2015

Single or Dual Resources: The Role of Working Memory in Syntactic Processing

Rebecca Ann Horn
Louisiana State University and Agricultural and Mechanical College

SINGLE OR DUAL RESOURCES: THE ROLE OF WORKING MEMORY IN SYNTACTIC PROCESSING

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Psychology

by
Rebecca A. Horn
B.A., University of Wisconsin Whitewater, 2005
M.S., University of Wisconsin Oshkosh, 2009
May 2015

ACKNOWLEDGMENTS

I would like to express my gratitude towards those who have helped me through the process of completing this project. First there is my advisor, Dr. Janet L. McDonald, whose thoughtfulness, patience, and scholarly passion have been my guiding light for the past six years and without whom the current work would have been infinitely more daunting to undertake. I would also like to thank the members of my committee, Drs. Jason L. Hicks, Emily M. Elliott, and Megan H. Papesh, for their valuable insights and critiques from the proposal through the defense of my dissertation, and my Dean's Representative, Dr. William Saas, for his insightful comments during my defense.

On a more practical level, I would like to thank Corey McGill (of the Elliott Working Memory Laboratory) and Kacie Mennie (of the Office of Applied Cognition) for collaborating with myself and my undergraduate research assistants in the administration of the working memory screening battery. Finally, I would like to thank my undergraduate research assistants, Kenzie Alexander, Jeanette Rimmer, Tiffany Rivers, Ryan Sartor, Laura Siqueiros, and Trey Toups, for the many hours they spent administering the working memory tasks, and to Ryan and Trey specifically for their help in proof-reading stimulus sentences in SuperLab.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Within the field of psycholinguistics there are those who argue for a close relationship between working memory capacity (WMC) and syntactic processing (Just and Carpenter, 1992) and those who argue that there is no such relationship (Waters and Caplan, 1996b; 2004). Despite years of research, empirical data has yet to settle this disagreement, perhaps because a number of methodological differences between studies from each side make direct comparisons of data nearly impossible. The current study was designed to partially replicate three previous studies using their own experimental sentence types in a self-paced word-by-word reading paradigm in order to examine the effects of several methodological factors, including judgment type and object-relative clause construction, on performance in the syntactic processing task. In Experiment 1 we used Just and Carpenter's (1992) true/false judgment and found data theoretically consistent with their main results which supported a relationship between WMC and syntactic processing, but only for sentence sets constructed in the manner of Waters and Caplan (1996b; 2004). In Experiment 2 we used Waters and Caplan's (1996b; 2004) acceptability judgment with the same stimulus sets and found no support for a WMC-syntactic processing relationship. Finally, in Experiment 3 we used a grammaticality judgment with the same stimulus sets and once again found no support for a WMC-syntactic processing relationship. Together, the results of all three experiments suggest that a number of methodological factors that have been previously considered irrelevant in the syntactic processing task in fact produce significant changes in the results that in turn alter the conclusions drawn. In particular, we found evidence that judgment type can alter overall reading times, suggesting that task demands may cause participants to alter their processing strategies in the task. Our results also illustrate that the pattern of results can depend largely upon the method of object-relative clause construction used for the task.

CHAPTER 1. OVERVIEW OF THE SYNTACTIC PROCESSING CONTROVERSY

1.1 Syntactic Processing

Human language is an incredibly rich and complex code used for the storage, organization, and communication of ideas. Successful production and comprehension of human language requires the capacity to process and integrate a number of interconnected types of information (e.g., phonology, semantics, syntax, pragmatics, etc.) within a short period of time. The question that has long plagued language researchers is: To what extent is the capacity to process and integrate linguistic information domain-specific? In other words, is it the case that special cognitive mechanisms and neural substrates are devoted exclusively to the processing and integration of linguistic information, or is it the case that domain-general mechanisms and resources perform these functions for linguistic information just as they do for other types of information?

The proposed answers to this major theoretical question are closely linked with theories of language acquisition, which are of necessity also concerned with issues of domain-specific and domain-general skills and knowledge, as well as external influences (e.g., linguistic input) and natural laws that govern things such as anatomy and stimulus properties (Berwick, Pietroski, Yankama, & Chomsky, 2011). Disagreements amongst various acquisition theories typically boil down to the extent to which each of these factors is believed to be responsible for both commonalities and differences in language acquisition. In other words, while it is the case that all typically developing individuals who are not artificially deprived of linguistic input will acquire a language, they do not all acquire the same language, and thus any comprehensive theory of acquisition must explain how this is so.

One theory of language acquisition, popular with linguists, is the nativist or generativist perspective. According to generativists, the basis of language proficiency is the possession of a grammar (i.e., a set of rules that determines what is considered to be a well-formed sentence and what is considered ill-formed). Moreover, generativists argue that possession of a grammar is necessarily innate because linguistic input is too impoverished to provide sufficient opportunity to learn what must be learned in order to produce an infinite variety of utterances from the finite pool of components (Chomsky, 1965, 1986). To put it another way, generativists argue that generalization from directly experienced linguistic exemplars is insufficient to account for the development of the full expressive potential of a language. This "poverty of the stimulus" (POS) argument leads to the logical conclusion that innate factors must weigh more heavily in the acquisition of a language than do external factors. In particular, generativists argue for an inborn set of potential parameters for language, known as Universal Grammar, and a domain-specific mechanism by which any given individual is guided to appropriate parameter settings by their linguistic experience, known as the Language Acquisition Device (Chomsky, 1965). That is, Universal Grammar is argued to contain a finite set of potential grammatical rules for human languages, and by means of exposure to a language in childhood, an individual's Language Acquisition Device becomes set on the particular rules appropriate for that language.

The generativist view is thus a theory that relies heavily upon innate domain-specific skills and knowledge to explain how an individual comes to acquire a language, and this has important implications for the generativist view of language processing in adults. In particular, the generativist view assumes a fairly invariant level of linguistic competence in native speakers of a particular language and discounts factors that might lead to individual differences in language use, such as memory constraints (Seidenberg & MacDonald, 1999).

Support for the generativist view typically comes from linguistic arguments explaining transformed grammatical structures such as Chomsky's (1968) "polar interrogatives." To illustrate this sort of argument, Berwick et al. (2011) use the example of how native English speakers invariably know that the answer to 5a below is 5b rather than 5c, even though technically the auxiliary verb *can* could be modifying either *fly* or *eat* because both 5b and 5c are grammatical utterances.

(5a) Can eagles that fly eat?

(5b) Eagles that fly can eat.

(5c) Eagles that can fly eat.

According to Berwick et al. (2011) and other generativists, the fact that English speakers do not select 5c as the answer to 5a is due to rules in the grammar about the location in which words of certain grammatical classes can be generated and how those words can be deleted or moved within the general grammatical framework (these rules are referred to as the deep structure of a sentence). Generativists thus argue that the reason that 5b is the unambiguous answer to 5a is that the auxiliary verb *can* is generated under 'S' rather than in the verb phrase of 5b, and because of this it can be moved in order to form the question 5a, while in 5c *can* is generated within a complementizer phrase modifying the main subject, *eagles*, a location from which it cannot be moved to create question 5a. The Figure 1.1.1 below demonstrates this difference in the generation position of *can*.

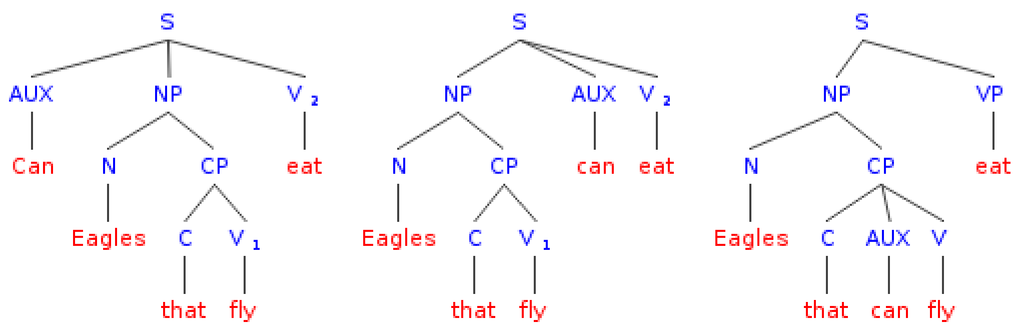


Figure 1.1.1. The (simplified) deep structures of question 5a (left) and sentences 5b (middle) and 5c (right).

The generativist argument is that movement rules such as those demonstrated above are part of the innate, abstract grammar and cannot possibly be learned from the imperfect and incomplete input that human children are exposed to while acquiring their native languages. Furthermore, such rules are similar across a variety of languages, suggesting some common core of an abstracted form of language that persists across wide variation in personal experiences with specific languages.

In contrast to the generativist view is the probabilistic constraints framework which sees the task of language acquisition as a matter of learning to use a language rather than a matter of grammatical parameter setting (Seidenberg, 1997). According to this framework, language is governed by probabilistic constraints at all levels of representation and these constraints are learnable through the engagement of domain-general statistical learning mechanisms that are active when an individual is attempting to comprehend and produce language (Seidenberg & MacDonald, 1999). In this functionalist approach, language acquisition begins with the encoding of the basic form-meaning associations required for producing and comprehending utterances into neural networks. As these neural networks are used and elaborated, other properties such as morphology and syntax emerge (e.g., Allen & Seidenberg, 1999). The probabilistic constraints framework thus explains the ability to make judgments about the grammaticality of a sentence as the result of differing outputs of the neural network rather than as dependent upon abstract, innate knowledge of a grammar (Seidenberg & MacDonald, 1999).

According to Seidenberg and MacDonald (1999), the appeal of the probabilistic view is that it incorporates domain-general factors, such as memory limitations, that have demonstrable impacts on language performance, in contrast to the generativist view's focus on the idealized speaker that abstracts away from language in use to language in theory. By focusing on the competence model of an idealized speaker, the generativist view separates the causes of errors in production and comprehension from domain-specific linguistic processes. Thus in the generativist view, errors are caused by factors extrinsic to the grammar. The probabilistic constraints view, on the other hand, incorporates the causes of errors into the process of producing and comprehending language, and thus embeds domain-specific skills and knowledge within the domain-general influences that are present in the real-world linguistic context. Because of this, the probabilistic constraints view limits itself to providing an explanation of utterances that people can and do produce, rather than the unusual and complicated constructions that are theoretically possible within a grammar, but that are rarely actually produced outside of the context of linguistic studies (e.g., The cat that the dog that the girl who Willard loves bought chased is now sitting on the windowsill.; Seidenberg & MacDonald, 1999).

While the support for the generativist view comes from looking at how native speakers of a language judge grammaticality and extrapolates backwards from this data to propose the POS argument, support for the probabilistic constraints view moves in the opposite direction by looking at how properties of linguistic stimuli can support the acquisition of complex and abstract linguistic knowledge. Specifically, linguistic stimuli contain statistical information such as transitional probabilities that have been shown to be useful in segmenting words from continuous speech streams in both infants (Saffran, Newport, &Aslin, 1996) and adults (Mirman, Magnuson, Graf Estes, & Dixon, 2008; Thiessen, 2010). Statistical information in language input has also been demonstrated to support the acquisition of syntax (i.e., the rules for arranged words and phrases into well-formed sentences) in children (Kidd, 2012) because it provides access to distributional cues in the language (Thiessen & Erickson, 2013) that results in the same ability to judge the grammaticality of sentences that generativists attribute to the abstract grammar (Seidenberg & MacDonald, 1999). Furthermore, statistical learning is not specific to language, but rather is a domain-general learning strategy (Yang, 2004) subject to domain-general constraints (Thiessen, 2011).

Returning to the example of polar interrogatives above, probabilistic constraints theorists would argue that it is not the case that generation and movement rules in an abstract grammar account for the finding that sentence 5b is unambiguously the answer to question 5a (i.e., the generativist argument), but rather that such decisions are based upon differential outputs of neural networks trained on statistical and distributional regularities in the linguistic input (Seidenberg & MacDonald, 1999). That is, when asked the question 5a, the neural network will produce an output consistent with 5b and not consistent with 5c based solely off the combined calculations of distributional properties at multiple level of representation.

Sensitive to the criticism that their theory does not do enough to account for individual differences in language production and comprehension, linguists from the generativist perspective have proposed separations amongst linguistic information that allow the different types (e.g., morphology, syntax, pragmatics) to be processed in different ways. These exact divisions vary somewhat depending upon the researcher presenting the argument, but there seems to be a certain amount of consensus for the position that the processing of syntactic information operates independently of the processing of things like semantics and pragmatics (Fodor, 1988). These linguistic arguments have in turn been incorporated into psycholinguistic theories of sentence processing and comprehension that make predictions about the relationship between the domain-specific processing of syntax and domain-general skills such as Working Memory Capacity (WMC). Specifically, Just and Carpenter (1992) have argued that syntactic processing requires recruitment of a general verbal working memory resource that is used for all types of linguistic information because it involves not only short term maintenance of information, but also manipulation of that information and storage of intermediate processing outputs. On the other hand, there are psycholinguists who maintain that syntactic processing, guided by innate grammar, is both automatic and obligatory (Waters & Caplan, 1996a) and thus occurs without recruiting the same resource that is used to process other types of linguistic information (Caplan, Waters, & Dede, 2007).

These two psycholinguistic theories make very different predictions about the outcome of a number of experimental manipulations that will be explained further below. Furthermore, the predictions of both sides have been tested a number of times over the years. Yet somehow it remains unclear which theory more accurately describes human syntactic processing and its relationship to domain-general skills. In the current work we propose that this lack of clarity is largely an effect of simple, yet profoundly important methodological choices that have delineated studies from the two perspectives. Furthermore, by closely examining and testing these methodological differences, the current work is able to demonstrate that the answer you get when asking which resources are involved in syntactic processing depends largely upon how you choose to manipulate and measure processing and how you choose to look at the data produced.

1.2 Measurement of Syntactic Processing

In order to fully describe the two competing psycholinguistic theories of syntactic processing, it is necessary to describe the predictions they make about performance on syntactic processing tasks. While the particular methodologies used to create and present these tasks vary, as will be discussed in more depth in Chapter 2, a brief introduction to the general syntactic processing task will be of use during the discussion of the theories.

In a syntactic processing task, participants are asked to read or listen to a variety of sentences and then make judgments about each (variations in judgment type will be discussed further in Chapter 2). Depending upon how the task is set up, there are three dependent variables that may be of interest: reading/listening time, judgment reaction time, and judgment accuracy. The set of sentences presented to participants is manipulated so that some sentences are simple and some are complex, and differences in the three dependent measures are expected as complexity increases.

There are a number of ways to manipulate sentence complexity. Main clause sentences such as “The woman went out last night.” are the simplest type of sentence, involving as they do a straight-forward word order. One method of increasing the complexity level of a sentence involves embedded clauses of various types into such sentences. For example, a subject-relative clause is one that is inserted into a main clause sentence to modify the main subject of the sentence (e.g., “The woman that called a friend went out last night,” where “that called a friend” is modifying “the woman”). A second type, more complex type of clause is known as an object-relative clause. Object-relative clauses modify the subject of the sentence by making it the object of a subordinate clause. An example of an object-relative clause would be something like "The woman that I phoned went out last night." In this sentence, "the woman" is simultaneously the subject of the main verb, "went out," and the object of the clause "that I phoned." Using one element to fulfill two roles in a sentence increases the complexity of a subordinate clause, making object-relative clauses more complex than subject-relatives (which in turn are more complex than main clauses). King and Just (1991) argued that object-relatives are also more complex than subject-relatives because they interrupt the main clause, requiring that the individual temporarily stop processing the main clause in order to process the relative clause, thus necessitating temporary storage of main clause information while the relative clause is processed, whereas subject-relative clauses do not require such a pause due to the head noun's role as the subject of both clauses.

In addition to embedding clauses of varying types, sentence complexity can also be increased by introducing syntactic ambiguity. One common way to accomplish this is by reducing embedded clauses. Reduction of a clause simply means removing any pronouns or complementizers (that, which, whom, etc.) that typically introduce the clause (e.g., The woman I phoned went out last night.). How ambiguous a reduced clause is depends upon its structure. The previous example might still be easily understood after being reduced, but something like "The horse raced past the barn fell," is typically far more difficult to parse. This difficulty arises because the verb within the relative clause, "raced", could be the main verb of the sentence (e.g., The horse raced.) or it could be a past participle used in the relative clause to modify the subject of the main clause (i.e., the horse that did race). In fact, this second interpretation is the accurate one, but a reader does not know this until encountering the final word of the sentence (which also happens to be the main verb), "fell."

A simple main effect of increasing complexity on the three dependent measures of the syntactic processing task (reading/listening time, judgment reaction time, and judgment accuracy) is not in dispute. It is generally accepted that more complex sentences take longer to read/listen to, longer to react to, and result in lower accuracy than simple sentences, whatever the theoretical background of the researcher. Essentially, it is not a matter of debate that complex sentences are harder than simple ones. What is in dispute is whether complex sentences are disproportionately harder for some people than for others, and if they are, why this should be so. This issue will be explained in more depth below.

1.3 Single Resource Theory

The first psycholinguistic model of syntactic processing is the capacity theory proposed by Just and Carpenter (1992) (to be referred to here as "single resource theory"). In this model, it is assumed that each element (i.e., stored words or phrases, syntactic structures, computations, etc.) of a sentence requires a certain level of activation to be processed successfully. Some elements require more activation and others require less, but all contribute to the overall demand of the sentence. Because sentences vary in the number of elements they contain and the activation required by each element, they also vary in the amount of overall activation required to successfully process them. Thus, sentences with more complex syntactic structures or with a greater number of words will have higher activation demands than syntactically simple, short sentences.

In addition to variability in the activation demands imposed by sentences, there is natural variability amongst individuals in the maximum amount of activation they have available to meet these demands, according to single resource theory. Furthermore, the theory holds that this individual variability should interact with sentence demands. When the activation demands of sentential elements are low, even individuals with lower levels of maximum available activation should have enough of it on hand to enable fast, parallel processing of the sentence as a whole, resulting in high levels of judgment accuracy as well as relatively fast reading/listening and reaction times. However, as sentential demands increase and begin to tax available activation, trade-offs between storage and processing elements are required to avoid exceeding the maximum. One trade-off that may occur is that between speed and accuracy. This means that processing of all sentential elements may be slowed down significantly so that all elements still consume a share of the available activation, but complete processing of each element requires longer than it would if there were more activation available. The results of this trade-off would be slowed reading/listening and reaction times accompanied by high comprehension accuracy.

A second trade-off that could occur when overall demands are high could result in high-demand elements (e.g., pragmatic cues, complex syntactic structures, etc.) being dropped from the processing queue entirely in favor of processing less demanding elements. This type of trade-off would not necessarily affect reading/listening and reaction times on a task, but would certainly result in decreases in comprehension accuracy. This pattern of results, according to Just and Carpenter (1992) could create the appearance of modular syntactic processing. In other words, if a sentence's activation demands are high enough relative to the individual's available activation to cause a trade-off where elements such as pragmatic cues are dropped entirely, then comprehension data for that sentence would appear to support claims that pragmatic cues and syntactic structures cannot be processed in parallel (i.e., because they were not being processed in parallel in that particular sentence). Just and Carpenter (1992) would argue, however, that this support would be faulty in that it is not the case that pragmatic cues can never be processed in parallel with syntactic structure, only that the individual must have enough available activation to do so.

The important theoretical takeaway from this model of demands and trade-offs in syntactic processing is that all of the elements of a sentence (e.g., syntax, semantics, pragmatics, etc.) are thought to be handled by the same resource, which we will refer to as general verbal working memory (gvWM). In their original paper, Just and Carpenter (1992) theorized that this resource was located specifically in the domain-general central executive component of Baddeley's (1986) model of working memory. While it is beyond the scope of the current work to delve deeply into the various models of working memory, and even more so to determine which model is more accurate, the most relevant aspect of Just and Carpenter's (1992) proposed location of the gvWM resource is that it is domain-general. Because of this feature of the resource, the single resource model can make a number of testable predictions about the relationship between performance on syntactic processing measures and measures of domain-general working memory capacity (WMC) such as the Daneman and Carpenter (1980) Reading-span task.

The most obvious prediction is that, if syntactic processing and WMC tasks both rely on the same resource, then scores on the tasks that measure them should be strongly positively correlated. Just and Carpenter (1992) report that some studies have found correlations of .5 to .6 between sentence comprehension tasks and the Reading-span task (Daneman & Carpenter, 1980; Masson & Miller, 1983). A second prediction is that when the activation demands of a sentence are high in relation to the activation available to the individual, performance on sentence comprehension tasks will be lower in terms of accuracy and slower in terms of reading/listening and reaction times. This means that performance should decline when complexity is increased for everyone, but also that one should find that individuals with less available activation (i.e., lower working memory capacity) should differ in terms of judgment accuracy as well as reading/listening and reaction times from those with more activation available (i.e., higher working memory capacity) when asked to process sentences with high activation demands (i.e., grammatically complex or long sentences). Several studies have found just such effects when participants were divided into span groups based on reading span scores and their performance on sentence comprehension tasks was compared (e.g., Just & Carpenter, 1992; King & Just, 1991; MacDonald et al., 1992). Furthermore, taxing the activation available to high capacity individuals by imposing extra load during sentence processing (e.g., by requiring concurrent maintenance of extra information) should decrease accuracy and cause increases to reaction and reading times. This too has been observed experimentally (e.g., King & Just, 1991; MacDonald et al., 1992).

Although there is empirical support for the single resource view, not all theorists agree that this support is particularly convincing. Waters and Caplan (1996a) in particular take issue with the interpretation of these data for several reasons, among them a seemingly consistent failure to find three-way interactions between sentence structure, phrase, and working memory capacity in the data that would support the prediction that increasing complexity differentially impairs the performance of low-span individuals where processing load is highest. For this and other reasons, Waters and Caplan (1996a) proposed their own model of syntactic processing, here to be referred to as the "dual resource theory".

1.4 Dual Resource Theory

In contrast to single resource theory, Waters and Caplan's (1996a; see also Caplan et al., 2007) dual resource theory holds that verbally mediated tasks such as sentence comprehension and the Daneman and Carpenter (1980) Reading-span task require the recruitment of two separate cognitive resources, one to handle interpretive processes and one for post-interpretive processes. Interpretive processes are defined by Waters and Caplan (1996a) as any processes involved in the initial assignment of word meaning and syntactic structure during sentence processing, while post-interpretive processes are those that involve verbal retrieval and explicit reasoning. The authors argue that interpretive processes occur automatically and obligatorily in a domain-specific resource referred to as syntactic verbal working memory (svWM) and that post-interpretive processes require the recruitment of a consciously controlled domain-general resource referred to as general verbal working memory (gvWM).

Two important conclusions arise from the nature of these proposed resources. The first conclusion is that there must be a strict separation between svWM and gvWM, rooted in their automatic versus controlled natures and coupled with a strict non-separation between gvWM and resources used for information in other domains. If dual resource theory is accurate, then interpretive processes never recruit gvWM because the automatic nature of svWM means it is always sufficient to handle them (and only them; svWM never handles post-interpretive processing), but gvWM resources can be depleted by demands from other domains (e.g., visual information). The second, related conclusion is that svWM capacity does not vary significantly between individuals (an idea that harkens back to the competence assumption of the generativists discussed earlier) because of its automatic nature, but gvWM does because of its effortful nature.

The distinction between automatic (svWM) and controlled (gvWM) processing has important methodological implications in that it stands to reason that one must measure the two processes differently. Almost by definition, interpretive processes must be measured online to capture their automatic nature, while post-interpretive processes are necessarily reflected in offline measures due to their controlled nature. Waters and Caplan (1996a) thus argue that comprehension questions such as those used in single resource studies (King & Just, 1991; MacDonald et al., 1992; Just & Carpenter, 1992), which require further processing of the question and reasoning to complete, are not in fact measuring the svWM resource as those researchers would claim, but rather the gvWM resource. This overlap, they argue, is the reason single resource researchers found any relationship between sentence comprehension tasks and the Reading-span task. While this argument is quite valid in terms of questioning what comprehension questions are actually measuring, it assumes that the single resource studies used only comprehension accuracy as the critical measure of syntactic processing. This is not the case. Just and Carpenter (1992), MacDonald et al. (1992), and King and Just (1991) also used first-pass or per-word reading times to support their hypotheses. It seems difficult to argue that reading time, an online measure, is actually capturing the gvWM resource rather than the svWM resource. Regardless of the appropriateness of this criticism of single resource studies, it is the basis of Waters and Caplan's (1996b) decision to use a different type of judgment, acceptability/plausibility, in their measure of syntactic processing. This judgment, though still offline because it comes at the end of each sentence, is thought to require less processing than comprehension judgments and thus is considered more likely to be a measure of svWM. For online measurement, these studies have relied either on mean per word reading times (calculated by dividing total reading and reaction time by the number of words in the sentence), reading/listening times for phrases, and judgment reaction times. The implications of these modifications to both offline and online measures in the core syntactic processing task will be discussed further in Chapter 2.

Given the theoretical disparities between the two models, it logically follows that dual resource theory makes predictions that directly oppose those of the single resource theory. Specifically, dual resource theory would predict that there should be no significant correlation between sentence comprehension tasks and WMC tasks, provided that the sentence comprehension tasks appropriately exclude offline measures that allow for the use of post-interpretive processing. Indeed, Waters and Caplan (1996b), Waters, Caplan, and Yampolsky (2003), and Waters and Caplan (2004) all failed to find significant correlations between the tasks. In terms of complexity, the dual resource theory does predict that complex sentences will be associated with lower accuracy and longer reading times generally, but does not predict performance differences between those who score low and high on measures of verbal working memory capacity that would demonstrate a disproportionate decline in performance for low scoring individuals wherever processing demands are highest (i.e., a three-way interaction between capacity group, sentence type, and critical phrases). Waters and Caplan (1996b), Waters et al. (2003), and Waters and Caplan (2004) all found that complex sentences result in lower accuracy and longer reading times generally, and that low-span individuals are less accurate and slower than high-span individuals overall, but they all failed to find the three-way interaction (structure x phrase x span) predicted by single resource theory. Finally, dual resource theory would predict no effect of cognitive load (i.e., storage of words or digits) on accuracy, reading times, or reaction times during sentence comprehension tasks because cognitive load should tax gvWM but not svWM. Waters et al. (2003) did not find an effect of digit load on syntactic processing in their acceptability/plausibility task.

In summary, the single and dual resource perspectives make incompatible claims about the recruitment of resource(s) during syntactic processing that should be easily distinguishable with the use of syntactic processing and WMC tasks, yet use of these tasks has generated contradictory empirical evidence. This is quite clearly problematic for both theories.

One explanation for the contradictory empirical results may be that the cumulative effects of a number of methodological differences have rendered them incapable of being directly compared with each other. For example, single resource studies typically utilize a comprehension judgment (e.g., Just and Carpenter, 1992), but dual resource studies typically use an acceptability/plausibility judgment (e.g., Waters and Caplan, 1996b). While Waters and Caplan (1996b) justify their judgment type by arguing it is more online than comprehension, no study has yet directly compared the two judgment types to test this argument, nor is it clear what effects this difference may have on the very nature of the syntactic processing task.

In addition to judgment types, single and dual resource studies vary in the types of sentences included in the syntactic processing task and in how those sentences are presented. In dual resource theory (and its studies), all complex structures are considered roughly equivalent in how much harder they are than simple structures, so it should not matter precisely how one constructs an object-relative sentence. In single resource theory, however, there is reason to believe that things like word order, subject animacy, and verb bias affect the processing load of the sentence, and so it stands to reason that such factors must be taken into account when constructing object-relatives sentences for any syntactic processing task designed to fairly distinguish between the two theories. Furthermore, careful consideration needs to be given to how this fair syntactic processing task should be presented. As will be discussed in more depth below, single and dual resource studies have varied in how precisely (if at all) they have measured online processing both in terms of the tools used (e.g., eye-tracker vs. a computer recording reaction times) and in terms of how they have broken up sentences for presentation (e.g., whole sentences, phrases, word-by-word). It stands to reason that this variability is an impediment to any definitive comparisons, most especially when one is trying to compare first-pass reading times directly measured (e.g., Just and Carpenter, 1992) with per-word reading times estimated from judgment reaction times (e.g., Waters and Caplan, 1996b).

A second issue to consider is that of how the data from syntactic processing tasks is examined. For one thing, it is typically the case that reading times are only examined for correctly judged items. Because low-spans have much lower accuracy than high-spans, this means more of their data gets dropped. This could be one reason for the dearth of significant three-way interactions cited by Waters and Caplan (1996a) as a criticism of single resource theory (inaccurately so in the case of MacDonald et al., 1992). Indeed King and Just's (1991) finding that 11 of 22 low-spans were 'non-comprehenders' of complex sentences suggests exactly the disproportionate effect of syntactic complexity on low-span individuals that the three-way interaction is designed to test for. Another problem is that studies vary in exactly what measures they use (accuracy, proportion, A', etc.), an issue to be revisited below.

Collectively, these concerns regarding methodological and data examination choices suggest that it is premature to conclude that the empirical evidence for each theory is equally strong. Rather, it seems to be the case that the jury must remain out on the question of what resource(s) are required for syntactic processing until such time as empirical evidence is provided that speaks to the effects, both individual and cumulative, of these issues. Thus, the current work serves two purposes. The first purpose is to explain and critically examine the methodological differences between core studies from the single and dual resource perspectives. The second purpose is to use this examination to partially replicate key findings of previous studies from both perspectives, thus allowing for direct comparisons of the resulting data. While these comparisons do not definitively settle the core theoretical question at hand, they are at the very least a solid first step in doing so.

CHAPTER 2. METHODOLOGICAL PROBLEMS IN THE LITERATURE

2.1 Single Resource Studies

The single resource theory was proposed based upon the results of three core studies: King and Just (1991), MacDonald et al. (1992), and Just and Carpenter (1992). All three studies used a syntactic processing measure with a comprehension judgment as well as the original Daneman and Carpenter (1980) reading span task, and found that those low and high in verbal working memory capacity differed in their accuracy and reading times on the syntactic processing task. In this section we will review the methodologies of these studies in some depth in order to later highlight the many ways in which single resource studies differ from dual resource studies.

In their first experiment, King and Just (1991) manipulated syntactic complexity by creating target sentences with either a subject-relative clause (e.g., “The reporter that attacked the senator admitted the error publicly after the hearing.”), or an object-relative clause (e.g., “The reporter that the senator attacked admitted the error publicly after the hearing.”). Storage load was also manipulated, by having participants read sets of one, two, or three sentences in which the target sentence was always the final sentence of the set (i.e., storage load of 0, 1, or 2 words). After each target sentence, participants answered a true/false comprehension question about it and then reported all sentence-final words for the set. All sentences were presented visually, one word at a time. Participants clicked a handheld switch in order to advance to the next word, giving a measure of per-word reading time.

In addition to the sentence comprehension task, participants completed the Daneman and Carpenter (1992) reading span task and were classified into span groups (low and high) based upon their scores. As mentioned previously, in this task participants read sets of two to six sentences before attempting to recall all sentence-final words at the end of the set. Span scores are the highest set size for which participants can correctly recall all words for three of five sets.

In line with single resource theory, the researchers hypothesized that when the storage and processing demands of sentences were too high for an individual’s capacity, performance would suffer. In other words, they expected that low-span participants would have lower comprehension accuracy on complex object-relative sentences as well as longer reading times in critical areas where syntactic processing demands are highest (i.e., at the relative clause ending and at the main verb). Additionally, they expected to find that extra storage load would impair the performance of high-span individuals such that their accuracy and reading times would look like low-span participants not under load. Load was also expected to impact low-span participants, spurring even greater declines in performance up to the point of complete failure to process.

Results showed that low-span participants were significantly lower in comprehension accuracy than high-spans for both sentence types, but especially so for the complex object-relative sentences. In fact, low-span participants’ comprehension was so low for object-relative sentences that the researchers divided participants into two groups: comprehenders and non-comprehenders. Non-comprehenders were defined as those participants who scored within a 95% confidence interval of chance performance. Eleven of the 22 low-span participants were classified as non-comprehenders according to this cut-off, whereas only two of the 24 high-span participants were. Load also affected comprehension performance, with accuracy decreasing as load increased for all participants in complex object-relative sentences, and for low-span participants even in simple subject-relatives.

Per-word reading times in target sentences where a correct comprehension judgment was made were analyzed at four points (the beginning of the relative clause, the end of the relative clause, the main verb, and the final words of the sentence) in order to examine potential differences between span groups in critical areas. Critical areas are those points at which greater load is placed on processing resources by the demands of the syntactic structure. The relative clause-ending area is critical in both subject- and object-relative sentences because it is the point of clause resolution. Additionally, the main verb is a critical area for object- but not subject-relative sentences because extra processing must occur in order to correctly assign the verb to its subject. Overall, low-span comprehenders had longer reading times at all points of the target sentences, but these differences were especially large (and statistically significant) in the critical areas, with low-span comprehenders taking 60-ms longer at the critical area of subject-relative sentences and 100-ms longer at each of the two critical areas in object-relative sentences. Low-span non-comprehenders did not show these differences in the critical areas, suggesting they did not spend the time required to process these demanding sentences, resulting in chance comprehension performance.

Taken together, the results of the first experiment support the single resource theory of syntactic processing by showing accuracy declines and reading times increase as the processing and storage demands of sentences are increased by syntactic complexity and storage load, and this is especially so for participants who are already low in available resources (i.e., low-span).

In their second experiment, King and Just (1991) were interested in determining whether or not low-span individuals could take advantage of an additional cue within sentences to compensate for their limited resources and thereby improve comprehension performance for complex sentences. Object-relative sentences were created in which either both verbs (e.g., “The robber that the fireman rescued stole the jewelry.”), the main verb only (e.g., “The robber that the fireman detested stole the jewelry.”), the relative verb only (e.g., “The robber that the fireman rescued watched the program.”), or neither verb (e.g., “The robber that the fireman detested watched the program.”) were pragmatically biased by their likelihood of being associated with their subjects (i.e., “rescue” is more likely to be associated with “fireman”, while “stole” is more likely to be associated with “robber”).

Each sentence was presented in the same self-paced word-by-word manner as in the first experiment, but this time target sentences were always presented individually, eliminating the extra storage load factor. Participants once again answered true/false comprehension questions following each sentence. For half the sentences, the true/false question tested comprehension of the relative clause while for the other half, the main clause was tested. Participants also completed the reading span task (Daneman & Carpenter, 1980) and were classified into low- and high-span groups based upon their scores.

The researchers reasoned that low-span participants ought to be able to use pragmatically biased verbs in place of complex syntactic computations to improve their comprehension accuracy because simple retrieval of pragmatic associations between nouns and verbs should be less resource demanding, especially when the cue was provided at exactly that point in the sentence where it would be most useful, eliminating any extra storage demands. Accordingly, it was expected that low-span participants’ comprehension accuracy would be equivalent to that of high-spans when both verbs were biased, but lower when neither verb was biased. In sentences where only one verb was biased, it was predicted that only the relative verb would improve performance because it, unlike the main verb, occurred at exactly the point it where its pragmatic association could be most useful (i.e., the end of the relative clause). This pattern of results would contradict the dual resource theory, because this theory would predict that even complex syntactic processing should be automatic and obligatory, and therefore it should not be the case that low-span individuals can forgo it and use pragmatic cues instead. In terms of reading time, it was not necessarily expected that the pragmatic cue would have much impact in terms of speeding processing in the critical areas because it might not be incorporated rapidly enough to show online effects. A lack of such an online effect may be seen as compatible with the dual resource theory in that it could be argued that any comprehension increases are due to offline post-interpretive processes rather than a trade-off between syntactic and pragmatic processing online. Finally, no effect of the pragmatic cue was expected in the high-span group because these participants should not need to use it, as they have sufficient capacity to handle the syntactic computations of the complex sentences.

The comprehension accuracy results showed that overall, low-span participants performed worse than high-span participants, though all participants showed high comprehension accuracy when the true/false question was testing comprehension of the main clause. This high accuracy was somewhat lessened when neither verb was pragmatically biased (around 80%). When the relative clause was tested, the accuracy of high-span participants remained high in all verb biasing conditions, but low-span participants' accuracy was high only when the relative verb or both verbs were pragmatically biased. In fact, when pragmatic bias was either absent or placed on the main verb, low-span individuals' accuracy dropped significantly (by 10 to 15 percent).

Per-word reading times were examined for the same four sentence areas as in the previous experiment and reported collapsed across verb biasing conditions. The only significant difference between span groups occurred at the second critical area (i.e., the main verb) where low-span participants took 116-ms longer than high-span participants. There was no longer a significant difference at the clause-ending area. Recall from the first experiment that the clause-ending area was considered to be critical in both subject- and object-relative sentences, whereas the main verb area is only critical in the object-relative sentences. This suggests that low-span participants were able to utilize the pragmatic cue to reduce the load associated with the relative clause, without affecting the load associated with processing at the main verb. This finding contradicts the dual resource theory, which would explain the increased comprehension performance as a result of pragmatic cues being used in post-interpretive processing rather than online. Thus, both the comprehension and reading time results support the single resource theory.

Taken together, the results of both experiments show that low-span individuals are reliably slower than high-spans at processing sentence areas with high syntactic demands as well as less accurate in comprehension performance, especially when the relative clauses is being tested and there are no readily available pragmatic cues to assist in interpretation. These results are consistent with the single resource theory of syntactic processing. However, because there was no three-way interaction between complexity, phrase, and span group, Waters and Caplan (1996a) argue that these results are actually compatible with the dual resource view.

In a second study testing the single resource model, MacDonald et al. (1992) manipulated syntactic complexity in the sentence comprehension task by crossing two factors: clause resolution and ambiguity. Clause resolution refers to the fact that for half the sentences, the first verb was part of the main clause while for the other half, the first verb was part of the relative clause. Of these two resolution types, main clause resolution is the more frequently used and is thus the interpretation that is more likely to be initially activated when processing a sentence. The second factor, ambiguity, refers to whether or not the first verb was amenable to both a main clause and a relative clause resolution. In unambiguous sentences, the first verb was either clearly part of the main clause (e.g., "The experienced soldiers spoke about the dangers before the midnight raid.") or clearly part of the relative clause (e.g., "The experienced soldiers who were told about the dangers conducted the midnight raid."). In ambiguous sentences, however, the first verb could be interpreted as either a past tense main clause verb (e.g., "The experienced soldiers warned about the dangers before the midnight raid.") or a past participle introducing a reduced relative clause (e.g., "The experienced soldiers warned about the dangers conducted the midnight raid."). Which interpretation is correct is unclear until one encounters either the prepositional phrase "before the midnight raid" or the second, unambiguous verb "conducted." Because they allow for two interpretations, ambiguous sentences should be generally more difficult to process than unambiguous sentences.

All four sentence types (unambiguous main clause, unambiguous relative clause, ambiguous main clause, and ambiguous relative clause) were presented in a self-paced, word-by-word manner and participants answered a yes/no comprehension question at the end of each. In addition to the sentence comprehension task, working memory capacity was measured using the reading span task (Daneman & Carpenter, 1980).

The researchers were most interested in performance differences between low- and high-span participants on ambiguous, main clause resolved sentences. Because main clause resolution is the preferred initial interpretation of a sentence, it should be the case that unambiguous versions of these sentences are processed quickly and with high accuracy. However, the researchers hypothesized that when main clause resolution sentences contained ambiguous first verbs, individuals with sufficient capacity (i.e., high-span) would also activate and maintain the less preferred, relative clause interpretation until disambiguating information was provided, whereas those without sufficient capacity (i.e., low-span) would not. This activation of alternative interpretations in high-span participants was expected to be associated with a cost in terms of longer reading times for the disambiguating region of these sentences, but both span groups were expected to be highly accurate in terms of comprehension judgments because of the main clause resolution.

High-span participants were also expected to have longer reading times for ambiguous, relative clause resolved sentences; however, because the less preferred resolution turns out to be the correct one in these sentences and low-span individuals do not have this interpretation at the ready, the increased reading times were expected to be accompanied by significantly higher accuracy rates in the high-span group. No span group differences were predicted for unambiguous versions of relative clause resolved sentences.

The comprehension accuracy results showed that all participants had higher error rates on ambiguous than on unambiguous sentences, as expected. In main verb resolved sentences, accuracy did not differ between span groups on either unambiguous or ambiguous sentences. There was also no effect of span for unambiguous sentences with relative clause resolution. For ambiguous relative clause resolved sentences, however, there was a significant difference between low- and high-spans, with the latter showing significantly lower error rates.

Per-word reading times were examined for three sentence areas: the first verb, the disambiguating region, and the last word of the sentence. The critical region was defined as the region where the disambiguating information was provided, as this area should trigger the processing required to choose an interpretation for the sentence. High-span reading times were approximately 160-ms longer than low-span times for ambiguous sentences with main clause resolution and approximately 130-ms longer for ambiguous relative clause resolved sentences.

However, these increased reading times were found at the last word of the sentence rather than at the area defined as critical by the researchers. Despite their location, it is important to note that these reading times did show the critical three-way interaction (complexity x phrase x span group) that supports the single rather than dual resource theory.

The accuracy rates clearly support the single resource theory by showing lower performance on complex structures (i.e., ambiguous and relative clause resolved sentences) for those low in capacity. Additionally, although reading times did not increase exactly where expected, these results do support the single resource model by supporting the contention that those with the capacity to do so (i.e., high-span participants) activate multiple syntactic interpretations in order to correctly process ambiguous sentences, whereas those without such capacity (i.e., low-span) do not, and by exhibiting the critical three-way interaction.

A second experiment was conducted to rule out the possibility that the high-span performance was due to either a task-specific strategy activated by the knowledge that at least some sentences in the task did have a relative clause resolution or to linguistic characteristics of the verbs, such as permitted argument structures (i.e., the number of elements that may be involved in the action of the verb; MacDonald, et al., 1992). Strategy use was controlled by presenting high-span participants with only main clause resolution sentences. Verb characteristics were controlled by constraining argument structures with the inclusion of proper nouns in half the sentences, thereby ruling out relative clause interpretations (e.g., "Colonel Wilson spoke/warned about the dangers before the midnight raid."). High-span participants still spent more time reading at the end of ambiguous sentences than at the end of unambiguous sentences, though this effect was attenuated in sentences with proper nouns. The authors interpret these results as ruling out a strategy development explanation as well as indicating that it is the possibility of a relative clause interpretation (provided by the common noun) that explains performance changes for high-span participants, rather than specific properties of the ambiguous verbs used in the first experiment.

Finally, MacDonald et al. (1992) conducted a third experiment to determine whether or not participants would activate and maintain dual interpretations for ambiguous sentences when overall storage load between the verb and disambiguating information was increased by increasing sentence length. Stimulus sentences from the first experiment were modified to create long versions by adding three words to each (e.g., Short version: "The experienced soldiers warned about the attacks conducted the midnight raid."; Long version: "The experienced soldiers warned about the surprise enemy guerilla attacks conducted the midnight raid."). The sentence comprehension task was otherwise identical to that in the first experiment. As before, participants were divided into span groups based upon their scores on the reading span task (Daneman & Carpenter, 1980).

The results of the first two experiments were replicated for the short version sentences. That is, high-span participants had longer reading times for both main and relative clause resolved ambiguous sentences, as well as higher comprehension accuracy for relative clause resolved ambiguous sentences. The effect of ambiguity on reading times for ambiguous sentences virtually disappeared in the long version sentences, such that high-span participants no longer reliably differed from low-span participants, suggesting that high-spans were no longer maintaining the alternative interpretation that had previously slowed processing of these sentences. As expected, accuracy for long main clause sentences was equivalent across groups, yet it was still the case that high-spans showed higher comprehension accuracy for ambiguous relative clause sentences. This finding is difficult to reconcile, if the assumption is that participants are only more accurate on relative clause sentences if they can maintain both interpretations, yet the main clause reading time results suggest that high-span individuals were no longer doing so. One possible explanation for this may be that high-span participants could more quickly re-activate the relative clause interpretation because it had previously been activated before being abandoned, and that this was what allowed accuracy to remain high. Despite this potential anomaly in the results, the researchers concluded on the basis of the overall reading time data that this experiment provided evidence that that high-span individuals do abandon the alternative interpretation when their capacity is further taxed by the storage demands of the sentence, a conclusion in line with predictions of the single resource model.

Taken together, the results of the three experiments of MacDonald et al. (1992) support the single resource view of syntactic processing by suggesting that high-span participants engage in more demanding syntactic processing (i.e., activating dual representations) in order to achieve higher comprehension accuracy than low-span individuals.

The studies by King and Just (1991) and MacDonald et al. (1992) both support the single resource theory by demonstrating that span groups formed using scores on the standard reading span task (Daneman & Carpenter, 1980) differ in their performance on comprehension tasks that manipulate the activation demands during sentence processing through syntactic complexity and storage load, in terms of both accuracy and reading times. This relationship between the two tasks contradicts the idea that syntactic processing is performed by an isolated resource (svWM) (Caplan et al., 2007). To further demonstrate this point, Just and Carpenter (1992) conducted an experiment to replicate a prior study by Ferreira and Clifton (1986) on the use of pragmatic cues during garden path sentence processing. In the original study, participants read garden path sentences (i.e., sentences with reduced relative clauses and ambiguous first verbs, similar to those found in MacDonald et al., 1992) containing head nouns that were either animate or inanimate (e.g., “The defendant/evidence examined by the lawyer shocked the jury.”). Animacy is a pragmatic cue (i.e., a non-syntactic cue) that constrains the potential interpretations of syntactically ambiguous sentences, and thus should enable participants to avoid being led down the garden path (i.e., to an incorrect interpretation of the sentence) during the comprehension task if they are able to use it during online processing. Specifically, inanimate head nouns, in combination with the verbs used in the task sentences, should trigger the object-relative clause interpretation from the start of the sentence, resulting in faster and more accurate processing. Ferreira and Clifton (1980) were surprised to discover that their participants did not make use of the animacy cue to disambiguate sentences, instead showing evidence of following the garden path in the form of longer first-pass reading times for later portions of sentences where disambiguating syntactic information was provided. This result supports the core idea of the dual resource theory that syntactic processing is modular with respect to other types of linguistic information. At issue, however, is the fact that Ferreira and Clifton (1980) did not measure working memory capacity and thus had no way of determining if individual differences in capacity affected performance. In their replication, Just and Carpenter (1992) addressed this issue by having participants complete the standard reading span task (Daneman & Carpenter, 1980) and dividing participants into low- and high-span groups based upon their scores. Just and Carpenter (1992) also made slight alterations to sentence comprehension task, by adding sentences with unreduced relative clauses as a simpler, control structure, and by excluding any sentences from the original task where the grammatical subject of the sentence could also be interpreted as an instrument (i.e., sentences where the subject could be the one doing the action or the means by which the action is accomplished). All of the sentences were presented all at once on a computer screen, with an eye-tracker recording fixations during reading. For each sentence, participants answered a true/false comprehension question.

The primary variable of interest in this study was first-pass reading times. Overall, it was expected that reading times would be longer for syntactically more complex garden path sentences than for non-garden path sentences for both span groups. It was also expected that high- but not low-span individuals would be able to take advantage of the animacy cue in garden path sentences and thus show faster initial reading times on the disambiguating portion of these sentences (i.e., the high span individuals would have used the inanimacy of the first noun to correctly parse the sentence as a reduced relative on the first pass).

Fixation durations were examined for three sentence areas: the *by* phrase (e.g., "by the lawyer", the initial verb (e.g., "examined"), and the main verb (e.g., "shocked"). The *by* phrase was considered the most critical area, as this is where disambiguating information is provided and thus where processing demands should be highest for those who are unable to use the animacy of the head noun to initially construct a correct interpretation for the sentence. Overall, non-garden path sentence reading times were faster than those for garden path sentences. First-pass reading times were also faster for sentences containing an inanimate head noun than for those with animate head nouns, but only for high-span participants; low-span participants did not show an effect of animacy. In subsequent reading times, this interaction of span and animacy at the *by* phrase was no longer present.

These results support the single resource theory by demonstrating that high-span participants can use non-syntactic pragmatic cues to correctly select a complex relative clause interpretation during first-pass reading, whereas low-span individuals cannot use these cues until later in processing. In other words, syntactic information and pragmatic information may only appear modular when one does not have the capacity required to process both in parallel. This conclusion is also supported by the second experiment of MacDonald et al. (1992), where proper nouns were shown to also speed the processing of syntactically ambiguous sentences for high-span participants. However, there is an apparent contradiction when one compares the King and Just (1991) experiment to the Just and Carpenter (1992) experiment, as pointed out by Waters & Caplan (1996a), that must be discussed further. In King and Just (1991), low- but not high-span participants used the pragmatic cue online, whereas in Just and Carpenter (1992) the high- but not low-span participants did so. One explanation may be that high-span participants did actually use the pragmatic cue in King and Just (1991), but their performance was already so high (above 80% regardless of the phrase being tested or the presence of pragmatic cues) that its use was not reflected in their accuracy rates. Another explanation might be that the two types of pragmatic cues are not directly comparable due to differences in the demands of the sentences containing them. Both studies used object-relative sentences, which are more complex than subject-relatives, but while Just and Carpenter (1992) further increased complexity by reducing the relative clauses, King and Just (1991) decreased complexity by using unreduced clauses. Thus, it could be the case that the combined results of these studies simply show that low-span participants can use pragmatic cues in moderately complex (i.e., unreduced object-relative) but not very complex (i.e., reduced object-relative) sentences because of the increased demands. A second stimulus factor to consider is the placement of the pragmatic cues. King and Just (1991) used pragmatically biased verbs placed near the ends of their sentences deliberately so that they might contribute to processing exactly when needed, while Just and Carpenter (1992) manipulated animacy of head nouns that were found at the very beginning of sentences, necessitating maintenance of the animacy cue for the full duration of the sentence. This difference in storage load may have made the verb bias cue less demanding to use than the animacy cue, and thus explain why low-spans could use one but not the other.

