleft, version 1: It was the psychic that tricked the snake-charmer by using a spell.
Subject cleft, version 2: It was the snake-charmer that tricked the psychic by using a spell.

Sub-experiment 3: Relative Clauses in Null Contexts

19. Object relative clause, version 1: The decorator that the planner telephoned intended to leave the company.
Object relative clause, version 2: The planner that the decorator telephoned intended to leave the company.
Subject relative clause, version 1: The decorator that telephoned the planner intended to leave the company.
Subject relative clause, version 2: The planner that telephoned the decorator intended to leave the company.

20. Object relative clause, version 1: The dog that the cat bit was taken to a veterinarian.
Object relative clause, version 2: The cat that the dog bit was taken to a veterinarian.
Subject relative clause, version 1: The dog that bit the cat was taken to a veterinarian.

Subject relative clause, version 2: The cat that bit the dog was taken to a veterinarian.

21. Object relative clause, version 1: The sister that the brother kissed spent two years in the Peace Corps in South Africa.
Object relative clause, version 2: The brother that the sister kissed spent two years in the Peace Corps in South Africa.
Subject relative clause, version 1: The sister that kissed the brother spent two years in the Peace Corps in South Africa.
Subject relative clause, version 2: The brother that kissed the sister spent two years in the Peace Corps in South Africa.

22. Object relative clause, version 1: The farmhand that the rancher yelled to fell off a horse last month and broke his leg.
Object relative clause, version 2: The rancher that the farmhand yelled to fell off a horse last month and broke his leg.
Subject relative clause, version 1: The farmhand that yelled to the rancher fell off a horse last month and broke his leg.
Subject relative clause, version 2: The rancher that yelled to the farmhand fell off a horse last month and broke his leg.

23. Object relative clause, version 1: The colonel that the sergeant embraced had received many medals and honors.
Object relative clause, version 2: The sergeant that the colonel embraced had received many medals and honors.
Subject relative clause, version 1: The colonel that embraced the sergeant had received many medals and honors.
Subject relative clause, version 2: The sergeant that embraced the colonel had received many medals and honors.

24. Object relative clause, version 1: The nurse that the therapist embarrassed used to have an alcohol problem.
Object relative clause, version 2: The therapist that the nurse embarrassed used to have an alcohol problem.
Subject relative clause, version 1: The nurse that embarrassed the therapist used to have an alcohol problem.
Subject relative clause, version 2: The therapist that embarrassed the nurse used to have an alcohol problem.

Sub-experiment 4: Plausibility Violations

1. Plausibility Violation: The pilot is flying the plane and the cart is pushing the flight attendant.
Control: The pilot is flying the plane and the flight attendant is pushing the cart.

2. Plausibility Violation: The chickens are pecking the ground and the mud is sniffing the pigs.
Control: The chickens are pecking the ground and the pigs are sniffing the mud.

3. Plausibility Violation: The technicians are solving the problems and the situation is assessing the engineers.
Control: The technicians are solving the problems and the engineers are assessing the situation.

4. Plausibility Violation: The orphan is shining the shoes and the audience is fooling the conmen.
Control: The orphan is shining the shoes and the conmen are fooling the audience.

5. Plausibility Violation: The journalists are reporting the story and the article is correcting the editor.
Control: The journalists are reporting the story and the editor is correcting the article.

6. Plausibility Violation: The carpenters are sanding the tables and the pipes are unclogging the plumber.
Control: The carpenters are sanding the tables and the plumber is unclogging the pipes.