2.2 Dual Resource Studies

The evidence reviewed so far has supported Just and Carpenter's (1992) single resource theory by demonstrating that when activation demands during syntactic processing strain or exceed the individual's available capacity, comprehension accuracy declines while reading and reaction times slow (although in some cases slowing can be associated with higher comprehension, e.g., MacDonald et al., 1992). Still, not everyone agrees with the single resource explanation for these results. One problem raised by Waters and Caplan (1996a) is that these studies have failed to find statistically significant three-way interactions that would show a disproportionate effect of syntactic complexity on low-span participants' reading times in critical sentence areas. This argument is technically accurate in regards to King and Just (1991) and Just and Carpenter (1992),though it could be argued that the finding in King and Just (1991) that half the low-span participants were so-called "non-comprehenders", while only two in the high-span group were, could be classified as a disproportionate effect despite the fact that reading times for correctly comprehended sentences only did not show a significant three-way interaction, and MacDonald et al. (1992) did find a significant three-way interaction of ambiguity, phrase, and WMC.

Aside from statistical concerns about the single resource data, Waters and Caplan (1996a) have also argued on theoretical grounds that any relationship found between sentence comprehension tasks and the original reading span task (Daneman & Carpenter, 1980) do not necessarily reflect a shared resource for interpretive processing (svWM), but rather could be attributed to recruitment of a more general resource used for post-interpretive processing (gvWM) because of the nature of the offline comprehension questions, which presumably require further processing to complete. While this argument merits further exploration, in that it does seem likely that the demands of syntactic processing task judgments could affect results, it should be noted that it does not account for the fact that online measurement of per-word/first pass reading times in the Just and Carpenter (1992) showed an effect of working memory capacity, as measured by the reading span task.

In support of these criticisms of single resource theory, dual resource theorists have conducted their own studies to test the relationship between measures of syntactic processing and measures of WMC by modifying both types of tasks in several potentially important ways. As with the single resource studies, we will review three core dual resource studies in some depth in order to highlight the methodological contrasts between studies from the two perspectives. Waters and Caplan (1996b), Waters, Caplan, Alpert, and Stanczak (2003), and Waters and Caplan (2004) all used modified versions of the sentence comprehension and reading span tasks to examine syntactic processing resource(s). Although these studies all found main effects of both syntactic complexity and working memory span, none demonstrated the vital three-way interaction that would support the single resource prediction of a disproportionate effect of complexity in low-span groups, nor did they find that additional storage load harmed syntactic processing, as one would expect under single resource theory.

In the first of the core dual resource studies, Waters and Caplan (1996b) were interested in replicating the results of MacDonald et al. (1992), but with certain changes in methodology. Recall that in MacDonald et al. (1992) high-span participants had longer reading times, but also higher comprehension rates than low-span participants when sentences were temporarily syntactically ambiguous because they contained first verbs that could be interpreted either as a past tense main verb or as a past participle introducing a reduced relative clause. Waters and Caplan (1996b) also included reduced relative clause sentences (e.g., “The horse raced past the barn fell.”) in their syntactic processing measure, as well as two additional types of syntactically ambiguous sentences, reduced sentential complements (e.g., “The defendant confided to the lawyer he admired the judge was his brother.”) and reduced embedded clauses (e.g., “The picture books were lying beside was a landscape.”). Non-garden path sentences were created for these same clause types by including either unreducing clauses (referred to as “Non-garden path A”; e.g., “The farmer who was leased a tractor ploughed.”) or non-ambiguous first verbs (referred to as “Non-garden path B”; e.g., “The boy given a cookie was hungry.”).

In addition to these changes to syntactic structure, Waters and Caplan (1996b) modified the syntactic processing measure by having participants complete acceptability judgments (i.e., responding “yes” when sentences made sense and “no” when they did not) for each sentence rather than answer a comprehension question. Unacceptable sentences were created for each of the clause types described above by inserting semantically incongruent words into garden path (e.g., “The cookie fried the oven ate.”) and non-garden path A (e.g., “The waiter who was tipped a wig quit.”) sentences.

Finally, Waters and Caplan (1996b) also modified the method of presentation for the syntactic processing task. In their first experiment, participants between 50 and 80 years of age were presented with the sentence comprehension task in one of two ways: whole sentence or RSVP (rapid serial visual presentation). In the whole sentence condition, each sentence remained displayed on the computer screen in full until participants made their acceptability judgments. In the RSVP condition, each successive word appeared for 250-ms and participants made their judgments after the last word of each sentence had disappeared.

Participants also completed a modified version of the reading span task consisting of semantically acceptable (e.g., “It was the man that clenched the pillow.”) and unacceptable (e.g., “It was the toy that clenched the man.”) cleft subject sentences divided into sets from two to six. Each set size was presented five times. For each sentence in a set, participants made an acceptability judgment. After the last sentence of each set, participants were prompted to recall the last word of each sentence in the set. Participants were then divided into low-, medium-, and high-span groups based upon their scores on the modified reading span task.

In the whole sentence condition, the researchers expected that acceptability judgment accuracy would be lower for garden path than for non-garden path sentences, and that high-span participants would have higher accuracy than low-span participants on garden path sentences. These predictions are compatible with both the single and dual resource theories. Predictions for the RSVP condition are generally the same, that garden path performance should be less accurate than non-garden path and that low-spans should be less accurate than high, but if single resource theory is accurate, then it should be the case that the effect of working memory span is much larger in the RSVP condition than in the whole sentence condition. This is because it has been argued that the speed of the RSVP condition decreases processing resources (Miyake, Carpenter, & Just, 1994) while also increasing storage load by disallowing reviewing of stimuli after the first pass (Waters & Caplan, 1996b).

For acceptability judgments, A' scores were calculated. A' is a nonparametric sensitivity index calculated by estimating the area under the receiver-operating characteristic (ROC) curve. An A' value of .5 would indicate chance performance, while a value of 1 would indicate perfect discrimination on a forced-choice test (Pollack & Norman, 1964). A' values were analyzed with a 3 (span) x 2 (presentation condition) x 3 (clause structure) x 3 (garden path status) ANOVA.

There were main effects for each factor: low-span participants were less sensitive than medium- or high-spans, sensitivity was lower in the RSVP than in the whole sentence condition, sensitivity was also lower for garden path than for non-garden path sentences, and sensitivity was lower for the complement clause structure than for the embedded and reduced relative structures. These main effects were qualified by significant two-way interactions. The structure by span interaction showed that only low-span participants were less sensitive on complement clause structures, whereas medium- and high-span participants were equally sensitive on all structures. The structure by garden path status interaction showed that sensitivity on non-garden path A and non-garden path B sentences did not differ in sentences with complement or embedded clause structures, but that in sentences with reduced relative structures, sensitivity was higher for non-garden path A. All other interactions were non-significant.

In addition to analyses of acceptability sensitivity, reaction times for sentences correctly judged as acceptable were examined within presentation conditions. In the whole sentence condition reaction times, which included the reading time for the entire sentence, were divided by the number of words in each sentence to calculate a mean per-word reading time. Overall, mean reading times were longer for garden path sentences than for both types of non-garden path sentences in this condition. For embedded clause structure sentences, there was also a difference between non-garden path sentence types such that mean reading times were longer for non-garden path A sentences. In the RSVP condition, reaction times to the final word of each sentence, which included the reading time for that word, were examined. Reaction times for garden path sentences were longer than for non-garden path sentences for all clause types. For reduced relative clause sentences only, reaction times for non-garden path A sentences were longer than those for non-garden path B sentences. There were no effects of working memory span group for either condition.

Together, the accuracy and reaction time results confirmed that garden path sentences are more difficult to process than non-garden path, that the RSVP condition was more difficult than the whole sentence condition, and that low-span participants perform lower overall than medium- and high-span participants. However, because the low-span participants were not differentially impaired in the RSVP condition, the authors conclude that their lower performance could not be attributed to differences in syntactic processing and thus the results are consistent with the dual rather than single resource view.

Anticipating one potential criticism of the first experiment, that it was conducted with older adults, Waters and Caplan (1996b) conducted two further studies with college-aged adults. In these experiments, participants completed the same sentence comprehension task described above, but only under RSVP presentation conditions. In one of these experiments, RSVP presentation rate was set to 250-ms per word for one condition and to 170-ms per-word for another. In the next experiment, RSVP presentation rate was set to 120-ms per word. These speeded up rates were used in order to rule out the possibility that the RSVP condition was not fast enough in the first experiment to produce the interaction predicted by the single resource theory. In both follow-up experiments, participants were divided into high- and low-span groups based upon scores of the standard reading span task (Daneman & Carpenter, 1980) rather than the modified version from the first experiment.

Acceptability results were similar to those in the first experiment. Low-span participants' sensitivity scores were lower than those of high-span participants, meaning they had a more difficult time discriminating acceptable from unacceptable sentences. In terms of structure, performance was lower for garden path than for non-garden path sentences, and for complement clause sentences than for the other two types of clauses. Interestingly, performance on garden path sentences was lower for the 250-ms than for the 170-ms RSVP condition.

Reaction times were also analyzed. At slower presentation rates (250-ms and 170-ms), reaction times were longer for garden path sentences than for non-garden path sentences with all clause structure types, and high-span participants had longer reaction times to garden path sentences than did low-spans. In the 250-ms condition only, reaction times to complement structures were longer than those to embedded and reduced relative clauses. In the fastest presentation rate (120-ms) condition, garden path sentence reaction times were longer than non-garden path reaction times for both reduced relative and embedded clause structures, but not complements. Also in this condition (120-ms), high-span participants showed longer reaction times to garden path sentences than to both types of non-garden path sentences, but low-span participants showed longer reaction times to garden path sentences only in comparison to non-garden path B sentences.

The fact that high-spans took longer, but were more accurate on garden path sentences than low-spans is consistent with MacDonald et al.'s (1992) finding that high-span participants maintain dual representations when possible in order to correctly parse ambiguous sentences. However, because low-span participants in this study were not disproportionately impaired by the faster RSVP rates, the author argue that the results are in fact consistent with the dual rather than single resource view.

One potential problem in comparing Waters and Caplan (1996b) to the single resource studies discussed above is that their measure of per-word reading time in the whole sentence condition was actually a mean created by dividing reaction times by the number of words in a sentence. Likewise in the RSVP condition, times were in fact reaction times recorded after presentation of the last word of each sentence. These methods are problematic for direct comparison in that they do not allow the measurement of individual differences in reading times at specific points within sentences and because there is a well-known wrap-up effect that is independent of syntactic complexity, whereby processing times at the ends of sentences are longer (Balogh, Zurif, Prather, Swinney, & Finkel, 1998). In a later study, Waters and Caplan (2004) did allow participants to control the rate at which stimuli the syntactic processing task were presented, though sentences were in phrases rather than word-by-word.

In the first experiment of Waters and Caplan (2004), the syntactic processing task contained items with four types of unreduced subordinate clauses: cleft-subject (e.g., "It was the food that nourished the child."), cleft-object (e.g., "It was the woman that the toy amazed."), subject-relative (e.g., "The father read the book that terrified the child.", and object-relative (e.g., "The man that the fire injured called the doctor."). Half the sentences were semantically plausible and the other half were not (e.g., "It was the car that drove the woman."). Sentences were presented auditorily, spliced into phrases. Participants had to push a button to hear successive phrases until the complete sentence was heard, at which time they made a plausibility judgment (plausible? yes/no). Participants also completed a modified reading span task, a subset of the task from Waters and Caplan (1996c). In this version of the reading span task, all sentences contained cleft-subject clauses and semantic plausibility was varied. Sentences were arranged in sets from two to six sentences long. Participants made plausibility judgments about each sentence and recalled sentence-final words after the last sentence in each set. In the task instructions, speed and accuracy on the plausibility judgment were stressed over storage of sentence-final words. Participants were divided into low-, medium-, and high-span groups based upon their scores on the reading span task.

According to Waters and Caplan (2004), single resource theory would predict a three-way interaction in this task, such that low-span participants should show longer listening times for computationally demanding portions of complex sentences than high-spans (i.e., a complexity x phrase x span group interaction). Dual resource theory would predict no such interaction.

Listening times for trials where correct plausibility judgments were made were analyzed for two syntactic complexity comparisons: cleft-subject vs. cleft-object and subject-relative vs. object-relative (i.e., simple vs. complex comparisons). For both comparisons, listening times were longer for the complex structure (i.e., cleft-object and object-relative) than for the simple structure (i.e., cleft-subject and subject-relative) at the most capacity-demanding portion of the sentences (i.e., the verb). There was also evidence of the wrap-up effect, where listening times were longer for phrases at the end of sentences. In terms of group differences, low-span individuals had longer listening times for the complex cleft-object sentences than either medium- or high-span participants, but no such difference existed for the complex object-relative sentences. Low-span participants also showed longer listening times for the third noun phrase of simple cleft-subject sentences.

Accuracy rates on the plausibility judgments were also examined. These results showed that low-span participants were significantly less accurate than medium- and high-span individuals for both of the complex structures (cleft-object and object-relative) but not for the simple structures (cleft-subject and subject-relative).

Although this experiment did find differences on the syntactic processing task that were related to working memory span group, they argue that these differences do not support single resource theory for several reasons. The first reason is a criticism Waters and Caplan (1996a) have applied to Just and Carpenter (1992) data: that there was not a significant three-way interaction (Group x Sentence Type x Phrase). In other words, Waters and Caplan (2004) maintain that if the data were to support single resource theory, there should be evidence that syntactic complexity affects low-span individuals disproportionately more than it does medium- or high-span individuals. Another reason these data are interpreted as supporting dual rather than single resource theory is that the low-span participants showed longer reading times for sentences with both complex (cleft-object) and simple (cleft-subject) structures, albeit in different regions. It is argued that this result indicates low-span individuals found the task harder, but not syntactically harder. (This point is a confusing one to parse out, considering that the task was to process syntax. If processing the syntax was not the difficult portion, then one wonders what was difficult about the task, but this argument is not explained further.) Waters and Caplan (2004) also considered the possibility that better measures of working memory capacity might be required to find the relationship predicted by single resource theory. A second experiment was conducted to address this issue.

In the second experiment of Waters and Caplan (2004), participants completed the same syntactic processing task as in the first, with the exception that another simple structure was added: subject-subject (e.g., "The law that favored the millionaire frustrated the workers."). These types of sentences were used by others (e.g., King & Just, 1991) and are considered a better comparison structure for object-relative sentences (e.g., "The law that the millionaire favored frustrated the workers.") than the subject-relative structure from the first experiment (e.g., "The millionaire favored the law that frustrated the workers.") sentences because both require the verb to be associated with the head noun across an intervening relative clause.

In addition to the modified reading span task from the first experiment, three more working memory measures were administered. A second reading span task was added that differed from the first only in that it contained all complex rather than all simple sentences. The third working memory task was an alphabet span task (Craik, 1986) in which participants hear lists of words and are required to recall them in alphabetical order. The final working memory task was a subtract 2 span task (Salthouse, 1988) in which participants are required to repeat random sequences of digits after subtracting two from each. Participants were not separated into span groups in this experiment, instead scores on the individual tasks and a composite of the tasks were used to calculate correlations with performance on the syntactic processing task.

The predictions for this experiment were the same as the first. That is, if single resource theory is accurate, there was expected to be a three-way interaction. A lack of such an interaction would support the dual resource theory.

Listening times were examined only for those trials in which participants made correct plausibility judgments. Once again, listening time results confirmed that participants listened longer to complex than simple sentences (cleft-object longer than cleft-subject and subject-object longer than both object-subject and subject-subject). As participants were not divided into span groups, there were no data reported on specific listening times within groups. Instead, correlational analyses are reported. Scores on all four working memory measures as well as a composite score combining the four were analyzed for correlations with difference scores that were obtained by subtracting listening times at specific regions (i.e., verbs or noun phrases) in simple sentences from listening times in the same regions of complex sentences. None of these correlations were significant, ranging from .008 to .16 in magnitude.

Taken together, Waters and Caplan (2004) argue that these two experiments support a dual rather than single resource view. Again, the main reason for this is a lack of a three-way interaction (span group x sentence type x phrase) showing disproportionate declines in performance on complex sentences for low-span individuals in Experiment 1, but the conclusion is also drawn from the lack of significant correlations in Experiment 2.

Recall that comparing performance on simple and complex sentences is but one test of single resource theory. Just and Carpenter's (1992) theory also predicts that increasing storage load should affect performance during syntax processing. MacDonald et al. (1992) found that increasing the number of words in sentences caused high-span participants to abandon their strategy of holding alternate interpretations of ambiguous sentences in mind and King and Just (1991) found that performance on the comprehension task declined as the number of sentence-final words to be recalled increased. Waters, Caplan, and Yampolsky (2003) also decided to test this prediction of single resource theory, using digit load. In this experiment, participants completed the same syntactic processing task as that in Waters and Caplan (2004), presented in the same auditory moving window paradigm (i.e., self-paced by phrases). All participants completed the task under all three load conditions (no load, three digit load, and five digit load) in three counterbalanced sessions. Working memory capacity was not measured, so participants could not be divided into span groups. However, all participants have verbal working memory capacity, measured or not, and presenting each participant with each level of load should still provide a way to capture any potential effects of varying storage load on syntactic processing. On the other hand, some effects may be hidden by the lack of measurement, because low-span and high-span individuals might perform differently under each level of load.

The first analysis was on listening times for sentences that were correctly judged as plausible. For all load conditions, there was a main effect of complexity such that listening times were longer at the verb area of complex (i.e., cleft-object and object-relative) sentences than in simple (i.e., cleft-subject, subject-relative, and subject-subject) sentences. There was also a main effect of load for all sentences, such that listening times were longer under five digit load than under three digit load and longer in both load conditions than in the no-load condition. The three-way interaction of sentence type, phrase, and load condition was only significant when comparing object-relative to subject-relative sentences by subjects, but not by items. In terms of plausibility judgment accuracy, the only significant result was that accuracy was higher for the simple cleft-subject sentences than for the complex cleft-object sentences. There was no effect of load on plausibility judgments.

In light of the lack of evidence of a disproportionate effect of load on critical phrases (i.e., the verb area), Waters et al. (2003) argue that these results are consistent with a dual rather than single resource theory. One problem with this conclusion is that working memory capacity was not measured. Individuals vary in the capacity they have available. Someone with a lower capacity should show greater effects of three digit load than someone with a higher capacity according to single resource theory, but we have no way to tell if this occurred in the Waters et al. (2003) data. Furthermore, given that all sentences were unreduced, it could be the case that sentences were not sufficiently complex to produce three-way interactions in critical areas, even under load.

2.3 Neurological Evidence

The fact that all of the evidence reviewed to this point has been purely behavioral in nature, regardless of whether it has been produced by single or dual resource theorists, should not be taken as an indication that neurological predictions have not been made and tested by each theory. Indeed, studies of brain function should be at least as useful in distinguishing between the two theories as behavioral studies are because it should be possible to determine which brain area(s) are involved in syntactic processing. However, as with behavioral studies, the neurological evidence produced so far is somewhat mixed.

Functional dissociations exposed by brain disorders or trauma are one source of information about brain areas involved in a variety of tasks. Several studies from the dual resource perspective have suggested that patients with disorders which significantly impair domain general resources such as WMC (e.g., aphasia, Alzheimer's, etc.) retain their ability to process complex syntactic structures, suggesting that there is indeed a dissociation between domain-general resources (gvWM) and syntactic resources (svWM) (e.g., Rochon, Waters, & Caplan, 1994; see Waters and Caplan, 1996a for a review). However, other researchers have found that such patients do show deficits in syntactic processing if the items they are asked to process are of sufficient complexity (Almor, Kempler, MacDonald, Anderson, & Tyler, 1999), suggesting that the methodological issues here highlighted may have been at play in these investigations as well.

Aside from patient data, brain activation during syntactic processing can also be assessed with a variety of neuroscience techniques. For example, Martín-Loeches, Muñoz, Casado, Melcón, and Fernández-Frías (2005) looked at ERP waveforms in response to grammatical violations and complex grammatical structures presumed to tax WMC. Both manipulations elicited identifiable negativities in the waveform, and these negativities differed qualitatively from each other. However, the negativity associated with grammatical violations disappeared when complexity was highest, suggesting that processing of violations and structure are separable when demands are low but can interfere with each other when demands are high. This conclusion is supportive of the single resource theory, but is not particularly convincing on its own, as it does not take into account individual differences in either syntactic processing or WMC.

King and Kutas (1995) did assess individual differences by classifying participants as ‘poor’ or ‘good’ comprehenders on the basis of true/false comprehension item accuracy before looking at ERP waveforms in response to simple and complex sentences. The authors found a prolonged left anterior negativity (LAN) localized on the verbs of complex sentences (i.e., where processing load should be highest) that was exaggerated in ‘poor comprehenders’ compared to ‘good comprehenders.’ Multi-word analysis of three sentence regions ruled out the possibility that the LAN was related to lexical factors and instead supported the notion that complex grammatical structures load on WMC. Although this study did not assess working memory directly, the observed LAN was similar to that attributed to the effects of load on WMC in previous studies (e.g., Kluender & Kutas, 1993). Furthermore, we know that low-span individuals tend to show lower accuracy than high-spans overall (e.g., King & Just, 1991), which at least suggests that the ‘poor comprehenders’ were likely to also be low-spans.

Another ERP study by Vos, Gunter, Schriefers, and Friederici (2001) looked the relationship between WMC (measured by a reading span task) and the processing of complex grammar under concurrent load (monitoring for specific words) in native German speakers. The behavioral results showed that low-spans and high-span individuals under concurrent load slowed down and were less accurate on complex sentences. The ERP waveforms also differentiated between span groups in the windows from 200- to350-ms and 500- to 800-ms after the presentation of disambiguating information. In the earlier window, high-spans showed a posterior positivity while processing object-relative sentences that has previously been argued to reflect the recognition of a garden path sentence (Friederici, 1997), while in the later window there was a frontal positivity while processing object-relative sentences that has been argued to reflect the assignment of correct syntactic structure (Friederici, 1997). This frontal positivity was smaller for high- than for low-spans, suggesting that the high-spans were more efficient at using disambiguating information to assign the correct syntactic structure.

In the final ERP study to be discussed here, Vos, Gunter, Kolk, and Mulder (2001) measured WMC with an auditory version of a reading span task and manipulated syntactic complexity, grammaticality, and working memory load during a sentence comprehension task. Their behavioral results were in line with dual resource theory in that they found main effects of grammaticality and working memory load, but no significant interactions that would indicate span differences on syntactically complex sentences. ERP waveforms, on the other hand, presented a different picture. High-span participants showed an anterior negativity in the 250- to 450-ms window after the presentation of a critical verb that was not present in low-span waveforms. Furthermore, this negativity disappeared when high-spans processed the complex sentences under working memory load. A bit later, in the 500- to 800-ms window, a centroparietal positivity appeared in both low- and high-span waveforms; however, this positivity came later for both low-spans and for high-spans under high load. These results support the single resource theory first by showing that span groups differed and second by showing that working memory load caused high-spans to look like low-spans. Even more interestingly, the fact that span effects were seen in the ERP waveforms but not the behavioral data suggests that behavioral data are less sensitive to the theoretical relationship in question, suggesting one explanation for conflicts between single and dual resource studies that only utilize behavioral measures.

Other researchers have used other imaging techniques to examine brain activation during syntactic processing. For example, Makuuchi, Bahlmann, Anwander, and Friederici (2009) had participants read main clause and subject-relative sentences that varied in length (to manipulate storage load) in an fMRI study. They found two distinct areas of activation in the left inferior frontal gyrus, one associated with syntactic structure and the other associated with length, which seems to support the dual resource theory. However, the authors also found evidence of extensive connectivity between the 'distinct' areas, especially when processing the more complex sentences, suggesting the two interact when the task demands it. This connectivity, in combination with the fact that their 'complex' subject-relative sentences are actually commonly used as simple sentences in other studies (e.g., Just and Carpenter, 1992) and the fact that they did not actually assess individual differences in WMC, suggests that caution should be exercised when attempting to claim these results as support for the dual resource theory.

Finally, PET imaging was used by Waters, Caplan, Alpert, and Stanczak (2003) to examine changes in regional cerebral blood flow (rCBF) during processing of simple and complex sentences. Participants were classified into span groups on the basis of scores on four WMC measures and into 'speed of processing' groups based upon plausibility judgment response times during a syntactic processing task separate from the one in which complexity was manipulated. None of the behavioral results showed an effect of span group, nor were there any significant differences in rCBF between span groups, though low-spans did show a significant correlation between their reaction times and their rCBF in both the inferior frontal lobe and the right cingulate that was not seen in high-spans. Examined in terms of speed of processing, rCBF data showed that fast processors exhibited bilateral activation in the inferior frontal lobe while slow processors showed activation in the left superior temporal lobe and structures in the right hippocampus. Waters et al. (2003) argue that these results support the dual resource theory by demonstrating that performance on the syntactic processing task is related to speed of processing in general, but not to WMC. However, it should be noted that the syntactic processing stimuli used in this study are the same as those used in Waters and Caplan (2004). These sentences are less complex than those that were used to elicit span group differences in Just and Carpenter (1992) at least by virtue of being unreduced, and perhaps by virtue of being constructed in an entirely different manner (an issue to be explored in the experiments reported in Chapters 3, 4, and 5).

While this brief review of the neurological evidence is certainly not comprehensive, it should be sufficient to illustrate that the imaging literature is almost as contradictory on the issue of the resource(s) involved in syntactic processing as the behavioral literature reviewed before it. One could argue that this is perhaps due to the fact that imaging results are only as sound as the design of the behavioral task they are measuring brain activation for, and thus the same issues raised in relation to those tasks are thus a factor in the imaging studies as well.

2.4 Summary and Conclusions

In principle, providing empirical evidence that clearly distinguishes between single and dual resource theories of syntactic processing is a fairly straightforward task, considering the opposing predictions of the two models. All of the studies discussed thus far have set out to do just this, yet there remains disagreement about which theory more accurately describes the resources(s) required for the processing of syntax, with some studies supporting the single resource view (MacDonald et al., 1992; King & Just, 1991; Just & Carpenter, 1992) and others supporting the dual resources view (Waters & Caplan, 1996b; Waters & Caplan, 2004; Waters et al., 2003). Determining which set of studies provides stronger evidence for which theory is difficult, as there are many methodological differences that make direct comparisons questionable. These differences could be argued to individually affect results to varying degrees, in addition to producing a cumulative effect. Before discussing how to resolve these conflicts, methodological differences in both the syntactic processing and working memory tasks will be reviewed and their potential effects discussed.

One of the first comparison issues in the syntactic processing task is the type of judgment used. The single resource studies described above all use yes/no or true/false comprehension questions, whereas the dual resource studies use acceptability/plausibility judgments. The purpose of modifying the sentence comprehension task with acceptability judgments was to ensure that the task did not allow for the use of post-interpretive processes, as when comprehension questions are completed, in order to increase confidence that the task was measuring svWM rather than gvWM. There is reason to think that acceptability judgments should accomplish this goal because initial assignment of word meaning is assumed to be an interpretive process, under control of svWM. However, it is also important to note that there is psychophysiological evidence to suggest that semantic and syntactic incongruities are processed differently by the brain (e.g., Kolk, Chwilla, Van Herten, & Oor, 2003), suggesting that these acceptability judgments may not ensure measurement of purely syntactic processing. Given that the primary interest of these studies is to measure syntactic processing, it stands to reason that the most desirable measure is one that taps syntactic rather than semantic violations. For this reason, it seems plausible that an acceptable alternative to both comprehension and acceptability judgments is the grammaticality judgment. In grammaticality judgment, participants simply indicate whether or not a given sentence is grammatically correct. While such judgments are somewhat similar to acceptability, in that agrammatical sentences may not appear to make sense, they do not rely on sentences containing semantically implausible words. Rather, grammatical violations often consist of subject-verb disagreements (e.g., “They goes to the store.”). Previous studies have found that grammaticality judgments are related to performance on measures of working memory capacity and that performance on them suffers under conditions of cognitive load (McDonald, 2008a and 2008b).

A second methodological issue is the way in which syntactic processing items were presented to participants. In the single resource studies sentences were presented either word-by-word (King & Just, 1991; MacDonald et al, 1992) or all at once with an eye-tracker (Just and Carpenter, 1992), both allowing for precise online measurement. In Waters and Caplan (1996b), presentation was either controlled by the experimenter (RSVP conditions) or the participant (whole sentence condition), but there was never an online measure of reading time. Per-word times were estimated by dividing reading-plus-reaction times by the number of words in the sentence, but this method assumes a priori that reading time is equal for all words in the sentence and thus does not allow for analysis of any differences in critical areas. Waters and Caplan (2004) and Waters et al. (2003) did allow participants to self-pace presentation of sentences, however, sentences were presented auditorily rather than visually and in phrases rather than word-by-word. Artificial chunking on its own may wipe out span group differences by changing natural stimulus chunking strategies of low- and high-span individuals (e.g., Swets, Desmet, Hambrick, & Ferreira, 2007), but these methods are also problematic in that they do not provide the same level of precision in the measurement of per-word processing time as that found in Just and Carpenter (1992). Waters and Caplan (1996a) themselves argue that online measurement is vital to distinguishing between single and dual resource theories, so it stands to reason that this type of measurement is the one that should be compared across single and dual resource studies, and yet we cannot do so with any confidence because of the widely varying methods that have been used in the previous studies. Indeed, we have even seen evidence that offline behavioral and online ERP data can directly contradict each other in terms of which theory they support (e.g., Vos et al., 2002).

Another methodological issue with the syntactic processing tasks is that the items presented in dual resource studies differ from single resource studies in how their complex syntactic structures are formed. Complex object-relative sentences can be formed in several ways that may differ in terms of the demands they impose. Table 2.4.1 provides examples of comparable object-relative sentences from each of the six studies. The first thing to note about the sentences in Table 1 is that reducing relative clauses makes them more difficult to process. MacDonald et al. (1992), Just and Carpenter (1992) and Waters and Caplan (1996b) did reduce object-relative clauses, but the other three studies (King & Just, 1991; Waters et al., 2003; Waters and Caplan, 2004) did not. Also of note is the fact that the reduced relative clause sentences in Waters and Caplan (1996b) are quite a bit shorter than those in MacDonald et al. (1992), which would decrease storage load in these sentences.

Table 2.4.1. Comparable syntactic processing items by study.

	Study	Example Sentence	Length
Unreduced object-relative clauses	King & Just (1991)	The reporter that the senator attacked admitted the error publicly.	12-17
	MacDonald et al. (1992)	The experienced soldiers who were told about the dangers conducted the midnight raid.	11-13
	Just & Carpenter (1992)	The defendant/evidence that was examined by the lawyer shocked the jury.	9-11
	Waters & Caplan (1996b)	The farmer who was leased the barn ploughed. The ocean that the fish swam in was polluted.	7-8 8-9
	Waters & Caplan (2004)/ Waters et al. (2003)	The law that the millionaire favored frustrated the workers.	9 (mean)
Reduced object-relative clauses	MacDonald et al. (1992)	The experienced soldiers warned about the dangers conducted the midnight raid.	11-13
	Just & Carpenter (1992)	The defendant/evidence examined by the lawyer shocked the jury.	9-11
	Waters & Caplan (1996b)	The horse raced past the barn fell. The picture books were lying beside was a landscape.	7-8 8-9

Focusing on just the reduced clause sentences in Table 2.4.1, note that in both MacDonald et al. (1992) and Just and Carpenter (1992) these clauses always begin with an ambiguous verb. In the first example from Waters and Caplan (1996b), the clause also begins with an ambiguous verb, but they differ in terms of where the disambiguating information is presented. MacDonald et al. (1992) and Waters and Caplan (1996b) introduce disambiguating information with second, non-ambiguous verbs that occur near or at the end of sentences, whereas Just and Carpenter (1992) provided disambiguating information in the form of *by* phrases that were then followed by further information. This location difference is important because sentence wrap-up effects (Balogh et al., 1998) may obscure span group differences when disambiguating information is sentence-final, unless perhaps the sentence also contains sufficient load (i.e., more words) such as those of MacDonald et al. (1992). As for the second example sentence from Waters and Caplan (1996b), the reduced object-relative clause begins with a noun rather than an ambiguous verb, a difference which could affect complexity and by affecting the potential for ambiguity.

A final point about the examples found in Table 2.4.1 concerns the animacy of the head noun in ambiguous sentences. Specifically, Just and Carpenter (1992) only found span group differences in first-pass reading times when sentences had inanimate head nouns (i.e., only the high-span participants could use this cue to speed processing in ambiguous sentences). The Waters and Caplan (1996b) sentences that were formed in the same was as those of Just and Carpenter (1992; i.e., ambiguous verb first) only contain animate nouns (as far as can be determined from examples, full lists are not presented in the article nor is animacy discussed). Thus, it may not be at all surprising that Waters and Caplan (1996b) did not find effects of span group on these sentences.

All of these syntactic processing task stimuli differences may affect complexity in various ways. Precise control over syntactic complexity is key to distinguishing between the single and dual resource views because syntactic processing is such a well-practiced skill for all typically developing individuals, and such well-practiced skills may become automatized over time, requiring fewer resources to complete as expertise grows. Indeed, developmental evidence does suggest that automatization occurs for frequently used syntactic structures. In childhood, syntactic structures are generally acquired in order of increasing complexity. After years of experience, adults far outperform children in grammaticality judgments for all structures, but McDonald (2008b) showed that taxing the working memory capacities of adults with cognitive load resulted in declines in performance on more complex, later acquired structures (e.g., past tense and third-person plural) but not on simpler, earlier acquired structures (e.g., word order). In other words, finding syntactic processing differences predicted by single resource theory in adults requires the use of appropriately complex/rare structures because simpler/common structures are so well practiced that they can be processed accurately even under resource demanding circumstances.

Caplan, Waters, and Dede (2007) argue against the automatization of syntactic processing argument because of previously mentioned data from their own studies showing that even patients with disorders that significantly impair domain general resources (e.g., aphasia, Alzheimer's, etc.) can process complex syntactic structures (e.g., Caplan, et al., 1985; Rochon, et al., 1994), which they say suggests that the syntax specific resource (svWM) must be fully automatic and sufficient to process even the most complex sentences without recruitment of domain general resources. But as previously mentioned, others have found that such patients do show deficits in syntactic processing, if the items are of sufficient complexity (Almor, et al., 1999).

Aside from differences in the syntactic processing task, the previous studies have also showed significant variation in how they measure WMC. The single resource studies above all used the standard reading span task of Daneman and Carpenter (1980) while most dual resource experiments use a modified version. This modified version was developed by Waters and Caplan (1996c), who report that the task has higher internal consistency (between .92 and .95) as well as higher test-retest reliability (.65-.73) than the original task. However, they also report that the modified task only moderately correlates (.52-.63) with the original task, which is problematic when trying to compare across studies. In their second experiment Waters and Caplan (2004) did use other measures (i.e., alphabet span and subtract 2 span) in addition to a more complex version of their modified reading span task, and reported no significant correlations with their syntactic processing task. Using multiple working memory measures is a good idea both because it helps to ensure a more accurate score for individuals and because if the syntactic processing task really recruits a domain general resource, then it should be related to a variety of tasks measuring that resource. It is important, though, to fully analyze performance on the syntactic processing task in relation to working memory scores by using working memory measures to categorize individuals into groups, which Waters and Caplan (2004) did not do.

Taking into account all of the methodological differences between single and dual resource theory experiments, it seems clear that making any final conclusion about which side presents a more compelling argument for the resource(s) involved in syntactic processing would be premature. Before this important theoretical question can be answered, the effects of these methodological differences need to be thoroughly evaluated. This is precisely the purpose of the three current experiments.

In the following three experiments, we attempt to partially replicate the work of Just and Carpenter (1992), Waters and Caplan (1996b), and Waters and Caplan (2004). One set of stimulus sentences was created in the style of each of these three studies. Each set contained sentences representing four levels of complexity: main clause, subject-relative clause, unreduced object-relative clause, and reduced object-relative clause. We also manipulated animacy at each level of complexity, such that half of each sentence type contained an inanimate subject and the other half contained an animate one. Finally, we also manipulated the type of judgment made about the sentences. In Experiment 1, participants answered true/false items after all sentences in the manner of Just and Carpenter (1992). In Experiment 2, participants made acceptability judgments about sentences, in the manner of Waters and Caplan (1996b) and Waters and Caplan (2004). In Experiment 3, we test a judgment type not previously utilized by single or dual resource researchers, the grammaticality task. For all experiments, sentences were presented in a self-paced word-by-word paradigm, thus we have three types of dependent variable to look at: reading times, judgment response times, and judgment accuracy.

In addition to measuring syntactic processing, we also administered measures of WMC. First, we included the original Daneman and Carpenter (1980) Reading-span task because it has been used in so many single resource studies. However, because Waters and Caplan (1996c) report that this task is not strongly correlated with other measures they have administered, and because the single resource theory predicts that the resource used for syntactic processing is domain general, we also included three automated working memory span tasks to increase the confidence in our measurement of WMC. The three tasks used, Operation-span (Unsworth, Heitz, Schrock, & Engle, 2005), Symmetry-span (Shah & Miyake, 1996), and Reading-span (another modified version of Daneman & Carpenter, 1980), have been administered to over 6,000 participants over eight years and show high test-retest reliability and internal consistency, as well as high convergent, divergent, and criterion-related validity(Redick, Broadway, Meier, Kiriakose, Unsworth, Kane, & Engle, 2012).

By analyzed data from these two types of tasks in the convention of the original studies (e.g., looking at critical areas, A' values, etc.), we are able to comment on the effects of the various methodological manipulations on the dependent variables of interest, thus providing a first step toward coming to some resolution within the literature.

2.5 Power Analysis

Before undertaking the following experiments, we attempted to determine the number of participants needed to find the three-way interaction of complexity, animacy, and WMC (or of complexity, phrase, and WMC with the Waters and Caplan, 2004 stimulus set) by conducting a power analysis. According to Bakeman (2005), generalized eta squared is the best effect size measure for repeated measures ANOVAs. The formula given by Olejnik and Algina (2003) for calculated generalized eta squared for mixed designs where the between-subjects factor is manipulated requires the Sums of Squares for each of the factors individually as well as their interaction (and is only illustrated for two factors total). Unfortunately this information is not available in the three studies (Just and Carpenter, 1992; Waters and Caplan, 1996b, 2004) that we are attempting to replicate, nor is it calculable from the data that is provided in them. Instead, using the methods recommended by Tahlheimer and Cook (2002), we calculated Cohen's *d* using the data from MacDonald et al. (1992) and Just and Carpenter (1992). Only these two studies were used because they were the only ones from Chapter 2 that had data for the kinds of interaction that we would predict to find if single resource theory is accurate (i.e., the only interactions that distinguish between the predictions of the single and dual resource theories).

In a 2 (ambiguity) x 3 (region) x 2 (WMC) ANOVA, MacDonald et al. (1992) found a significant three-way interaction. Note that here ambiguity refers to whether or not the sentence was reduced and region refers to the area of the sentence. Using their sample sizes, means (estimated from their figures), and MS_e, we calculated a Cohen's *d* of 2.16 for this interaction, which is a very large effect.

Just and Carpenter (1992) conducted a 2 (animacy) x 2 (WMC) ANOVA on unreduced object-relative sentences and another on reduced object-relative sentences. In both analyses, the interaction of animacy and WMC was significant. Using their sample sizes, means (estimated from their figures), and MS_e, we calculated a Cohen's *d* between .56 and .60 for these interactions, which is a medium effect size.

Note that neither of these interaction matches precisely with what we will be testing in each of the current experiments. Our complexity manipulation in this case will be structure (main clause, subject-relative clause, unreduced object-relative clause, and reduced object-relative clause. We will also use the animacy manipulation of Just and Carpenter (1992) and we will have two WMC groups. Because of this, our ANOVAs will be 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between analyses.

Based upon the above calculations and the fact that we are using extreme groups rather than a median split, we chose to assume a medium effect size of .5 for our power analysis for a repeated measures ANOVA with a within-between interaction, performed in G*Power 3.1.2. The results of this analysis called for 16 participants per experiment, eight low-span and eight high-span. However, this number intuitively seems low. For this reason, we chose to double the recommended *N* to 32 participants per experiment (16 low-span, 16 high-span), for a total of 96 participants across all three. Given our calculations from the available information, our use of extreme groups, and the fact that previous experiments have used as few as 10 participants in one span condition (MacDonald et al., 1992), we feel that this *N* is appropriate.

CHAPTER 3. EXPERIMENT 1

3.1 Introduction

The primary purpose of Experiment 1 was to replicate the results of Just and Carpenter (1992) using their stimuli and judgment in a self-paced word-by-word reading paradigm. The secondary purpose was to examine the effects of this judgment type and presentation paradigm on results obtained using the stimuli of Waters and Caplan (1996b) and Waters and Caplan (2004). Experimental sentences in Just and Carpenter (1992) consisted of reduced and unreduced object-relative sentences containing *by* phrases (e.g., The witness/evidence that was examined by the lawyer shocked the jury.). The main subject of these sentences were either animate (e.g., witness) or inanimate (e.g., evidence). Waters and Caplan (1996b) also used object-relative sentences that were either reduced or not, but instead of using a *by* phrase to resolve the ambiguity, resolution was held off until the presentation of a second verb in these sentences (e.g., The horse raced past the barn fell.). The reduced versions of these sentences are referred to as 'garden path sentences' because they tend to lead the reader to an incorrect interpretation before the ambiguity is resolved. Finally, Waters and Caplan (2004) used a third type of object-relative sentence where the subject is followed directly by an object. In this study, all object-relative sentences were presented in unreduced form (e.g., The law that the millionaire favored frustrated the workers.). Note that neither of the Waters and Caplan studies manipulated animacy in their experimental sentences. Using these details, three sets of experimental stimuli were created for testing syntactic processing in the current experiment (see Appendices B, C, and D for full lists of experimental sentences).

In the original Just and Carpenter (1992) experiment, sentences in the syntactic processing task were shown all at once, below a neutral filler sentence and an eye-tracker was used to measure first-pass reading times on each word. The primary dependent variable of interest was first-pass reading time of the *by* phrase of correctly judged experimental sentences. In the current experiment, sentences were presented in a self-paced word-by-word reading paradigm, a methodology that has been argued to be equivalent to eye-tracking measures (Just, Carpenter, & Woolley, 1982). In addition to the sentence comprehension stimuli, measures of WMC were administered and used to divide participants into low- and high-span groups for the analyses reported below.

3.2 Methods

Participants

Participants were 32 (19 female) undergraduate psychology students at Louisiana State University who received course credit for their participation. All participants reported normal or corrected to normal vision and hearing, and no history of language or speech impairments. Mean age was 19.28 (*SD* = 1.33). Sixteen participants were classified as low-span and 16 were classified as high-span, as explained below (descriptives can be found in Table 3.2.1). An additional 16 participants were run in Experiment 1, but dropped from analyses for the following reasons: 11 participants were medium-span (see explanation of cut-off criterion below), two participants made too many errors on at least one automated span task (see criterion below), one participant was a non-native English speaker, and two participants were dropped for experimenter errors (lost automated task data and repetition of counterbalanced set order).

Syntactic Processing Task

Three stimulus sets were created based upon example sentences found in Just and Carpenter (1992), Waters and Caplan (1996b), and Waters and Caplan (2004). Each of the three stimulus sets consisted of 80 experimental sentences containing an object-relative clause and 80 filler sentences (40 main clause or MC and 40 subject-relative or SR sentences). Object-relative clauses in Set 1 (the Just and Carpenter stimuli) contained a *by* phrase, while those in Set 2 were constructed to be garden path sentences (like Waters and Caplan, 1996b), and those in Set 3 were constructed with subject-object word order (like Waters and Caplan, 2004). Within each set, half of the object-relative clause sentences were unreduced (OR-U) and the other half were reduced (OR-R). Half of the sentences of all types had an animate subject and half had an inanimate one. Note for Set 1 and Set 2 OR sentences, this means that the first noun of the sentence was either animate or inanimate, but for Set 3 this manipulation was found on the second noun of the sentence, due to the subject-object construction. Example sentences can be found in Table 3.2.1 below, and full lists of the experimental object-relative sentences can be found in Appendices B (Set 1), C (Set 2), and D (Set 3).

Table 3.2.1. Example syntactic processing items by set.

Set	Structure	Example Sentence
All Sets	Main Clause (MC)	The doctors airlifted the patient to the large hospital in the city.
	Subject-relative (SR)	The door that locked automatically caused a lot of trouble.
Set 1	Unreduced Object-relative (OR-U)	The evidence that was examined by the lawyer turned out to be unreliable.
	Reduced Object-relative (OR-R)	The boy sniffed by the dog was sitting on the curb.
Set 2	Unreduced Object-relative (OR-U)	The suitcase that was searched for missing valuables vanished.
	Reduced Object-relative (OR-R)	The baby grabbed with both hands squealed.
Set 3	Unreduced Object-relative (OR-U)	The tree that the ordinance protected sheltered the squirrel.
	Reduced Object-relative (OR-R)	The law the millionaire favored frustrated the workers.

Within each set, four lists were created by including one version of each of the base experimental sentences found in the appendices. For example in Set 1: list 1 might include the animate reduced version of a sentence, list 2 might then have the inanimate reduced version, list 3 the animate unreduced version, and list 4 the inanimate unreduced version. Filler sentences were the same for all four lists within a set. Before beginning the syntactic processing tasks, participants were assigned a list number (1, 2, 3, or 4) and a serial order for set presentation (123, 231, 132, 213, 132, or 321). Participants would then be presented with the same list number from each of the three sets in the serial order assigned. This assignment ensured that four participants from each span group saw each of the four lists for each set, and that each set was seen in each serial position by between four and six participants (full counterbalancing of serial set orders would require 24 participants per span group). Within each list for each set, items were arranged into five blocks containing equal numbers of each type of experimental and filler sentences. Presentation within these blocks was randomized by the stimulus software.

All sentences were presented in a self-paced word-by-word manner, using SuperLab software on a Mac computer. At the beginning of each trial, the participant would see the first word of the sentence on the screen and dashed lines holding the places of the other words, as follows:

The _____ _____ ____ __________ ________ __________ ____ __________.

The participant would then press the spacebar to move through the sentence, with each button press causing each consecutive word to appear while the previous word would disappear. Word presentation moved forward only, so it was not possible for participants to review any part of the sentence after progressing past it. After the participant read the entire sentence, a response screen would appear. The response screen consisted of a shorter statement referring back to the sentence that had just been read, followed by the question “True?” Half of the shorter statements were true of the previous sentence (e.g., The millionaire favored the law.), while the other half were false (e.g., The law favored the millionaire.). Participants were asked to indicate whether the statement was true or false by pressing labeled keys on the keyboard.

This methodology was chosen because it has been used previously by single resource researchers (e.g., Just, Carpenter, & Woolley, 1982; King and Just, 1991), but it should be noted that this manner of presentation meant that the participant always knew the approximate length of the sentence from the start of the trial and could also determine how close she was to the end of it as she progressed through it. The relevance of this point will be discussed in the results below.

Working Memory Capacity Measures

All participants completed four measures of working memory capacity: the automated Reading-span, the automated Symmetry-span, the automated Operation-span (Unsworth, Heitz, Schrock, & Engle, 2005), and the Daneman and Carpenter (1980) Reading-span.

Automated Reading span (R-span).

In the R-span task, participants read sentences on the screen and make judgments about whether or not the sentences make sense (nonsensical sentences contain words that are semantically anomalous). After each judgment, participants see a letter on the screen. Judgment-letter pairs are presented in set sizes ranging in size from four to seven. When participants have completed all judgment-letter pairs in a set, they are asked to recall the order of the letters in the set. Three trials of each set size are presented in random order. Participants who do not perform with at least 85% accuracy on the sentence judgments are excluded to ensure both tasks are being equally attended to. Scores on the task are calculated by summing the total number of correct letters in the correct order.

Automated Symmetry span (S-span).

In the S-span task, participants view pictures and make judgments about whether or not the pictures are symmetrical about the vertical axis. After each judgment, participants view a red square within a grid. Judgment-square pairs are presented in sets sizes ranging from three to seven. When participants have completed all judgment-square pairs in a set, they are asked to recall the locations of all red squares in the set. Three trials of each set size are presented in random order. Participants who do not perform with at least 85% accuracy on the symmetry judgments are excluded in order to ensure both tasks are being equally attended to. Scores on the task are calculated by summing the total number of correct square locations in the correct order.

Automated Operation span (O-span).

In the O-span task, participants view one math problem and one word on the screen at a time. Participants read the math problem and solve it aloud, then read the word on the screen.

Problem-word pairs are presented in sets sizes ranging from two to five. When participants have completed all problem-word pairs in a set, they are asked to recall all the words from the set in order. Three trials of each set size are presented in random order. Participants who do not perform with at least 85% accuracy on the math problems are excluded in order to ensure both tasks are being equally attended to. Scores on the task are calculated by summing the total number of correct words in the correct ordinal positions.

Daneman and Carpenter (1980) Reading-span.

In the Reading-span task participants read sets of sentences while memorizing the last word of each until asked to recall the words. Sets were presented using PowerPoint software on a Mac computer, with the experimenter controlling the rate of presentation and hand scoring the task. The number of sentences in a set varies from two to six, and each set size is presented five times. The reading span score for each individual is commonly determined by the largest set size for which three of five sets are accurately recalled. In the current experiments, an alternate scoring method was also used, due to a lack of variability in our sample. This scoring method will be described in more detail in the results section below.

Procedure

The three automated span tasks were always administered before the syntactic processing tasks and the Daneman and Carpenter (1980) task was always administered after them, but for some participants the automated tasks were presented in an entirely separate 1-hour session, while for others they were presented at the beginning of one 2-hour long session. This difference occurred because we attempted to use the automated span tasks as a screening tool to select participants for all three experiments during the first semester of the project, but unfortunately found that only a small number of those invited back (36 out of almost 200) actually returned to participate. For this reason, in the second semester we ran the automated tasks immediately before administering the syntactic processing tasks (in all three experiments) and then examined results afterwards to determine whose data to keep.

Participants were included in the current study if they met two criteria. The first criterion was scoring 85% or higher on the secondary task in all three automated measures, as described above. The second criterion was that they fall in the low or high end of the distribution of scores. For this criterion, cut-off values were selected by finding the upper and lower 50 composite z-scores in a previous sample of 200 participants who had completed the same measures in another lab. (Participants in this sample had given permission to be contacted about future studies, and so were also invited to participate in the syntactic processing portion of the current experiment, but fewer than five actually did so.) Using this method, the lower cut-off value was set at -.53 while the upper value was set at .72 (centered on a mean of 0, with a standard deviation of 1).

3.3 Results

Both for the sake of thoroughness and for the sake of comparison with previous studies and later experiments, results of three types of dependent measure are reported in Experiment 1: true/false judgment accuracy, true/false judgment response time, and critical area reading times. However, it should be born in mind that the critical comparison between Experiment 1 and the study it is designed to replicate, Just and Carpenter (1992), involves the examination of the critical area reading times only, as this is the online measure of processing that the authors used to support their single resource theory. Figure 3.3.1 below shows the (recreated) critical area first-pass reading time results from Just and Carpenter (1992) in the left panel. These results are contrasted with the results that dual resource theory might predict for reading times in the same area in the right panel. Note that Waters and Caplan (1996a) would not predict the two-way interaction demonstrated in the left panel, but they would also argue that it is insufficient to support single resource theory. Instead, Waters and Caplan (1996a) argue that only a three-way interaction of WMC, complexity, and animacy that would demonstrate disproportionate processing load in the low-span group would sufficiently support the theory. For the purposes of clarity, we will hereafter refer to this theoretically required interaction by the abbreviation 'WCA$_{REQ}$' (i.e., the WMC x complexity x animacy interaction).

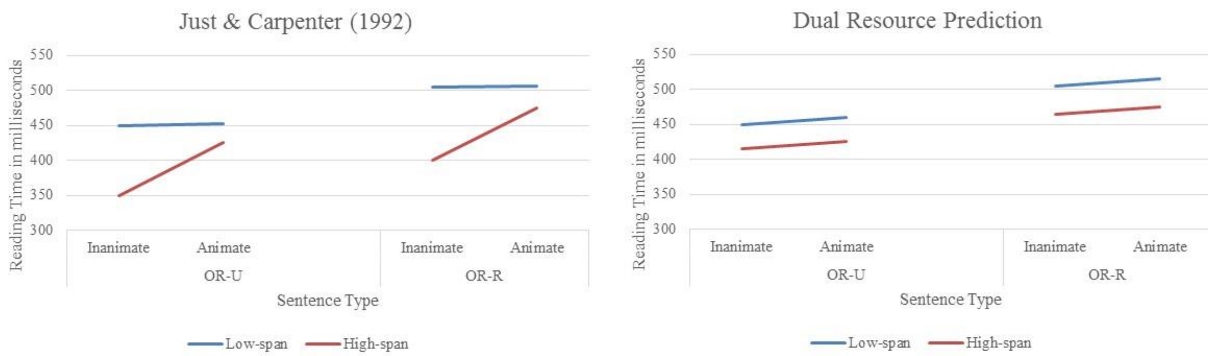


Figure 3.3.1. Recreated results for critical area reading times on experimental sentences in Just and Carpenter (1992) on the left compared with dual resource theory's predictions on the right.

Working Memory Capacity Measures

As already mentioned, the automated span tasks were used for participant selection based upon composite z-score cut-off values. For the Daneman and Carpenter (1980) task, we first scored it in the traditional method, by finding the highest set level for which the participant correctly recalled three of five sets. This scoring method for the task showed little variability in our sample, with most participants scoring 2, some scoring 3, and very few scoring 4, resulting in an imbalanced split between high- and low-span participants (11 to 21 in Experiment 1). Thus, this task was re-scored such that each participant's score was a proportion of words correctly recalled out of the number possible to recall (determined by the number of sets read). Table 3.3.1 below shows the descriptive statistics for all of the WMC measures in the current experiment.

Table 3.3.1. Experiment 1 WMC measure descriptives.

		Low-span		High-span	
		Mean	*SD*	Mean	*SD*
Automated Tasks	R-span	-1.01	0.76	0.92	0.27
	S-span	-0.53	0.53	0.77	0.62
	O-span	-1.55	0.90	1.01	0.28
	Composite	-1.28	0.59	1.12	0.31
D & C (1980)	Original	2.13	0.34	2.69	0.70
	Recode	0.70	0.10	0.79	0.07

Table 3.3.2 below shows the correlations amongst all WMC measures across all three experiments. Note that the Daneman and Carpenter (1980) task and the composite WMC score

show a modest (yet significant) correlation. This is in line with the findings of Waters and Caplan (1996b) that the Daneman and Carpenter (1980) task is not highly correlated with other measures of WMC. Also note that all analyses reported in all experiments were conducted first using the composite z-score and then using the Daneman and Carpenter (1980) recoded scores for the low/high split. Results did not differ by WMC task, so all results are reported using the first analyses run with the composite z-score as the split variable.

Table 3.3.2. Correlations amongst WMC measures across experiments.

		1	2	3	4	5	6
1	D&C original	-	-	-	-	-	-
2	D&C recode	.41**	-	-	-	-	-
3	Composite WMC	.40**	.41**	-	-	-	-
4	R-span Z-score	.37**	.38**	.90**	-	-	-
5	O-span Z-score	.38**	.38**	.91**	.74**	-	-
6	S-span Z-score	.30**	.30**	.83**	.63**	.60**	-

*** p < .01*

Just and Carpenter (1992) Stimuli – Set 1 (Appendix B)

Accuracy.

A 4 (complexity) x 2 (animacy) x 2 (answer) x 2 (WMC) mixed within-between repeated measures ANOVA on true/false judgment accuracy showed a main effect of complexity, *F* (3, 90) = 32.47, *p* < .001, η_p^2 = .52, a main effect of answer, *F* (1, 30) = 6.49, *p* = .02, η_p^2 = .18, and a main effect of WMC, *F* (1, 30) = 4.67, *p* = .04, η_p^2 = .14. These main effects were qualified by three significant two-way interactions. Complexity interacted with answer, *F* (3, 90) = 3.85, *p* = .01, η_p^2 = .11 such that accuracy was higher on true unreduced object-relatives than on false unreduced object-relatives, *t* (31) = 3.19, *p* < .01, but all other contrasts were nonsignificant. There was also an interaction of complexity and animacy, *F* (3, 90) = 3.99, *p* = .01, η_p^2 = .12, such that accuracy was higher for inanimate reduced object-relative clauses than for animate reduced object-relative clauses, *t* (31) = -2.74, *p* = .01. Finally, there was also an interaction between complexity and WMC, *F* (3, 90) = 3.71, *p* = .01, η_p^2 = .11, such that low-spans were significantly less accurate on unreduced object-relatives than high-spans, *t* (30) = -2.91, *p* < .01. Figure 3.3.2 below shows the mean accuracy scores (out of 10) collapsed across answer for each sentence type (answer is collapsed because it is an indicator of bias to respond ‘yes’ or ‘no,’ but is not part of the experimental manipulations of interest).

The finding that accuracy decreases with complexity (i.e., from left to right in Figure 3.3.2) is in line with the predictions of both single and dual resource theory. Both theories would also predict a main effect of WMC, such that low-spans would be less accurate than high-spans, but this was not the case, except for unreduced object-relative (OR-U) items. This finding suggests that the items in the current set were not particularly difficult, a conclusion confirmed by subjective reports of the participants that Set 1 was the easiest of the three in the experiment.

Judgment Response Time.

A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on true/false judgment response times showed a main effect of complexity, *F* (3, 90) =

21.30, $p < .001$, $\eta_p^2 = .42$, a result of longer response times on more complex sentences (see Figure 3.3.3 below). The main effects of animacy and WMC were not significant ($p = .07$ and .31 respectively), but there was a significant interaction between complexity and animacy, F (3, 90) = 3.34, $p = .02$, $\eta_p^2 = .10$, such that response time was faster for inanimate items with reduced and unreduced object-relative clauses, *t*s > 2.09, *p*s < .05 than for animate ones. Figure 3.3.3 below shows the judgment response times for each sentence type.

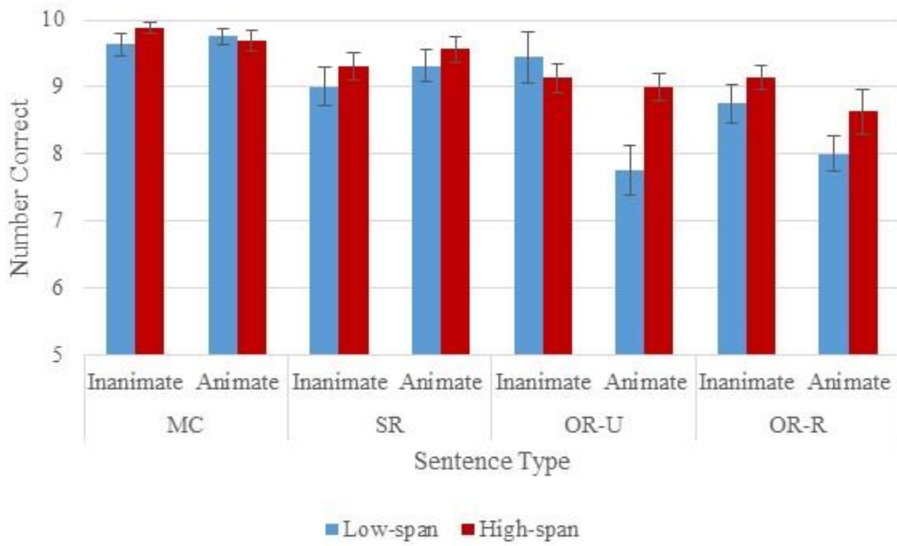


Figure 3.3.2. Experiment 1 Set 1 True/False judgment accuracy, with standard error bars.

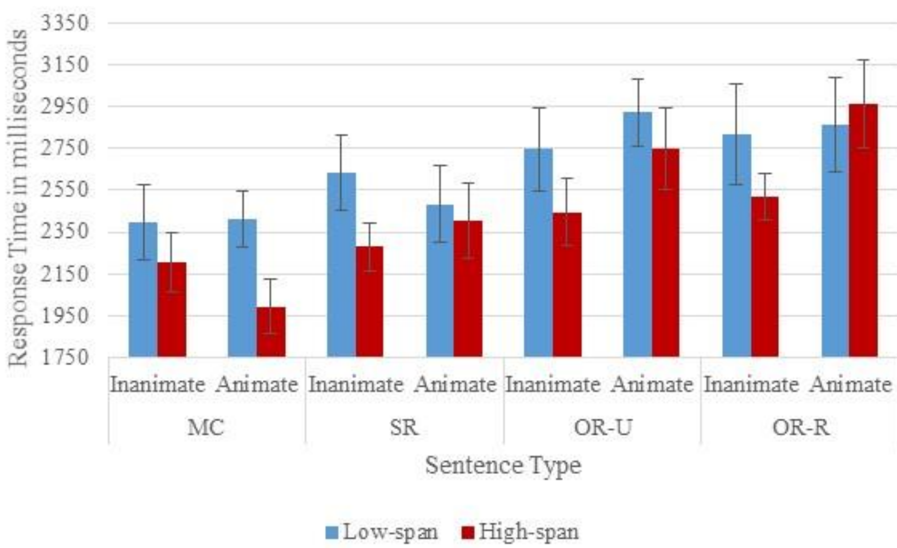


Figure 3.3.3. Experiment 1 Set 1 True/False judgment response times, with standard error bars.

Waters and Caplan (1996a) argued that true/false comprehension items required post-interpretive processing to complete. In the current data, the main effect of complexity on judgment response times seems to support this idea. Judgment response times include both the time it takes participants to read the true/false item and the time it takes them to judge whether that statement is true or false. Given that true/false statements were of the same length for all sentence types, it seems unlikely that the main effect of complexity was due to anything other than an increase in the time it took to make a judgment about the veracity of the statement, a post-interpretive process.

The significant interaction between complexity and animacy suggests a somewhat different conclusion. At first glance, this interaction seems to be compatible with the dual resource view in that it suggests that participants are using the pragmatic cue (animacy) post-interpretively in order to reason about the true/false comprehension items. Because it is generally the case that inanimate nouns are less likely to be the subjects of a main verb, participants are able to make judgments about those items more quickly than when the head noun is animate. This explanation does not, however, account for why accuracy was generally higher for items with an inanimate head noun because in the current set of stimuli, the inanimate noun in the true/false comprehension item was equally likely to be the subject or the object. If participants explicitly and rapidly decided that inanimate nouns were probably not the subjects of the main verbs, then they would be wrong half of the time and thus their accuracy scores would be around five instead of around eight or higher as they are in Figure 3.3.2. Furthermore, although it is not the area of primary interest, we can examine reading times for the last portion of the sentence (i.e., the area after the critical area but before the judgment item). A 2 (complexity) x 2 (animacy) ANOVA on just the experimental object-relative sentences showed that there was a marginally significant main effect of animacy in this area, F (1, 31) = 4.11, p = .051, such that reading times were faster for sentences with inanimate subjects than for those with animate ones. This suggests that the animacy cue was already being incorporated into the participants' interpretation of the sentences before the time of judgment, while the fact that judgment response times were longer for animate items suggests that the animate head noun was indeed more ambiguous during online processing and thus participants required extra time to determine if it was indeed the subject of the main verb or not.

Finally, although there was no significant effect of WMC, nor an interaction, we can see in Figure 3.3.3 that the effect of animacy appears roughly equal in magnitude for both WMC groups in the unreduced object-relative items, but in reduced object-relative items it almost disappears for low- but not high-spans, a trend that is consistent with the idea that low-spans lose the ability to incorporate the animacy cue when the activation demands of a sentence increase as they do when an object-relative clause is reduced. Together, this trend and the animacy effect in the last portion of the sentence are supportive of the single resource theory in that they suggest the animacy cue is being processed online, at least to a certain extent.

Critical Area Reading Time.

While the accuracy and judgment response time data are interesting, the main dependent measure of interest to Just and Carpenter (1992) was the time it took to read the critical area, defined in their stimulus set as the *by* phrase in object-relative sentences (both reduced and unreduced). In a reduced object-relative sentence, this area should place the highest demand on processing resources because it is where the ambiguity introduced by reducing the clause is resolved. In an unreduced object-relative sentence, this area should not be as demanding because the complementizer at the head of the clause has already signaled to the reader that a clause is indeed present. Thus, any differences in reading times between reduced and unreduced sentences in this area should be reflective of differential demands on processing resources.

Before any analyses were conducted, reading times were trimmed to reduce outliers. Trimming was accomplished by using each participant's own mean and standard deviation.

Original reading time values that were greater than three standard deviations from the participant's mean were replaced with the value of the mean plus three standard deviations. The same was done for values three standard deviations below the mean, with the exception that no replacement value was allowed to be less than zero. The vast majority of replaced data points were above rather than below the mean. For each of the three experiments, no more than 2% of data points were replaced overall, nor was the replacement rate any higher than 2% for any individual participant within each experiment.

After trimming was completed, reading times for the individual words in the *by* phrase were averaged together to get one reading time for the entire area of the sentence. Finally, items for which participants made a correct true/false judgment were extracted for analysis. We included both items for which the correct answer was true and for which the correct answer was false because participants did not know until they reached the comprehension statement if the item would be true or false, therefore, the actual correct answer should not affect reading times online. Note that this contrasts with Experiments 2 and 3, where only correctly judged acceptable and grammatical items could be used due to the nature of the stimuli.

Critical area reading times were first tested in a 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA which revealed a main effect of complexity, F (3, 90) = 21.94, $p < .001$, $\eta_p^2 = .42$, and an interaction of complexity and animacy, F (3, 90) = 2.72, $p < .05$, $\eta_p^2 = .08$, reflected by the fact that critical area reading times were longer for sentences with animate reduced object-relative clauses than for those with inanimate reduced object-relative clauses, t (31) = 2.43, $p = .02$, but there were no other significant differences. Figure 3.3.4 below shows mean *by* phrase reading times for the experimental sentences (OR-U and OR-R) along with reading times in the analogous areas of filler sentences (MC and SR).

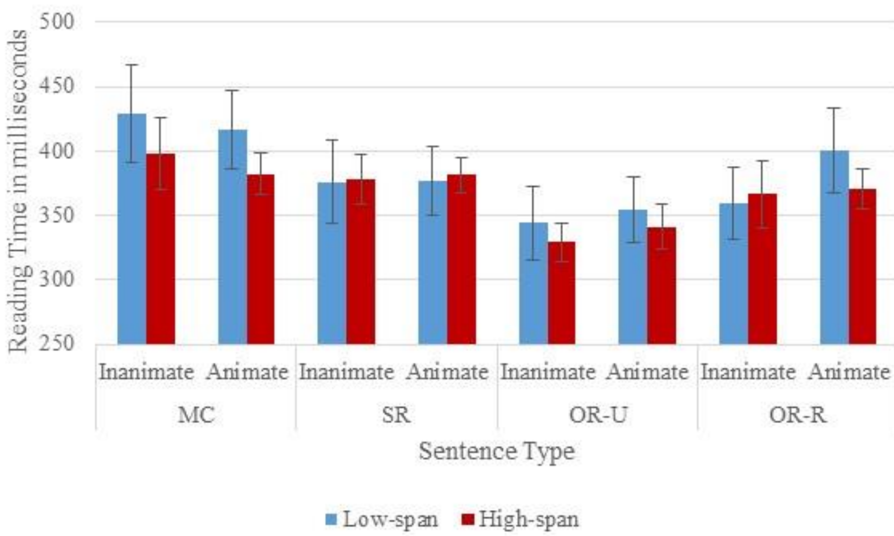


Figure 3.3.4. Experiment 1 Set 1 critical area (i.e., *by* phrase) reading times, with standard error bars.

One oddity about the critical area reading times in Figure 3.3.4 is that they are so long for main clause items regardless of animacy. An explanation for this may be that main clause sentences and reduced object-relative sentences are completely indistinguishable from each other for the first three words. That is, they both begin with a noun phrase and a verb in the past tense/participle form (e.g., The shelf contained). It is not until a *by* phrase appears that a participant knows the sentence is indeed a reduced object-relative. Given that reduced object-relative sentences were twice as frequent as main clause sentences in the set (40 vs. 20), it is possible that participants spent longer reading the noun phrase following the first three words (e.g., an entire encyclopedia) because they were surprised or confused at the lack of a *by* phrase.

While this oddity in the filler sentences is interesting, the more pressing contrast for the purposes of replicating Just and Carpenter (1992) is found in the experimental object-relative sentences only. The left panel of Figure 3.3.5 below shows the recreated critical area reading times from Just and Carpenter (1992) while the right panel shows critical area reading times from the current experiment for comparison. In the left panel, we can see that animacy did not matter for Just and Carpenter's (1992) low-span participants, but for their high-span participants an inanimate head noun led to significantly faster reading times in the critical area than when the head noun was animate. We see no such effects in the right panel (as expected, given the lack of significant interactions above).

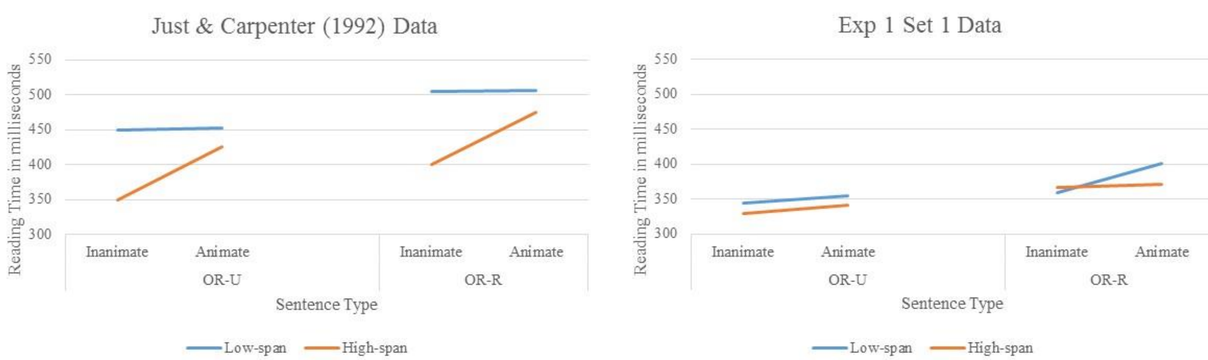


Figure 3.3.5. Recreated results of Just and Carpenter (1992) in the left panel, next to actual data for Set 1 in Experiment 1 on the right.

Clearly critical area reading times in Figure 3.3.5 do not replicate the principle finding of Just and Carpenter (1992). There were no differences associated with WMC, and certainly no WCA_{REQ} interaction called for by Waters and Caplan (1996a). One explanation for this might be found in the debrief questionnaire. Participants were asked specifically about any button pressing strategies they may have employed while reading sentences. The overwhelming majority of participants, both low- and high-span (and in all judgment conditions, a point to be revisited later), reported that as they became accustomed to the tasks, they began to realize it was easier to press as fast as they could in order to get all of the words of the sentence in their minds so that they could make the judgment, especially when the sentences were longer. As previously mentioned, participants knew from the beginning of each trial the approximate length of the sentence due to the fact that dashes held the places of the words that had not yet appeared. Participants could also tell approximately how close to the end of a sentence they were for the same reason. Sentences in Set 1 were longer than in the other two sets, and note that the reading times in the right panel are at the same level as the fastest, facilitated reading times found in Just and Carpenter (1992). Thus, although it has been argued that the self-paced word-by-word presentation paradigm is equivalent to first-pass reading times collected with eye-trackers (Just, Carpenter, & Woolley, 1982), this may not have been the case in the current study due to a strategy inadvertently introduced by the method of presentation (a method used, it should be noted, by both Just et al., 1982 and King & Just, 1991).

In order to determine whether or not this 'rushing strategy hypothesis' holds any weight, we attempted to separate participants by their overall reading speed. We first calculated mean per-word reading times for each participant on each type of object-relative sentence and found the median reading time for both types. We then classified participants as either 'fast,' 'mixed,' or 'slow.' Fast participants were below the median reading time on both true and false items, while slow participants were above for both. Mixed participants were above the median on one type and below it on the other. The majority of participants were classified as 'fast' (N = 20, 19, 18, and 14 respectively for each sentence type), while one quarter or less were 'slow' (N = 7, 8, 5, and 8 respectively). The remaining participants were classified as 'mixed.' While these imbalanced sample sizes violate the assumptions of standard statistical tests, we can still look at the trends in the data by plotting them for comparison with the data above. Figure 3.3.6 below shows the critical area reading times for 'fast' participants in the left panel and those for 'slow' participants on the right ('mixed' participants are not shown, but their reading times were similar to those of the 'fast' participants, just slightly slower overall when averaged together).

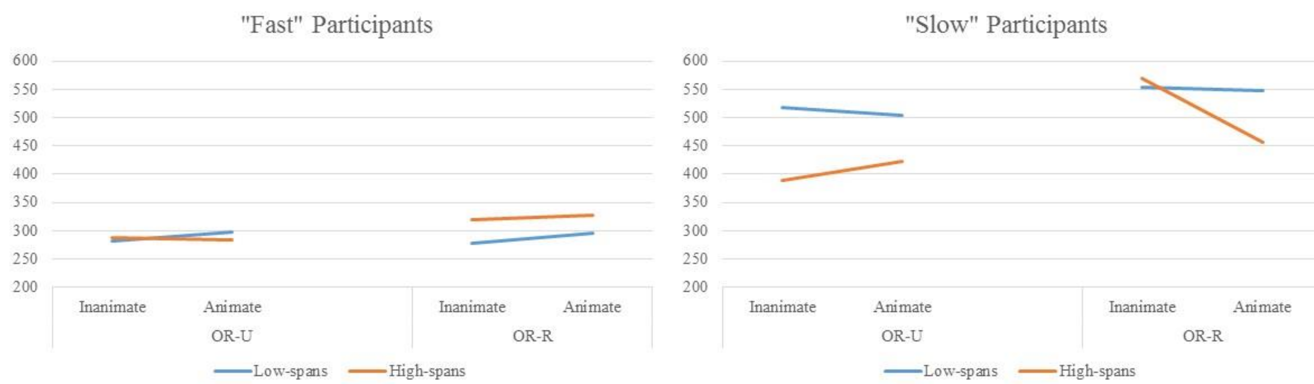


Figure 3.3.6. Fast vs. Slow participant critical area reading times, presented for comparison with the Just and Carpenter (1992) data recreated above.

The first thing to note about Figure 3.3.6 is that critical area reading times of 'fast' participants are all below 350ms, whereas in Just and Carpenter (1992) the fastest critical area reading time was 350ms, and this is true for both low- and high-span participants. Moving to the right panel, we can see that 'slow' participants show critical area reading times within the range of reading times observed by Just and Carpenter (1992) and that there is now some differentiation between low- and high-span participants (i.e., the blue and red lines have drawn apart from each other as they did not in the left panel). Though the reading times in the right panel do not show a pattern identical to that of the Just and Carpenter (1992) data, this is perhaps not overly troubling, given the extremely small sample sizes represented in this panel that also preclude the use of traditional statistical tests for group differences. Rather, the important takeaway from this qualitative examination is that participants who rushed through sentences all look the same, while those who slowed down showed some evidence of differences related to span group membership, suggesting that there is indeed something to our 'rushing strategy hypothesis' that explains the lack of the WCA_{REQ} interaction in the analyses reported above.

A second explanation for the lack of an effect of WMC may be related to the stimuli themselves. Subjectively speaking, participants reported that the Set 1 sentences were the easiest/least confusing of the all the sets they encountered in the study (no matter the order of set presentation, suggesting it was not just an order effect). Perhaps, for whatever reason, the sentences created for Set 1 were not challenging enough to elicit the WMC difference. Just and Carpenter (1992) themselves revised sentences from Ferreira and Clifton (1986) in order to find the effect, and these same sentences were used as a starting point for the sentences created for Set 1 in the current experiment (because that list was available, whereas the Just and Carpenter sentences were not). It is possible that there was some unknown element missing from the current set that inadvertently lessened the difficulty of the sentences. The fact that there was an effect of animacy in reading times for the last area of the sentence (i.e., the words following the critical area) that carried over into judgment response times (as described above) supports this argument because the effect was present for both low- and high-span participants (though showed a trend to disappear for low-spans in the reduced sentences). Just and Carpenter's (1992) explanation of their data was that high- but not low-span participants were taking the inanimacy of the head noun into account during online processing in order to read the critical area faster. According to single resource theory, this difference between span groups should occur because the complex object-relative sentences should be too demanding by virtue of their syntactic structures to allow low-span participants to use the pragmatic cue online. Thus, the data of the current experiment suggest that the syntactic structures of the Set 1 sentences were not demanding enough to preclude the online use of pragmatic cues by low-span participants. This conclusion would be strengthened by results demonstrating the WCA_{REQ} interaction with sets rated subjectively as more difficult. According to participant reports, the most difficult set was Set 2, while Set 3 fell somewhere in the middle.

Waters and Caplan (1996b) Stimuli – Set 2 (Appendix C)

Accuracy.

A 4 (complexity) x 2 (animacy) x 2 (answer) x 2 (WMC) mixed within-between repeated measures on ANOVA on true/false judgment accuracy showed a main effect of complexity, F (3, 90) = 99.29, $p < .001$, $\eta_p^2 = .80$, and of answer, F (1, 30) = 121.51, $p < .001$, $\eta_p^2 = .80$. The main effect of complexity was qualified by an interaction with animacy, F (3, 90) = 7.76, $p < .001$, $\eta_p^2 = .34$, which was in turn qualified by a three-way interaction of complexity, animacy, and WMC, F (3, 90) = 3.21, $p = .03$, $\eta_p^2 = .10$. This interaction is reflected by the fact that low-spans were significantly less accurate than high-spans on sentences with inanimate subject-relative clauses, t (30) = -3.51, $p < .01$. All other contrasts were non-significant.

The main effect of complexity was also qualified by an interaction with answer, F (3, 90) = 50.16, $p < .001$, such that accuracy was significantly higher for true items than for false for all sentence types, except reduced object-relatives, where accuracy was significantly higher for false items than for true, all ts > 4.62, ps < .001. Answer also interacted with animacy, F (3, 90) = 17.08, $p < .001$. For true items, accuracy was higher with an inanimate subject than with an animate subject, but for false items accuracy was higher for animate subjects than for inanimate. Figure 3.3.7 below shows the accuracy scores (out of 10) collapsed across answer.

The three-way interaction of complexity, animacy, and WMC on subject-relative items, where low-spans showed lower accuracy for inanimate items than high-spans, is unexpected and is certainly not the WCA_{REQ} interaction that would support single resource theory. One explanation may be that in these items both the subject and the object were inanimate, while animate versions had inanimate objects. Perhaps when it came time to judge the veracity of the true/false statement, low-span participants had a harder time distinguishing the subject from the object if both were inanimate.

Overall we can see in Figure 3.3.7 that the judgment accuracy drops significantly for the experimental sentences compared to the filler sentences. Much of this effect is due to a drop in accuracy on false items. Because of this drop on false items, 18 of the 32 participants did not have any correctly judged false items for at least one sentence type. This is important because accuracy was used as a criterion for including a trial in the critical area reading time analysis below. Critical area reading times are collapsed across answer, due to the fact that judgments come after the sentence has been read, but it is still worth noting that 18 of 32 participants contributed reading times from correctly judged true items only. More details will be provided about these dropped data points below.

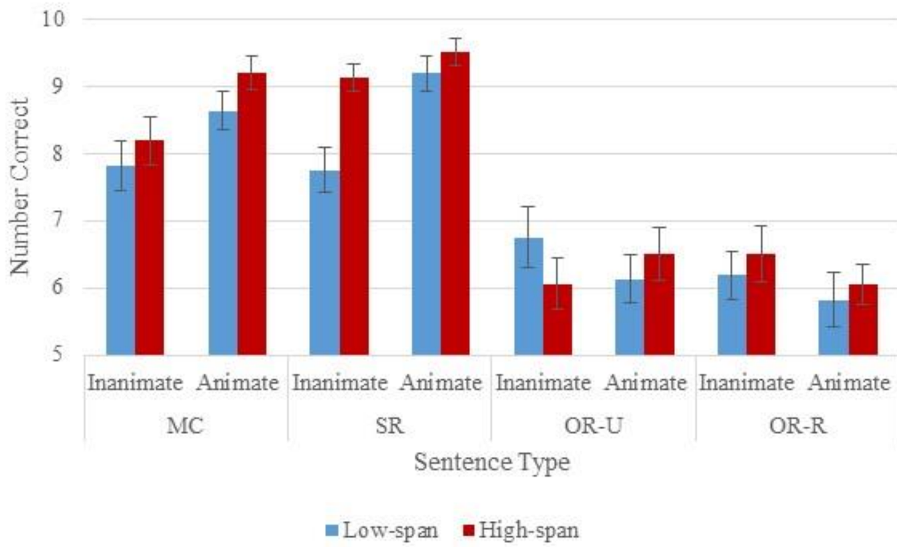


Figure 3.3.7. Experiment 1 Set 2 True/False judgment accuracy, with standard error bars.

Judgment Response Time.

A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on true/false judgment response time showed main effects of complexity, F (3, 90) = 34.11, $p < .001$, $\eta_p^2 = .52$, and animacy, F (1, 30) = 19.00, $p < .001$, $\eta_p^2 = 34$. These main effects were qualified by a two-way interaction of complexity and animacy, F 3, 90) = 4.32, $p < .01$, η_p^2 = .14. Figure 3.3.8 shows the means for each sentence type.

The interaction of complexity and animacy was due to the fact that response times were longer for inanimate main clause and subject-relative sentences than for animate main clause and subject-relative sentences, *t*s > 3.85, *p*s < .01, but there was no significant difference between inanimate and animate object-relative items of either type.

Critical Area Reading Time.

Unlike the Just and Carpenter's (1992) stimuli, the Waters and Caplan (1996b) experimental sentences do not contain a *by* phrase. For this reason, the critical area in Set 2 sentences is defined as the second verb (which was also the last word of the sentence) because this is where any ambiguity is resolved, and thus should place the greatest demand on processing resource(s).

As with Set 1 stimuli, critical area reading times analyses for Set 2 were confined to those sentences for which a correct true/false judgment was made (again, regardless of whether the correct answer was true or false) and reading times were trimmed using the participants' own means and standard deviations, as described above. However, because of the difficulty in correctly responding to false comprehension items, 18 of 32 participants never did so for at least one type of sentence. All told, 37 data points were lost, with five participants losing one type each, eight losing two types, four losing three types, and one losing four types. Put another way: 11 people did not contribute critical area reading times for false inanimate OR-U sentences, 7 did not contribute for false animate OR-U items, 7 did not contribute for false inanimate OR-R items, 11 did not contribute for false animate OR-R items, and 1 did not contribute for false inanimate MC items. Because of this amount of missing data, all analyses reported below were also conducted including items with incorrect true/false judgments. The results did not change, and so are reported here with the incorrect items excluded to be consistent with the previous literature. Figure 3.3.9 below shows the mean reading time for each sentence type by span group.

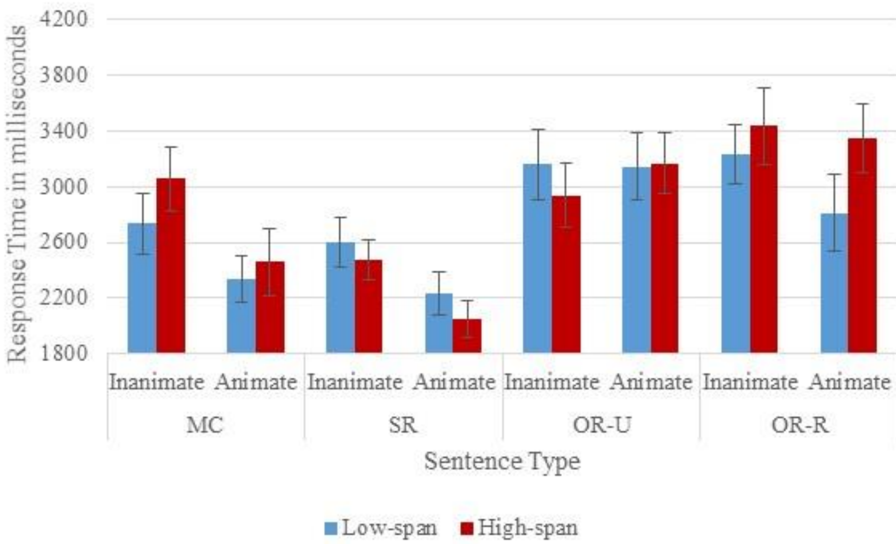


Figure 3.3.8. Experiment 1 Set 2 True/False judgment response times, with standard error bars.

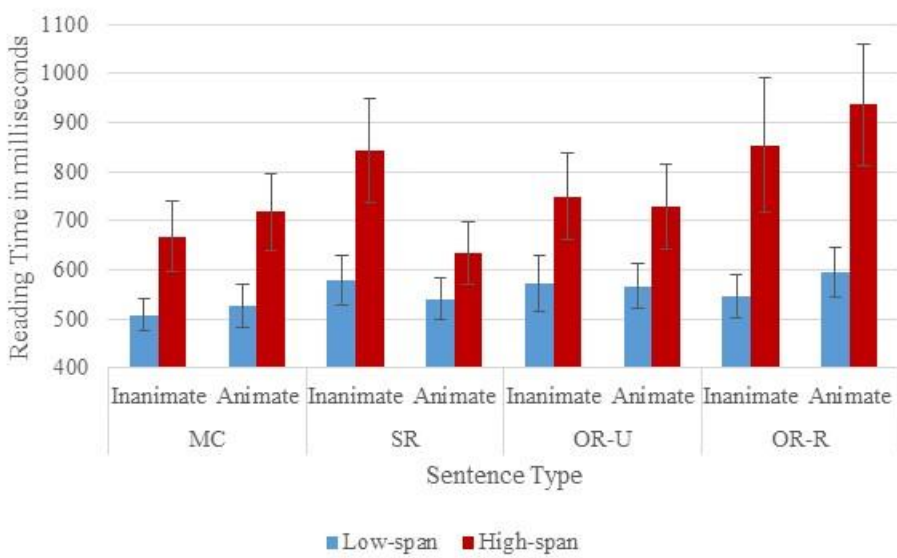


Figure 3.3.9. Experiment 1 Set 2 critical area reading times, with standard error bars.

A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA revealed a main effect of complexity, F (3, 90) = 7.35, p < .001, η_p^2 = .20, a main effect of WMC, F (1, 30) = 4.92, p = .03, η_p^2 = .14, an interaction between complexity and animacy, F (3, 90) = 4.99, p < .01, η_p^2 = .14, and an interaction between complexity and WMC, F (3, 90) = 3.68, p < .02, η_p^2 = .11. The interaction of complexity and animacy was reflected in the fact that critical area reading times were significantly longer for inanimate subject-relatives than for animate subject-relatives, t (31) = 3.35, p < .01, but no other contrasts were significant. The interaction of complexity and WMC reflected the fact that low-spans were significantly faster than high-spans for main clause items, subject-relative items, and reduced object-relative items, ts > 2.41, p < .04. Figure 3.3.10 below shows the critical area reading times for just the experimental object-relative sentences in Set 2, displayed for comparison with Just and Carpenter's (1992) findings.

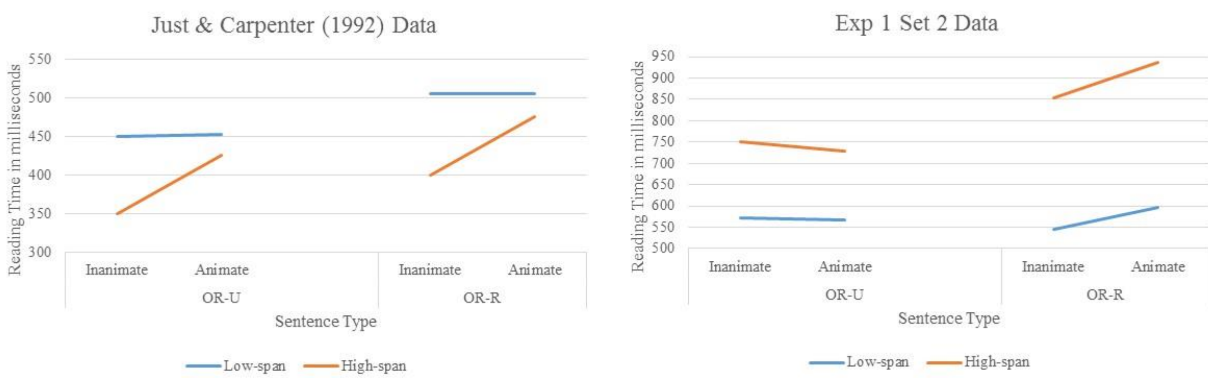


Figure 3.3.10. Recreated results of Just and Carpenter (1992) in the left panel, next to actual data for Set 2 in Experiment 1 on the right.

In the left panel of Figure 3.3.10 we can see that Just and Carpenter (1992) found an interaction of animacy and WMC on their *by* phrase sentences. The right panel of Figure 3.3.10 clearly shows that we did not replicate this interaction with Waters and Caplan's (1996b) garden path sentences. Data from the current study do, however, show an interaction of complexity and WMC that was not found by Waters and Caplan (1996b). The question is why the nature of the interaction in our results differs from the results of both of the previous studies.

It is perhaps easiest to explain why our results differ from those of Waters and Caplan (1996b). In that study, reading time for individual words or phrases was not recorded directly. Rather, the authors estimated per word reading time by dividing the time it took to read and respond to sentences by the number of words in the sentence. This methodology assumes a priori that reading times are uniform across all areas of the sentence and does not allow any differences based upon sentence area to be seen. In the current experiment, we are able to examine differences in reading speed in different areas. Indeed, if we look at mean reading times for all words leading up to the critical area we see a marginal effect of complexity (p = .051), no effect of WMC, and no interaction of the two; This is exactly what Waters and Caplan (1996b) found with their per word reading time estimate. It is only when we are able to separate the last word/critical area from the rest of the sentence that the effect of WMC, including its interaction with complexity, emerges.

Next, we must consider why it could be that we find a different two-way interaction than that found by Just and Carpenter (1992). In the first place, our reading times differ from those in the previous study both because they are longer overall and because it is the high- and not low-spans who show the longest reading times. The fact that our reading times are all equal to or greater than the longest reading times found in Just and Carpenter (1992) is likely attributable to the well-known sentence wrap up effect discussed by Balogh et al. (1998), where reading times just generally increase at the ends of sentences, because the critical area in this set of sentences is also the last word. The flip between low- and high-spans too is explainable by the construction of the sentence as it is in line with the finding of MacDonald et al. (1992) that high-spans show longer reading times at the ends of garden path sentences than low-spans. Those authors attributed the difference to a cost that high-spans pay for maintaining dual interpretations of a sentence during processing.

The second issue to examine is why the two-way interaction in the current data is between WMC and complexity rather than WMC and animacy, as it was in Just and Carpenter (1992). Comparing this set to the first, accuracy was significantly lower while judgment response times were significantly slower (see section 3.4 below for statistics across all three sets), suggesting that these items were quite difficult. Two factors could account for this difficulty. The first factor is the structure of the sentences themselves. Although short, reduced object-relatives in Set 2 are quite confusing due to the abrupt onset of a verb at the end. For most people, their first reading of the sentence "The horse raced past the barn fell" is extremely confusing and we would expect lower performance on such sentences. However, we would also expect performance to be better when these sentences are unreduced, and yet it is not. This suggests a second explanation for the difficulty: the nature of the true/false comprehension items.

Unlike in the previous set, true/false comprehension items in the current set could not be created simply by reversing the roles of the subject and object of the relative clause in the stimulus sentence because the subject is not explicitly present in the sentence (i.e., the entity that raced the horse past the barn is never named). Instead, true/false comprehension items were formed by making the object of the relative clause the actor or object of the verb. In other words, for Set 1 participants had both nouns in the true/false item and simply had to judge if the relationship between them was accurate (i.e., did the person race the horse or did the horse race the person), while in Set 2 they had to determine if the object did the action or had the action done to it by some unnamed entity (i.e., did the horse race itself past the barn or did someone else race it past the barn). This second judgment is arguably a far more subtle distinction, especially in the case of inanimate objects.

This difference also raises the question of whether or not the animacy of the noun was as relevant to the task in Set 2 as it was in Set 1. In Set 1 items, most of the subjects of the *by* phrase were animate as a matter of course (see Appendix B). We did not purposely set out to manipulate this factor as previous studies did not discuss doing so, but note that this would mean that true/false judgments for animate items would most often involve choosing which of two animate nouns performed an action while inanimate items would involve choosing between nouns that contrasted in animacy. This contrast may account for the facilitation effect of inanimate nouns on critical area reading times in reduced object-relatives in Set 1 because storing nouns with similar properties (e.g., both common, both proper, etc.) while processing a complex sentence can impair comprehension performance (Gordon, Hendrick, & Levine, 2002). In Set 2, on the other hand, there was no contrast of animacy for any items because true/false items only ever contained the head noun. Furthermore, because the head noun was animate and inanimate equally often, animacy itself was not a strong clue as to the head noun's role in the sentence, and thus participants would gain little by incorporating the animacy cue online during sentence processing. If it is the case that participants can alter their reading strategies to best meet task demands, then it would be unsurprising to find little effect of animacy on the critical area reading times of Set 2 sentences.

Finally, it must be noted that although we found a two-way interaction of WMC and complexity that Waters and Caplan (1996b) did not, we did not find the WCA_{REQ} interaction that they would argue is necessary to support single resource theory. We have already argued that animacy may not have been the strongest cue in Set 2 sentences, due both to the nature of the sentences themselves and by extension to the nature of the true/false comprehension items, and that this may have altered reading strategies. However, in Figure 3.3.9, we can see a non-significant trend for animacy to affect the critical area reading times of reduced object-relative sentences for high-spans only. The difference in these reading times, 83-ms, roughly equal to the significant 75-ms difference found by Just and Carpenter (1992) on the same type of sentence. One reason this this difference does not reach significance in the current data may be due the large degree of variability in the high-spans' reading times (see standard error bars in Figure 3.3.9). Low-spans, on the other hand, do not show the slightest trend for an animacy affect and also have much less variability in their reading times. One explanation for this pattern of results may be that high-spans differed more in the extent to which they altered their reading strategies in response to task demands, whereas low-spans were more uniform in their reading strategy shifts.

Waters and Caplan (2004) Stimuli – Set 3

Accuracy.

A 4 (complexity) x 2 (animacy) x 2 (answer) x 2 (WMC) mixed within-between repeated measures ANOVA on true/false judgment accuracy showed significant main effects of complexity, F (3, 90) = 9.12, p < .001, η_p^2 = .23, and animacy, F (1, 30) = 16.57, p < .001, η_p^2 = .36, which were qualified by a significant interaction between the two, F (3, 90) = 5.68, p < .01, η_p^2 = .16. Examination of this interaction showed that accuracy was higher for animate object-relative items (both unreduced and reduced) than for inanimate object-relative items, ts > 2.62, ps ≤ .01, did not differ for subject-relative items, and was lower for animate main clause items than for inanimate ones, t (31) = -2.78, p < .01.

The analysis also revealed a significant main effect of WMC, F (1, 30) = 11.12, p < .01, η_p^2 = .27, such that accuracy was generally greater for high-spans than for low-spans. Finally, there was a significant main effect of answer, F (1, 30) = 6.41, p = .02, η_p^2 = .18, such that accuracy was greater for true items than for false, suggesting a bias to response 'true' to items. Figure 3.3.11 shows the accuracy scores (out of 10) for each sentence type, collapsed across answer.

The fact that accuracy was higher for animate than inanimate reduced object-relatives is interesting in that it is a reversal of the animacy effect found by Just and Carpenter (1992), and suggests that inanimate items are the more ambiguous item type in Set 3 sentences. This reversal makes sense in light of the structure of Set 3 sentences (see Appendix D), where the subject and object of the relative clause are found at the beginning of the sentence and must be stored while the rest of the sentence is processed. As previously mentioned, Gordon et al. (2002) found that storing nouns that shared characteristics (in that case, the characteristic was common or proper noun status) while processing complex structures impaired performance on the comprehension task. Based on this result, it would make sense that having two animate nouns at the head of Set

3 sentences would impair performance on the task, whereas having a contrasting pair (i.e., one animate and one inanimate) would facilitate it.

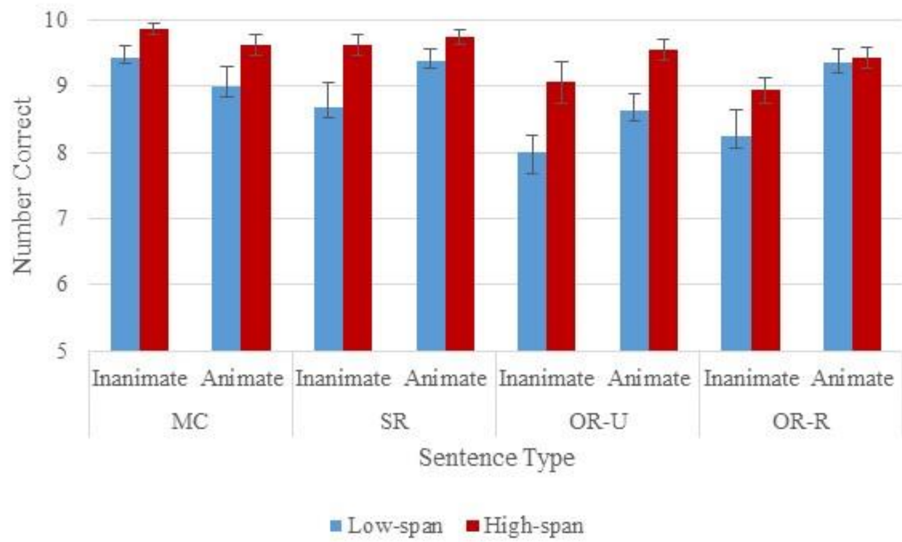


Figure 3.3.11. Experiment 1 Set 3 True/False judgment accuracy, with standard error bars.

Judgment Response Time.

A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on true/false judgment response time showed significant main effects of complexity, F (3, 90) = 29.68, $p < .001$, $\eta_p^2 = .50$, and animacy, F (1, 30) = 22.84, $p < .001$, $\eta_p^2 = .43$. These main effects were qualified by two-way interactions between: complexity and WMC ($\eta_p^2 = .10$), complexity and animacy ($\eta_p^2 = .16$), animacy and WMC ($\eta_p^2 = .16$), all Fs > 3.48, ps < .02, which in turn were qualified by a significant three-way interaction of complexity, animacy, and WMC, F (3, 90) = 3.39, $p = .02$, $\eta_p^2 = .10$. Examination of this interaction revealed that true/false judgment response times were longer for inanimate unreduced object-relative items than animate unreduced object-relative items for high-spans t (15) = 4.39, $p < .01$, and for inanimate reduced object-relative items than for animate reduced object-relative items for high-spans, t (15) = 6.89, $p < .001$. Figure 3.3.12 below shows the true/false judgment response times for each sentence type, collapsed across answer.

Note that the significant three-way interaction is the WCA_{REQ} called for by Waters and Caplan (1996a) and as such is support for the single resource theory of syntactic processing, as it would not be predicted in the dual resource view. However, there are two difficulties with this Set 3 interaction being interpreted as a replication of the Just and Carpenter (1992) data. The first problem is that the interaction is not in the critical area, as predicted by the single resource theory, but rather is found at the time of the true/false judgment. One could argue that this represents a carry-over effect in these sentences, where the differential processing is delayed until the time of judgment. However, this is problematic in that it does not go far in rebutting the Waters and Caplan (1996b) argument that an interaction with WMC in the syntactic processing task is a result of post-interpretive processing triggered by the use of true/false judgments rather than a reflection of online syntactic processing (even though that argument would not explain an effect of animacy, since pragmatic cues should not affect complexity in the dual resources theory). One way to address this comment is to look at whether or not true/false reaction times in Set 3 are significantly longer than those in the previous sets, which would suggest more post-interpretive processing occurring. This comparison will be presented below in section 3.4, but to preview: Set 3 judgment response times were longer than those in Set 1 only for inanimate object-relative items, and for no items in Set 2.

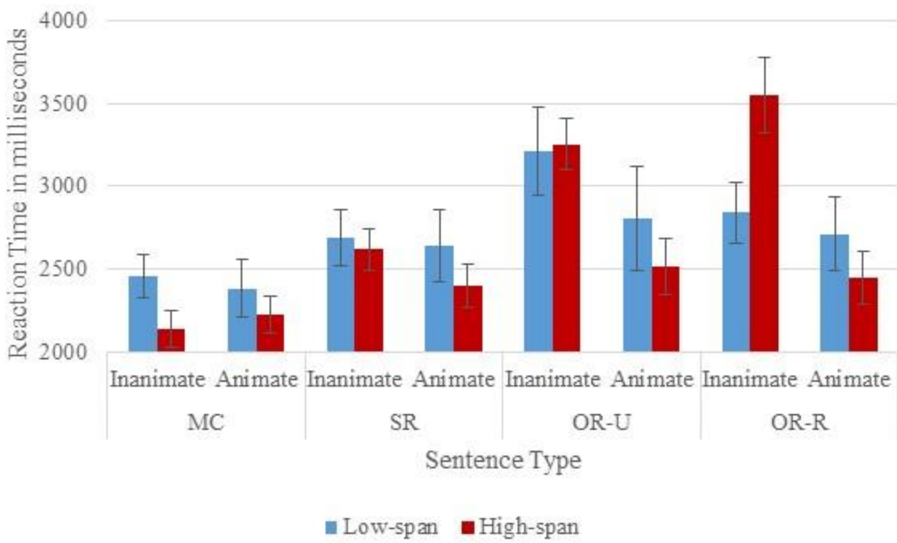


Figure 3.3.12. Experiment 1 Set 3 True/False judgment response times, with standard error bars.

Figure 3.3.13 below more clearly illustrates the WCA_{REQ} interaction by displaying true/false judgment response times for the object-relative sentences only. We can see that animacy has no effect on the low-spans, but for high-spans true/false judgment responses are faster when the subject is animate than when it is inanimate. This pattern of results suggests that animate items are only less ambiguous for high-spans (because they can take into account the pragmatic cue), and that any extra time spent in post-interpretive processing is only necessary when the subject is inanimate and therefore more ambiguous. For low-spans, animacy is not taken into account online, and so they spend an equal amount of time on the true/false judgment for inanimate and animate items.

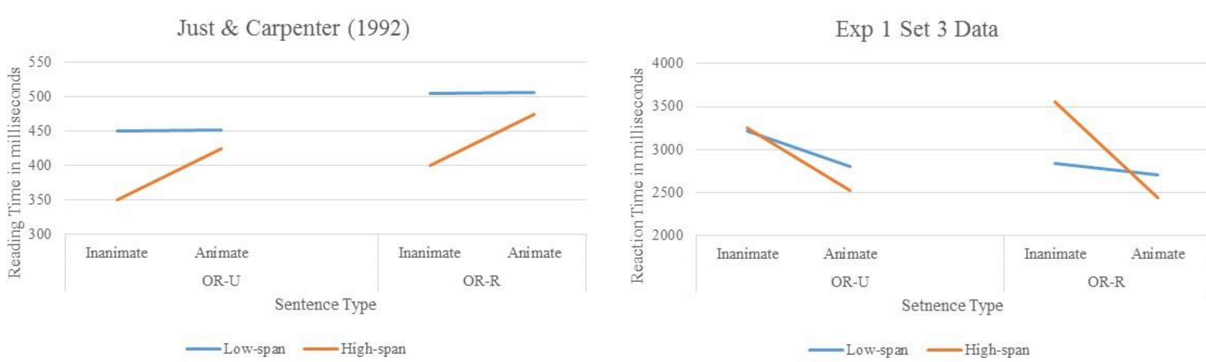


Figure 3.3.13. Recreated critical area reading times from Just and Carpenter (1992) in the left panel, next to judgment response time data for Set 3 in Experiment 1 on the right.

Critical Area Reading Time.

Like Waters and Caplan (1996b), the Waters and Caplan (2004) stimuli lack the *by* phrase of the Just and Carpenter (1992) stimuli, and thus the critical area in these sentences was defined as the second verb, where any ambiguity is resolved. However, unlike Waters and Caplan (1996b), the second verb in Set 3 is followed by a short noun phrase (e.g., 'the workers') rather than being the last word of the sentence itself. As with the previous sets, critical area reading times were trimmed using individual means and standard deviations and items for which participants made correct judgments were used for the following analyses.

A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA showed a significant main effect of complexity, F (3, 90) = 4.69, p < .01, η_p^2 = .14, such that critical area reading times in main clause and subject-relative sentences were significantly faster than those in unreduced and reduced object-relative sentences, ts > -2.08, ps < .05. There were no other significant main effects or interactions. Figure 3.3.14 below shows the means for each sentence type, while Figure 3.3.15 shows means in experimental sentences only for comparison with Just and Carpenter (1992) data.

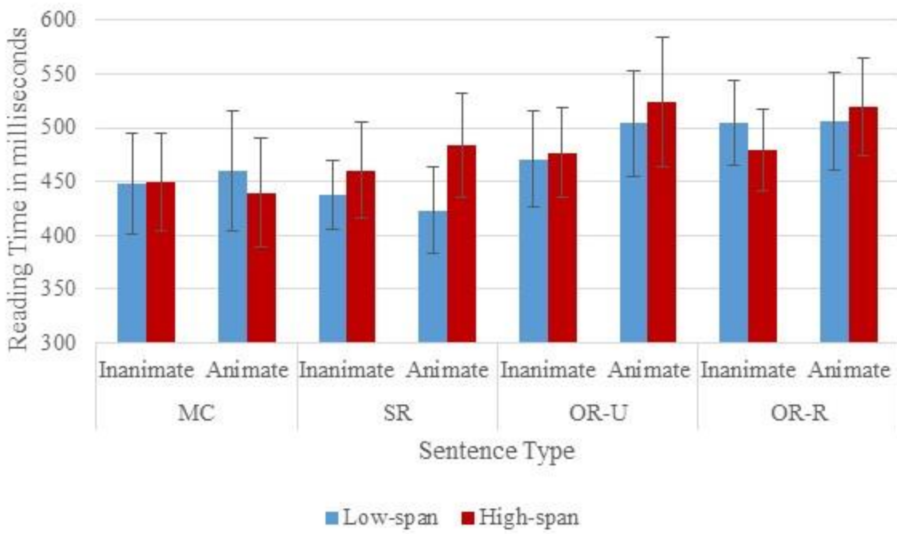


Figure 3.3.14. Experiment 1 Set 3 critical area reading times, with standard error bars.

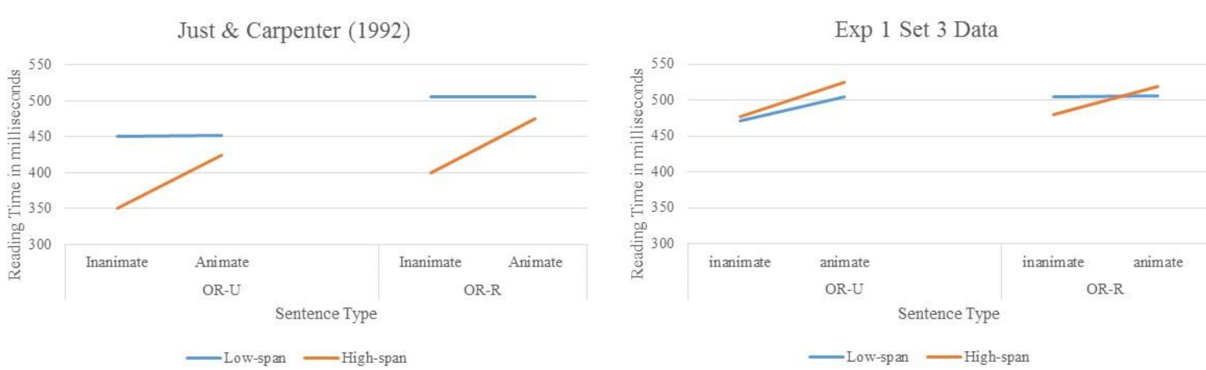


Figure 3.3.15. Recreated results of Just and Carpenter (1992) in the left panel, next to actual data for Set 3 in Experiment 1 on the right.

The data in Figure 3.3.15 clearly show no replication of the Just and Carpenter (1992) three-way interaction in the critical area. The same analyses were also conducted on the noun phrase immediately following the second verb to see if any effect might be carried over into this region, but this did not change the results. If one accepts the argument above that the animacy cue was still at play in Set 3, albeit in the opposite direction from the first two sets, one would still expect to find differences in the critical area reading times. These differences would just be the opposite of those found in the Just and Carpenter (1992) data (i.e., high-spans showing longer reading times for inanimate items than for animate ones). However, recall that we previously argued that three factors might serve to wash out interactions in the critical area. The first factor was the fast button pressing strategy employed by both low- and high-span participants, particularly in longer sentences. Just as in the previous sets, participants knew how long Set 3 sentences were from the beginning of the trial. Set 3 sentence lengths (8 to 12 words) happen to be intermediate between Set 1 and Set 2 (6 to 9 and 10 to 14 words respectively), suggesting that rushing might still play a role in critical area reading times of these sentences.

The second, related factor mentioned above is the location of the critical area. In Set 1, the critical area is right in the middle of the sentence, while in Set 2 it is the last word. For Set 3 the critical area is near the end of the sentence, but not the last word. The shifting of the three-way interaction from the critical area to the judgment response time could indicate that ambiguous inanimate items were not resolved on the second verb as predicted, but rather could only be properly interpreted after the noun phrase accompanying the verb had been read. Taking a look at the deep structure of reduced inanimate object-relative sentences from each set supports this argument because it can be seen that the final noun phrase is actually part of the same verb phrase as V2, just as how in Set 1 and Set 2 the *by* phrase and V2 respectively are components of verb phrases.

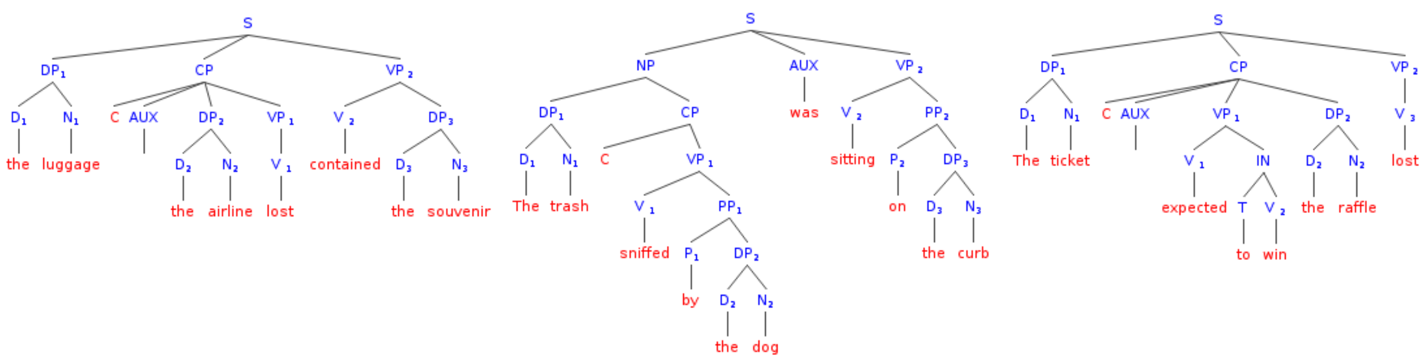


Figure 3.3.16. (Simplified) Deep structure sentence trees for inanimate reduced object-relative sentences in Set 3 (left), Set 1 (middle) and Set 2 (right).

Finally, recall that we argued earlier that sentences must be of sufficient complexity to elicit the three-way interaction. Set 1 did not appear to be complex enough, while Set 2 sentences appeared to be quite difficult. Set 3 sentences were subjectively rated as having a difficulty level somewhere between these two sets (see section 3.4 for statistical comparisons). Thus, it stands to reason they might be of sufficient complexity to elicit the differences predicted by single resource theory, albeit not in the critical area due to the other factors described above.

3.4 Cross-Set Comparisons

Our explanation of the pattern of results in Experiment 1 depends upon several arguments. First, we argue that Set 1 sentences may have been too easy, while Set 2 and Set 3 sentences were more difficult. Difficulty level matters because span group differences are not likely to emerge when sentences are too easy. Secondly, we argue that the WCA_{REQ} interaction seen with Set 3 sentences in judgment response times is not entirely an artifact of post-interpretive processing, despite occurring after entire sentence has been read. Instead, we argue that the interaction has shifted to judgment times due to the nature of Set 3 sentences. Finally, we argue that the length of sentences impacted the speed with which participants pressed to continue reading. Specifically, we argue that longer sentences (i.e., Set 1) led to more rushing than shorter ones (i.e., Set 2). In order to examine these arguments, we compared all three dependent variables (accuracy, judgment response time, and critical area reading time) across sets.

Accuracy.

Comparing Set 1 to Set 2, paired-samples *t*-tests revealed that accuracy was significantly higher on Set 1 than Set 2 for all types of sentences, *t*s > 6.63, *p*s < .001. Comparing Set 1 to Set 3, paired-samples *t*-tests showed that accuracy was significantly lower on Set 1 animate items (both unreduced and reduced), *t*s > -2.96, *p*s < .01. Finally, comparing Set 2 to Set 3, paired-samples *t*-tests revealed that accuracy was significantly lower in Set 2 than in Set 3 for all sentence types, *t*s > -6.07, *p*s < .001. Means are shown in Figure 3.4.1.

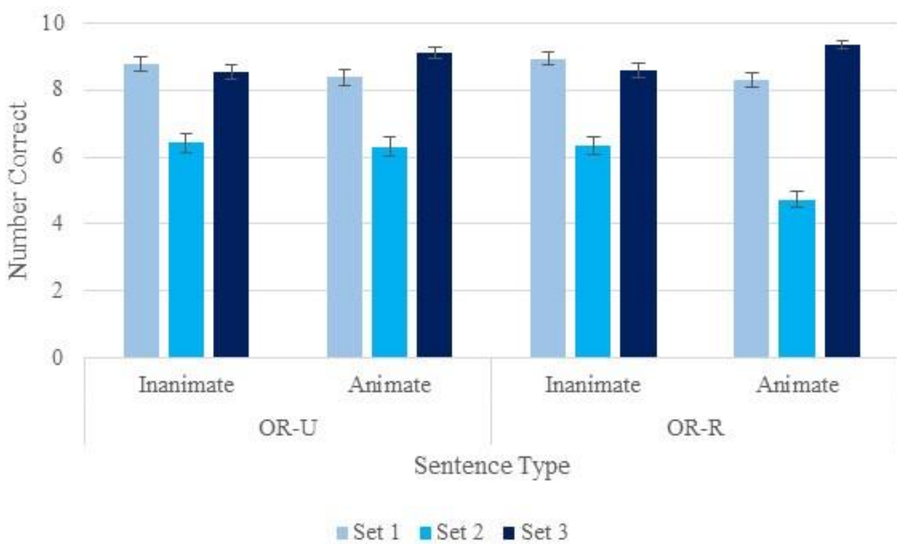


Figure 3.4.1. Judgment accuracy across sets, with standard error bars.

These accuracy results support the subjective reports of participants by showing that Set 2 sentences were indeed more difficult than either Set 1 or Set 3. The fact that animate items in Set 3 were easier than those in Set 1 does not support the subjective reports that Set 3 was of medium difficulty. It does, however, support the idea that animate items in Set 3 were not the most difficult items because of the flip in the effects of the animacy cue in these types of sentences.

Judgment Response Times.

Comparing Set 1 to Set 2, paired-samples *t*-tests revealed that judgment response time was significantly longer in Set 2 than in Set 1 for all sentence types except animate reduced, *t*s > -2.37, *p*s < .05. Comparing Set 1 to Set 3, paired-samples *t*-tests showed that judgment response time was significantly longer in Set 3 than in Set 1 for inanimate items (both types), *t*s > -3.30, *p*s < .01, but the reverse was true for animate reduced items, *t* (31) = 2.56, *p* < .05. Finally, comparing Set 2 to Set 3, paired-samples *t*-tests revealed that judgment response time was significantly slower in Set 2 than in Set 3 for animate items of both types, *t*s > 2.33, *p*s < .05. Mean number judgment response time for each set by sentence type can be seen in Figure 3.4.2.

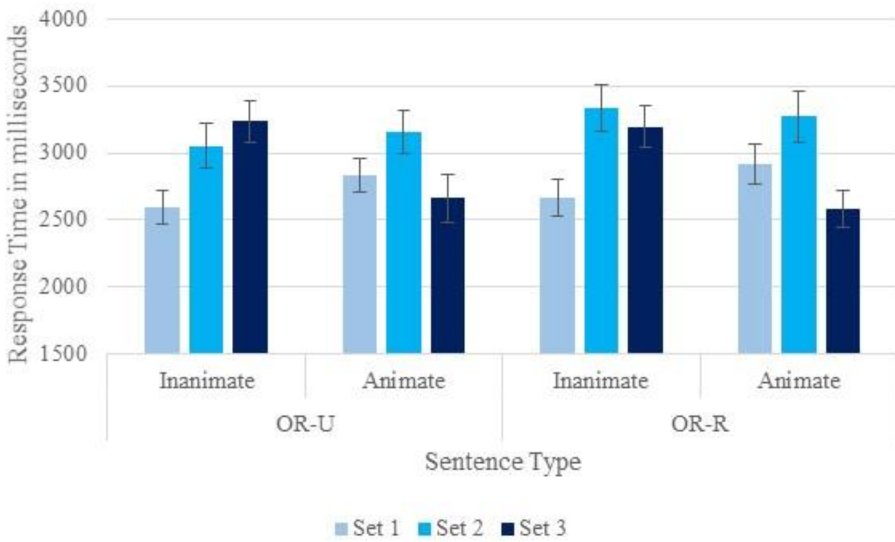


Figure 3.4.2. True/False judgment response times across the three stimulus sets of Experiment 1, with standard error bars.

The fact that Set 3 judgment response times are longer for inanimate than for animate items does suggest that such items require more post-interpretive processing at the time of judgment. However, the fact that animacy matters at all suggests that participants are able to take animacy into account during sentence processing. It is only in the more ambiguous case, the inanimate items, that extra post-interpretive processing is required to make a true/false comprehension judgment. This supports the idea that the animacy cue operates differently in Set 3 than in Set 1 and Set 2. It also supports the single resource theory by showing that the pragmatic cue can facilitate online processing, as does the fact that the cue only does this for those participants with sufficient capacity (i.e., high-spans).

Critical Area Reading Times.

Comparing Set 1 to Set 2, paired-samples *t*-tests revealed that critical area reading times were faster in Set 1 for all sentence types, *t*s > -4.81, *p*s < .001. Comparing Set 1 to Set 3, paired-samples *t*-tests revealed that critical area reading times were faster in Set 1 for all sentence types, *t*s > -4.46, *p*s < .001. Finally, comparing Set 2 to Set 3, paired-samples *t*-tests revealed that critical area reading times were longer in Set 2 for all sentence types, *t*s > 2.79, *p*s < .01.

One explanation for the longer critical area reading times in Set 2 is that the critical area is also the last word of the sentence, and thus subject to sentence wrap-up effects (Balogh et al., 1998). However, sentence wrap-up effects do not explain why there is also an interaction of complexity and WMC in the Set 2 critical area. Finding that interaction in Set 2 but not in Set 1 supports the argument that rushing in longer sentences (i.e., Set 1) may have contributed to a lack of significant interactions with WMC in Set 1. When participants slowed down in the critical area of the more difficult Set 2 sentences, this interaction emerged. Rushing may still have been present in Set 3, but to a lesser extent than in Set 1 because of the intermediate length of sentences in Set 3.

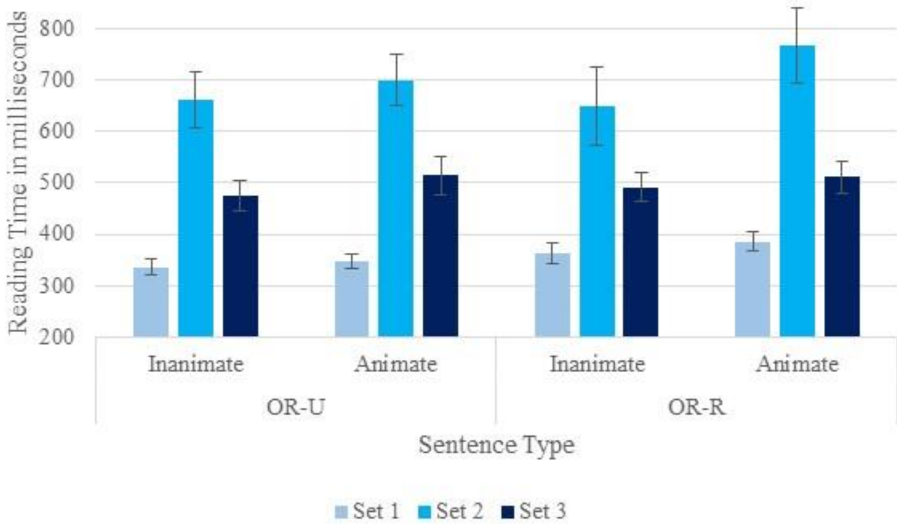


Figure 3.4.3. Critical area reading times across sets, with standard error bars.

3.5 Discussion

The goal of Experiment 1 was to produce conditions under which syntactic processing task performance on three types of experimental sentences formed in the manner of three previous studies (Just and Carpenter, 1992; Waters and Caplan, 1996b; Waters and Caplan, 2004) could be directly compared with each other using the same presentation method and syntactic processing task judgment type.

Presentation method (self-paced word-by-word reading paradigm) and task judgment type (true/false comprehension items) were chosen specifically to replicate, if possible, the results of Just and Carpenter (1992). Using *by* phrase object-relative sentences, Just and Carpenter (1992) found a two-way interaction of WMC and animacy that supported single resource theory by suggesting that high- but not low-span participants could use the animacy pragmatic cue online to facilitate processing in the critical area of complex sentences. Waters and Caplan (1996b) and Waters and Caplan (2004) used garden path and subject-object object-relative sentences (among others) in a syntactic processing task with an acceptability/plausibility judgment to refute the results of Just and Carpenter (1992). Both studies failed to find a significant relationship between WMC and performance on the syntactic processing task. However, comparison of these results to those of Just and Carpenter (1992) is somewhat questionable, given that the studies also differed in presentation method and syntactic processing task judgment type.

Using three sets of sentences, one for each of the previous experiments, we attempted to determine the effects of these methodological differences on the three main dependent variables of the syntactic processing task (accuracy, judgment response time, and critical area reading time) by adding animacy to Sets 2 and 3, creating reduced versions of Set 3 experimental sentences, and presenting all sentences in the same method and with the same type of judgment. Given the data and arguments in the previous literature, we hypothesized that if single resource theory were true we would find significant interactions with WMC for the critical area reading times in all three sets. In particular, we were interested in whether or not these methods would produce a three-way interaction of WMC, complexity, and animacy (i.e., the WCA_{REQ} interaction) that would satisfy Waters and Caplan's (1996a) criticisms of single resource theory.

Our results for Set 1 failed to replicate those of Just and Carpenter (1992), despite the fact that this set contained the same type of *by* phrase sentences and we used the same judgment type they did. Critical area reading times did show an interaction of complexity and animacy, but there was no effect of WMC. Closer examination of the data and subjective participant reports suggested that the failure to replicate may have been due to insufficient item difficulty and a rushing reading strategy. Although we attempted to recreate the stimuli from Just and Carpenter (1992), accuracy levels, subjective reports, and evidence of online use of the pragmatic cue by low-span participants all suggest that our experimental sentences were less difficult than those in the previous study for unknown reasons. Furthermore, subjective reports, comparisons of 'fast' and 'slow' readers, and comparisons of reading times across sets support the idea that the majority of participants adopted a rushing reading strategy that may have obscured any effects of WMC on critical area reading times. This is in contrast to previous literature that used the same presentation methods and found per word reading times measured in this manner to be equivalent to first-pass reading times measured with an eye-tracker (e.g., Just et al., 1982).

With Set 2 sentences we found a two-way interaction of WMC and complexity that was not found by Waters and Caplan (1996b) using the same type of sentences, most likely because the previous study used mean per word reading times rather than directly measuring critical area reading times as in the current study. This two-way interaction is not, however, a direct replication of Just and Carpenter's (1992) interaction of WMC and animacy nor is it the WCA_{REQ}. Examination of standard error bars on these means suggests that this difference did not reach significance in the current study due to high variability. Based upon accuracy data and consideration of the nature of the Set 2 experimental sentences and their corresponding true/false comprehension statements, we argue that this high variability in the high-span group may be a result of differing reading strategies prompted by task demands. Specifically, we argue that the animacy cue was not as relevant to making true/false judgments in Set 2 as in Set 1 because of the lack of an explicitly named subject of the object-relative clause, and that because of this decreased relevance, at least some high-span participants adopted a reading strategy that ignored animacy during processing of the complex sentences.

With Set 3 sentences we found the WCA_{REQ} interaction, but it was located in the judgment response times rather than in the critical area of the sentence. It is not entirely clear why this was the case, but one possibility suggested by cross-set comparisons is that the Set 3 sentences were long enough to produce a rushing reading strategy similar to that seen in Set 1. It could also be argued, however, that the effects of WMC are seen in judgment response times and not the critical area because judgments in this set required more post-interpretive processing. We cannot entirely rule out this possibility, though it seems unlikely based upon cross-set comparisons showing that Set 3 judgment times were only longer than those in Set 1 for the most difficult items (inanimate reduced object-relatives in this set). In other words, the results suggest that high-spans do incorporate the animacy of nouns online, but require extra time to make judgments when the noun's animacy status is not helpful for answering the true/false item.

Together, the results of Experiment 1 do not fully support either single or dual resource theory, but they do suggest that several factors that have been previously overlooked can significantly impact results and should thus be considered carefully in future research.

CHAPTER 4. EXPERIMENT 2

4.1 Introduction

Experiment 2 was an attempt to replicate Waters and Caplan (1996b, 2004) using their judgment type (sense judgments) on a subset of their stimuli (garden path and subject-object sentences respectively) as well as the *by* phrase sentences of Just and Carpenter (1992) in order to determine what effects, if any, judgment type has on syntactic processing task performance.

In addition to modifying the judgment type, data analyses are somewhat different in Experiment 2, to match the studies we are attempting to replicate. Waters and Caplan (1996b) did not consider critical area reading times to be the primary dependent variable because they did not measure it. Rather, they first analyzed A' scores and response times for the sentence judgments. Beyond that, they did calculate mean per word reading times by dividing the total time to read the sentence and respond to the judgment by the number of words in the sentence. These same measures will be examined in the current experiment. In addition to these analyses, we will also look at our data in the manner of Waters and Caplan (2004). In this study judgment accuracy was looked at as a mean proportion correct, judgment response times were not considered of interest, and critical area listening times were reported and compared to listening times in other areas of the sentences. Finally, these data will also be looked at in the manner of Just and Carpenter (1992) for ease of comparison with the first experiment.

4.2 Methods

Participants

Participants were 32 (27 female) undergraduate psychology students at Louisiana State University who received course credit for their participation. All participants reported normal or corrected to normal vision and hearing, and no history of language or speech impairments. Mean age was 19.50 (SD = 1.32). Sixteen participants were categorized as low-span and 16 as high-span (descriptives for WMC tasks can be found in Table 4.2.1 below). An additional 19 participants were run in Experiment 2, but 15 were dropped because they were medium-span and the remaining four did not achieve the 85% performance criterion on at least one automated task.

Table 4.2.1. Experiment 2 WMC measure descriptives.

		Low-span		High-span	
		Mean	*SD*	Mean	*SD*
Automated Tasks	R-span	-1.01	0.66	0.89	0.30
	S-span	-0.94	0.84	0.90	0.52
	O-span	-1.03	1.14	0.90	0.28
	Composite	-1.23	0.54	1.11	0.18
D & C (1980)	Original	2.06	0.25	2.63	0.50
	Recode	0.72	0.08	0.79	0.06

Syntactic Processing Task

The three stimulus sets of Experiment 2 are identical to those used in Experiment 1, including the main clause (MC) and subject-relative (SR) filler sentences, with the exception that items which were previously followed by a false statement for the true/false judgment were modified to contain a word that did not make sense in the context of the rest of the sentence. Items that were previously followed by a true comprehension statement were not modified. After each sentence, participants were asked to judge whether or not it made sense by pressing keys marked "Yes" or "No" on the keyboard.

Because of the very different constructions of object-relative sentences in the three sets, it was the case that the semantically anomalous words were found in different locations within each type of sentence. For the Waters and Caplan (1996b) garden path sentences (both OR-U and OR-R), the semantically anomalous words of necessity preceded the final verb sentence (e.g., The hair that was rinsed in hot *bacon* dried.), while in the Waters and Caplan (2004) subject-object sentences, semantically anomalous words were always last, located in the noun phrase following the critical verb (e.g., The document that the machine copied shocked the *cassette*.), and in the Just and Carpenter (1992) sentences the anomalous word was the last in the by phrase (e.g., The pedestrian that was followed by the *potato* amused the gathered spectators.)

All other presentation methods were the same as in Experiment 1, including counterbalancing of set presentation and randomization of blocks within sets, thus we have once again have judgment accuracy, judgment response time, and critical area reading time data for analyses.

Working Memory Capacity Measures

The same four working memory capacity (WMC) measures from Experiment 1 were used in Experiment 2, in precisely the same way. Once again, composite z-scores for the automated tasks were used to select participants for the low- and high-span groups, and results were also calculated using the recoded Daneman and Carpenter (1980) Reading-span task to split participants. Results were the same for both splits, so the first analyses with the composite z-scores are the ones reported here.

4.3 Results

Waters and Caplan (1996b) Stimuli – Set 2 (Appendix C)

The Set 2 stimuli in Experiment 2 were identical to those in Experiment 1, with the exception that items which were followed by a false comprehension item in Experiment 1 were modified in Experiment 2 to contain a semantically anomalous word. This word always came before the critical area (i.e., the second verb) in the OR-U and OR-R sentences because the critical area coincides with the last word in these garden path sentences.

The primary variable of interest to Waters and Caplan (1996b) was sense judgment accuracy, and specifically A' scores. In their whole sentence condition, Waters and Caplan (1996b) also calculated mean per word reading times by dividing the time between initial sentence presentation and participant response by the number of words in the sentence. In order for more direct comparisons, these measures will be added to the reported results for Set 2.

Accuracy.

A 4 (complexity) x 2 (WMC) within-between repeated measures ANOVA showed a main effect of complexity, F (3, 90) = 47.18, $p < .001$, $\eta_p^2 = .47$, but no effect of WMC and no interaction between the two. Note that animacy was not included in this analysis because Waters and Caplan (1996b) did not utilize this factor and we are attempting to determine if our data resembles theirs. A' were significantly lower on reduced object-relative items than all other items for both low- and high-span participants, *t*s > 5.74, *p*s < .001. The left panel of Figure 4.3.1 below shows the A' scores reconstructed from the results of the relevant sentences in Waters and Caplan (1996b), while the right panel shows the A' scores for Set 2 (calculated according to Stanislaw and Todorov, 1999). Comparing the panels, it can be seen that reduced object-relative performance was much lower in Set 2 than in the Waters and Caplan (1996b) data, though the unreduced object-relative scores are comparable. Despite this difference, the Set 2 data represents a replication of Waters and Caplan (1996b) in that there is no effect of WMC and no interaction with complexity.

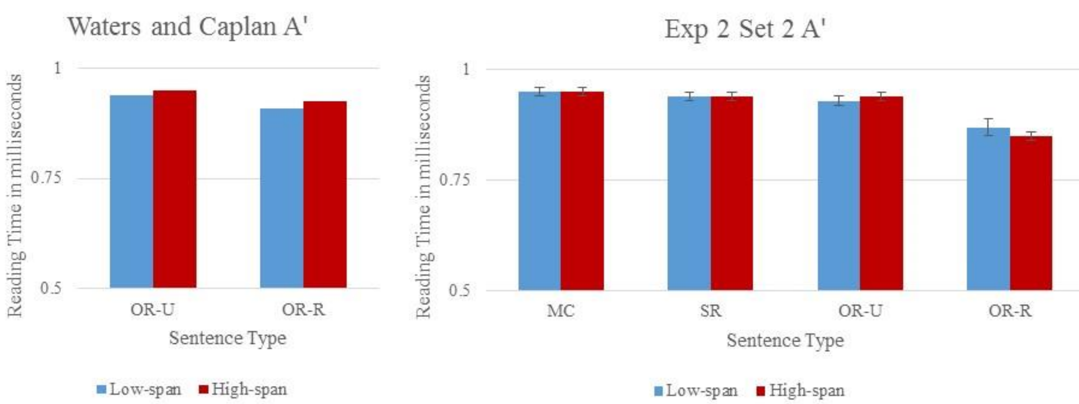


Figure 4.3.1. Recreated A' scores from Waters and Caplan (1996b) in the left panel compared to Set 2 A' scores in Experiment 2 on the right, with standard error bars.

For the sake of comparison, and because Waters and Caplan (1996b) did not manipulate animacy, sense judgment accuracy in terms of number of correctly judged items was also examined in the manner of Experiment 1, with a 4 (complexity) x 2 (animacy) x 2 (answer) x 2 (WMC) mixed within-between repeated measures ANOVA. This analysis showed main effects of complexity, F (3, 90) = 37.84, p < .001, η_p^2 = .56, animacy, F (1, 30) = 16.97, p < .001, η_p^2 = .36, and answer, F (1, 30) = 30.25, p < .001, η_p^2 = .50. These main effects were qualified by two-way interactions of complexity and animacy, F (3, 90) = 5.14, p < .01, η_p^2 = .15, of complexity and answer, F (3, 90) = 47.04, p < .001, η_p^2 = .61, and of animacy and answer, F (1, 30) = 4.41, p < .05, η_p^2 = .13. These two-way interactions were in turn qualified by three-way interactions of complexity, animacy, and answer, F (3, 90) = 13.19, p < .001, η_p^2 = .31, of complexity, answer, and WMC, F (3, 90) = 4.45, p = .01, η_p^2 = .13, and of complexity, animacy, and WMC, F (3, 90) = 3.31, p = .02, η_p^2 = .10. Figure 4.3.2 displays the means by sentence type.

Examination of the three-way interaction of complexity, animacy, and answer revealed that acceptable item accuracy was higher for inanimate main clauses than for animate main clauses, t (31) = -3.52, p < .01, for inanimate reduced object relative clauses than for animate reduced object-relative clauses, t (31) = -3.35, p < .01, and for animate unreduced object-relative clauses than for inanimate unreduced object-relative clauses, t (31) = 2.20, p < .05. For unacceptable items, accuracy was higher for animate main clauses than for inanimate main clauses, t (31) = 3.04, p < .01, and for inanimate subject-relative clauses than for animate subject-relative clauses, t (31) = -4.06, p < .001. Examination of the three-way interaction of complexity, answer, and WMC revealed that the low-spans were significantly more accurate than high-spans on acceptable reduced object-relative items, t (30) = 2.21, p = .04, but no other contrasts were significant.

Finally, examination of the three-way interaction of complexity, animacy, and WMC revealed low-spans were significantly more accurate than high-spans on the animate reduced object-relative items, t (30) = 3.17, p < .01, but no other contrasts were significant. Note that this three-way interaction includes the variables of the WCA_{REQ} , but is not of the form predicted by single resource theory in that the theory would not predict that low-spans would have higher accuracy than high-spans. However, this odd result does not necessarily represent a serious challenge to the single resource theory because accuracy is an offline measure and Just and Carpenter's (1992) key finding rested upon support from the online measure of critical area reading time.

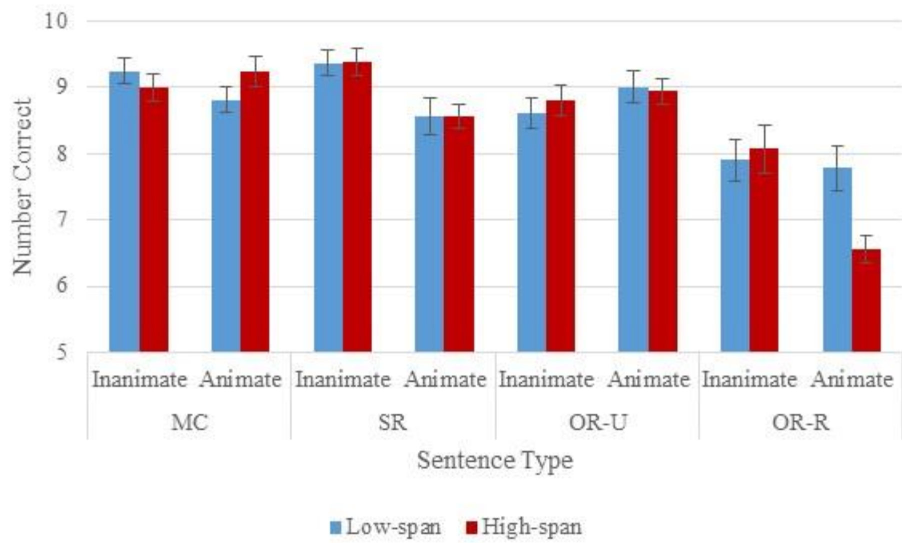


Figure 4.3.2. Experiment 2 Set 2 Sense judgment accuracy, with standard error bars.

As in Experiment 1, performance on unacceptable items in Set 2 was apparently significantly more difficult than in other sets. Ten of 32 participants did not have any accurate unacceptable items for animate reduced object-relative sentences. Unlike Experiment 1, however, this does not affect the analysis of critical area reading times because only correctly judged acceptable items will be used, but there was one person out of 32 that did not have any correctly judged acceptable animate reduced object-relative items.

Judgment Response Time.

Waters and Caplan (1996b) did not analyze sense judgment response times independently in their whole sentence condition, which is the condition of their study most analogous to the current experiment. Rather, they took the entire time from the initial presentation of a sentence until the sense judgment response (i.e., reading time and response time) and divided by the number of words in the sentence to estimate a mean per word reading time. This measure is more analogous to the critical area reading time measure than to the sense judgment response time, so here we will look at sense judgment response times exactly as we did in Experiment 1. Figure 4.3.3 displays the mean sense judgment response times for Set 2.

A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on sense judgment response times showed an interaction between complexity and animacy, F (3, 90) = 3.34, p = .02, η_p^2 = .10, and another between animacy and WMC, F (1, 30) = 5.57, p = .03, η_p^2 = .16. The interaction between complexity and animacy is due to significantly longer sense judgment response times for animate subject-relatives than for inanimate subject-relatives, t (31) = 2.54, p = .02. The interaction between animacy and WMC is due to low-spans having significantly faster sense judgment response times than high-spans for inanimate subject-relatives, inanimate unreduced object-relatives, and inanimate reduced object-relatives, all *t*s > -2.21, *p*s < .05. These two interactions do not represent a replication of key finding of Just and Carpenter (1992).

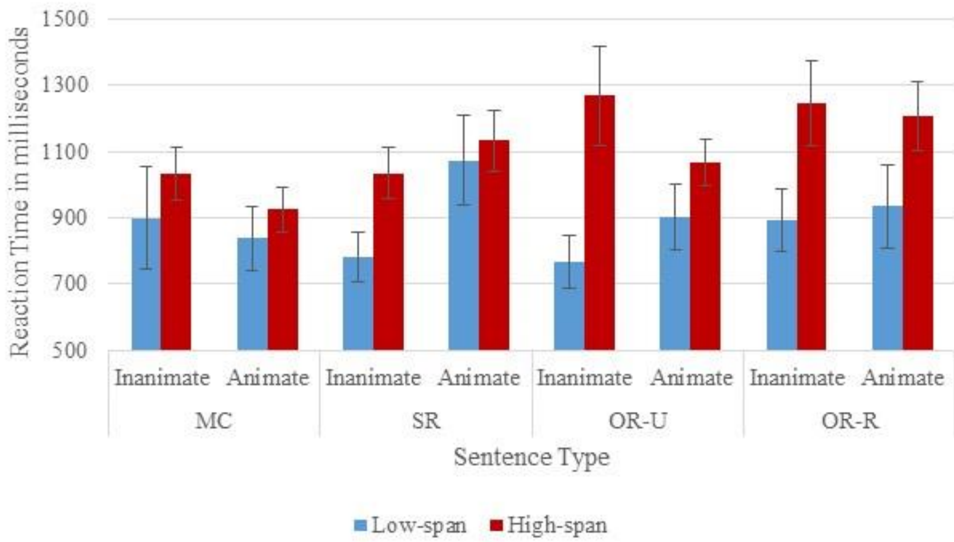


Figure 4.3.3. Experiment 2 Set 2 Sense judgment response times, with standard error bars.

Critical Area Reading Time.

As already mentioned, Waters and Caplan (1996b) did not examine critical area reading times but rather divided reading-plus-judgment reaction times by the number of words in each sentence to arrive at mean per word reading times. In the current experiment, sense judgment reaction times were already separated out and examined above, but we did calculate mean word reading times by summing all individual word reading times and dividing by the number of words in the sentence. The left panel of Figure 4.3.4 below shows the mean per word reading times as recreated from the Waters and Caplan (1996b) data, while the right panel shows the mean per word reading times calculated on the Set 2 data. A 4 (complexity) x 2 (WMC) within-between repeated measures ANOVA showed no significant main effects of complexity or WMC, and no interactions. The lack of an effect of WMC and an interaction is consistent with Waters and Caplan (1996b), though the lack of an effect of complexity is not.

The problem with the Waters and Caplan (1996b) method of calculating per word reading times is that it assumes reading times remain steady throughout sentences. This assumption is contrary to the prediction of the single resource theory that reading times will increase when processing demands are high (i.e., in the critical area of sentences). Calculating mean per word reading times from reading times for the whole sentence cannot possibly address any variability attributable to region. Thus, we also looked at critical area reading times in Set 2 sentences in the manner of Experiment 1. However, unlike in Experiment 1, it is not sound to include all items in this analysis because the presence of an anomalous word is likely to disrupt processing independently of any effects of syntactic complexity, and in Set 2 garden path sentences the anomalous word always came before the critical area, due to the critical area coinciding with the last word of the sentence. Thus, the current analysis was conducted only on correctly judged acceptable items rather than on all correctly judged items. It should also be noted that one person did not have any correct acceptable animate reduced items, so that person was dropped from the analysis.

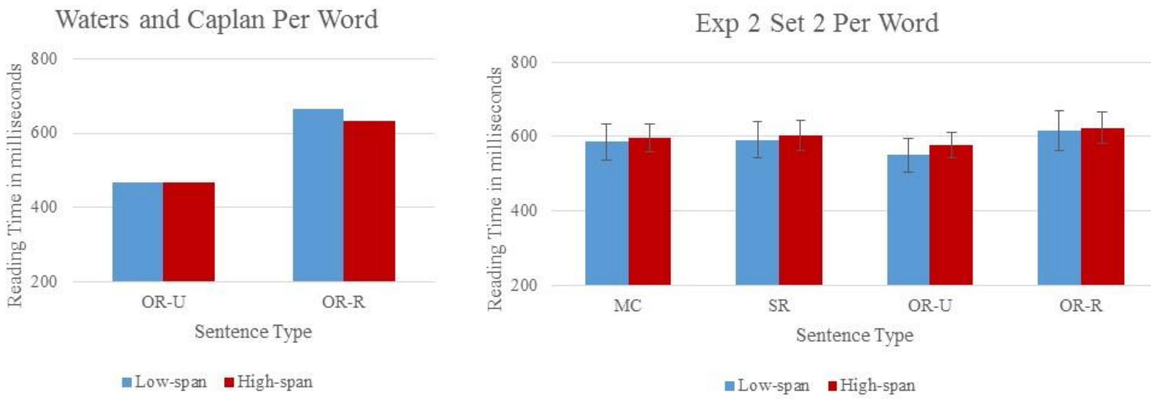


Figure 4.3.4. Recreated per word reading times from Waters and Caplan (1996b) in the left panel compared to per word data from Set 2 in Experiment 2 on the right, with standard error bars.

Figure 4.3.5 below shows mean critical area reading times for each sentence type for acceptable items only. These times were trimmed as described in Experiment 1, using individual means and standard deviations to replace less than 2% of outlying data points overall and less than 2% for any individual. A 4 (complexity) x 2 (animacy) x 2 (WMC) within-between repeated measures ANOVA showed a main effect of complexity, F (3, 87) = 5.09, $p < .01$, $\eta_p^2 = .15$, such that critical area reading times were faster in both main clause and subject-relative clause sentences than in reduced object-relatives, *ts* > -2.77, *ps* < .01. There was also a main effect of animacy, F (1, 29) = 6.05, $p = .02$, $\eta_p^2 = .21$, such that critical area reading times were faster for inanimate than for animate items. The lack of a two-way interaction of WMC and animacy does not replicate the findings of Just and Carpenter (1992), while the lack of a two-way interaction of WMC and complexity fails to replicate Experiment 1 results.

Figure 4.3.6 below shows the comparison of Just and Carpenter (1992) and Set 2 critical area reading times for object-relative sentences only. Note that the trend in the Set 2 data for high-spans is similar to that in the Just and Carpenter (1992) data; the high-spans show faster reading times for the critical area in the object-relative sentences when the head noun is inanimate than when it is animate. However, unlike Just and Carpenter (1992), this trend is also present to some extent in the low-span data, resulting in a lack of a significant three-way interaction. Finally, note the difference in scale between the two panels. As in Experiment 1, Set 2 critical area reading times are significantly longer overall than those found in the Just and Carpenter (1992) data, likely due to the same sentence wrap-up effects discussed earlier (Balogh, et al., 1998).

The bigger question in the context of the current study is why the Set 2 results of Experiment 1 were not replicated in Experiment 2, when the only thing that changed was the type of judgment being made about the items. One reason may be that the Experiment 2 judgment type required the insertion of an anomalous word into the Set 2 sentences. These anomalous words had to be located before the critical area of the experimental sentences because the critical area was the very last word. Although we only included acceptable items in the current analysis, the presence of the anomalous word before the critical area in half the experimental trials may have encouraged the use of a different task strategy for all sentences because participants could easily have learned that the key to answering the sense judgment was the second-to-last word of the majority of the sentences in the task. A change in reading strategy could very well have washed out the interaction found in Experiment 2. One way to determine if this explanation holds weight is to look at reading times on the anomalous words in Experiment 2 sentences (left panel of Figure 4.3.7 below) compared to reading times on the non-anomalous words in the same location in Experiment 1 sentences (right panel of Figure 4.3.7 below).

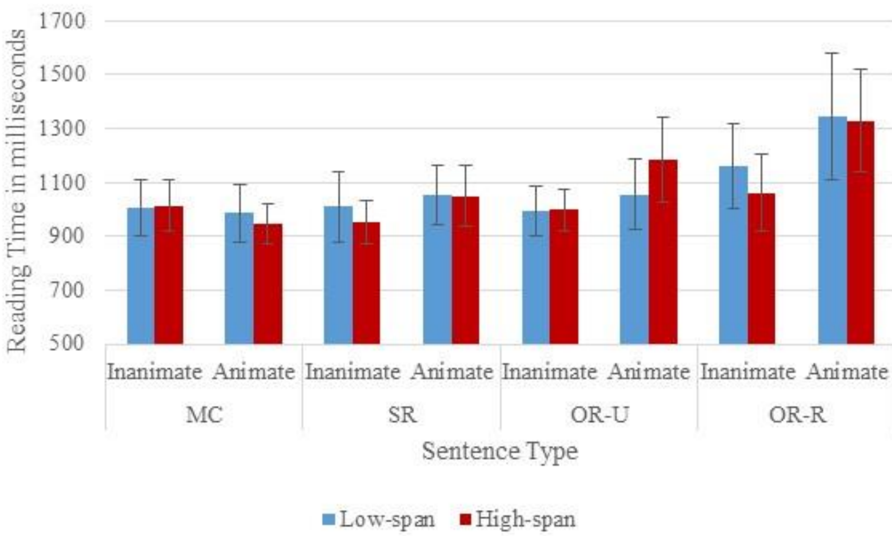


Figure 4.3.5. Experiment 2 Set 2 critical area reading times, with standard error bars.

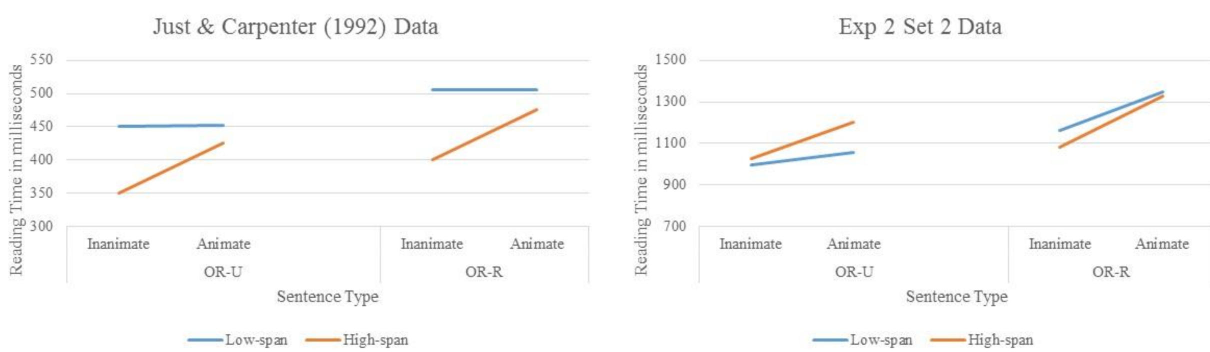
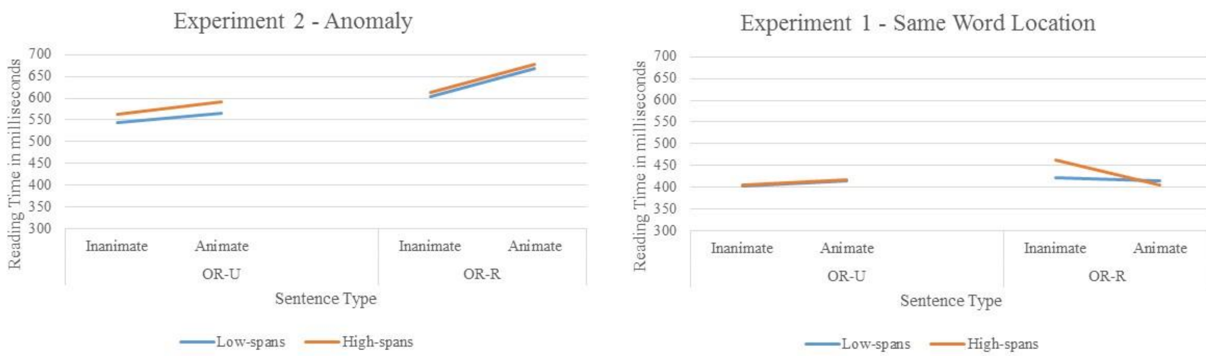


Figure 4.3.6. Recreated Just and Carpenter (1992) data in the left panel compared to Experiment 2 Set 2 critical area reading times on the right.

As we can see in Figure 4.3.7, reading times at the anomalous word in Experiment 2 show an upward trajectory for both low- and high-span individuals that continues into the last word of the sentence, which is also the critical area seen in Figure 4.3.6. In Experiment 1, on the other hand, reading times are fairly flat just before the critical area. The interaction of complexity and WMC does not emerge until the final word. These differing patterns at the same location in Set 2 sentences do seem to suggest that there is a different strategy at play in the two judgment conditions. Thus, although Waters and Caplan (1996a) argued that the sense judgment was a better syntactic processing measure because it was a less offline measure than the comprehension item, it seems likely that this judgment type is problematic, at least for use with object-relative sentences constructed in the manner of Waters and Caplan (1996b) that require the anomaly to appear before the critical area of the sentence because it changes the way participants read the sentences and the nature of this change is the same in both low- and high-spans.



Figuer 4.3.7. Experiment 2 anomalous word reading times in the left panel compared to Experiment 1 same-location word reading times in the right panel.

Waters and Caplan (2004) Stimuli – Set 3 (Appendix D)

The Set 3 stimuli in Experiment 2 were identical to those in Experiment 1, with the exception that items which were followed by a false comprehension item in Experiment 1 were modified in Experiment 2 to contain a semantically anomalous word. Anomalous words were deliberately placed after the critical area (i.e., the second verb). Because the critical area was towards the end of the sentence, this meant that the anomalous word always ended up being the last word of the sentence. Note that this differs from Waters and Caplan (2004) in that (as far as can be determined from their example sentences) their anomalous words appeared before the critical area (e.g., The secretary that the camera met drove the car.).

The results of Waters and Caplan (2004) were presented somewhat differently than those of Waters and Caplan (1996b). First, their sense judgment accuracy measure was proportion correct rather than A'. Secondly, because they presented their stimuli auditorily in phrases, with participants controlling the rate of presentation, Waters and Caplan (2004) were able to analyze listening times across sentence regions. The results of Set 3 are presented in accordance with these methods and those of the previous experiments for the sake of comparison.

Accuracy.

A 4 (complexity) x 2 (WMC) mixed within-between repeated measures ANOVA on mean proportion correct was conducted in a manner analogous to Waters and Caplan (2004). This analysis showed a main effect of complexity, F (3, 90) = 3.33, p = .02, η_p^2 = .10, such that proportion scores were higher for main clause than for subject-relative clause sentences, t (31) = -2.90, p < .01, and lower for subject-relative clause than for reduced object-relative clause sentences, t (31) = -2.07, p < .05. There was no main effect of WMC nor an interaction. This is compatible with the findings of Waters and Caplan (2004). The left panel of Figure 4.3.8 below shows mean proportion correct recreated from the Waters and Caplan (2004) data, while the right panel shows the Set 3 data.

Waters and Caplan (2004) did not manipulate animacy, but this factor was present in the Set 3 data, so we also ran a 4 (complexity) x 2 (animacy) x 2 (answer) x 2 (WMC) mixed within-between repeated measures ANOVA on the mean proportion correct. This analysis showed a main effect of complexity, F (3, 90) = 3.33, p = .02, η_p^2 = .10, as well as an interaction of complexity and answer, F (3, 90) = 8.60, $p < .001$, $\eta_p^2 = .22$, that was then qualified by a three-way interaction of complexity, answer, and WMC, F (3, 90) = 3.91, $p = .01$, $\eta_p^2 = .12$. Examination of this interaction showed that low-spans had higher accuracy on acceptable animate main clause sentences than on unacceptable animate main clause sentences, t (15) = 2.15, $p < .05$, while high-spans had higher accuracy on acceptable inanimate subject-relative sentences than on unacceptable inanimate subject-relatives, t (15) = 4.57, $p < .001$, and lower accuracy on acceptable animate and inanimate unreduced object-relative items than unacceptable animate and inanimate unreduced object-relatives, *t*s < -2.61, *p*s = .02, as well as lower accuracy on acceptable animate reduced items than on unacceptable animate reduced items, t (15) = - 2.61, $p = .02$. These results suggest a bias to call object-relative items unacceptable.

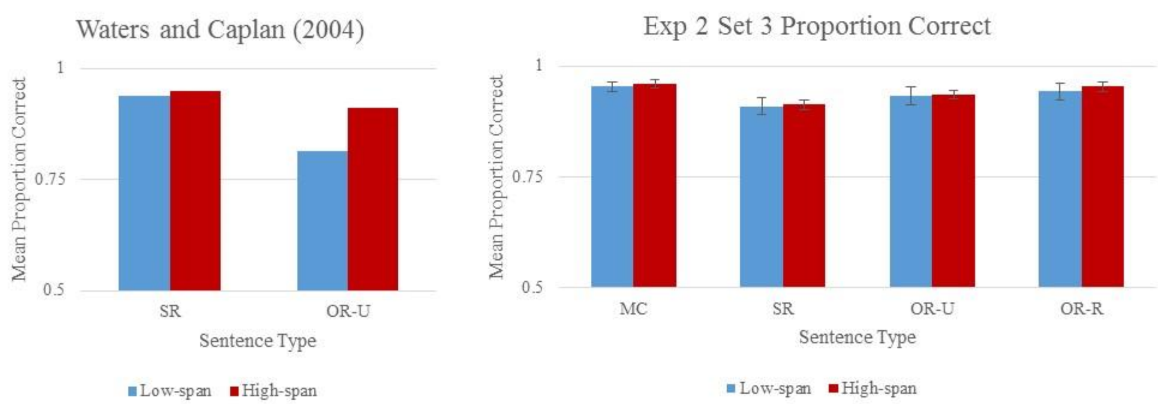


Figure 4.3.8. Recreated proportion correct data from Waters and Caplan (2004) in the left panel compared to data from Experiment 2 Set 3 on the right, with standard error bars.

There was also a three-way interaction of complexity, animacy, and answer, F (3, 90) = 2.87, $p = .04$, $\eta_p^2 = .09$. Examination of this interaction revealed that accuracy was significantly better for acceptable animate main clause items than for acceptable inanimate main clause items, t (31) = 2.78, $p < .01$, and for acceptable inanimate subject-relative items over acceptable animate subject-relative items, t (31) = -2.15, $p = .04$, but no other contrasts were significant.

Finally, there was a three-way interaction of complexity, animacy, and WMC, F (3, 90) = 4.54, $p < .01$, $\eta_p^2 = .13$. Examination of this interaction showed that high-spans were significantly more accurate on inanimate unreduced object-relative clauses than on animate unreduced object-relative clauses, t (15) = 2.72, $p = .02$. No other contrasts were significant. Figure 4.3.9 below shows the mean proportion correct by sentence type. This three-way interaction is not quite the WCA_{REQ} single resource theory would predict because high-spans do not also show a significant effect of animacy on the more complex reduced object-relatives, but as we have mentioned previously, this is not a fatal flaw because accuracy is not the primary variable upon which Just and Carpenter (1992) base their arguments.

Recall from Experiment 1 that we argued that the animacy cue operated differently in Set 3 sentences, with animacy actually being the cue that would make the sentence less ambiguous due to the subject-object word order, which might lead one to expect accuracy to be higher on animate object-relatives than on inanimate ones. However, this is a different type of judgment, in which the participant need only know if the last noun of the sentence (which is not the subject) makes sense in the context of the rest of the sentence, so the meaning of this difference is unclear.

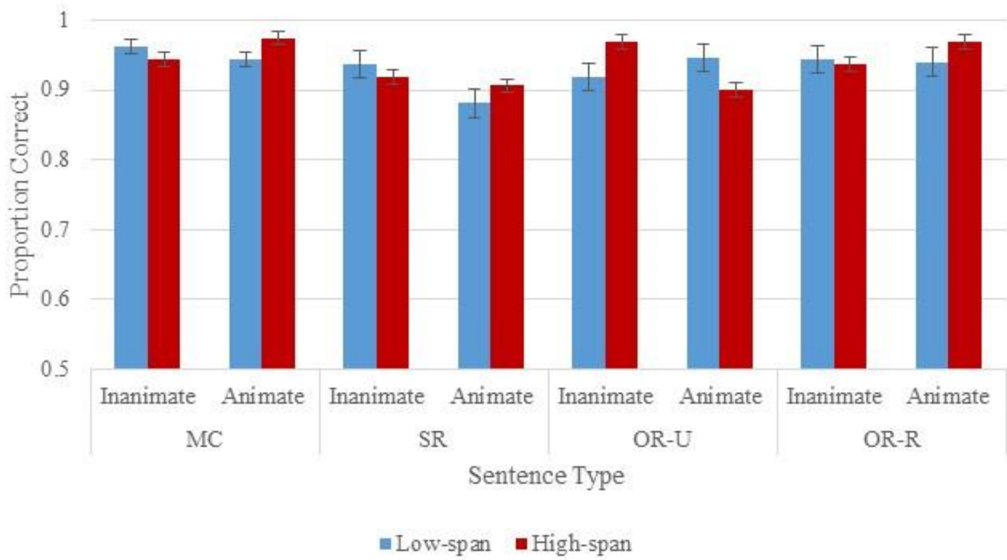


Figure 4.3.9. Experiment 2 Set 3 Sense judgment proportion correct, with standard error bars.

Judgment Response Time.

Unlike Waters and Caplan (1996b), Waters and Caplan (2004) did examine sense judgment response times in their paper. Figure 4.3.10 below shows a reconstruction of the Waters and Caplan (2004) data in the left panel and the analogous mean sense judgment reaction times in the Set 3 data, collapsed across answer. A 4 (complexity) x 2 (WMC) mixed within-between repeated measures ANOVA on the Set 3 data revealed a marginal main effect of WMC, F (1, 30) = 4.18, p = .05, η_p^2 = .12, such that high-spans showed longer response times than low-spans.

Because the Set 3 data also included animacy as a factor, we also conducted a 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on judgment response times. This analysis showed a main effect of WMC, F (3, 90) = 4.29, $p < .05$, η_p^2 = .13, such that high-spans had longer judgment response times than low-spans. No other main effects were significant, nor were there any significant interactions. Figure 4.3.11 below shows the mean sense judgment reaction times collapsed across answer.

These results do not replicate the WCA_{REQ} interaction found in Set 3 judgment response times in Experiment 1. In Experiment 1 we argued that the WCA_{REQ} interaction may have been located in response times rather than the critical area because of the deep structure of the sentences (i.e., that the last noun phrase was part of the verb phrase, so perhaps verb two was not the sole word in the critical area). If this argument is accepted, it could also explain why the WCA_{REQ} interaction is not found in Experiment 2, where the anomalous word necessary for the sense judgment is found in the noun phrase after the second verb and is thus part of the syntactic component (i.e., the verb phrase) that constitutes the critical area. As we argued with Set 2 above, it is possible that the location of the anomaly may change the strategy employed while reading sentences, even when an anomalous word is not actually present in the item (i.e., acceptable items), either because it disrupts processing of the syntactic component altogether (e.g., in unacceptable items) or because it's predictable location changes how participants read the sentences in general (i.e., in both unacceptable and acceptable items).

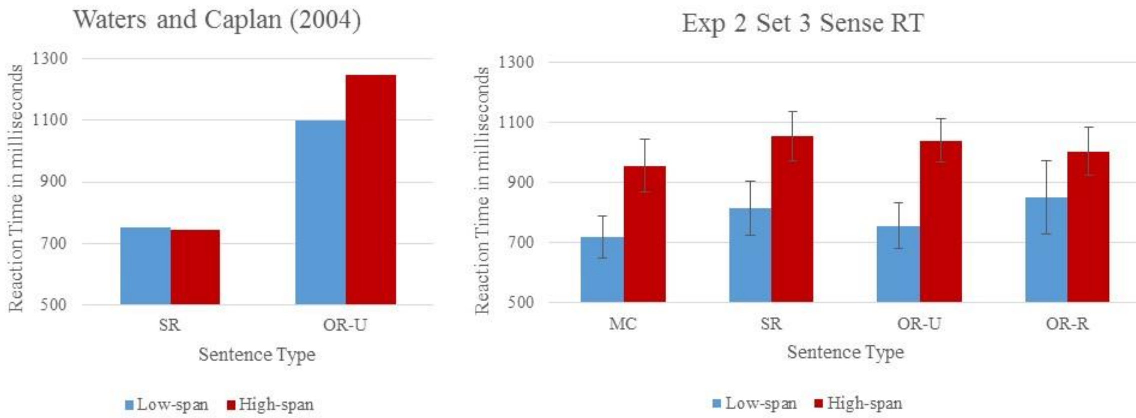


Figure 4.3.10. Recreated Sense judgment response times from Waters and Caplan (2004) in the left panel compared to data from Experiment 2 Set 3 on the right, with standard error bars.

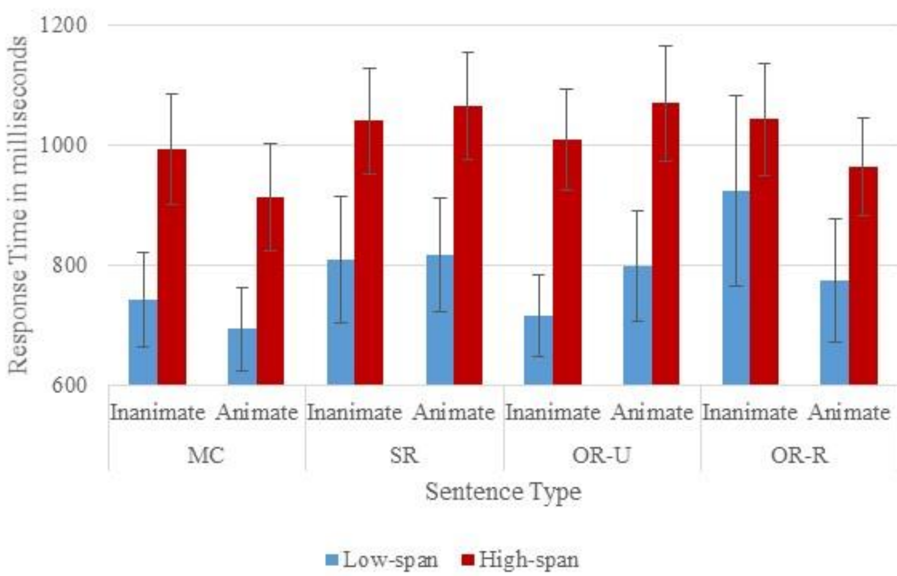


Figure 4.3.11. Experiment 2 Set 3 Send judgment response times, with standard error bars.

As with Set 2 above, one way to determine whether or not the location of the anomalous word changed how participants read sentences is to look at reading times in the anomaly location of Experiment 2 sentences (left panel of Figure 4.3.12 below) compared to reading times on the same word location in Experiment 1 (right panel of Figure 4.3.12 below). Figure 4.3.12 shows these reading times for acceptable experimental sentences only.

As can be seen in Figure 4.3.12, reading times in the anomalous word location are noticeably longer in Experiment 2 than in the same area in Experiment 1, and this is despite the fact that the figure shows only acceptable sentences that did not actually contain an anomalous word. This suggests that the presence of the anomalous words, and perhaps especially its predictable location, altered participants' reading strategies during the task. This alteration of strategy may explain the lack of the WCA_{REQ} interaction in this data.

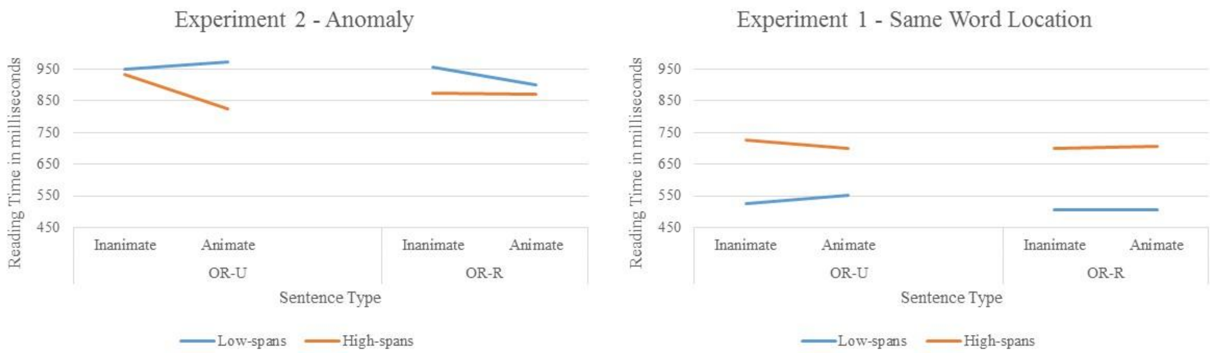


Figure 4.3.12. Reading times for anomalous words in Experiment 2 in the left panel, compared to reading times for the same word location in Experiment 1 in the right panel.

Critical Area Reading Time.

Waters and Caplan (2004) examined their critical area listening times in a different manner than Just and Carpenter (1992). Rather than looking for a three-way interaction of complexity, animacy, and WMC, Waters and Caplan (2004) looked for a three-way interaction of complexity, phrase, and WMC. One reason for this difference was that the authors did not manipulate animacy. A second reason for this difference was that the authors argued that if single resource theory is correct, then the area of the sentence where processing load is highest (i.e., the critical area) should place a disproportionate burden on the resources of low-span individuals, and thus you should see a significant increase in their reading times that is not seen for high-spans (or is seen, but to a lesser extent). Figure 4.3.13 below is a reconstruction of the data from Waters and Caplan (2004). Note that the contrast is between only two types of sentences: simple subject-relatives (which they formed in a different manner than those in the current study) and the more complex unreduced object-relatives. Also note that the listening times were calculated by subtracting the duration of the auditory phrase from the length of time between when it began to play and when the participant pressed the button to hear the next phrase. Because listening times were calculated in this manner, the values dip below zero on some phrases.

Figure 4.3.14 below shows the Set 3 reading time data presented in the same manner, with phrase on the X-axis. (These times were trimmed as described in Experiment 1, using individual means and standard deviations to replace less than 2% of outlying data points overall and less than 2% for any individual.) The left panel shows the comparison between SR and OR-U sentences, which is directly comparable to the Waters and Caplan (2004) data, while the panel on the right shows the comparison between the SR and OR-R sentences. Recall that Waters and Caplan (2004) did not present OR-R sentences in their study.

A 2 (complexity) x 5 (phrase) x 2 (WMC) mixed within-between repeated measures on ANOVA on the Set 3 data in the left panel of the figure below showed a main effect of complexity, F (1, 30) = 16.60, $p < .001$, $\eta_p^2 = .36$, and a main effect of phrase, F (4, 120) = 19.98, $p < .001$, $\eta_p^2 = .40$. These main effects were qualified by an interaction between the two, F (4, 120) = 3.69, $p < .01$, $\eta_p^2 = .11$. Examination of this interaction revealed that reading times in subject-relative sentences were faster than those in unreduced object-relative sentences for areas

V1, V2, and NP3, ts > -2.34, ps < .05. We also conducted independent samples t-tests for just the V2 area. These results confirmed that low- and high-span reading times did not differ for any sentence type.

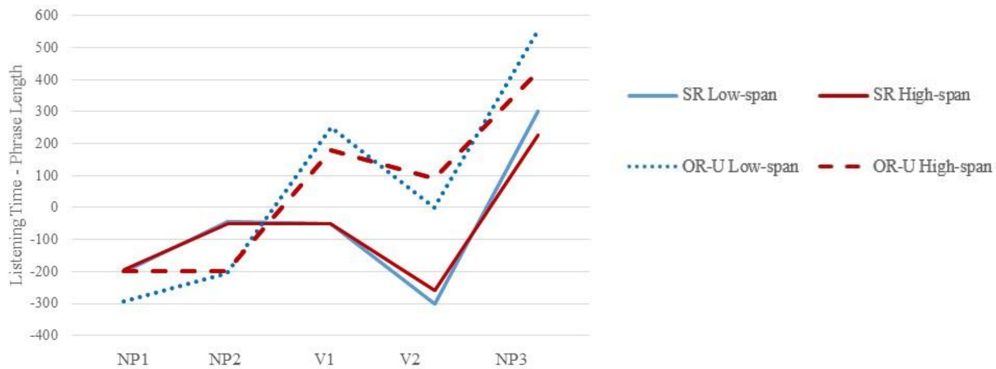


Figure 4.3.13. Recreation of Waters and Caplan (2004) phrase listening times.

A 2 (complexity) x 5 (phrase) x 2 (WMC) mixed within-between repeated measures on ANOVA on the data in the right panel of the figure below showed a main effect of complexity, F (1, 30) = 14.86, p < .01, η_p^2 = .33, and a main effect of phrase, F (4, 120) = 20.60, p < .001, η_p^2 = .41. These main effects were qualified by an interaction, F (4, 120) = 4.26, p < .01, η_p^2 = .12. Examination of this interaction showed that reading times in subject-relatives were faster than those in reduced object-relatives for areas V2 and NP3, ts = -3.13, ps < .01. We also conducted independent samples t-tests for just the V2 area. These results confirmed that low- and high-span reading times did not differ for any sentence type.

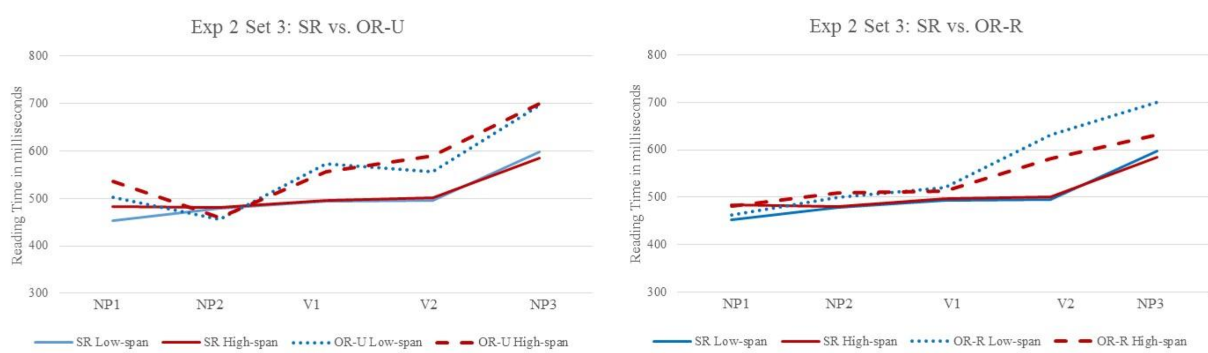


Figure 4.3.14. Experiment 2 Set 3 phrase reading time data.

For the sake of comparison, we also looked at critical area reading times in the Set 3 data in the same manner as Just and Carpenter (1992). In the Set 2 data above, this analysis could only be conducted on correctly judged acceptable items because of the disruption in processing caused by the presence of the anomalous word before the critical area. Set 3 does not (in theory) have this same problem because the anomalous word was always the last word of the sentence, and therefore did come after the critical area (i.e., defined in the literature as the second verb only). Despite this, the analyses reported below were run two times: once with only correct acceptable items and once with all correct items. The pattern of results did not differ. The data reported here are for correct acceptable items only. Figure 4.3.15 shows mean reading times in the critical area (V2) for Set 3 experimental sentences in the right panel and data for the same sentence types recreated from Just and Carpenter (1992) on the left.

A 2 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on just the object-relative sentences showed no significant main effects or interactions, despite the fact that there appears to be a trend for low-span participants to spend less time reading the critical areas of animate sentences than inanimate ones.

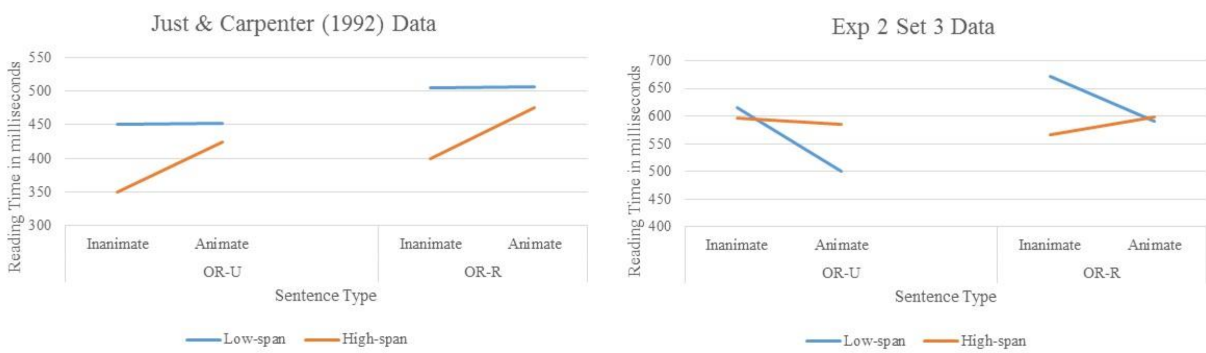


Figure 4.3.15. Recreation of the Just and Carpenter (1992) data in the left panel compared to Experiment 2 Set 3 data on the right.

Note that overall the critical area reading times seen in Figure 4.3.15 are much longer than those found in Just and Carpenter (1992), and even those for the same set in Experiment 1. Cleary all participants are spending longer in the critical phrase. One possible explanation is that participants may spend longer checking their interpretation of the sentence thus far before they encounter the potential anomaly. Recall that participants had a rough idea of the length of the sentence and knew when they were approaching it. It was also the case that the anomaly was always on the last word, right after the critical area, giving participants and expectation of when it could be encountered (and this was also reinforced by the location of the anomalies in the unacceptable filler items). If one knows the anomaly might be coming up, it makes sense to pause and be sure one has a handle on the sentence before proceeding. It would also make sense for the low-spans to spend more time doing this on inanimate items than on animate ones, if the argument that animacy works differently in Set 3 (from the previous experiment) holds. This is because the inanimate sentences would actually be the ambiguous ones in this case, and thus it would take longer to check their interpretations for low-spans who have not already used the pragmatic cue in their processing of the sentence. However, this would mean that the low- and not high-spans were the ones gaining an advantage from the animacy cue, which is the opposite of what was found in Experiment 1 and what would be predicted by single resource theory. This may not be problematic, however, if the pause here actually reflects a pause for offline reasoning about the sentence so far before encountering the possible anomaly. If this were the case, it would not be surprising to see no animacy-based differences for high-spans, yet still see them for low-spans.

Just and Carpenter (1992) Stimuli – Set 1 (Appendix B)

The Set 1 stimuli in Experiment 2 were identical to those found in Experiment 1, with the exception that items previously followed by a false comprehension item were modified to contain a semantically anomalous word. In this case, the anomalous word was found in the *by* phrase (The pedestrian followed by the *potato* amused the gathered spectators.; see Appendix B).

By now it should be clear from the discussion of the results of the other sets why this may have been problematic, but in actuality this feature potentially makes the stimulus sentences more comparable to those used in Waters and Caplan (1996b, 2004) because stimulus sentence in those experiments also contained anomalous words before the critical area, at least as far as can be determined from the example sentences provided. Because the overall purpose of Experiment 2 was to replicate Waters and Caplan (1996b, 2004) if possible, results are reported after their conventions where possible as well as in a manner comparable to Experiment 1.

Accuracy.

The right panel of Figure 4.3.16 below shows the mean A' scores in the Set 1 data by sentence type, while the left panel shows the reconstructed Waters and Caplan (1996b) data for their object-relative sentences. A 4 (complexity) x 2 (WMC) mixed within-between repeated measures ANOVA on the Set 2 A' scores showed no significant main effects and no interactions, in line with the findings of Waters and Caplan (1996b).

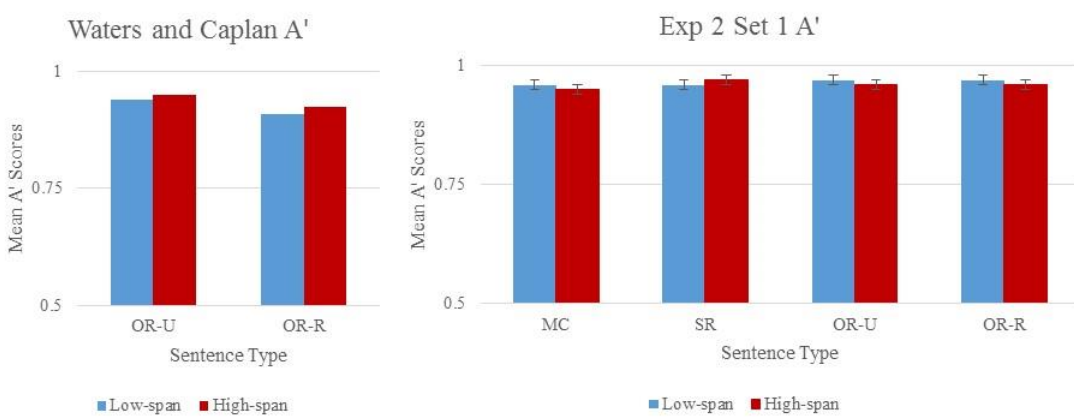


Figure 4.3.16. Recreated A' scores from Waters and Caplan (1996b) in the left panel compared to Set 1 A' scores in Experiment 2, with standard error bars.

Sense judgment accuracy was also analyzed in the manner of Just and Carpenter (1992). A 4 (complexity) x 2 (animacy) x 2 (answer) x 2 (WMC) mixed within-between repeated measures ANOVA on the accuracy scores revealed three two-way interactions. The interaction of complexity and animacy, F (3, 90) = 7.49, p < .001, η_p^2 = .20, was such that accuracy was greater for inanimate main clause items than for animate ones, t (31) = 2.05, p < .05, but no other contrasts were significant. There was also an interaction of complexity and answer, F (3, 90) = 6.66, p < .001, η_p^2 = .18, in which accuracy was higher for acceptable main clause items than for unacceptable, t (31) = 4.10, p < .001, but no other contrasts were significant. Finally, there was an interaction of animacy and answer, F (1, 30) = 14.70, p < .01, η_p^2 = .33, such that accuracy was higher on acceptable inanimate items, t (31) = 3.32, p < .001, but there was no difference for animate items. Figure 4.3.17 below shows accuracy scores (out of 10) collapsed across answer.

These results not only show no evidence of the WCA_{REQ} interaction, they do not even show an effect of WMC. The overall high level of performance, whether measured in A' or in sums of scores, suggests once again that these items were quite easy to judge. That being the case, there is no reason to expect there to be any effect of WMC, much less an interaction. It is still unclear why the Set 1 sentences were not difficult enough to elicit an interaction, when the same type of sentences did do so in Just and Carpenter (1992), but the data so far suggests that this is the case.

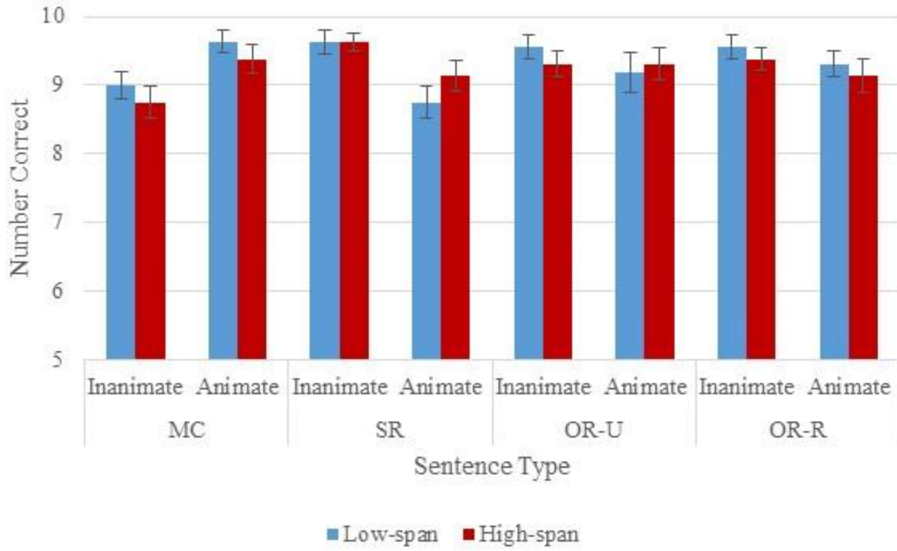


Figure 4.3.17. Experiment 2 Set 1 Sense judgment accuracy, with standard error bars.

Judgment Response Time.

A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on sense judgment response times showed a significant main effect of WMC, *F* (1, 30) = 8.84, $p < .01$, $\eta_p^2 = .23$, such that high-spans has significantly longer response times than low-spans. There was also a two-way interaction of complexity and animacy, *F* (3, 90) = 3.01, $p = .03$, $\eta_p^2 = .09$, due to longer reaction times on inanimate main clause items than on animate ones *t* (31) = 2.07, $p < .05$. Figure 4.3.18 below shows the means for all sentence types.

Clearly these results do not show the WCA_{REQ} interaction that one might use to support the single resource theory. However, it is also the case that single resource theory would not really predict this interaction in the judgment response times anyway. The critical areas of Set 1 items are embedded near the beginning of the sentences and there is no compelling reason to think any effect in that area would carry over into the judgment response times. What is apparent is an effect of WMC, such that high-spans have longer response times than low-spans. This may be a simple effect of the high-spans putting in more of an effort to reconsider sentences before making a final judgment, despite the overall ease of the sentences.

Critical Area Reading Time.

Because the anomalous word in the Set 1 sentences occurred in the *by* phrase, critical reading time analyses include only correctly judged acceptable items, and because the purpose of Experiment 2 was to replicate the findings of Waters and Caplan (1996b), we calculated mean per word reading times for Set 1 as described above in the Set 2 results. Critical area reading times were trimmed as in all other sets and experiments, with less than 2% of data replaced. Figure 4.3.19 below shows the Set 2 data next to the recreated data from Waters and Caplan (1996b).

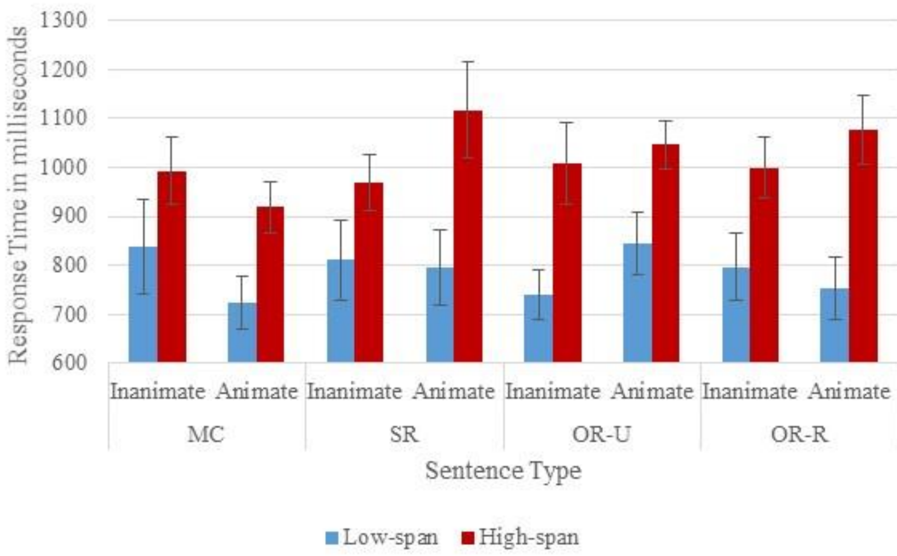
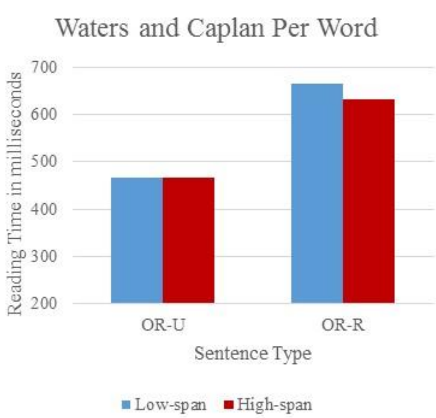


Figure 4.3.18. Experiment 2 Set 1 Sense judgment response times, with standard error bars.



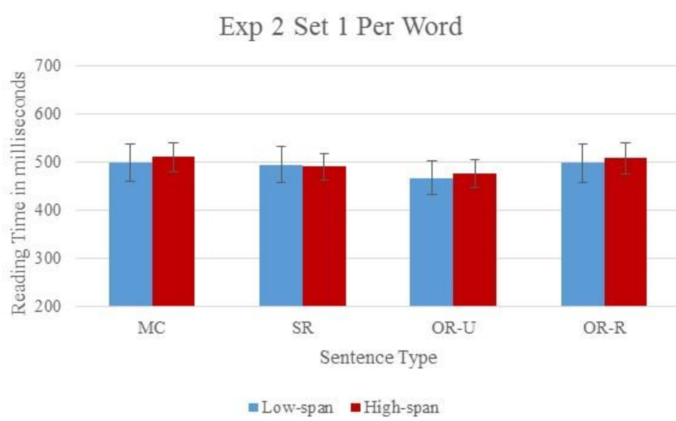


Figure 4.3.19. Recreated per word reading times from Waters and Caplan (1996b) in the left panel compared to Experiment 2 Set 1 data on the right, with standard error bars.

A 4 (complexity) x 2 (WMC) within-between repeated measures ANOVA on per word reading times revealed a significant main effect of complexity, F (3, 90) = 16.59, $p < .001$, $\eta_p^2 = .14$, but no effect of WMC and no interaction. Examination of this effect showed that mean per word times were longer for main clauses than for both subject-relatives and unreduced object-relatives, and both subject-relative and reduced object-relatives showed longer times than unreduced object-relatives, all ts > 2.31, ps < .05.

By phrase reading times were also analyzed according to the methods of Just and Carpenter (1992). Figure 4.3.20 below shows the critical area reading times for correctly judged acceptable items of all sentence types. These times were trimmed as described in Experiment 1, using individual means and standard deviations to replace less than 2% of outlying data points overall and less than 2% for any individual. A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on these reading times revealed a main effect of complexity, F (3, 90) = 3.56, $p = .02$, $\eta_p^2 = .11$. There were no other significant main effects and no significant interactions. Examination of the effect of complexity showed that critical area reading times were longer in main clause sentences than in subject-relatives and in unreduced object-relatives, *t*s > 2.49, *p*s < .05. No other contrasts were significant.

Once again for comparison purposes, the critical area reading times for experimental sentences only are pulled out of the data and displayed in Figure 4.3.21 below. The left panel of 4.3.21 shows the comparable recreated data from Just and Carpenter (1992). As can be clearly seen, the interaction found in the Just and Carpenter (1992) data is not in evidence in the Set 1 data. Both low- and high-spans showed no difference between inanimate and animate object-relative items of either type.

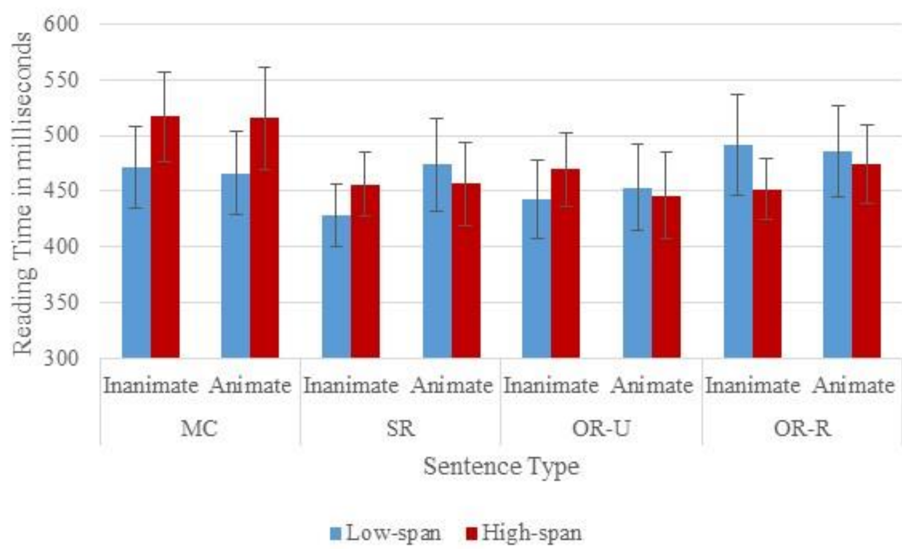


Figure 4.3.20. Experiment 2 Set 1 critical area reading times, with standard error bars.

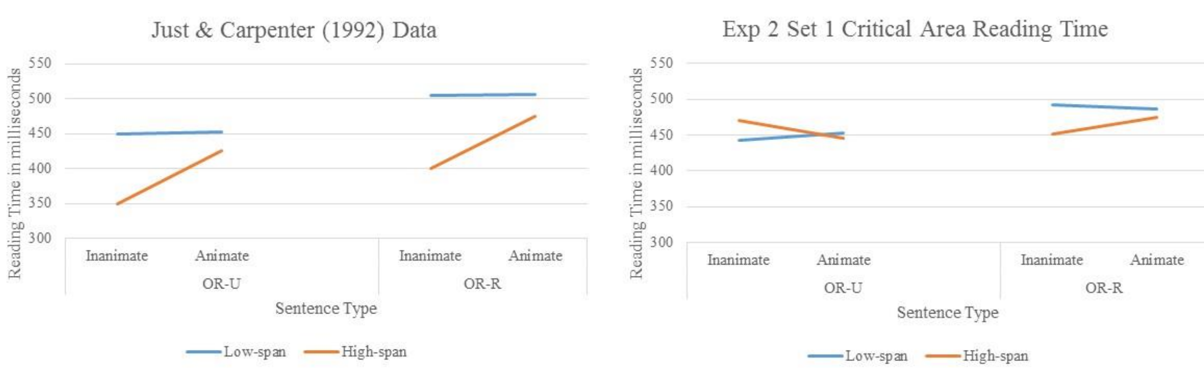


Figure 4.3.21. Recreated critical area reading times from Just and Carpenter (1992) in the left panel compared to Experiment 2 data for Set 1 on the right.

The critical area reading time results for Set 1 are in line with the predictions of the dual resource theory in that they certainly provide no evidence of the WCA_{REQ} interaction called for by Waters and Caplan (1996a), nor even the two-way interaction of WMC and animacy that was observed by Just and Carpenter (1992). However, this may not be that surprising, given the variety of factors that have already been argued to disrupt the interaction. First is the fact that the sentences in this set are the easiest complex sentences in the current experiment, as confirmed by both subjective participant reports and objective assessment of accuracy rates. Second is the fact that participants quite likely rushed through reading these rather long sentences as they did in Experiment 1. Next is the fact that the anomalous word was found in the critical area, and participant knew this. This may have changed the task significantly for the participants. Comparing critical area times from Experiment 1 (that are at the lowest level seen in Just and Carpenter, 1992) to those from Experiment 2 (that are at the highest level seen in Just and Carpenter, 1992), it appears as though overall all participants were encouraged by the setup of the task to pause longer at the *by* phrase to check for the semantic anomaly, but this strategy was independent of both animacy and WMC.

4.4 Discussion

The purpose of Experiment 2 was to determine the effect of altering the type of judgment used in the syntactic processing task on the various performance variables. To accomplish this, we administered the same stimulus sets as in Experiment 1, but with slight alterations. These alterations involved inserting anomalous words into half of the sentences and were required to turn the previously false items into the unacceptable items needed for a sense judgment modeled after that used in Waters and Caplan (1996b; 2004).

The overall results of Experiment 2 are in line with the findings of Waters and Caplan (1996b; 2004). We did not find significant effects of WMC anywhere they would be predicted by single resource theory, and certainly no WCA_{REQ} interaction. This is in direct contrast to Experiment 1, where we found some support for single resource theory, particularly in the critical area reading times of Set 2 (the WMC by complexity interaction) and the judgment response times of Set 3 (the three-way interaction of WMC, complexity, and animacy). We argued that one reason the use of sense judgments eradicated the effects of WMC seen in Experiment 1 might be that anomalous words were in predictable locations before the critical area and thus induced the use of a different reading strategy than that used in Experiment 1. Examination of reading times for anomalous words in Experiment 2 compared to reading times for words in the same locations in Experiment 1 supported this idea, by showing that reading times were longer overall even in sentences were an anomalous word was not actually present (i.e., acceptable items).

Waters and Caplan (1996a) argued that the acceptability/plausibility judgment (which is the same as what we have called the sense judgment here) should be preferred over the true/false comprehension judgment because it represents a more online measure of syntactic processing than the latter type of judgment. The data from Experiments 1 and 2 support the idea that these two judgments are getting at different types of processing. However, the data do not necessarily comment on which judgment type is to be preferred. What is clear is that task demands induced by judgment type impact online reading strategies and that this is not a factor that can be dismissed when attempting to test the single and dual resource theories of syntactic processing.

CHAPTER 5. EXPERIMENT 3

5.1 Introduction

As the results of Experiment 1 and Experiment 2 have demonstrated, the type of judgment being made about sentences can have important consequences for the dependent measures of interest. Comprehension questions are the most obvious choice of judgment to use in a syntactic processing task, but have been criticized on the grounds that they may involve post-interpretive processing and thus are not capable of capturing variation in online syntactic processing (Waters & Caplan, 1996a). The most obvious remedy for this criticism is to not use accuracy as the primary dependent variable at all, but rather to use the more online measure of critical area reading times. Another remedy proposed by Waters and Caplan (1996a) was to use sense judgments because the detection of semantic anomalies is arguably more reflective of online processing. However, as the previous experiments have demonstrated, sense judgments can be problematic in their own right. One reason this may be so is that there is some evidence to suggest that detection of semantic and syntactic errors occurs independently (Koch et al., 2003). Furthermore, semantic anomalies that appear before the critical area of complex sentences may terminate processing of the complex syntactic structure altogether. Finally, the mere presence of anomalies appears to change the manner in which participants read sentences, as we saw when comparing anomalous word reading time to reading time in the same area of sentences without anomalies. If it is the case that sense judgments change the very nature of the syntactic processing task, as these arguments suggest, then sense judgments would arguably be essentially useless for testing between single and dual resource hypotheses.

The purpose of Experiment 3 was to examine an alternate judgment type, the grammaticality judgment, in the syntactic processing task in order to compare it to the previous types of judgment. The grammaticality judgment addresses the complaint lodged against comprehension questions by Waters and Caplan (1996b) because, like sense judgment, it should not involve post-interpretive processing. Furthermore, detecting grammatical errors is arguably an even more direct measure of syntactic processing than detection of semantic anomalies. Thus, it stands to reason that the grammaticality judgment may represent an acceptable compromise in the disagreement over the most appropriate judgment type. One problem that will remain, however, is that the grammatical error for Set 2 will always come before the critical area. This is unfortunate, but also unavoidable when the critical area is the last word of the sentence as it is in Set 2 sentences.

Because Experiment 3 is not a direct replication of any study, results will be presented in numerical set order (i.e., Set 1, Set 2, Set 3) and in the manner used in the previous experiments wherever possible for direct comparisons.

5.2 Methods

Participants

Participants were 32 (22 female) undergraduate psychology students at Louisiana State University who received course credit for their participation. All participants reported normal or corrected to normal vision and hearing, and no history of language or speech impairments. Mean age was 19.34 (*SD* = 1.49). Sixteen participants were classified as low-span and 16 as high-span on the basis of composite z-scores on the automated span tasks (descriptives for the WMC tasks are in Table 5.2.1 below). An additional 18 participants were run in Experiment 3: 10 were dropped because they were medium-span, five were dropped for not meeting the 85% performance criterion for at least one automated task, one was dropped for missing too many items on the syntactic processing task, one was dropped for reporting a history of hearing impairment, and one was dropped for exhibiting odd behavior during the experiment such as insisting on reading all sentences aloud and providing commentary on almost all of them.

Syntactic Processing Task

The three stimulus sets for Experiment 3 were identical to those in Experiment 1, with the exception that items that were followed by a false comprehension item in Experiment 1 were modified in Experiment 3 to contain a small grammatical error (e.g., subject-verb agreement errors; See Appendices B through D). For the Just and Carpenter (1992) sentences, the grammatical error was found after the *by* phrase (e.g., The pedestrian that was followed by the mime *were* amus*ing* the gathered spectators.), while for both the Waters and Caplan (1996b) and Waters and Caplan (2004) sets it was found in the critical area (e.g., “The hair that was rinsed in hot water *dry*.” and “The document that the machine copied *shock* the public.” respectively) because of the limited location options provided by the construction of the stimuli.

Working Memory Capacity Measures

The same four WMC measures from the previous two experiments were administered in Experiment 3, with the composite z-score from the automated tasks used to classify participants into low- and high-span groups. All results reported here use the composite as the grouping variable because results were no different when split by the recoded Daneman and Carpenter (1980) task. Table 5.2.1 below shows the means and standard deviations for each of the WMC measures in the current experiment.

Table 5.2.1. Experiment 3 WMC measure descriptives.

		Low-span		High-span	
		Mean	*SD*	Mean	*SD*
Automated Tasks	R-span	-0.82	0.85	0.97	0.34
	S-span	-0.89	0.93	0.71	0.55
	O-span	-1.23	0.72	1.05	0.25
	Composite	-1.22	0.61	1.13	0.27
D & C (1980)	Original	2.50	0.63	2.81	0.40
	Recode	0.75	0.07	0.78	0.08

5.3 Results

Just and Carpenter (1992) Stimuli – Set 1 (Appendix B)

Accuracy.

A 4 (complexity) x 2 (animacy) x 2 (answer) x 2 (WMC) mixed within-between repeated measures ANOVA on grammaticality judgment accuracy showed main effects of complexity, F (3, 90) = 14.45, $p < .001$, $\eta_p^2 = .33$, and answer, F (1, 30) = 8.43, $p < .01$, $\eta_p^2 = .22$ (the effect of WMC was marginal, F (1, 30) = 3.80, $p = .06$, $\eta_p^2 = .11$). These main effects were qualified by two two-way interactions of complexity and animacy, F (3, 90) = 4.63, $p < .001$, $\eta_p^2 = .13$, and of complexity and answer, F (3, 90) = 11.60, $p < .01$, $\eta_p^2 = .28$. These two-way interactions were in turn qualified by a three-way interaction of complexity, animacy, and answer, F (3, 90) = 12.32, $p < .001$, $\eta_p^2 = .29$. Examination of this interaction revealed that grammatical inanimate main clause items showed lower accuracy than grammatical animate main clause items, t (31) = -4.28,

$p < .001$, while grammatical inanimate subject-relatives and reduced object-relatives had higher accuracy than grammatical animate subject-relatives and object-relatives respectively, *t*s > 3.70, *p*s < .01. For agrammatical items, there were no significant differences in animacy for any sentence type. Figure 5.3.1 below shows mean grammaticality accuracy (out of 10) for each sentence type, collapsed across answer.

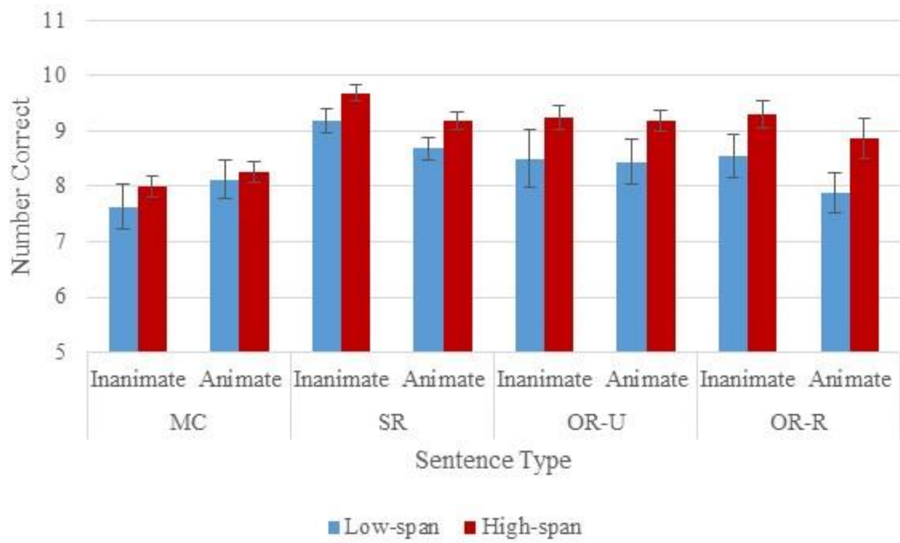


Figure 5.3.1. Experiment 3 Set 1 Grammaticality judgment accuracy, with standard error bars.

Although the effect of WMC is not significant, we can see in Figure 5.3.1 high-spans tend to show higher accuracy than low-spans, and that this differences appears somewhat larger in the more complex sentences than it is in the simple ones. One odd finding is that accuracy is lowest for main clause sentences for both low- and high-span participants. As has been argued previously, this oddity may be an artifact of the method of construction and lower frequency of main clause sentences in the current experiment that result in some confusion for these items. Regardless, what is clear from Figure 5.3.1 is that there is no WCA_{REQ} interaction in accuracy scores.

For comparison with Waters and Caplan (1996b), we also calculated A' scores for grammaticality judgments. A 4 (complexity) x 2 (WMC) mixed within-between ANOVA showed only a main effect for complexity, F (3, 90) = 7.94, $p < .001$, $\eta_p^2 = .21$, such that A' scores were lower for main clause sentences than for subject-relatives and object-relatives (both reduced and unreduced), *t*s > -2.03, *p*s ≤ .05, and higher for subject-relatives than for reduced object-relatives, t (31) = 2.19, $p = .04$. Figure 5.3.2 below shows mean A' scores by sentence type in Set 1 next to A' values recreated from Waters and Caplan (1996b).

In order to explore the effects of animacy, we also conducted a 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on A'. This analysis revealed a main effect of complexity, F (3, 90) = 7.94, $p < .001$, $\eta_p^2 = .21$, and an interaction of complexity and animacy, F (3, 90) = 4.14, $p < .01$, $\eta_p^2 = .12$. Examination of this interaction revealed that A' scores were higher for inanimate subject-relatives than for animate subject-relatives and for inanimate reduced object relatives than for animate reduced object-relatives, *t*s > 2.49, *p*s <.02.

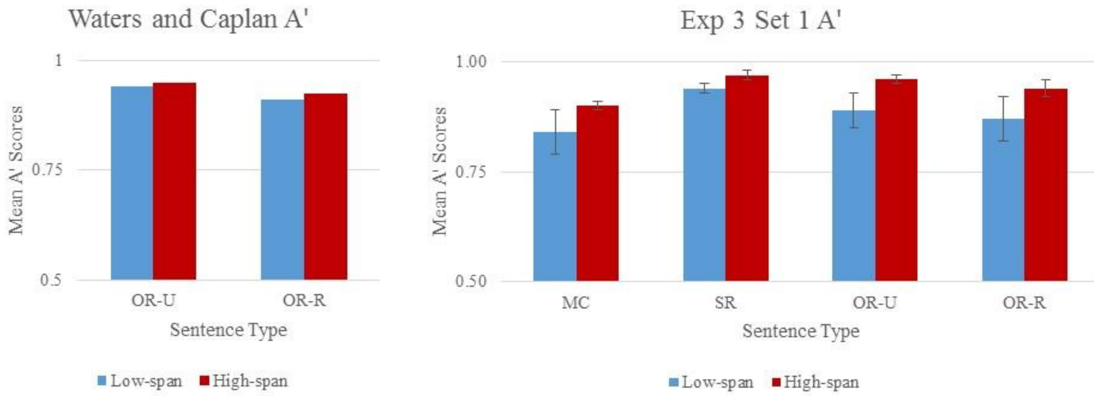


Figure 5.3.2. Recreated A' scores from Waters and Caplan (1996b) in the left panel compared to Set 1 A' scores in Experiment 3 on the right, with standard error bars.

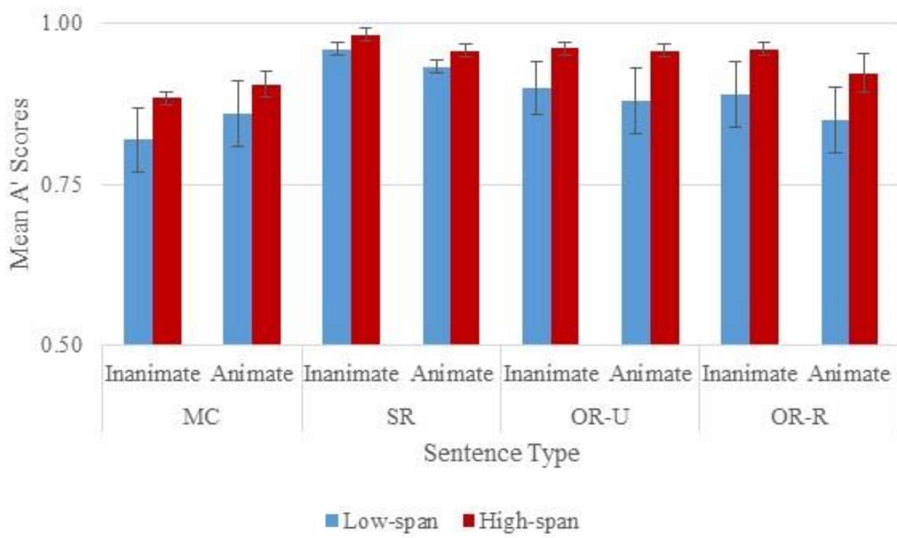


Figure 5.3.3. Experiment 3 Set 1 A' scores, with standard error bars.

Altogether, these analyses on judgment response accuracy suggest that main clause sentences may have confused participants due to their initial similarity to reduced animate clauses (discussed in Experiment 1 above). Aside from this, performance is generally better for simpler sentences, and high-spans tend to have better performance than low-spans. These findings are compatible with both the single and dual resource views of syntactic processing. As in both previous experiments, the current data show no WCA_{REQ} interaction, but this is not particularly supportive of either theory, since the WCA_{REQ} wouldn't necessarily be expected in accuracy performance even by single resource theory.

Judgment Response Time.

A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on grammaticality judgment response times showed no main effects and no interactions. Mean response times are displayed in Figure 5.3.4 below.

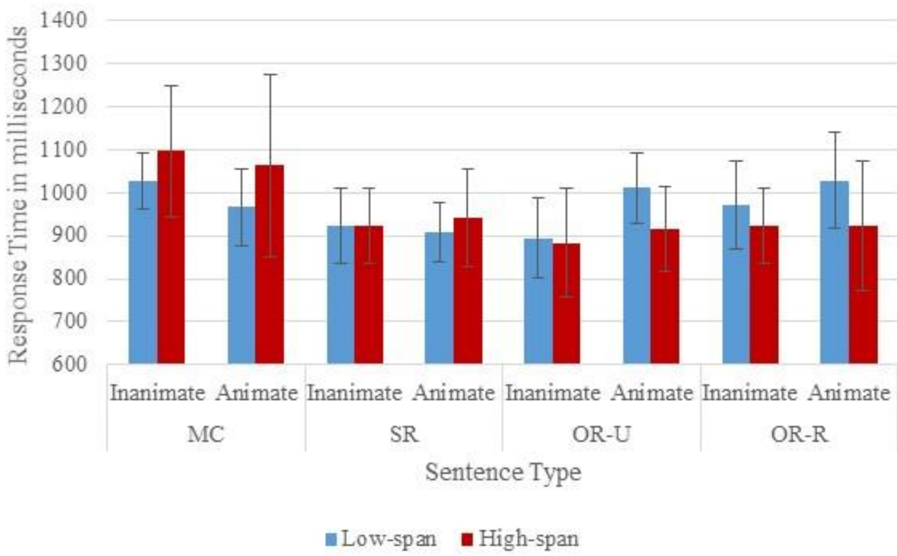


Figure 5.3.4. Experiment 3 Set 1 Grammaticality judgment response times, with standard error bars.

As in the previous experiments, judgment response times do not show any significant effects of WMC and certainly no interaction with complexity and animacy. Again, this is not necessarily surprising given that the critical area of Set 1 sentences is embedded in the items and there is no reason to expect any difference that might manifest itself there to carry over into the response times. Furthermore, single resource theory would not predict the interaction at the time of response. There is a slight trend in the experimental sentences for low-spans to show longer judgment times for more complex sentences, but there is also a large amount of variance in the data. Even if this difference were significant, it would not distinguish between theories.

Critical Area Reading Time.

In Experiment 1, we combined Set 1 true and false items for critical area reading time analysis because the question came after the critical area. In Experiment 2 the semantic anomaly was found in the critical area of Set 1 sentences, so critical area reading time analysis was done on acceptable items only. In Experiment 3, the grammatical anomaly was found immediately after the critical area in Set 1, so in theory we could include agrammatical items in our analysis. However, this is not possible for Set 2 and Set 3 data due to the location of their grammatical errors, so analyses are reported for grammatical Set 1 items only for the sake of consistency within the experiment. Finally, these times were trimmed as described in Experiment 1, using individual means and standard deviations to replace less than 2% of outlying data points overall and less than 2% for any individual.

A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on critical area reading times showed a main effect of complexity, F (3, 90) = 8.15, p < .001, η_p^2 = .21, as well as a two-way interaction of complexity and animacy, F (3, 90) = 4.62, p = .03, η_p^2 = .13. Examination of this interaction revealed that inanimate main clause items had longer reading times in the critical area than animate main clause items, t (31) = 2.23, p = .03, but no other contrast was significant.

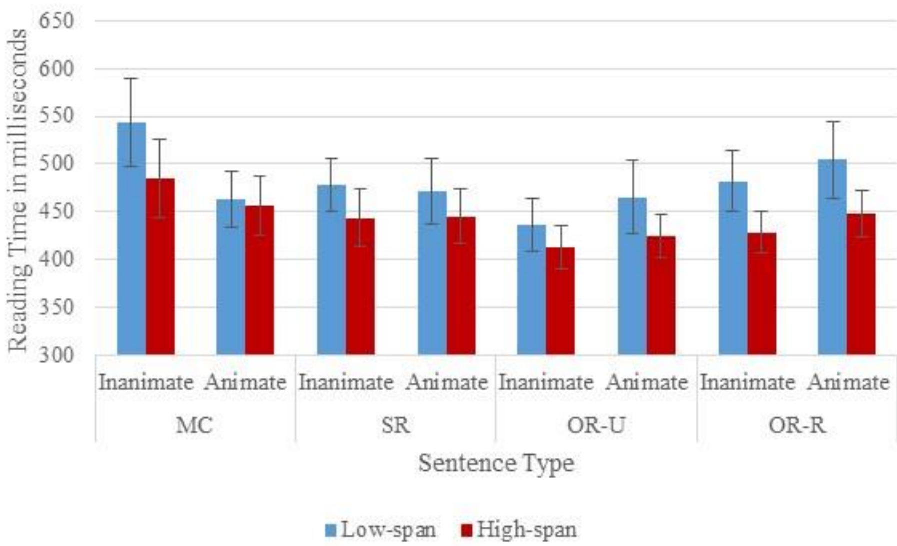


Figure 5.3.5. Experiment 3 Set 1 critical area reading times, with standard error bars.

Figure 5.3.4 below shows mean critical area reading times for just the experimental sentences of Set 1 in the right panel, compared to the recreated critical area reading time data from Just and Carpenter (1992) on the left.

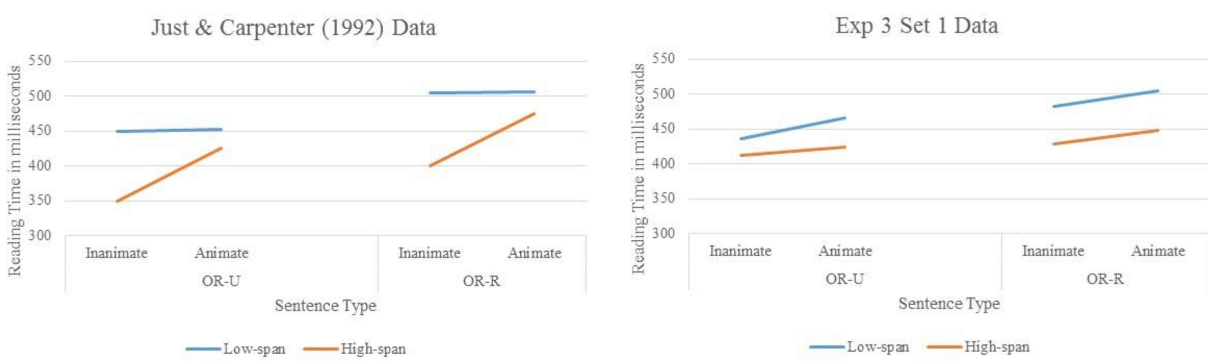


Figure 5.3.6. Recreated critical area reading times from Just and Carpenter (1992) in the left panel compared to Experiment 3 data for Set 1 on the right.

As in Experiment 2, the critical area reading times in Experiment 3 are flat near the uppermost values found in Just and Carpenter (1992) for all participants, whereas the times from Experiment 1 were near the lowest values. As with the sense judgment, the grammatical error in the current experiment was predictably located, though in this case it was right after rather than in the critical area. The overall longer times in the critical area suggest a strategy of slowing down in preparation of detecting the potential error. Low-spans appear to have paused somewhat longer than high-spans, but it appears to be a strategy applied equally in inanimate and animate sentences and at least partially explains the lack of an interaction. Another explanation is the same that has been argued previously: That Set 1 sentences were too easy to elicit the interaction and their length encouraged a rushing button-pressing strategy.

In addition to these analyses, we also calculated mean per word reading time as in Experiment 2 for the sake of comparison. A 4 (complexity) x 2 (WMC) mixed within-between repeated measures ANOVA showed a main effect of complexity, F (3, 90) = 6.54, p < .001, η_p^2 = .15. Per word reading times were higher in main clause sentences than in both subject-relatives and unreduced object-relatives, *t*s > 2.50, *p*s ≤ .02, higher in subject-relatives than in unreduced object-relatives, *t* (31) = 2.52, *p* = .02, and higher in reduced object-relatives than unreduced object-relatives, *t* (31) = -2.31, *p* = .03. Figure 5.3.7 below shows mean per word reading times for Set 1 sentence types in the right panel and the recreated mean per word reading times from Waters and Caplan (1996b) for comparison in the left panel.

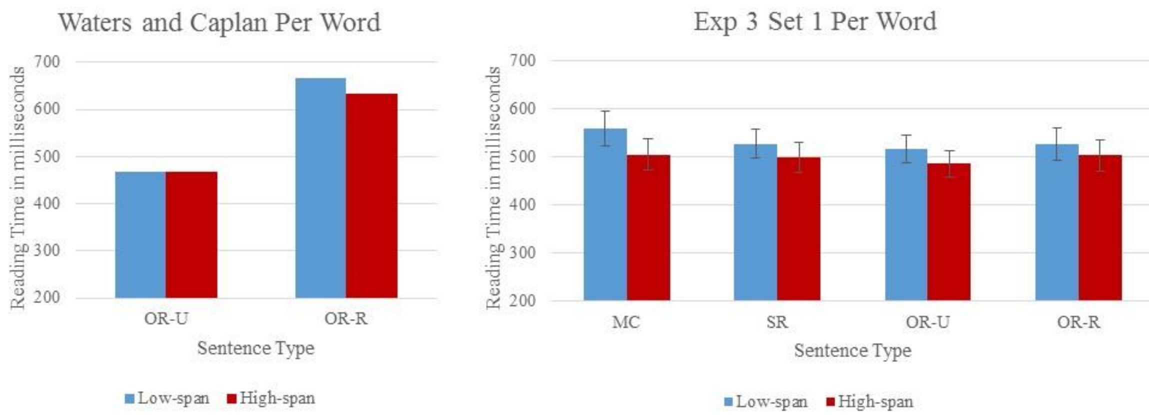


Figure 5.3.7. Recreated per word times from Waters and Caplan (1996b) in the left panel compared to Experiment 3 Set 1 data on the right, with standard error bars.

Because animacy was manipulated in the current experiment, we also analyzed per word reading times in a 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA. This analysis showed a main effect of complexity, *F* (3, 90) = 11.63, *p* < .001, η_p^2 = .28, that was qualified by a two-way interaction of complexity and animacy, *F* (3, 90) = 5.09, *p* < .01, η_p^2 = .15, which in turn was qualified by a three-way interaction of complexity, animacy, and WMC, *F* (3, 90) = 3.17, *p* = .03, η_p^2 = .10. Examination of the three-way interaction revealed that mean per word reading time was significantly lower in inanimate reduced object-relatives than in animate reduced object-relatives for low-spans, *t* (31) = -3.29, *p* < .01, while for high-spans mean per word reading time was lower in inanimate unreduced object-relatives than in animate unreduced object-relatives, *t* (31) = -2.56, *p* = .02. Note that this three-way interaction is not quite the WCA_{REQ} because animacy had an effect for low-spans on one type of sentence and for high-spans on another, whereas we would expect only the high-spans to show the animacy effect in the ideal version of the WCA_{REQ}. Figure 5.3.8 below displays the mean per word reading times for each sentence type.

The per word reading times do not change the picture presented by the critical area reading times because neither measure resulted in the exact WCA_{REQ} interaction. However, their inclusion is useful for the sake of comparing the methodology of computing mean times as opposed to actually recording critical area times. In the current experiment, we can see that the method of measuring and assessing reading times can change the statistical picture, from a simple two-way interaction with one measure to a more complex three-way with the other. Generally speaking, we would argue that the estimated per word times should be the least preferred, as they do not allow for measurement of differences related to sentence area.

Another issue highlighted by our analysis of per word reading times is that something is missed when animacy is not taken into account in the syntactic processing task. Although we did not have full stimulus lists from Just and Carpenter (1992) or Waters and Caplan (1996b; 2004), the fact that the latter two studies do not mention it either as a controlled or manipulated variable suggests that it was not accounted for. This is problematic when combining inanimate and animate items may obscure the effects of WMC, though it makes sense if one ascribes to the idea that pragmatic cues cannot be used in online processing (e.g., Fodor, 1988) and thus should not matter in the syntactic processing task.

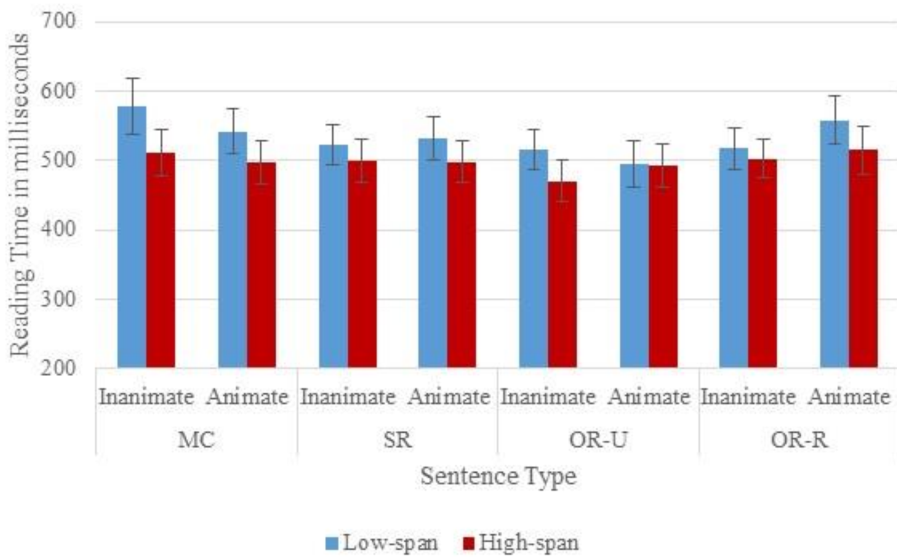


Figure 5.3.8. Experiment 3 Set 1 per word reading times, with standard error bars.

Waters and Caplan (1996b) Stimuli – Set 2 (Appendix C)

The Set 2 stimuli in Experiment 3 are identical to those in Experiment 1, with the exception that items that were followed by a false comprehension item in Experiment 1 were modified in Experiment 3 to contain a small grammatical error.

Unlike Set 1, it was not possible to place the grammatical error after the critical area in Set 2 sentences because the critical area coincides with the end in this set. Furthermore, because we wanted participants to at least process the ambiguity, it did not make sense to have the error near the beginning of the sentence. Thus, the grammatical error in Set 2 was placed on the last word/critical area/verb (e.g., The hair that was rinsed in hot water *dry*.).

Accuracy.

A 4 (complexity) x 2 (animacy) x 2 (answer) x 2 (WMC) mixed within-between repeated measures ANOVA on accuracy revealed main effects of complexity, F (3, 90) = 35.20, p < .001, η_p^2 = .54, animacy, F (1, 30) = 36.73, p < .001, η_p^2 = .55, and answer, F (1, 30) = 10.58, p < .01, η_p^2 = .26. These main effects were qualified by two-interactions of complexity and animacy, F (3, 90) = 16.35, p < .001, η_p^2 = .35, complexity and answer, F (1, 30) = 63.55, p < .001, η_p^2 = .68¸and animacy and answer, F (1, 30) = 7.13, p = .01, η_p^2 = .19. These two-way interactions were in turn qualified by a three-way interaction of complexity, animacy, and answer, F (3, 90) = 28.40, p < .001, η_p^2 = .49. Examination of this three-way interaction revealed that accuracy was higher for grammatical inanimate reduced object-relative items than for grammatical animate reduced object-relative items, t (31) = 11.36, $p < .001$, and for agrammatical inanimate main clause items than for agrammatical animate main clause items, t (31) = 3.05, $p < .01$. Figure 5.3.9 below shows mean number correct by sentence type.

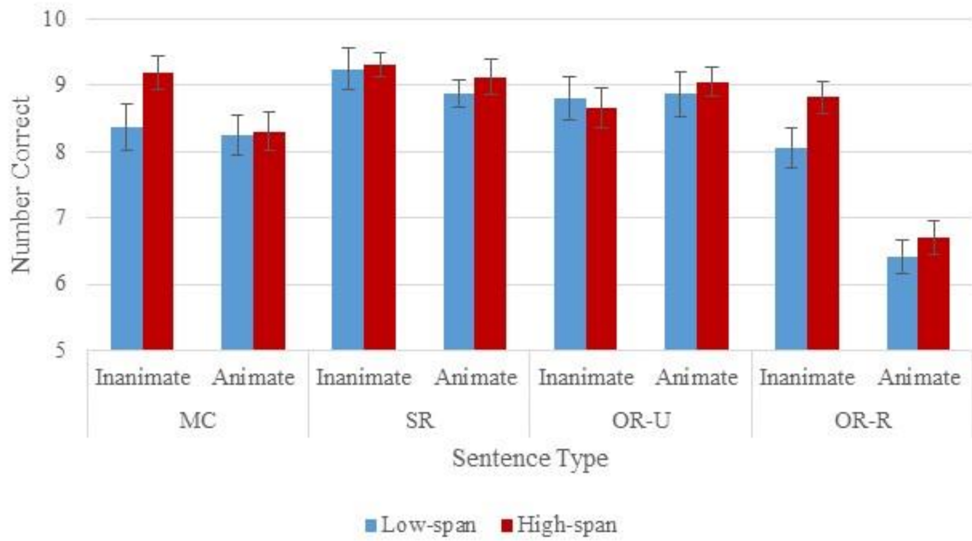


Figure 5.3.9. Experiment 3 Set 2 grammaticality accuracy, with standard error bars.

As in Experiments 1 and 2, there were participants in Experiment 3 who did not have any correctly judged agrammatical items (nine of the 32), and there were two who did not have any correctly judged grammatical animate reduced object-relatives. Only these latter two missing data points affect the critical area reading time analyses below, as only correctly judged grammatical items were included.

As before, we also calculated A' scores for comparison with Waters and Caplan (1996b). A 4 (complexity) x 2 (WMC) mixed within-between repeated measures ANOVA showed a main effect of complexity, F (3, 90) = 17.72, $p < .001$, $\eta_p^2 = .30$, such that A' scores were lower for main clause items than for subject-relative items, t (31) = -4.26, $p < .001$, and lower for reduced object-relatives than for main clause sentences, subject-relatives, and unreduced object-relatives, ts > 3.27, ps < .01. The left panel of Figure 5.3.10 below shows the recreated Waters and Caplan (1996b) A' data, while the right panel shows mean A' scores in the Set 2 data.

Because we manipulated animacy in the current experiment, we also analyzed A' scores in a 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA. This analysis revealed main effects of complexity, F (3, 84) = 15.95, $p < .001$, $\eta_p^2 = .37$, and animacy, F (1, 28) = 12.78, $p < .01$, $\eta_p^2 = .32$, and an interaction between the two, F (3, 84) = 10.12, $p < .001$, $\eta_p^2 = .24$. Examination of this interaction revealed that A' scores were higher for inanimate reduced object-relatives than for animate reduced object-relatives, t (30) = 5.67, $p < .001$, but no other contrasts were significant. Figure 5.3.11 below shows these means.

As already discussed, the lack of the critical three-way interaction in accuracy scores does not differentiate between the predictions of single and dual resource theory. Effects of complexity were in evidence in the Set 2 data, but these would be expected by both theories.

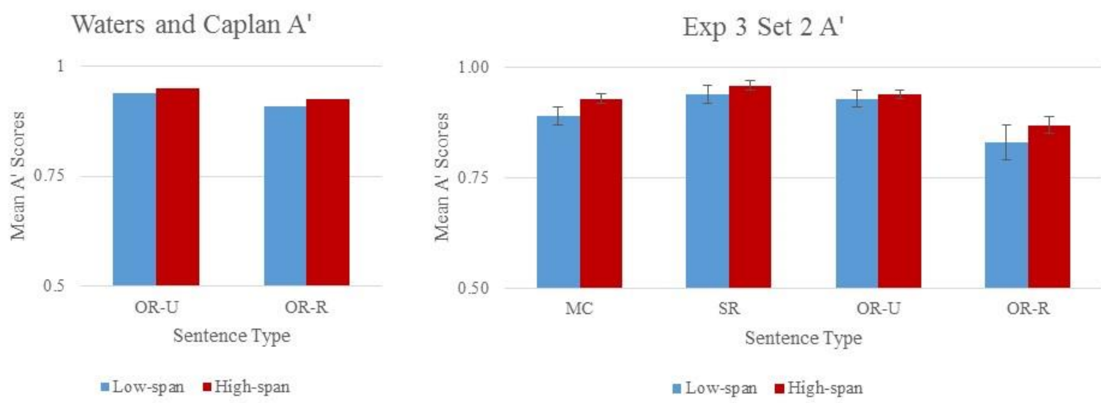


Figure 5.3.10. Recreated A' scores for Waters and Caplan (1996b) in the left panel for comparison with Experiment 3 Set 2 data on the right, with standard error bars.

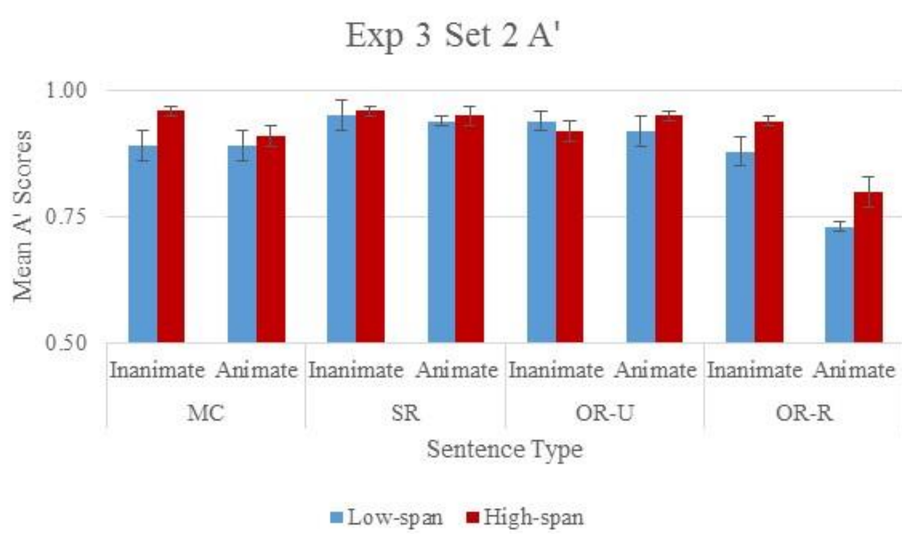


Figure 5.3.11. Experiment 3 Set 2 A' scores, with standard error bars.

Judgment Response Time.

A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on grammaticality response time showed only a main effects of complexity, F (3, 90) = 5.05, $p < .01$, $\eta_p^2 = .14$. Judgment response times were longer for main clause and reduced object-relative items than for unreduced object-relatives, *t*s > 2.41, *p*s < .05.

As already discussed, the single resource theory does not predict a three-way interaction in judgment times and one was not found here. There was a trend for low-spans to have longer reaction times for reduced object-relative items (both inanimate and animate) than high-spans, which would be in line with both single and dual resources because these are the most difficult sentences, but otherwise WMC does not seem to have much effect on the judgment response times in the current set. There is a bit of a trend of judgment times being faster for inanimate than for animate items in reduced object-relatives. The fact that this trend is seen for both low- and high-spans and is found here rather than in the critical area suggests it is the result of some post-interpretive processing of the pragmatic cue.

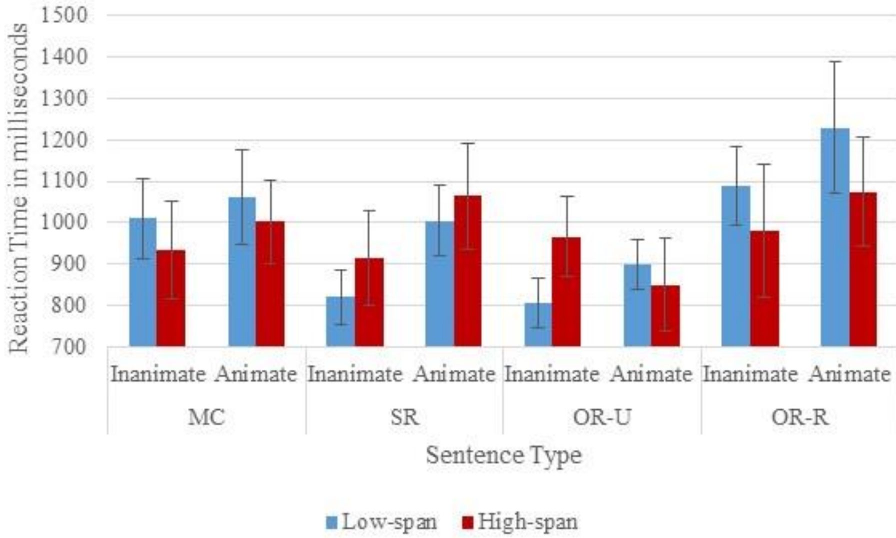


Figure 5.3.12. Experiment 3 Set 2 Grammaticality judgment response times, with standard error bars.

Critical Area Reading Time.

A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on correctly judged grammatical items revealed a main effect of complexity, F (3, 90) = 9.93, $p < .001$, $\eta_p^2 = .26$, such that critical area reading times were faster in main clause, subject-relative, and unreduced object-relative sentences than in reduced object-relative sentences, ts > -3.66, ps < .01. Figure 5.3.13 below shows mean critical area reading time for each sentence type. These times were trimmed as described above, with less than 2% of data replaced.

These results do not show any interaction of complexity and animacy, nor is there an interaction with WMC. However, it can be seen that there is a non-significant tendency that critical area reading times are longer for high-spans than for low-spans on the most difficult sentences (i.e., the reduced object-relatives), a trend in line with that of Experiment 1 and MacDonald et al.'s (1992) results with garden path sentences. However, the fact that the trend is for reading times to be higher on inanimate than animate items does not fit with these results and suggests that something different is at play in the current experiment. This may be related to the fact that grammatical anomaly is found before the critical area in Set 2 sentences.

For the sake of comparison with Just and Carpenter (1992), the left panel of Figure 5.3.14 below shows the recreated Just and Carpenter (1992) data for critical area reading times in the experimental sentences while the right shows the Set 2 data. The data on the right suggests that both low- and high-spans paused longer at the critical area of inanimate reduced object-relatives than at the critical area of animate reduced object-relative clauses. Although this trend is not significant, it is the opposite of the findings in the previous experiments and of the Just and Carpenter (1992) data. Overall (and in hindsight) the fact that there is no significant interaction of complexity, animacy, and WMC is not surprising mainly because the grammatical error occurs before the critical area. If error detection can occur before the complex structure is processed, it makes sense that the participant would proceed directly to the response screen rather than continue to struggle with what were subjectively and objectively the hardest sentences in the experiment.

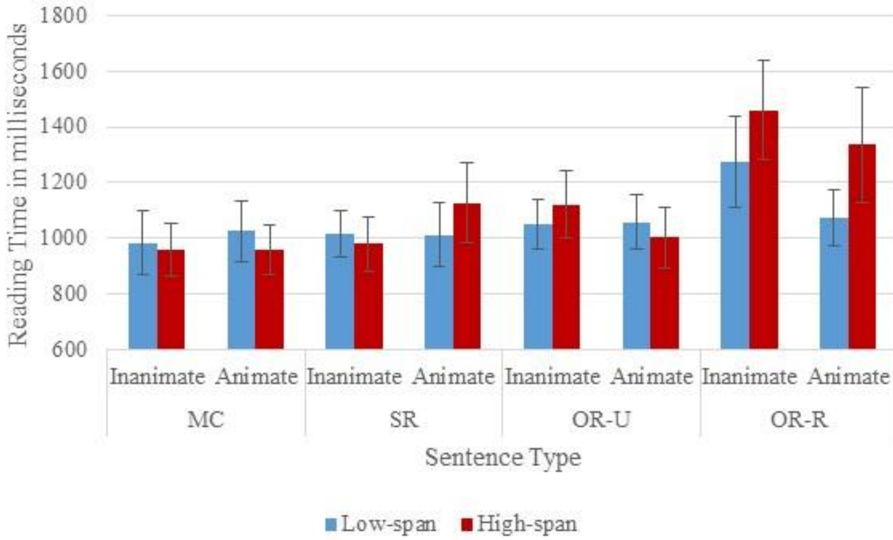


Figure 5.3.13. Experiment 3 Set 2 critical area reading times, with standard error bars.

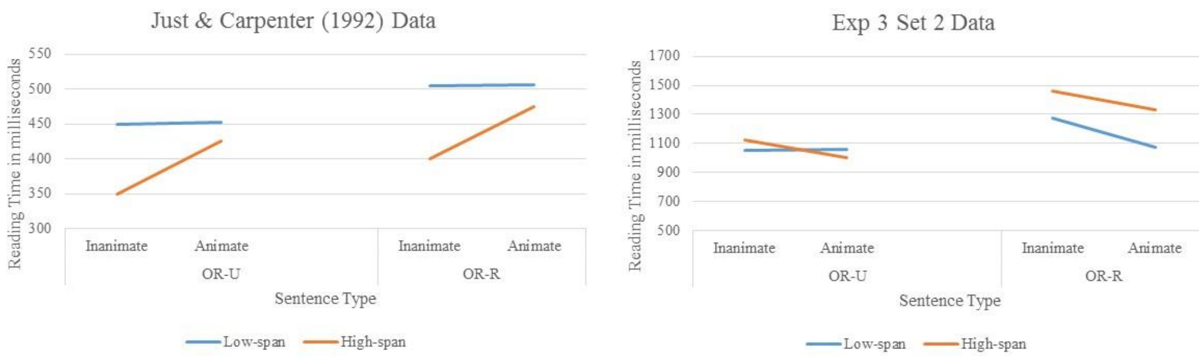


Figure 5.3.14. Recreated critical area reading times from Just and Carpenter (1992) in the left panel compared to Experiment 3 data for Set 2 on the rigth.

For comparison with Waters and Caplan (1996b), we also calculated mean per word reading time on Set 2 data. A 4 (complexity) x 2 (WMC) mixed within-between repeated measures ANOVA showed a main effect of complexity, F (3, 90) = 25.81, p < .001, η_p^2 = .46. Per word reading times were longer for main clause items than for subject-relatives and unreduced object-relatives, *t*s > 3.88, *p*s < .01, and longer for subject-relatives than for unreduced object-relatives, t (28) = 3.31, *p*s < .01, but shorter for subject-relatives and unreduced object-relatives than for reduced object-relatives, *t*s > -4.27, *p*s < .001. Figure 5.3.15 below shows the recreated mean per word reading times from Waters and Caplan (1996b) and mean per word times for Set 2.

As before, we also looked at mean per word reading times in light of our animacy manipulation, as displayed in Figure 5.3.16 below. A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA. This analysis showed main effects of complexity, F (3, 90) = 25.81 < .001, η_p^2 = .46, and animacy, F (1, 30) = 5.11, p < .03, η_p^2 = .15, that were qualified by an interaction between the two, F (3, 90) = 2.91, p = .04, η_p^2 = .08.

Examination of this interaction showed that per word reading times were longer for inanimate subject-relative items than for animate subject-relatives, t (31) = 3.00, p <.01, and for longer for inanimate reduced items than for animate reduced items, t (31) = 2.40, p = .02.

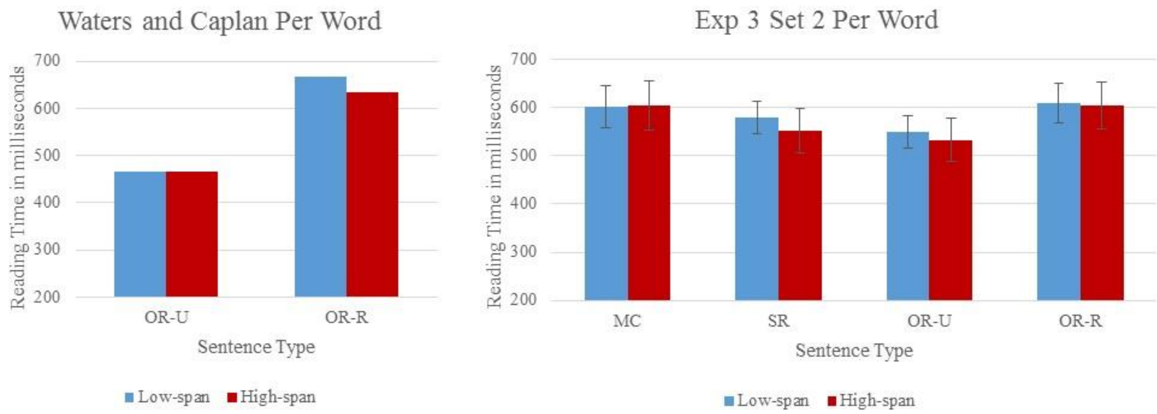


Figure 5.3.15. Recreated per word reading times from Waters and Caplan (1996b) in the panel of the left compared to Experiment 3 Set 2 data on the right, with standard error bars.

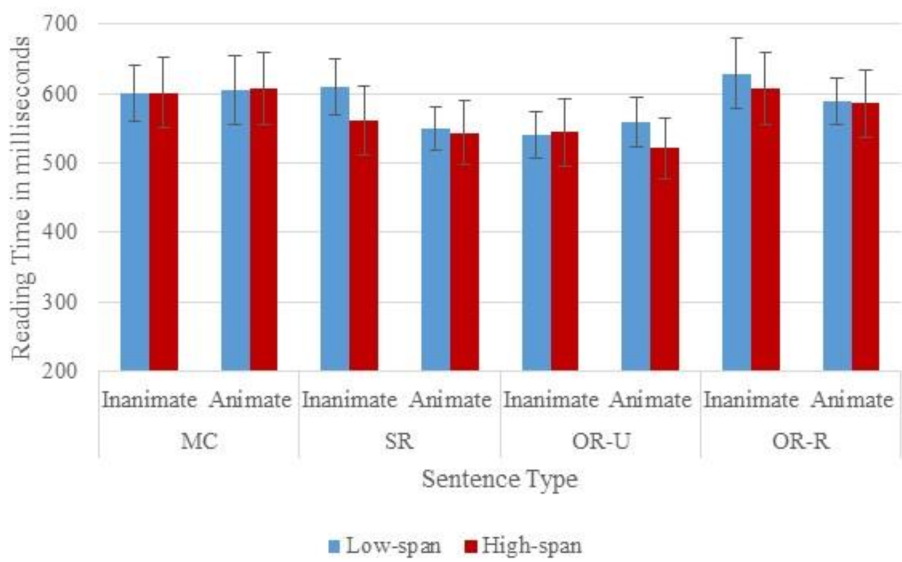


Figure 5.3.16. Experiment 3 Set 2 per word reading times, with standard error bars.

Waters and Caplan (2004) Stimuli – Set 3 (Appendix D)

The Set 3 stimuli in Experiment 3 are identical to those in Experiment 1, with the exception that items that were followed by a false comprehension item in Experiment 1 were modified in Experiment 3 to contain a small grammatical.

Unlike Set 1, it was not possible to place the grammatical error after the critical area in Set 3 because sentences ended with a noun phrase immediately following the critical area (i.e., the second verb). Therefore, the grammatical error was found on the critical area itself as in Set 1 sentences in Experiment 2 (e.g., The lock that the thief opened *secure* the safe.)

Accuracy.

A 4 (complexity) x 2 (animacy) x 2 (answer) x 2 (WMC) mixed within-between repeated measures ANOVA on grammaticality judgment accuracy showed main effects of complexity, *F* (3, 90) = 13.67, $p < .001$, $\eta_p^2 = .31$, animacy, *F* (1, 30) = 12.44, $p < .01$, $\eta_p^2 = .29$, answer, *F* (1, 30) = 19.62, $p < .001$, $\eta_p^2 = .40$, and WMC, *F* (1, 30) = 4.96, $p < .05$, $\eta_p^2 = .14$. These main effects were qualified by a two-way interaction of complexity and answer, *F* (3, 90) = 24.03, $p < .001$, $\eta_p^2 = .45$, which in turn was qualified by a three-way interaction of complexity, animacy, and answer, *F* (3, 90) = 2.73, $p < .05$, $\eta_p^2 = .08$. Examination of this interaction showed that accuracy was significantly higher for grammatical inanimate main clause items than for grammatical animate main clause items, *t* (31) = -2.78, $p < .01$, and that for agrammatical items, inanimate subject-relatives had higher accuracy than animate subject-relatives, *t* (31) = -4.19, $p < .001$. There was also an interaction of complexity and WMC, *F* (3, 90) = 3.08, $p = .03$, $\eta_p^2 = .09$, such that low-spans showed significantly higher accuracy on main clause items than on reduced object-relatives, *t* (15) = 2.62, $p = .02$, but high-spans did not differ in accuracy for any sentence type. Figure 5.3.17 below shows accuracy scores for each type of sentence.

In addition to simple accuracy, A' scores were calculated for the Set 3 data for comparison with Waters and Caplan (1996b). A 4 (complexity) x 2 (WMC) mixed within-between repeated measures ANOVA revealed a main effect of complexity, *F* (3, 90) = 9.12, $p < .001$, $\eta_p^2 = .23$, such that main clause A' scores were lower than those for subject-relatives, unreduced object-relatives, and reduced object-relatives, *t*s > -2.41, *p*s < .01, and A' scores for subject-relatives were lower than those for reduced object-relatives, $t = 2.07$, $p < .05$. The main effect of WMC was not significant ($p = .07$, $\eta_p^2 = .11$), nor was the interaction. Figure 5.3.18 below shows the mean A' scores by sentence type for Set 3 on the right and the A' scores recreated from Waters and Caplan (1996b) on the left.

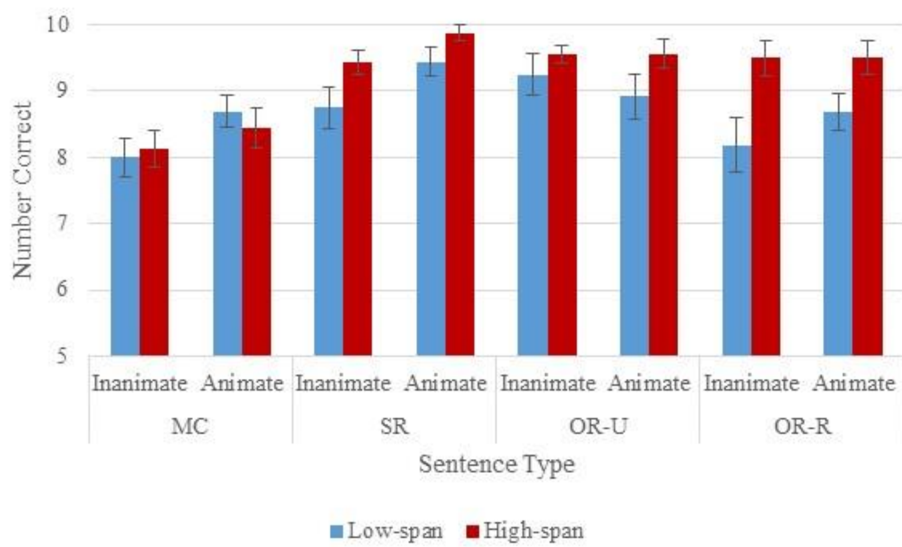


Figure 5.3.17. Experiment 3 Set 3 Grammaticality judgment accuracy, with standard error bars.

We also examined A' scores in a 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA to examine the effects of animacy. This analysis showed main effects of complexity, *F* (3, 90) = 9.12, $p < .001$, $\eta_p^2 = .23$, and animacy, *F* (1, 30) = 6.61, *p*

= .02, η_p^2 = .18, and an interaction between the two, F (3, 90) = 3.62, p = .02, η_p^2 = .10, such that A' were lower for inanimate main clause and subject-relative clause sentences than for animate main clause and subject-relative sentences, respectively, *t*s > -2.87, *p*s < .01. Figure 5.3.19 below displays the A' means by sentence type.

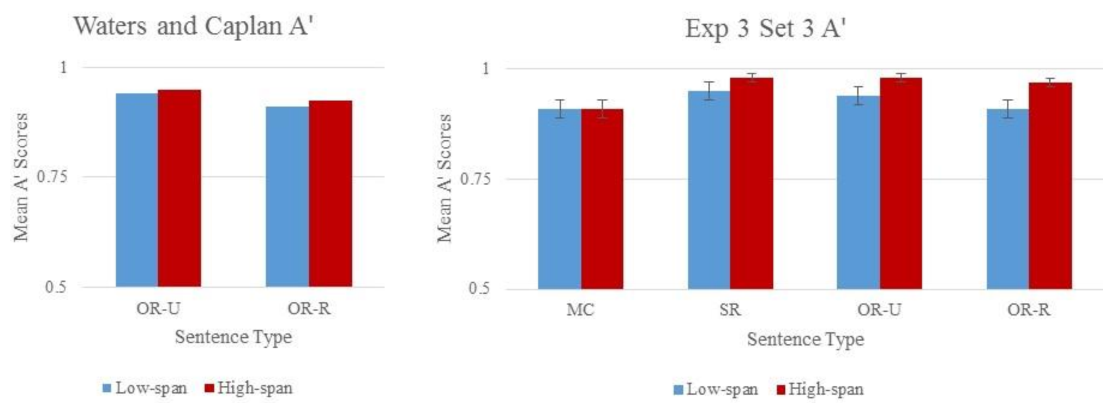


Figure 5.3.18. Recreated A' scores from Waters and Caplan (1996b) in the left panel compared to Set 3 A' scores in Experiment 3 on the right, with standard error bars.

The results of the accuracy analysis are in line with what one would expect from both the single and dual resource perspectives in that low-spans showed a significant decrease in performance between the simplest (i.e., MC) and most complex (i.e., reduced object-relative) sentences, but the high-spans showed no such decline. Both theories would agree that low-spans should decline in accuracy as complexity increases. The lack of three-way interaction is also congruent with both theories, as already discussed.

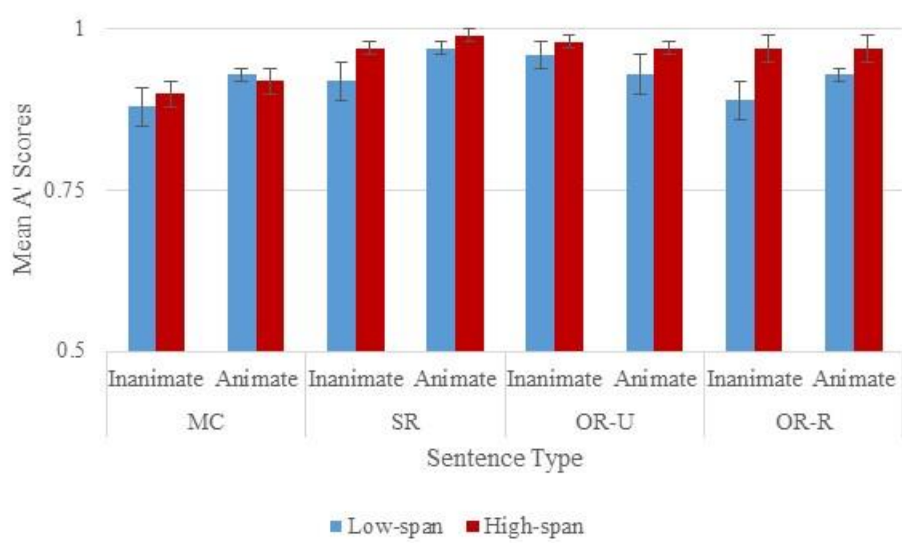


Figure 5.3.19. Experiment 3 Set 3 A' scores, with standard error bars.

Judgment Response Time.

A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on grammaticality judgment response time showed a main effect of animacy, F (1, 30) = 4.11, p = .05, η_p^2 = .12, that was due largely to longer response times for inanimate main clause items than for animate ones. Figure 5.3.20 below displays the mean grammaticality judgment response time for each sentence type.

In Experiment 1 we saw a three-way interaction of complexity, animacy, and WMC. In the current experiment, neither animacy nor WMC interact with complexity individually or together. This may be an effect of the grammatical error being placed in the critical area in the current experiment (an unfortunate limitation of the sentence structure). It is possible that not only the presence (in agrammatical items) but the possibility of a grammatical error (in grammatical items) in the critical area changed processing in such a way as to cancel out the interaction found previously. Indeed, judgment response times in the current experiment are significantly faster than those found in Experiment 1 (which were between 2500 and 3500ms), and more in line with those from Experiment 2 (which were between 700 and 1100ms), suggesting something about the task changed response strategies. It could be argued (likely by Waters and Caplan, 1996a) that what changed was that the judgments in Experiments 2 and 3 were more online than the true/false judgment in Experiment 1 and thus more reflective of syntactic processing. The evidence in the current experiment cannot refute this argument, though the effect of animacy for high-spans in Experiment 1 is suggestive that the interaction was due to online processing rather than offline (i.e., that the high-spans were taking animacy into account when it was encountered early in the sentence and this advantage was transferred over to judgment times). Thus it seems at least possible that the change was related to an error detection mechanism that when activated in the critical area erases the carry-over effect seen in Experiment 1.

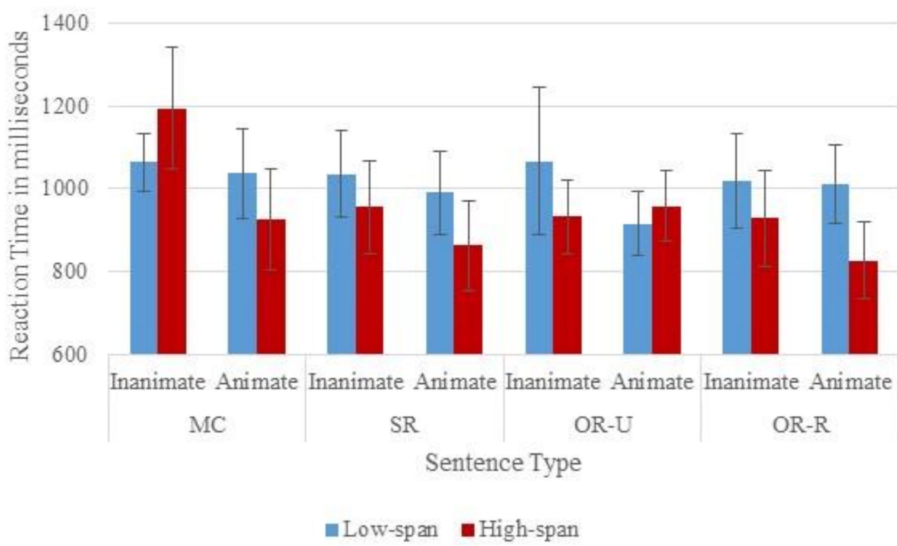


Figure 5.3.20. Experiment 3 Set 3 Grammaticality judgment response times, with standard error bars.

Critical Area Reading Time.

Critical reading times for Set 3 sentences could only be analyzed for correctly judged grammatical items, due to the presence of the grammatical anomaly in the critical area of the sentence. A 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA on these critical area reading times revealed a main effect of complexity, F (3, 90) = 4.07, $p < .01$, $\eta_p^2 = .12$, that was qualified by an interaction of complexity and animacy, F (3, 90) = 4.50, $p < .01$, $\eta_p^2 = .13$, such that reading times were faster in inanimate subject-relative items than in animate subject-relatives, t (31) = -3.85, $p < .01$, but no other contrasts were significant. Figure 5.3.21 below shows the mean critical area reading times for each sentence type. These times were trimmed as described above, and less than 2% replaced.

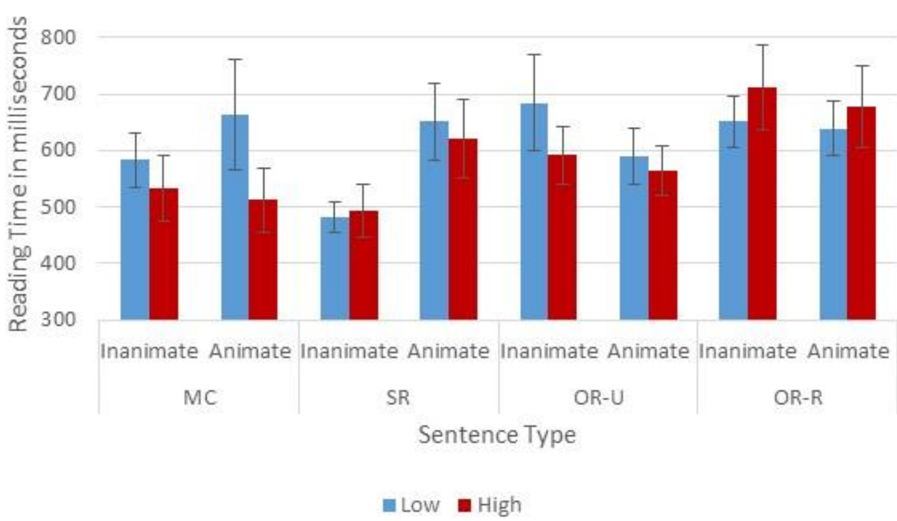


Figure 5.3.21. Experiment 3 Set 3 critical area reading times, with standard error bars.

Figure 5.3.22 below displays the critical area reading times from the Set 3 object-relative sentences in the right panel for comparison with the Just and Carpenter (1992) data on the left. While nothing is significant, three things are noticeable in this comparison. First, there is a trend for reading times to be longer for inanimate items than for animate ones for both low- and high-spans. Next, notice that reading times are shorter for high-spans on unreduced items but longer on reduced items. Finally, the reading times for Set 3 are all above the maximum values seen in the Just and Carpenter (1992) data. These differences are likely effects of locating the grammatical error on the critical area.

Mean per word reading times were also calculated for Set 3 (Figure 5.3.23 below), for ease of comparison with the results of Waters and Caplan (1996b). A 4 (complexity) x 2 (WMC) mixed within-between repeated measures ANOVA on per word reading times revealed a main effect of complexity, F (3, 90) = 12.49, $p < .001$, $\eta_p^2 = .29$, such that per word times were longer for main clause than for subject-relative items, t (31) = 3.26, $p < .01$, longer for reduced object relatives for main clauses, subject-relatives, and unreduced object-relatives, ts = -3.35, ps < .01.

We also examined per word reading times in a 4 (complexity) x 2 (animacy) x 2 (WMC) mixed within-between repeated measures ANOVA in order to explore any effect of the animacy manipulation. The analysis showed a main effect of complexity, F (3, 90) = 12.49, $p < .001$, η_p^2 =

.29, and one of animacy, animacy F (1, 30) = 4.55, p = .04, η_p^2 = .13, but no interaction of the two. The main effect of complexity was such that per word times were longer for main clause items than for subject-relatives, t (31) = 3.26, p < .01, and longer for reduced object-relative items than for main clauses, subject-relatives, and unreduced object-relatives, *t*s > 3.35, *p*s < .01. Figure 5.3.24 below displays the mean per word reading times for each sentence type.

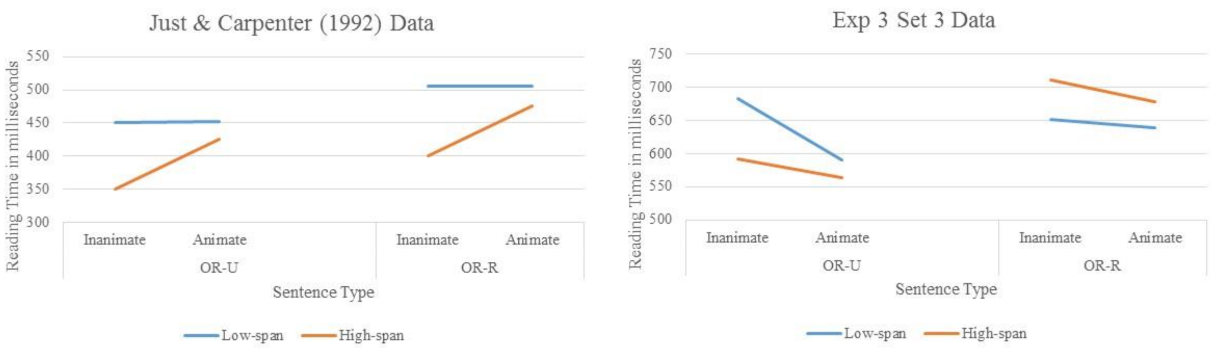


Figure 5.3.22. Experiment 3 Set 3 critical area reading times.

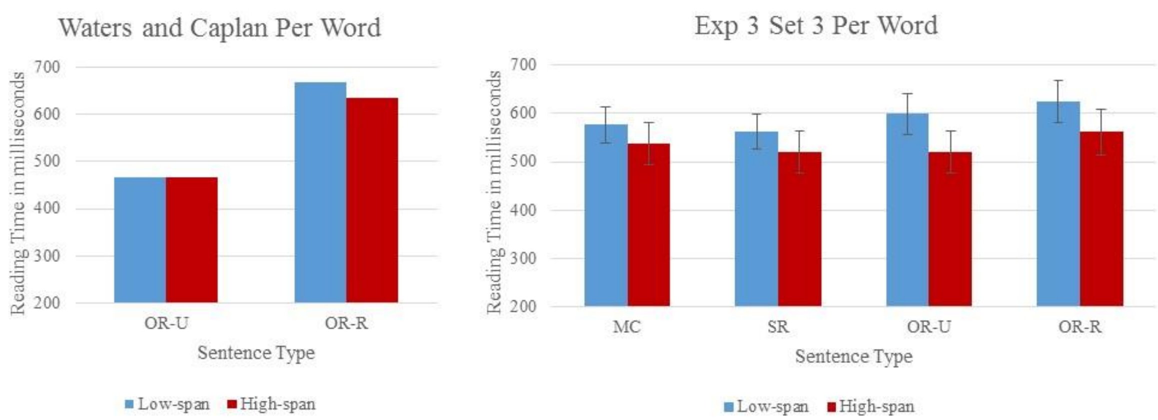


Figure 5.3.23. Recreated per word reading times from Waters and Caplan (1996b) in the panel of the left compared to Experiment 3 Set 3 data on the right, with standard error bars.

Finally, in line with Waters and Caplan (2004), we analyzed reading times in two 2 (complexity) x 5 (phrase) x 2 (WMC) mixed within-between repeated measures ANOVAs. Figure 5.3.25 below is a recreation of the Waters and Caplan (2004) data, shown for comparison with the Set 3 data to follow.

In the first 2 (complexity) x 5 (phrase) x 2 (WMC) mixed within-between ANOVA, we looked at subject-relatives compared to unreduced object-relatives (i.e., the same structures used in Waters and Caplan, 2004). This analysis showed main effects of complexity, F (1, 30) = 6.89, p = .01, η_p^2 = .19, and phrase, F (4, 120) = 12.31, p < .001, η_p^2 = .29, that were qualified by an interaction between the two, F (4, 120) = 5.81, p < .001, η_p^2 = .16. For subject-relative items reading times were faster: for NP1 than for NP3, t (31) = -2.51, p = .02, for NP2 than for V1 and V2, *t*s > -2.66, *p*s < .01, and for V1 than for V2 and NP3, *t*s > -2.43, *p*s ≤ .02. For unreduced object-relative items reading times were shorter for: NP1 than for V1, V2, and NP3, *t*s > -3.71, *p*s < .01, and for NP2 than for V1, V2, and NP3, *t*s > -4.33, *p*s < .001, and for V1 than NP3, t

(31) = -2.58, *p* = .02. We also conducted independent samples *t*-tests for just the V2 area. These results confirmed that low- and high-span reading times did not differ for any sentence type. The left panel of Figure 5.3.26 below displays the phrase reading times in this analysis.

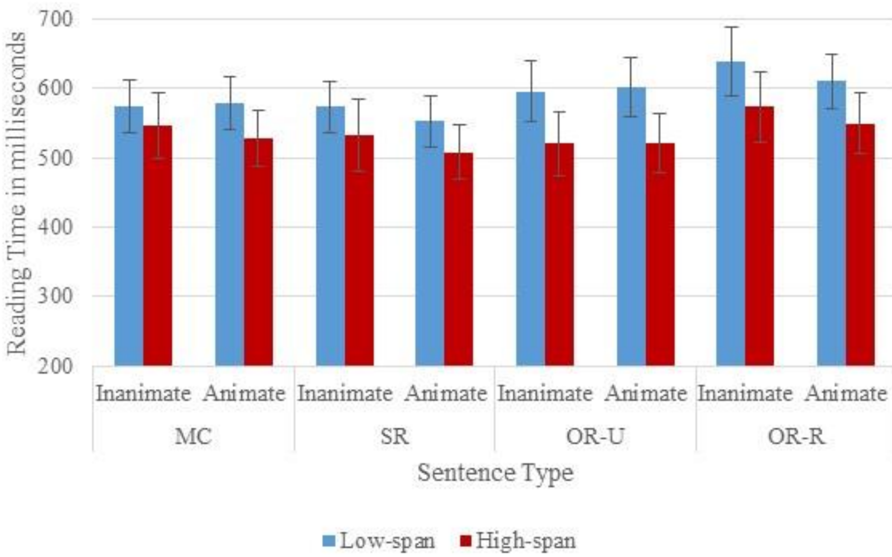


Figure 5.3.24. Experiment 3 Set 3 per word reading times, with standard error bars.

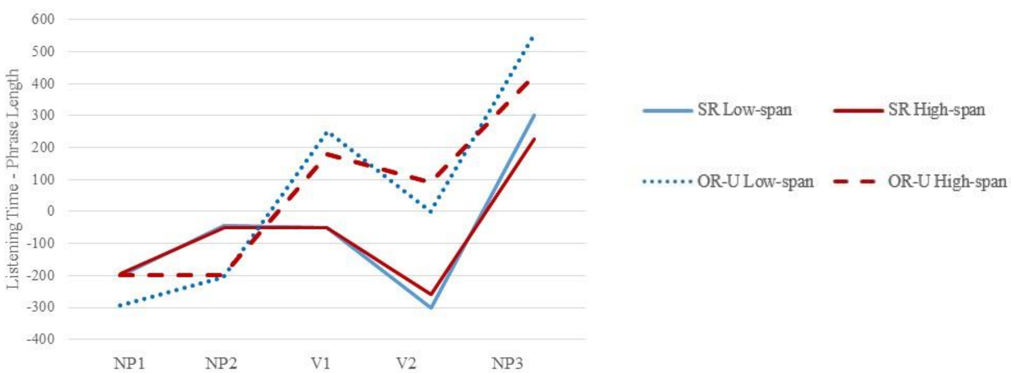


Figure 5.3.25. Recreation of Waters and Caplan (2004) phrase listening times.

In the second 2 (complexity) x 5 (phrase) x 2 (WMC) mixed within-between ANOVA, we looked at subject-relatives compared to reduced object-relatives (a type not presented by Waters and Caplan, 2004). This analysis showed main effects of complexity, *F* (1, 30) = 22.29, *p* < .001, η_p^2 = .43, and phrase, *F* (4, 120) = 11.14, *p* < .001, η_p^2 = .27, that were qualified by an interaction between the two, *F* (4, 120) = 8.98, *p* < .001, η_p^2 = .23. The results in this analysis for subject-relatives are the same as those in the previous analysis. For reduced object-relatives reading times were faster: for NP1 than for V1, V2, and NP3, *t*s > -3.31, *p*s < .01, and for NP2 than for V1, V2, and NP3, *t*s > -2.80, *p*s < .01. The right panel of Figure 5.3.26 below displays the phrase reading times in this analysis. We also conducted independent samples *t*-tests for just the V2 area. These results confirmed that low- and high-span reading times did not differ for any sentence type.

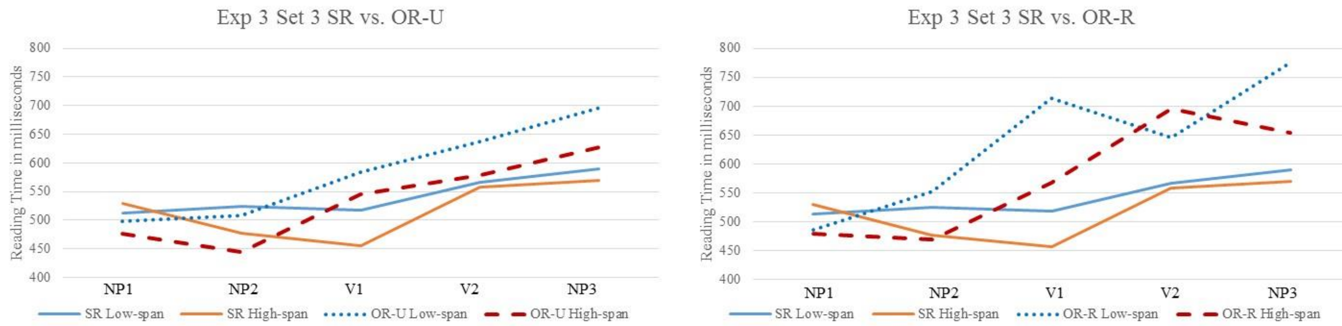


Figure 5.3.26. Experiment 3 Set 3 phrase reading time data.

Neither per word times nor phrase analyses showed critical three-way interactions (of complexity, animacy, and WMC in the first case, and of complexity, phrase, and WMC in the second). These findings are in line with those of dual resource studies such as Waters and Caplan (1996b, 2004), though this support for dual resource theories is qualified by the arguments already discussed for why the judgment type might have obscured any critical interactions in the current data.

5.4 Discussion

The purpose of Experiment 3 was to test an alternate type of judgment, grammaticality, using the same stimuli from the previous two experiments in a self-paced word-by-word reading paradigm in order to determine if this judgment might be more appropriate for use in the syntactic processing task than either true/false or sense judgments. The results suggest that the grammaticality judgment affects sentence reading strategies in a similar manner as the sense judgment used in Experiment 2. That is, the use of the grammaticality judgment causes all participants to spend longer reading sentences than they do when they are asked to make true/false comprehension judgments. This effect is one potential explanation for the fact that the significant effects of WMC seen in Experiment 1 did not re-appear in Experiment 3.

The lack of any significant interactions with WMC could be argued as providing support for the dual resource view. However, such a conclusion may be premature given the variety of stimulus characteristics noted above. For example, the fact that the grammatical error was predictable and came before the critical area in at least some experimental sentences (e.g., Set 2) could have altered reading strategies just as semantic anomalies seemed to in Experiment 2. Thus the main take-way from Experiment 3 is not support for either the single or dual resource theory, but rather a reinforcement of the argument that the type of judgment used in the syntactic processing task has significant effects on the major dependent variables. These effects should be fully considered when attempting to test between the theories.

CHAPTER 6. CROSS-EXPERIMENT COMPARISONS

In addition to the analyses within experiments, we were also interested in looking for cross-experiment differences within each of our three stimulus sets on the main dependent variables of accuracy, judgment response time, and critical area reading time, focusing exclusively on the experimental object-relative sentences. By looking at the same set across all three judgment types, we can more clearly determine how differing the type of judgment in the syntactic processing task affects the results.

6.1 Set 1 – Just and Carpenter (1992)

Accuracy.

We ran a series of one-way ANOVAs on accuracy rates for each of the four experimental object-relative sentences. There were no significant differences in accuracy for inanimate object-relative sentences of either type (i.e., unreduced or reduced). For animate unreduced object-relative sentences, there was a significant effect of experiment, F (2, 93) = 4.00, $p < .05$. Bonferroni post hoc comparisons revealed the accuracy was significantly higher in Experiment 2 than in Experiment 1, $p < .05$. There was also a significant effect of experiment in the animate reduced object-relative sentences, F (2, 93) = 5.53, $p < .01$. Bonferroni post hoc comparisons showed that accuracy was significantly higher in Experiment 2 than in both Experiment 1 and Experiment 3, $p < .05$. Mean accuracy rates in each experiment are displayed in Figure 6.1.1 below.

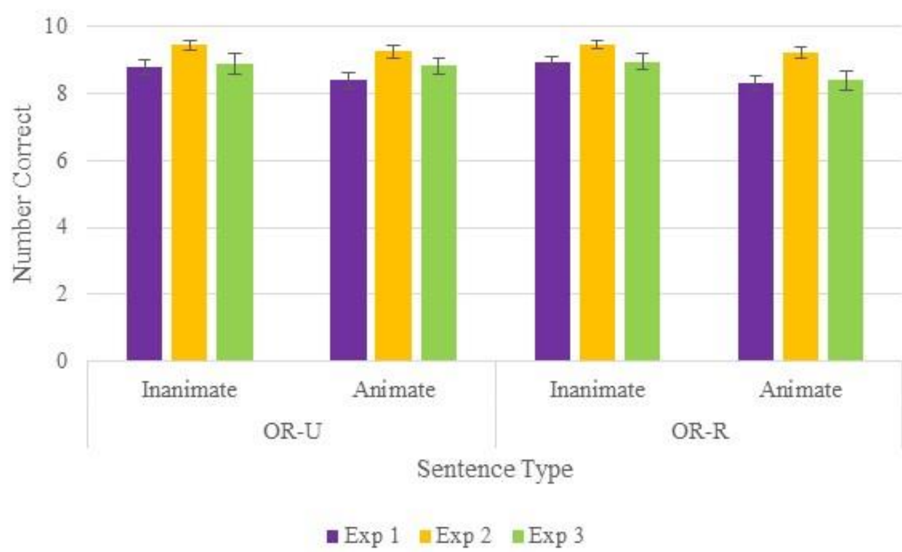


Figure 6.1.1. Judgment accuracy across Experiments for experimental object-relative sentences, with standard error bars.

Overall accuracy rates for experimental sentences are quite high, yet we still find a significant increase in performance on the most difficult sentences of the set (i.e., the animate reduced) using the sense judgment in Experiment 2. We have already argued, with some support, that placing an anomalous word in sentences changes the reading strategy and thus affects reading times. The significant change in accuracy also suggests the sense judgment task may be different from the true/false comprehension judgment in an important way. Importantly, this difference seems to make the sense judgment task easier than either the true/false comprehension task or the grammaticality task when complexity is greatest. This suggests that the sense judgment may not be the most appropriate one to use when assessing syntactic processing.

Judgment Response Time.

A series of one-way ANOVAs on judgment response times for each of the four experimental object-relative sentences showed significant group differences for all sentence types, *F*s > 111.41, *p*s < .001. Bonferroni post hoc comparisons showed that Experiment 1 judgment response times were significantly longer than those in Experiment 2 and Experiment 3, *p*s < .001. Figure 6.1.2 below displays the means by experiment.

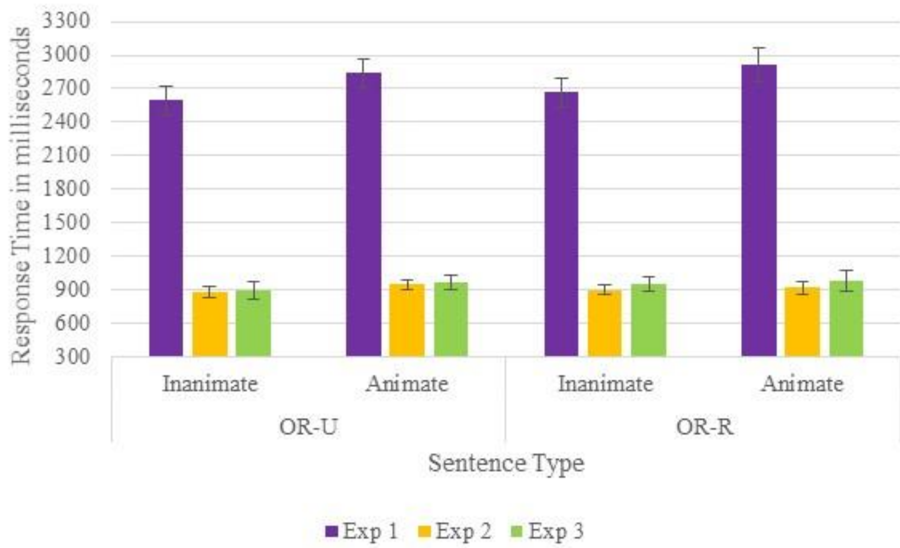


Figure 6.1.2. Judgment response times across Experiments for experimental object-relative sentences, with standard error bars.

Longer judgment response times for the true/false judgment are not surprising, given that the true/false comprehension judgment requires the participant to read a short statement and then make the judgment, while the sense and grammaticality judgments require only a response. Because of this, the comparison of judgment response times across experiments does not tell us much of interest about the influence of judgment type on performance.

Critical Area Reading Time.

A series of one-way ANOVAs on critical area reading times in true/acceptable/ grammatical items for each of the four experimental object-relative sentences showed significant group differences for all sentence types, *F*s > 4.68, *p*s < .01. Bonferroni post hoc comparisons showed that critical area reading times were significantly faster in Experiment 1 than in Experiment 2 and Experiment 3 for all sentence types, all *p*s < .05. Figure 6.1.3 below displays the means for each sentence type by experiment.

The finding that critical area reading times increased with sense and grammaticality judgments suggests that these types of judgment fundamentally change the nature of the syntactic processing task. If this effect were only found in Experiment 2, one might argue that it was strictly an artifact of placing the anomalous word in the critical area. However, the effect persists in Experiment 3 when the grammatical error is found just after the critical area. Still, one might argue that the effect in Experiment 3 is due to the anticipation of the error arriving just after the critical area. However, the same effect is present both before the critical area (*F*s > 5.86, *p*s < .01; Bonferroni *p*s < .05; see Figure 6.1.4) and after it (*F*s > 6.59, *p*s < .01; Bonferroni *p*s < .05; see Figure 6.1.5).

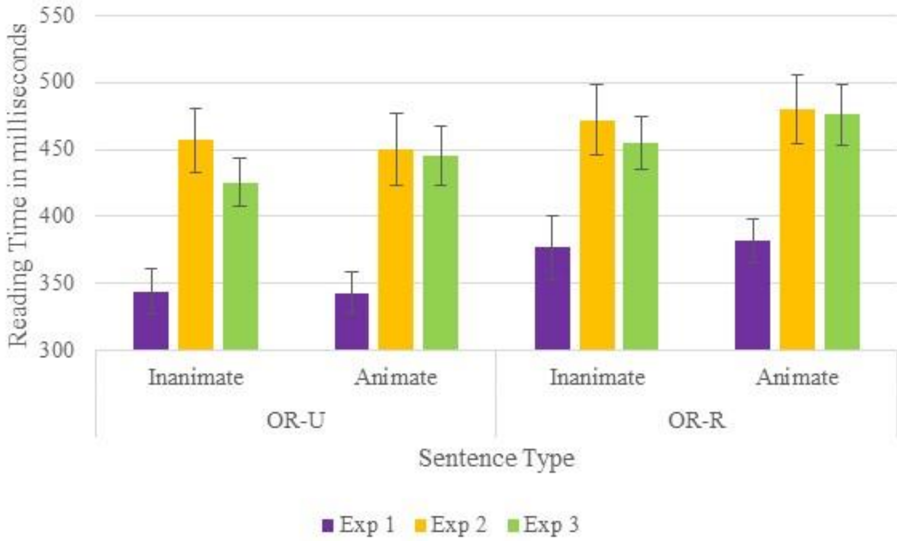


Figure 6.1.3. Critical area reading times by Experiment for experimental object-relative sentences, with standard error bars.

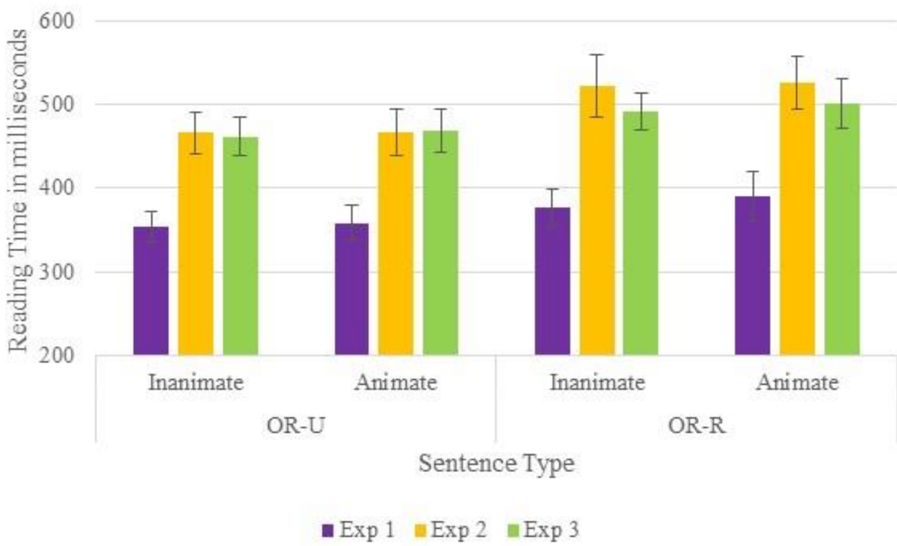


Figure 6.1.4. Area 1 reading times by Experiment for experimental object-relative sentences, with standard error bars.

Thus, analysis of each of the three areas of the sentence show an effect of judgment type, such that sense and grammaticality judgments slow down reading times, even when semantic or grammatical anomalies are not actually present in the acceptable and grammatical items. This supports the argument that participants are engaging in a different strategy while performing the task with instructions to look for semantic anomalies or grammatical errors than when they know they will be asked a true/false question about the sentence.

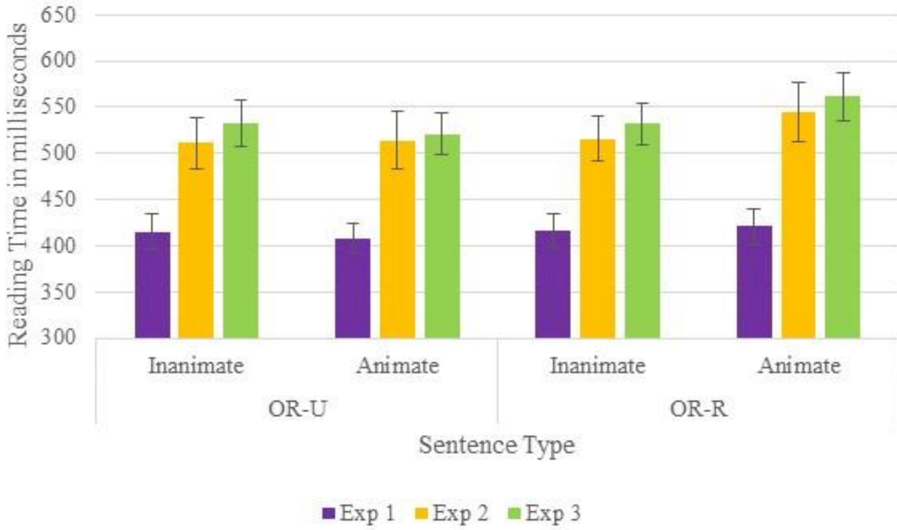


Figure 6.1.5. Area 3 reading times by Experiment for experimental object-relative sentences, with standard error bars.

6.2 Set 2 – Waters and Caplan (1996b; Appendix C)

Accuracy.

Figure 6.2.1 below shows mean accuracy on each of the four experimental sentence types by experiment for Set 2 stimuli. A series of one-way ANOVAs showed significant group differences for each of the four experimental sentence types, *F*s > 21.10, *p*s < .001. Bonferroni post hoc comparisons for inanimate and animate unreduced object-relatives and inanimate reduced object-relatives showed that accuracy was lower in Experiment 1 than in Experiment 2 and Experiment 3, *p*s < .001. For animate reduced object-relatives, the reverse was true – accuracy was higher in Experiment 1 than in Experiment 2 and Experiment 3, *p* < .001.

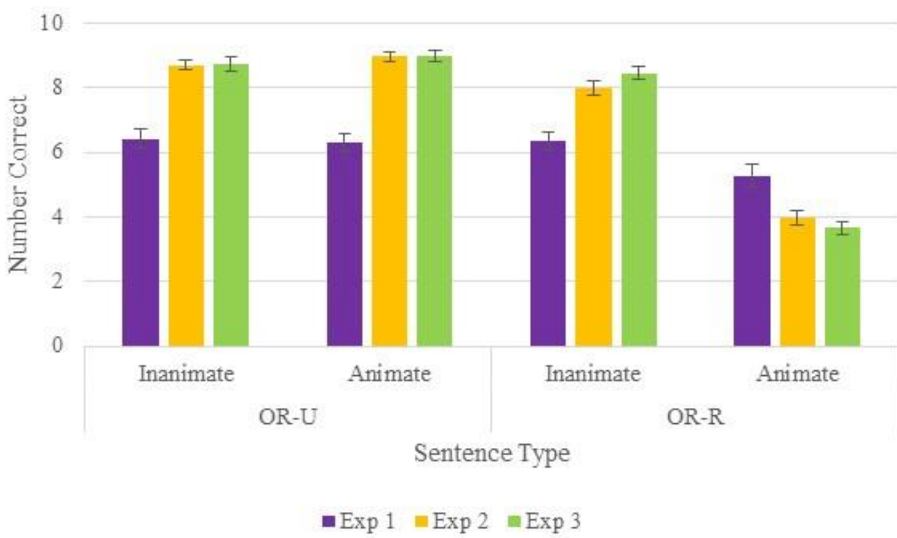


Figure 6.2.1. Judgment accuracy across experiments, with standard error bars.

These results suggest that both the sense and grammaticality manipulations reduce the difficulty of judging Set 2 sentences, except in the most ambiguous case (i.e., animate reduced object-relatives). This is likely due to the fact that the animate reduced object-relatives in this set are extremely difficult and unusual and thus appear nearly nonsensical as they are, regardless of deliberate anomalies or errors (e.g., The horse raced past the barn fell.)

Judgment Response Time.

A series of one-way ANOVAs on judgment response times for the four types of experimental sentences showed significant group differences for all types, *F*s > 90.49, *p*s < .001. As in the previous set, Bonferroni post hoc comparisons showed that Experiment 1 judgment times were significantly longer than those in Experiment 2 and Experiment 3, *p*s < .001. Figure 6.2.2 below displays the mean judgment response times for each sentence type.

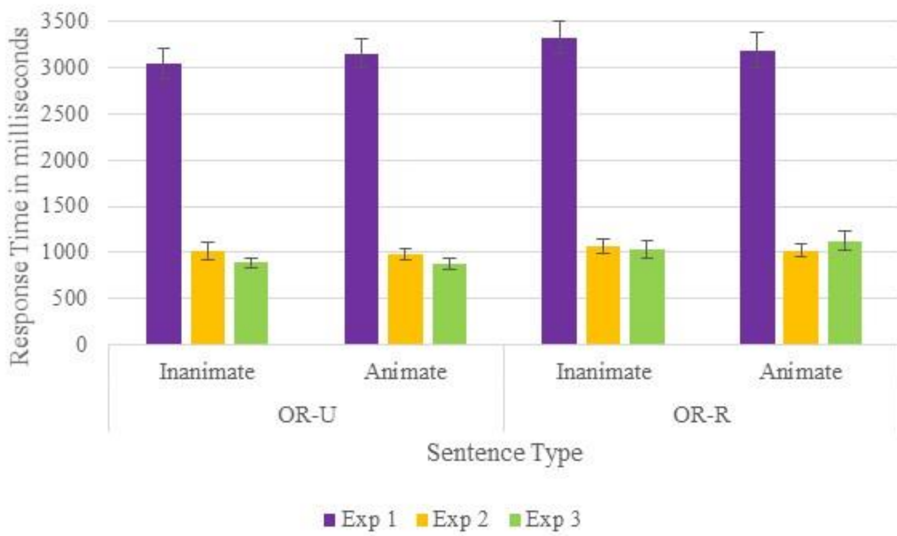


Figure 6.2.2. Judgment response times across experiments, with standard error bars.

As already discussed, the fact that judgment response times are longer in Experiment 1 is most likely due almost entirely to the fact that participants must read a true/false item and then make a judgment, as opposed to judging immediately when the response screen appears.

Critical Area Reading Time.

A series of one-way ANOVAs on critical area reading times for the four experimental sentences types revealed significant group differences for all types, *F*s > 6.82, *p*s < .01. Bonferroni post hoc comparisons revealed that this was due to faster critical area reading times in Experiment 1 than in Experiment 2 and Experiment 3, *p*s < .05. Figure 6.2.3 below shows the mean critical area reading times for each sentence type.

As with Set 1, we also looked at reading times in the pre-critical area of the sentence (the only other area in Set 2 sentences). As can be seen in Figure 6.2.4 below, reading times in this area also appear longer in the second and third experiments than in the first, though the difference is only significant for inanimate and animate unreduced object-relatives, *F*s > 3.51, *p*s < .05, and only between Experiment 1 and 3, Bonferroni post hoc *p*s < .05.

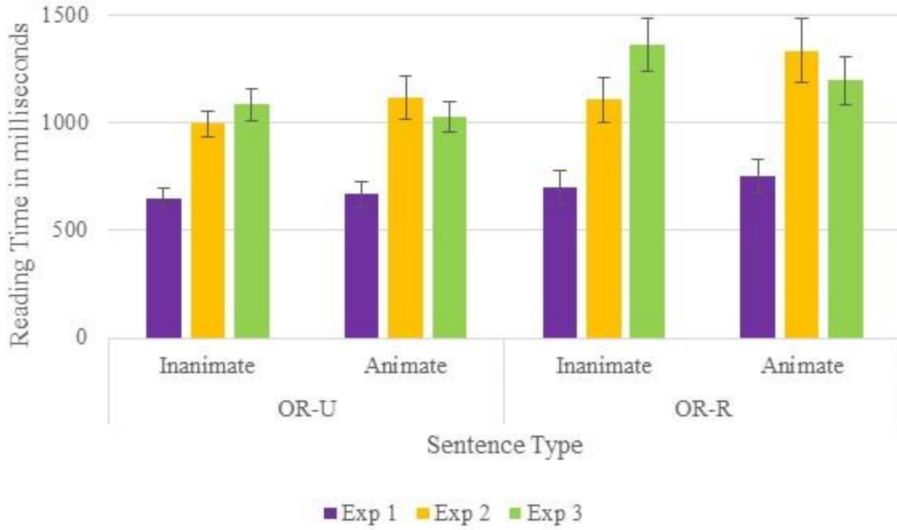


Figure 6.2.3. Critical area reading times across experiments, with standard error bars.

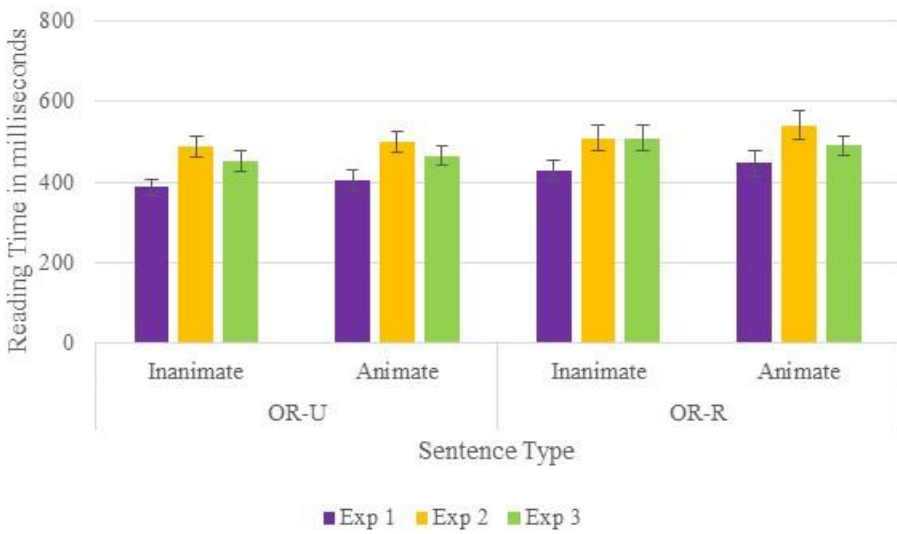


Figure 6.2.4. Pre-critical area reading times across experiments, with standard error bars.

Set 2 critical area reading time analyses line up with those seen in Set 1, that is that reading times increase in sense and grammaticality judgment conditions, compared to the true/false judgment condition. There was a smaller effect in the pre-critical area of Set 2, however, together these reading time results still support the argument of a fundamental change in the nature of the syntactic processing task caused by judgment type.

6.3 Set 3 – Waters and Caplan (2004)

Accuracy.

A series of one-way ANOVAs on judgment accuracy for the four experimental sentence types in Set 3 revealed significant group differences for inanimate unreduced object-relatives, F (2, 93) = 8.24, $p < .01$, and inanimate reduced object-relatives, F (2, 93) = 3.49, $p < .05$. Bonferroni post hoc comparisons revealed that accuracy on inanimate unreduced object-relatives was lower in Experiment 1 than in Experiment 2 and Experiment 3, $p < .01$. For inanimate reduced object-relatives, accuracy was lower in Experiment 1 than in Experiment 2, $p < .05$. Figure 6.3.1 shows the mean accuracy for each sentence type.

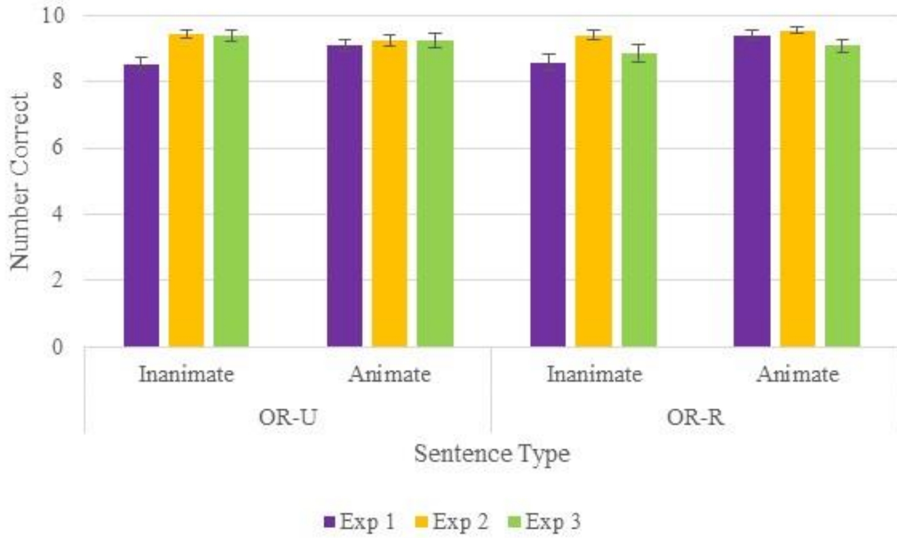


Figure 6.3.1. Judgment accuracy across experiments, with standard error bars.

Judgment Response Time.

A series of one-way ANOVAs on judgment response times for the four experimental sentences types revealed significant group differences for all types, *F*s > 77.11, *p*s < .001. Bonferroni post hoc comparisons showed that these differences were due to longer judgment response times in Experiment 1 than in Experiment 2 and Experiment 3, *p*s < .001. Figure 6.3.2 displays the mean judgment response time for each sentence type. As with the previous sets, the significantly longer times in Experiment 1 are not surprising, given the need for participants to read the true/false statement before making a judgment about it.

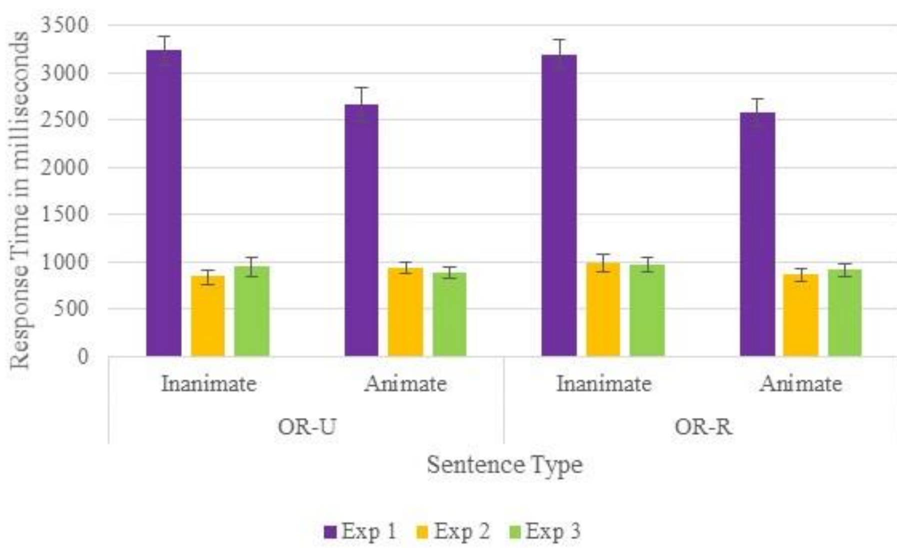


Figure 6.3.2. Judgment response time across experiments, with standard error bars.

Critical Area Reading Time.

A series of one-way ANOVAs on the critical area reading times for all four experimental sentence types revealed significant group differences for all but animate unreduced object-relatives, *F*s > 3.16, *p*s < .05. For inanimate unreduced and animate reduced object-relatives, Bonferroni post hoc comparisons showed that critical area reading times were shorter in Experiment 1 than in Experiment 3. For inanimate reduced object-relatives, Bonferroni post hoc comparisons showed that critical area reading times were shorter in Experiment 1 than in Experiment 2 and Experiment 3. Figure 6.3.3. shows the mean reading times for all types.

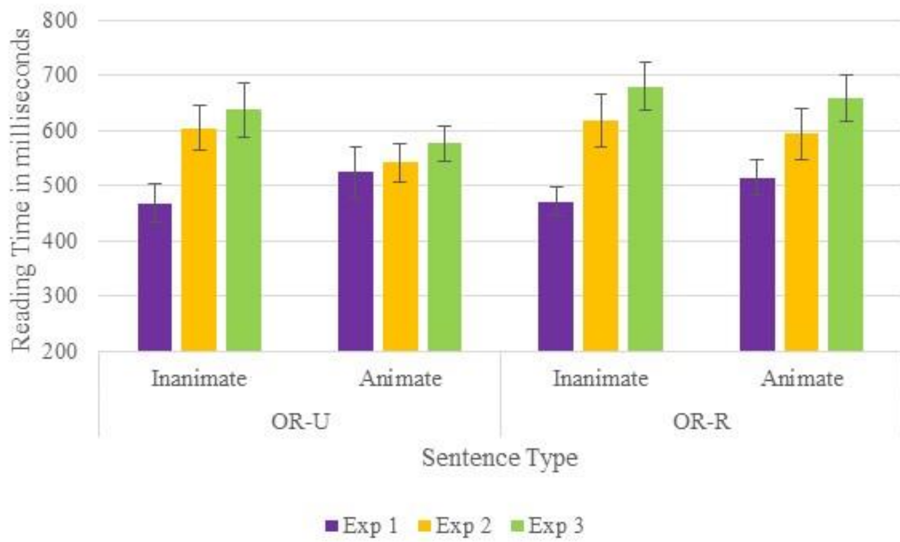


Figure 6.3.3. Critical area reading times across experiments, with standard error bars.

A series of one-way ANOVAs on the pre-critical area reading times showed significant group differences for both inanimate and animate reduced object-relatives, *F*s > 3.54, *p*s < .05. Bonferroni post hoc comparisons showed that reading times were faster in Experiment 1 than in Experiment 3 for both types of item, *p*s < .05. Means can be found in Figure 6.3.4.

A series of one-way ANOVAs on post-critical area reading times showed significant group differences for both types of inanimate item, *F*s > 4.48, *p*s <.01. Bonferroni post hoc comparisons for the inanimate unreduced items showed that post-critical area reading times were significantly faster in Experiment 1 than in Experiment 3, *p* < .05. For inanimate reduced items, post-critical area reading times were faster in Experiment 1 than in both Experiment 2 and Experiment 3, Bonferroni post hoc *p*s < .05. There was also a marginally significant group difference for animate reduced items, F (2, 93) = 3.06, p = .052. Means can be found in Figure 6.3.5 below.

Set 3 reading times thus show a pattern similar to that seen in Sets 1 and 2. Reading times throughout the sentence increase when the task is to detect a semantic anomaly or a grammatical error as compared to when the task is to make a true/false comprehension judgment.

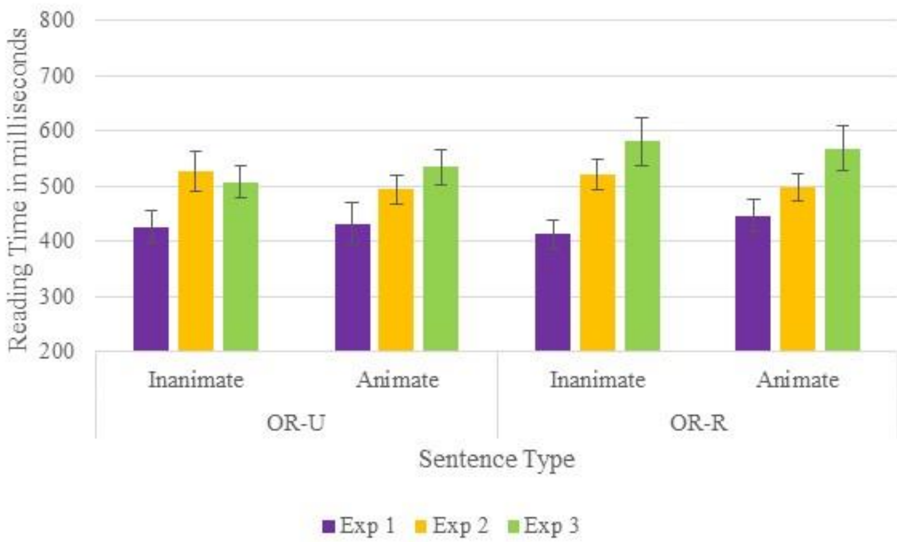


Figure 6.3.4. Pre-critical area reading times across experiments, with standard error bars.

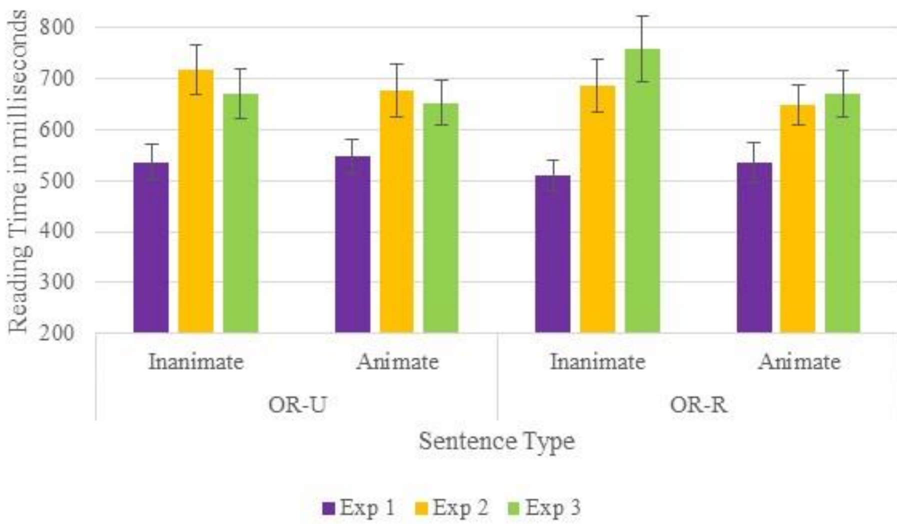


Figure 6.3.5. Post-critical area reading times across experiments, with standard error bars.

6.4 Discussion

The pattern of accuracy results across experiments for each of the three sets suggests that the sense and grammaticality judgments make it easier to judge complex sentences during the syntactic processing task (except in the case of Set 2 animate reduced sentences, which are just very confusing no matter what judgment type is used). In Set 1, accuracy increased on the animate items, which are more ambiguous than inanimate items. In Set 2, the most difficult set, accuracy increased on three of four sentence types. Finally, in Set 3, accuracy increased on the inanimate items, which we argued above are the more ambiguous items in this set. This change in task difficulty is problematic in terms of finding differences between span groups, because only a difficult task is likely to show such effects. It also supports the argument that the task is fundamentally different when sense and grammaticality manipulations are used than when true/false judgments are used.

In terms of judgment response time, our cross-experimental analyses tell us little about the effects of judgment type, given that the there is a very logical reason for the true/false comprehension judgment to take longer than either the sense or grammaticality judgment: the fact that it includes the time it takes participants to read the true/false statement, not just the time to make the judgment about it.

Finally, cross-experimental comparisons of sentence reading times, both in the critical area and out, support the argument that the sense and grammaticality judgment task demands produce a fundamental change in how participants read sentences because both judgment types produce slower reading times across the entire sentence. In Experiment 1 we argued that a rushing strategy in Set 1 sentences could be at least partially responsible for the lack of replication of Just and Carpenter's (1992) findings. Based upon this argument, we might expect that the slowing caused by sense and grammaticality judgments would result in replication with Set 1 in Experiments 2 and 3. We did not find this. We cannot entirely discount the possibility that the low difficulty level of Set 1 sentences, confirmed by cross-set accuracy comparisons, was sufficient in itself to continue to suppress any potential effects of WMC in these experiments. However, it is at least possible that the sense and grammaticality judgment types did not just slow reading, but in fact changed the very manner in which participants were processing the sentences in the task. The question remains as to whether this effect of judgment type would remain when the location of the anomaly or error was less predictable than it was in the current stimuli.

In summary, the current results provide clear evidence that the relationship between the syntactic processing task and WMC measures is different depending upon which judgment is used. Because of this, it is important that researchers determine the nature of the effect of judgment type on sentence reading times. If the effect is produced by a shift in processing strategy, this strategy must be assessed to determine if it is indeed reflective of interpretive processes, as Waters and Caplan (1996a) might argue, or if it is something else entirely. This determination will in turn elucidate which of the judgment types ought to be used in the syntactic processing task.

CHAPTER 7. GENERAL DISCUSSION AND CONCLUSIONS

One of the most enduring questions in linguistic research is the extent to which the production and comprehension of a language relies upon both domain-specific and domain-general resources. On the generativist side of the argument are those who extrapolate backward from linguistic data (e.g., judgments about polar interrogatives, as in Chomsky, 1968) to an abstract, innate, and domain-specific grammar that is believed to guide not only the production and comprehension of a native speaker, but also acquisition of a language in the first place because the language input to which children are exposed is considered too impoverished to be capable of allowing the child to learn the grammar from it. This account of language views the speaker as fully competent in the abstract grammar, and thus any errors in production or comprehension are attributed to external influences rather than to the domain-specific grammar.

On the other side of the debate are those who argue that linguistic input is constrained at all levels of representation by probabilities and distributional dependencies that are learnable through the use of domain-general statistical learning mechanisms. From a probabilistic constraints perspective, syntax and other higher order linguistic rules are in fact emergent properties of neural networks that are learning to map form to meaning in order to produce and comprehend language rather than evidence of an innate and abstract grammar. By placing linguistic acquisition and performance (i.e., production and comprehension) in the context of use, the probabilistic constraints framework is able to address domain-general influences where the generativists are not.

From these theories of what language is and how it is acquired, psycholinguists have developed models of how syntactic processing is accomplished. Just and Carpenter (1992) proposed the single resource theory of syntactic processing, arguing that processing of all types of linguistic information (e.g., morphology, syntax, pragmatics, etc.) recruits the same general verbal working memory resource, and therefore that measures of syntactic processing should be meaningfully related to measures of verbal working memory capacity. On the basis of this argument, they predicted that high- but not low-span participants would be able to take advantage of an animacy cue in complex object-relative sentences. Although their predictions were supported by their data, Waters and Caplan (1996a) disagreed with their arguments and conclusions, proposing instead that there are two separate resources used to process linguistic information. One resource is syntax-specific, and processes syntax obligatorily and automatically, while the other resource is used for everything else (e.g., pragmatics). To support these arguments, Waters and Caplan (1996b, 2004) also performed experiments in which they measured syntactic processing and verbal working memory. The results of these studies directly contradicted the findings of Just and Carpenter (1992), according to Waters and Caplan (1996a; 1996b; 2004).

Thus, both single and dual resource theories have empirical support, despite the fact that they make opposing predictions that should be able to determine which of them is more accurate. In Chapter 2 we discussed the empirical support in some detail in order to illuminate potentially important methodological differences between studies from the two perspectives. Among these differences are judgment type (comprehension vs. acceptability, due to Waters and Caplan's argument that comprehension items are offline measures), stimulus type (*by* phrase, garden path, or subject-object object-relative sentences), presentation method (eye-tracking, whole sentence presentation, listening times), and measurement of WMC (original and modified Daneman and Carpenter, 1980, tasks, etc.). The conclusion of this review was the idea that one reason there has not yet been a consensus in the literature is that the two studies have been comparing apples to oranges all along. This is because, cumulatively, all of the methodological differences between the studies from each perspective have resulted in seemingly contradictory data that actually fit together logically when external factors (i.e., more than just the object-relative structure type) are taken into account.

The purpose of the current experiments was therefore to provide a first step in comparing apples to apples by bringing together the stimuli from three representative studies (Just and Carpenter, 1992; Waters and Caplan, 1996b, 2004) in one presentation paradigm (self-paced word-by-word reading) in order to examine the effects of judgment type, stimulus type, and data analysis procedures on the relationship (if any) between syntactic processing tasks and measure of working memory capacity.

In the first experiment, we attempted to replicate the results of Just and Carpenter (1992) using their true/false judgments on three stimulus sets created based upon their own *by* phrase object-relatives, the garden path object-relatives of Waters and Caplan (1996b), and the subject-object relatives of Waters and Caplan (2004) in a self-paced word-by-word reading paradigm. Recall that Just and Carpenter's (1992) critical finding was that high- but not low-span participants could use the inanimacy of the subject of a sentence containing an object-relative clause to process the critical *by* phrase more quickly. Based upon the idea of Waters and Caplan (1996b, 2004) that all object-relatives are created equal, the overall prediction before the experiment was run was that whatever effects were found would be found in all three sets. That is, if single resource theory were to be supported by the presence of a three-way interaction of complexity, animacy, and WMC, it would be supported regardless of stimulus set, while if dual resource theory would to be supported by a lack of the same interaction, it would be lacking in all three sets. In order to ensure the direct comparability of the stimulus sets, the animacy of Just and Carpenter (1992) was added to the garden path (Waters and Caplan, 1996b) and subject-object (Waters and Caplan, 2004) stimulus sentences and half of the subject-objects sentences modeled after those of Waters and Caplan (2004) were reduced to increase their complexity.

The actual data from Experiment 1 showed several interesting results. For one thing, the results of Just and Carpenter (1992) were not replicated with their *by* phrase sentences. Several factors might explain this lack of replication. For one thing, participants subjectively reported that the Set 1 sentences were the easiest of the three sets. Objective accuracy data backed up these reports. Sentences for the set were created from an available list (Ferreira & Clifton, 1986) that served as a basis for the Just and Carpenter (1992) study because that was the list available. However, it is possible given the performance data in the current study that the Set 1 sentences in the current study were easier than those in Just and Carpenter (1992) due to some unknown factor, as we did not have the original materials (though this is hard to check, as Just and Carpenter, 1992, do not report accuracy data). Another explanation for the lack of replication with Set 1 was the participant-reported rushed button-pushing strategy, wherein participants chose to progress as quickly as possible through sentences, especially long ones such as those in Set 1, in order to get the entire sentence in mind before making the judgment. This strategy was reported by both low- and high-span participants, suggesting that even if the Set 1 sentences were difficult enough, it would still wash out any potential three-way interaction in the critical area.

For Set 2 sentences, which were subjectively and objectively the most difficult, we did find a two-way interaction of WMC and complexity in the critical area that Waters and Caplan (1996b) did not. This interaction did not represent an exact replication of Just and Carpenter

(1991), though we would argue that it is a theoretical replication. One problem was that the critical area in these sentences also happened to be the last word, meaning that these reading times also contain sentence wrap-up effects. Some might also argue that they could contain time spent on post-interpretive processing. However, the latter explanation was discounted on the basis that participants still took longer than they had in the first set to make judgments (though not all contrasts were significant). A second potential problem was that the WMC effect found in Just and Carpenter (1992) was reversed, such that it was the high-spans who took longer overall in the critical areas of object-relative sentences. However, this result makes more sense in the context of the results of MacDonald et. al. (1992). In this study, it was shown that high-spans take longer than low-spans on garden path sentences (such as those used in Set 2). MacDonald et al. (1992) argued that this difference reflected a processing cost for high-spans who maintained dual interpretations of the sentence in mind until disambiguating information was provided to indicate which interpretation was correct. Finally, the interaction in Set 2 data was between WMC and complexity rather than WMC and animacy as it was in Just and Carpenter (1992). We argued that this last difference was due to the nature of the garden path sentences which in turn limited the construction of their true/false comprehension items such that questions required a more subtle distinction of the relationship between the object and the verb of the relative clause rather than between two nouns as in Set 1 and Set 3. This slight change in the nature of judgment was argued to induce at least some high-span participants to alter their reading strategies so as to discount the animacy cue, resulting in high variability in critical area reading times for animate reduced object-relatives that may have obscured any effect of animacy. Thus, although the two-way interaction in Set 2 is not identical to that of Just and Carpenter (1992), we would argue that it is a theoretical replication nonetheless. Importantly, this finding demonstrates that the form of the object-relative sentences matters in terms of what results are found, something that seems to be discounted in the Waters and Caplan's (1996b) complexity manipulations.

Finally, in Set 3 we found the WCA_{REQ} interaction, but it was located in the judgment response times and in the opposite direction expected (i.e., animate subjects decreased response times). These results suggest that it was the inanimate nouns that were more ambiguous in our subject-object sentences, likely due the lack of animacy contrast with other nouns in the sentence. However, because the interaction was in the judgment times rather than in the critical areas, we cannot entirely rule out the argument that these results are due to post-interpretive processing. On the other hand, MacDonald et al. (1992) found a significant interaction in the last area of their sentences rather than in the critical area as they had predicted, suggesting that in at least some structures demanding processing may be somewhat delayed. The critical area of Set 3 sentences was followed by only a short noun phrase, making it at least possible that the interaction in the judgment response times reflects a similar sort of delay. Furthermore, the fact that the animacy effect was found only for high-spans suggests that they were able to take the animacy cue into account online when low-spans were not, and thus required less (potentially) post-interpretive processing time when items were less ambiguous (i.e., animate).

Overall, the evidence from Experiment 1 is more consistent with the single resource view than the dual resource view. However, there is evidence that post-interpretive processing does come into play when true/false comprehension judgments are used, as Waters and Caplan (1996a) argued. Moreover, it is clear from comparisons across sets that not all object-relative sentences are created equal and that factors such as word order, critical area placement, and animacy are important to consider when stimuli are developed for syntactic processing tasks.

In Experiment 2, we attempted to replicate the findings of Waters and Caplan (1996b) and Waters and Caplan (2004) using their sense judgment with the same three stimulus sets that were used in Experiment 1. Once again we used the self-paced word-by-word paradigm. We also introduced data examination methods (e.g., A', per word reading time estimations, etc.) of the previous studies in order to more directly compare our data to theirs. Results were compatible with those of Waters and Caplan (1996b; 2004) in that the significant interactions found in Experiment 1 were no longer present in Experiment 2, suggesting that use of the sense judgment was a methodological change sufficient to eradicate any relationship between the syntactic processing task and WMC measures. However, there are reservations in arguing that these data represent a ringing endorsement of dual resource theory. In particular, there is evidence that the sense judgment itself is problematic in ways that were not accounted for (or at least discussed) in the previous studies, and this is particularly true when semantic anomalies are found before critical areas due to limitations of the sentence structures (e.g., as in sentences like “The horse raced past the barn fell.”). Because of this, Waters and Caplan’s (1996a) argument that the sense judgment is superior to the comprehension judgment because it is a more online measure should be couched with the caveat that careful consideration must be applied when determining how to go about introducing semantic anomalies into sentences in the syntactic processing task and whether or not doing so changes participants’ sentence processing strategy in a desirable way (i.e., to a more interpretive one).

Finally, we also conducted a third experiment to test the grammaticality judgment using the stimuli and presentation methods from the previous two experiments. The purpose of this was to see if the grammaticality judgment might serve as an acceptable alternative to both of the previous judgment types, offering both a more online measure of syntactic processing (as opposed to the comprehension judgment) and a more direct measure of syntactic knowledge (as opposed to the semantic sense judgment). As in Experiment 2, we found no evidence of critical three-way interactions in Experiment 3. If one accepts that the grammaticality judgment is a good methodological choice for the syntactic processing task, then these results would seem to support the dual resource view of syntactic processing. However, it is not clear from the data that the grammaticality judgment was any less problematic than the sense judgment turned out to be. This is because limitations of sentence structure led to the grammatical errors being placed before or in the critical areas for Sets 2 and 3, making it unsurprising we did not find three-way interactions. In Set 1, this issue was not present, but we were still left with the issue that the items appeared to be too easy for our participants, allowing the rushed button-pressing strategy to once again rear its head and potentially disrupt any interaction in the *by* phrase.

On the whole, the results of the three experiments here seem to have accomplished their main goal. We have provided evidence that factors such as judgment type, animacy manipulations (or lack thereof), syntactic structure type, and presentation method matter when one is trying to measure the relationship, if any, between syntactic processing and measures of working memory capacity. The results of Experiment 1 are more supportive of the single resource model than the dual resource model, but this support is tentative, given the not altogether expected forms it took (e.g., being in the judgment response times rather than the critical area in Set 3) and the fact that it is unclear if the reading strategy employed by participants when the task requires true/false judgments is reflective of purely interpretive processes or if the slowed down reading strategy used when sense and grammaticality judgments are required is.

Future work exploring the predictions of the single and dual resource views should certainly explore the potential effects judgment type has on strategy, as it is not clear in these data if changes in dependent measures were due to the predictable location of the anomaly/grammatical error, or if the sense and grammaticality tasks themselves changed the way the participants processed items. If the changes are due to processing differences, then the question would be which type of processing, that done for true/false judgments or that done for sense and grammaticality judgments, is the more desirable for testing the single and dual resource theories. Furthermore, although we have not explored this issue deeply in the current work, we have touched on some evidence that the structure of filler sentences needs to be considered carefully as it may contribute to any expectancies developed in the syntactic processing task either by reinforcing things like the predictability of the location of the anomaly/grammatical error in the experimental sentences or by providing enough variability of structure that such predictability is impossible. It is also clear that future work should focus on online measures, preferably first-pass reading times measured with an eye-tracker, and should take into account things like the word order and the animacy of the subjects of complex object-relative clauses. There is an answer to which theory of syntactic processing is more accurate, but it cannot be found until appropriate methodologies are applied to control and/or measure the effects of factors that have traditionally been considered extraneous to the issue of syntactic structure.

REFERENCES

Allen, J., & Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist networks. In MacWhinney, B. (Ed.), *The emergence of language* (pp. 115-151). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Almor, A., Kempler, D., MacDonald, M. C., Anderson, E. S., & Tyler, L. K. T. (1999). Why do Alzheimer's patients have difficulty with pronouns? Working memory, semantics, and pronouns in Alzheimer's disease. *Brain & Language, 67 (3),* 202-227.

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

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

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37 (3)*, 379-384.

Balogh, J., Zurif, E. B., Prather, P., Swinney, D., & Finkel, L. (1998). Gap filling and end of sentence effects in real-time language processing: Implications for modeling sentence comprehension in aphasia. *Brain and Language, 61*, 169-182.

Berwick, R. C., Pietroski, P., Yankama, B., & Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science, 35*, 1207-1242.

Caplan, D., Baker, C. & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition, 21*, 117-175.

Caplan, D., Waters, G. S., & Dede, G. (2007). Specialized verbal working memory for language comprehension. In A. Miyake, A. R. A. Conway, C. Jarrold, M. J. Kane, & J. N. Towse (Eds.), *Variation in working memory* (pp. 272-302). New York, NY: Oxford University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1968). *Language and mind*. New York: Harcourt, Brace, Jovanovich

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

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 and Review, 12,* 769-786.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 72,* 656-669.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25,* 348-368.

Fodor, J. D. (1988). On modularity in syntactic processing. *Journal of Psycholinguistic Research, 17 (2),* 125-168.

Friederici, A. D. (1997). Diagnosis and reanalysis: Two processing aspects the brain may differentiate. In J. Fodor & F. Ferreira (Eds.). *Reanalysis in sentence processing,* (pp. 177-200). Dordrecht: Kluwer Academic.

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

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111 (2),* 228-238.

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

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology, 48 (1),* 171-184.

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

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7 (3),* 376-395.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5,* 196-214.

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85,* 1-36.

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

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *PNAS, 106 (20),* 8362-8367.

Martín-Loeches, M., Muñoz, F., Casado, P., Melcón, A., & Fernández-Frías, C. (2005). Are the anterior negativities to grammatical violations indexing working memory? *Psychophysiology, 42,* 508-519.

Masson, M. E. J., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology, 75,* 314-318.

McDonald, J. L. (2008a). Differences in cognitive demands of word order, plural, and subject-verb agreement constructions. *Psychonomic Bulletin & Review, 15 (5),* 980-984.

McDonald, J.L. (2008b). Grammaticality judgments in children: The role of age, working memory and phonological ability. *Journal of Child Language, 35,* 247-268.

Mirman, D., Magnuson, J. S., Estes, K. G., Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108,* 271-280.

Miyake, A., Carpenter, M. A., & Just P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple representations in neutral contexts. *Journal of Memory and Language, 33,* 175-202.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs.

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

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science, 1,* 125-126.

Redick, T. S., Broadway, J. M., Meier, M. E., Kiriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28 (3),* 164-171.

Rochon, E., Waters, G. S., & Caplan, D. (1994). Sentence comprehension in patients with Alzheimer's disease. *Brain & Language, 46,* 329-349.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35 (4),* 606–621.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275 (5306),* 1599–1603.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23 (4),* 569-588.

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

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection measures. *Behavior Research Methods, Instruments, & Computers, 31 (1)*, 137-149.

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General, 136 (1),* 64-81.

Talheimer, W., Cook, S. (2002). *How to calculate effect sizes from published research: A simplified methodology*. Retrieved from http://www.bwgriffin.com/gsu/coursesedur9131/content/Effect_Sizes_pdf5.pdf

Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cognitive Science, 34,* 1093-1106.

Thiessen, E. D. (2011). Domain general constraints on statistical learning. *Child Development, 82 (2),* 462-470.

Thiessen, E. D., & Erickson, L. C. (2013). Beyond word segmentation: A two-process account of statistical learning. *Current Directions in Psychological Science, 22 (3),* 239-243.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37 (3),* 498-505.

Vos, S. H., Gunter, T. C., Schriefers, H., & Friederici, A. D. (2001). Syntactic parsing and working memory: The effects of syntactic complexity, reading span, and concurrent load. *Language and Cognitive Processes, 16 (1),* 65-103.

Waters, G. S., & Caplan, D. (1996a). The capacity theory of sentence comprehension: Critique of Just & Carpenter (1992). *Psychological Review, 103 (4),* 761-772.

Waters, G. S., & Caplan, D. (1996b). Processing resource capacity and the comprehension of garden path sentences. *Memory & Cognition, 24 (3),* 342-355.

Waters, G. S., & Caplan, D. (1996c). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology, 49A (1),* 51-79.

Waters, G. S., & Caplan, D. (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *The Quarterly Journal of Experimental Psychology, 57A (1),* 129-163.

Waters, D., Caplan, G. S., Alpert, N., & Stanczak, L. (2003). Individual differences in rCBF correlates of syntactic processing in sentence comprehension: Effects of working memory and speed of processing. *NeuroImage, 19* , 101-112.

Waters, G.S., Caplan, D., & Yampolsky. (2003). On-line syntactic processing under memory load. *Psychonomic Bulletin & Review, 10 (1),* 88-95.

Yang, C. (2004). Universal grammar, statistics, or both? *Trends in Cognitive Sciences, 8, 10,* 451-456.

APPENDIX A: IRB APPROVAL

Application for Exemption from Institutional Oversight

LSU
Institutional Review Board
Dr. Robert Mathews, Chair
130 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.5983
irb@lsu.edu | lsu.edu/irb

Unless qualified as meeting the specific criteria for exemption from Institutional Review Board (IRB) oversight, ALL LSU research/projects using living humans as subjects, or samples, or data obtained from humans, directly or indirectly, with or without their consent, must be approved or exempted in advance by the LSU IRB. This Form helps the PI determine if a project may be exempted, and is used to request an exemption.

– Applicant, Please fill out the application in its entirety and include the completed application as well as parts A-F, listed below, when submitting to the IRB. Once the application is completed, please the completed application to the IRB Office or to a member of the Human Subjects Screening Committee. Members of this committee can be found at http://sites01.lsu.edu/wp/ored/human-subjects-screening-committee-members/

– A Complete Application Includes All of the Following:
(A) A copy of this completed form and a copy of parts B thru F.
(B) A brief project description (adequate to evaluate risks to subjects and to explain your responses to Parts 1&2)
(C) Copies of all instruments to be used.
*If this proposal is part of a grant proposal, include a copy of the proposal and all recruitment material.
(D) The consent form that you will use in the study (see part 3 for more information.)
(E) Certificate of Completion of Human Subjects Protection Training for all personnel involved in the project, including students who are involved with testing or handling data, unless already on file with the IRB. Training link: (http://phrp.nihtraining.com/users/login.php)
(F) IRB Security of Data Agreement: (https://sites01.lsu.edu/wp/ored/files/2013/07/Security-of-Data-Agreement.pdf)

1) Principal Investigator: Rebecca A. Horn Rank: Graduate Student

Dept: Psychology Ph: 808-449-3328 E-mail: rhorn2@lsu.edu

2) Co Investigator(s): please include department, rank, phone and e-mail for each
*If student, please identify and name supervising professor in this space
Janet L. McDonald, Ph.D., Psychology Department, 225-578-8273, psmcdo@lsu.edu

IRB# E8529 LSU Proposal # ____
☑ Complete Application
☑ Human Subjects Training
☑ IRB Security of Data Agreement

3) Project Title: Working memory capacity and Syntactic processing

STUDY EXEMPTED BY:
Dr. Robert C. Mathews, Chairman
Institutional Review Board
Louisiana State University
130 David Boyd Hall
225-578-8692 / www.lsu.edu/irb
Exemption Expires: 10/28/2016

4) Proposal? (yes or no) N If Yes, LSU Proposal Number ____

Also, if YES, either ○ This application completely matches the scope of work in the grant
OR ○ More IRB Applications will be filed later

5) Subject pool (e.g. Psychology students) Psychology students
*Circle any "vulnerable populations" to be used: (children <18; the mentally impaired, pregnant women, the ages, other). Projects with incarcerated persons cannot be exempted.

6) PI Signature ____ Date 10/19/13 (no per signatures)

** I certify my responses are accurate and complete. If the project scope or design is later changes, I will resubmit for review. I will obtain written approval from the Authorized Representative of all non-LSU institutions in which the study is conducted. I also understand that it is my responsibility to maintain copies of all consent forms at LSU for three years after completion of the study. If I leave LSU before that time the consent forms should be preserved in the Departmental Office.

Screening Committee Action: Exempted √ Not Exempted ____ Category/Paragraph 2
Signed Consent Waived?: Yes No
Reviewer Mathews Signature ____ Date 10/29/13

Consent Form *(non-eye-tracker conditions)*

1. Working memory capacity and syntactic processing
2. Performance Site: Louisiana State University
3. Investigators: The following investigators are available for questions about this study:
   Rebecca A. Horn rhorn2@lsu.edu
   Dr. Janet McDonald (225) 578-4116
4. Purpose of the Study: The purpose of this study is to examine the relationship between working memory capacity and syntactic processing.
5. Subject Inclusion: Adult speakers between 18 and 65 who do not report medical or neurological conditions and have normal or corrected to normal vision.
6. Number of Participants: 200
7. Study Procedures: Participants will complete four working memory span tasks and a sentence comprehension task, as well as a demographic questionnaire. Study time will be less than or equal to 1.5 hours.
8. Benefits: Participants will receive course credit or extra credit in exchange for participation.
9. Risks: Data will be kept secure in a location available only to the researchers and will not contain identifying information beyond a participant number.
10. Right to Refuse: Participants may choose not to participate or to withdraw from the study at any time without penalty or loss of any benefit to which they might otherwise be entitled.
11. Privacy: Results of the study may be published, but no names or identifying information will be included in the publication. Participant identification will remain confidential unless disclosure is required by law.
12. Signatures:
    The study has been discussed with me and all my questions have been answered. I may direct additional questions regarding study specifics to the investigators. If I have questions about subjects' rights or other concerns, I can contact Robert C. Mathews, Institutional Review Board, (225) 578-8692, irb@lsu.edu , www.lsu.edu/irb. I agree to participate in the study described above and acknowledge the investigators' obligation to provide me with a signed copy of this consent form.

Printed Name: ____________________

Signature: ____________________ Date: ____________

STUDY EXEMPTED BY:
Dr. Robert C. Mathews, Chairman
Institutional Review Board
Louisiana State University
130 David Boyd Hall
225-578-8692 / www.lsu.edu/irb
Exemption Expires: 10/28/2016

APPENDIX B: SET 1 STIMULUS SENTENCES

Sentence Type
Each base sentence below can be unpacked into 4 versions by choosing either the animate or inanimate subject (in purple) and by choosing the unreduced (*that was* in brackets) or reduced (omitting *that was*) forms. The critical reading area is also highlighted using **bold underlined yellow** text. Ignoring for the moment the red (Experiment 2) and green (Experiment 3) text, base sentences can thus be unpacked as follows:

The **doctor/medication** *[that was]* asked for **by the patient/toad** effectively treat**(ed)** the illness.

V1: The **doctor** that was asked for **by the patient** effectively treated the illness.
V2: The **doctor** asked for **by the patient** effectively treated the illness.
V3: The **medication** that was asked for **by the patient** effectively treated the illness.
V4: The **medication** asked for **by the patient** effectively treated the illness.

Judgment Type
In Experiment 1, each sentence was followed by a short statement that could either be true or false of the experimental sentence.

Exp 1 True Item example: The patient asked for the **doctor/medication**. True or False?
Exp 1 False Item example: The **doctor/medication** asked for the patient. True or False?

In Experiment 2 Unacceptable base sentences were modified to contain a **word** that did not make sense in the context of the rest of the sentence and were followed by the question "Makes Sense? Yes or No."

Exp 2 example: The **medication** asked for **by the toad** effectively treated the illness.

In Experiment 3 Agrammatical base sentences were modified to have an **error** of verb tense or agreement and were followed by the question "Grammatical? Yes or No."

Exp 3 example: The **medication** asked for **by the patient** effectively treat~~**(ed)**~~ the illness.

True/Acceptable/Grammatical Base Sentences
The **defendant/evidence** *[that was]* examined **by the lawyer** turned out to be unreliable.
The **visitor/ship** *[that was]* sighted **by the lookout** probably brought bad news.
The **child/letter** *[that was]* left **at the store** told a disturbing tale.
The **owner/bill** *[that was]* paid **by the tenant** had been owed for weeks.
The **singer/song** *[that was]* listened to **by the audience** was very beautiful.
The **man/car** *[that was]* reported **by mall security** had run over a traffic cone.
The **woman/sign** *[that was]* painted **by the artist** was very attractive to look at.
The **child/trash** *[that was]* sniffed **by the dog** was sitting on the curb.
The **player/name** *[that was]* dropped **by the team** had long offended many people.
The **worker/laundry** *[that was]* sprayed **by the hose** took a long time to dry out.
The **woman/trees** *[that was/were]* expected **by the gardeners** did not arrive on time.
The **model/dress** *[that was]* promoted **by the designer** was included in the fashion show.
The **man/ATM** *[that was]* robbed **by the thieves** actually had very little cash.
The **hiker/compass** *[that was]* found **by the gamekeeper** had been lost for weeks.
The **guest/gift** *[that was]* delivered **by the chauffer** delighted the host and hostess.
The **author/book** *[that was]* read **by the student** was very hard to understand.
The **player/ball** *[that was]* hit **by the bat** fell swiftly to the ground.

The **witness/report** *[that was]* dismissed **by the judge** contradicted the established facts.
The **artist/painting** *[that was]* admired **by the curator** caused an uproar at the auction.
The **diver/boat** *[that was]* rescued **by the Coast Guard** arrived in port.

False/Unacceptable/Agrammatical Items

The **doctor/medication** *[that was]* asked for **by the patient/toad** effectively treat**(ed)** the illness.
The **volunteer/donation** *[that was]* accepted **by the charity/waterfall** help**(ed)** at-risk youths.
The **tourist/rock** *[that was]* observed **by the geologist/skate** did**/do** not belong in the canyon.
The **chef/apple** *[that was]* cut **by the knife/balloon** was**/were** covered in sticky caramel.
The **man/meal** *[that was]* brought **by the nurse/loofah** was**/were** not very appealing.
The **spokesperson/policy** *[that was]* voted for **by the managers/desks were** greatly increase(ed**/ing)** sales.
The **executive/deal** *[that was]* exposed **by the whistleblower/moat** brought**/bring** down the company.
The **traveler/fruit** *[that was]* inspected **by Customs/banana** was**/were** not allowed into the country.
The **villager/shipment** *[that was]* seized **by the military/tissue** was**/were** not involved in the rebellion.
The **suspect/statement** *[that was]* released **by the police/porcupine** frighten**(ed)** many of the citizens.
The **children/packages** *[that were]* collected **by the social worker/brick** had**/has** been handled harshly.
The **parasailer/umbrella** *[that was]* carried **by the strong wind/lid** finally land**(ed)** on the beach.
The **wife/house** *[that was]* cared for **by the mother-in-law/dogwood** steadily improve**(d)** in appearance.
The **gunman/command** *[that was]* executed **by the soldier/toothpick** cause**(d)** several civilian fatalities.
The **reporter/message** *[that was]* recorded **by the tape recorder/coconut** could not be**/being** understood.
The **official/program** *[that was]* rewarded **by the state/skunk** had**/have** high status in the community.
The **farmers/towns** *[that was/were]* connected **by the new road/sponge** quickly became**/becomes** quite friendly.
The **philanthropist/project** *[that was]* funded **by the foundation/chandelier** made**/make** vaccines widely available.
The **pedestrian/concert** *[that was]* followed **by the mime/potato were** amus(ed**/ing)** the gathered spectators.
The **customer/merchandise** *[that was]*monitored **by the security guards/t-shirts** disappear**(ed)** without a trace.

APPENDIX C: SET 2 STIMULUS SENTENCES

Sentence Type

Each base sentence below can be unpacked into 4 versions by choosing either the animate or inanimate subject (in purple) and by choosing the unreduced (*that was* in brackets) or reduced (omitting *that was*) forms. The critical reading area is also highlighted using **bold underlined yellow** text. Ignoring for the moment the red (Experiment 2) and green (Experiment 3) text, base sentences can thus be unpacked as follows:

The **mechanic/branch** [*that was*] bent over the engine/knee/**stadium** **collapse(d)/crack(ed)**.

V1: The **mechanic** that was bent over the engine **collapsed**.

V2: The **mechanic** bent over the engine **collapsed**.

V3: The **branch** that was bent over the knee **cracked**.

V4: The **branch** bent over the knee **cracked**.

Judgment Type

In Experiment 1, each sentence was followed by a short statement that could either be true or false of the experimental sentence.

Exp 1 True Item example: The **branch** was bent over the knee. True or False?

Exp 1 False Item example: The **branch** did bend over the knee. True or False?

In Experiment 2 Unacceptable base sentences were modified to contain a **word** that did not make sense in the context of the rest of the sentence and were followed by the question "Makes Sense? Yes or No."

Exp 2 example: The **branch** bent over the **stadium** **cracked**.

In Experiment 3 Agrammatical base sentences were modified to have an **error** of verb tense and were followed by the question "Grammatical? Yes or No."

Exp 3 example: The **branch** bent over the knee **crack~~(ed)~~**.

True/Grammatical/Acceptable Items

The **staff/pudding** [*that was*] served in the cafeteria **ate/spoiled**.
The **reporter/crime** [*that was*]exposed in the newspaper **retired/shocked**.
The **man/ticket** [*that was*] expected to win the raffle **lost**.
The **coach/drink** [*that was*] tested for illegal drugs **fled/spilled**.
The **baby/toy** [*that was*] grabbed with both hands **squealed/squeaked**.
The **contestant/garden** [*that was*] registered for the competition **won**.
The **man/suitcase** [*that was*] searched for the hidden valuables **vanished**.
The **boy/field** [*that was*] measured with a yard stick **laughed/flourished**.
The **athlete/weight** [*that was*] lifted during the competition **waved/cracked**.
The **model/mannequin** [*that was*] posed under the lights **sweated/melted**.
The **resident/café** [*that was*] evacuated during the storm **returned/reopened**.
The **friend/question** [*that was*] answered in the letter **visited/rankled**.
The **spectators/clock** [*that were/was*] watched during the game **cheered/stopped**.
The **fireman/fire** [*that was*] hosed down with water **yelled/smoldered**.
The **patient/plane** [*that was*] recovered on the island **lived/rusted**.
The **camper/tree** [*that was*] chopped with the ax **bled/split**.
The **officer/detail** [*that was*] discovered during the investigation **cooperated/helped**.

The **candidate/bill** [*that was*] debated on television **campaigned/passed**.
The **nurse/bed** [*that was*] disinfected with the alcohol **stank**.
The **bystander/problem** [*that was*] noticed during the contest **left/resolved**.
EXTRA ANIMATE REDUCED OBJECT-RELATIVES
The **horse** raced past the barn **fell**.
The **knigh**t trained in the tent **collapsed**.
The **doct**or prescribed the medication **healed**.
The **seamstress** measured for the dress **twirled**.
The **terrorist** attacked in the market **exploded**.

False/Agrammatical/Unacceptable Items
The **mechanic/branch** [*that was*] bent over the engine/knee/**stadium collapse(d)/crack(ed)**.
The **retailer/ring** [*that was*] advertised in the magazine/**pineapple default(ed)/tarnish(ed)**.
The **prisoner/banner** [*that was*] paraded down the street/**outlet stumble(d)/rip(ped)**.
The **surfer/hair** [*that was*] rinsed in hot water/**bacon dried/dry**.
The **trucker/produce** [*that was*] headed to the warehouse/**omelet crash(ed)/spoil(ed)**.
The **acrobat/saucer** [*that was*] balanced on the ledge/**feather wobble(d)**.
The **golfer/secret** [*that was*] admitted to the club/**salmon play(ed)**/**were spreading**.
The **assistant/news** [*that was*] shared with the partner/**teapot were quitting/surprise(d)**.
The **brother/orchard** [*that was*] visited on the holiday/**nickel smile(d)/bloom(ed)**.
The **passengers/laundry** [*that were/was*] piled in the car/basket/**ear laugh(ed)/mildew(ed)**.
The **teacher/assignment** [*that was*] corrected in class/**chocolate blush(ed)/count(ed)**.
The **dancer/file** [*that was*] copied in the studio/office/**glove taught/teach/vanish(ed)**.
The **vendor/shipment** [*that was*] guaranteed punctual delivery/**curling move(d)/arrive(d)**.
The **pilot/crops** [*that was/were*] dusted with the fertilizer/**artichoke shower(ed)/grew(grows)**.
The **surgeon/bread** [*that was*] sliced with the knife/**pillow faint(ed)/steam(ed)**.
The **shoplifter/computer** [*that was*] returned to the store/**cake confess(ed)/overheat(ed)**.
The **passenger/package** [*that was*] hurried to the airport/**spinach complain(ed)./ship(ped)**.
The **spy/compound** [*that was*] surveilled from the van/**bulb grimace(d)/bustle(d)**.
The **protestors/clothes** [*that were/was*] agitated in the streets/washer/**burrow fought/fights/twist(ed)**.
The **trainer/resources** [*that was/were*] recruited for the project/**paperclip accept(d)/materialize(d)**.
EXTRA ANIMATE REDUCED OBJECT-RELATIVES
The **boyfriend** wooed with flowers **smell(ed)**.
The **psychiatris**t counseled at the new practice **recover(ed)**.
The **victim** reported to the police **lie(d)**.
The **bartender** served the drinks **spill(ed)**.
The **sniper** killed with the rifle **decay(ed)**.

APPENDIX D: SET 3 STIMULUS SENTENCES

Sentence Type

Each base sentence below can be unpacked into 4 versions by choosing either the animate or inanimate subject (in purple) and by choosing the unreduced (*that was* in brackets) or reduced (omitting *that was*) forms. The critical reading area is also highlighted using **bold underlined yellow** text. Ignoring for the moment the red (Experiment 2) and green (Experiment 3) text, base sentences can thus be unpacked as follows:

The towel *[that]* the **housekeeper/bleach** ruined **dried**/**dry** the spill/**lake**.

V1: The towel that the **housekeeper** ruined **dried** the spill.

V2: The towel the **housekeeper** ruined **dried** the spill.

V3: The towel that the **bleach** ruined **dried** the spill.

V4: The towel the **bleach** ruined **dried** the spill.

Judgment Type

In Experiment 1, each sentence was followed by a short statement that could either be true or false of the experimental sentence.

Exp 1 True Item example: The **bleach** ruined the towel. True or False?

Exp 1 False Item example: The towel ruined the **bleach**. True or False?

In Experiment 2 Unacceptable base sentences were modified to contain a **word** that did not make sense in the context of the rest of the sentence and were followed by the question "Makes Sense? Yes or No."

Exp 2 example: The towel the **bleach** ruined **dried** the **lake**.

In Experiment 3 Agrammatical base sentences were modified to have an **error** of verb tense and were followed by the question "Grammatical? Yes or No."

Exp 3 example: The towel the **bleach** ruined **dry** the spill.

True/Grammatical/Acceptable Items

The car *[that]* the **mechanic/trailer** transported **passed** the inspection.
The law *[that]* the **millionaire/courts** favored **frustrated** the workers.
The picture *[that]* the **illustrator/advertisement** included **depicted** the product.
The tree *[that]* the **activist/ordinance** protected **sheltered** the squirrel.
The fort *[that]* the **soldiers/cannons** assaulted **stored** the ammunition.
The boat *[that]* the **captain/barge** towed **won** the race.
The computer *[that]* the **scientists/experiment** used **recorded** the data.
The article *[that]* the **editor/magazine** published **satisfied** the author.
The door *[that]* the **worker/wind** closed **connected** the bedrooms.
The song *[that]* the **musician/radio** played **delighted** the patrons.
The luggage *[that]* the **tourist/airline** lost **contained** the souvenir.
The curtain *[that]* the **decorator/rod** held **filtered** the light.
The table *[that]* the **woman/cloth** covered **held** the vase.
The waste *[that]* the **man/truck** hauled **polluted** the landfill.
The painting *[that]* the **delinquent/graffiti** ruined **depicted** the president.
The statue *[that]* the **auctioneer/store** sold **pleased** the collector.
The rug *[that]* the **child/ink** stained **covered** the floor.

The yard *[that]* the **police/fence** surrounded **contained** toxic materials.
The museum *[that]* the **foreman/earthquake** demolished **held** the artefacts.
The screw *[that]* the **contractor/drill** removed **connected** the pieces.

False/Agrammatical/Unacceptable Items
The towel *[that]* the **housekeeper/bleach** ruined **dried/dry** the spill/**lake**.
The lock *[that]* the **thief/key** opened **secure(d)** the safe/**tangerine**.
The ball *[that]* the **player/racquet** hit **flew/fly** over the fence/**Pacific**.
The boulder *[that]* the **man/ice** cracked **block(ed)** the road/**pencil**.
The pan *[that]* the **chef/burner** heated **contain(ed)** the pasta/**farm**.
The fire *[that]* the **arsonist/match** ignited **burn(ed)** the forest/**rain**.
The dirt *[that]* the **maid/filter** missed **clog(ged)** the drain/**corn**.
The ribbon *[that]* the **crafter/scissors** cut **decorate(d)** the box/**pothole**.
The report *[that]* the **manager/county** filed **influence(d)** the decision/**camel**.
The novel *[that]* the **librarian/café** provided **inspire(d)** the student/ratio.
The lens *[that]* the **photographer/telescope** required **magnif(ied)** the view/**dignity**.
The purchase *[that]* the **mayor/budget** allowed **benefit(ed)** the community/**trampoline**.
The treatment *[that]* the **therapist/manual** recommend **cure(d)** the patient/**pebble**.
The button *[that]* the **soldier/lever** touched **launch(ed)** the missile/**Band-Aid**.
The glass *[that]* the **bully/rock** shattered **endanger(ed)** the child/**skateboard**.
The document *[that]* the **employee/machine** copied **shock(ed)** the public/**cassette**.
The correspondence *[that]* the **spy/server** saved **implicate(d)** the CEO/**cookie**.
The water *[that]* the **prankster/hose** sprayed **soak(ed)** the furniture/**river**.
The butter *[that]* the **girl/microwave** melted **flavor(ed)**. the popcorn/**mirror**.
The product *[that]* the **actress/commercial** promoted **dominate(d)** the market/**turtle**.

VITA

Rebecca A. Horn grew up in southern Wisconsin. She spent her junior year of high school studying abroad before graduating in 1999. After graduation she attended classes at the University of Wisconsin Rock County for two years. In the spring of 2004 she began work on her B.A. in Psychology with a minor in Philosophy at the University of Wisconsin Whitewater and graduated summa cum laude in December of 2005. She earned her M.S. in Experimental Psychology from the University of Wisconsin Oshkosh in 2009 and her Ph.D. in Cognitive/Developmental Psychology from Louisiana State University in 2015. For professional purposes (i.e., conference presentations and paper submissions), she has also been known as Rebecca Hammarlund since 2009.