Inference and Context: Individual Differences in Interpretation and Integration

By
Mary Michael

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology.

Chapel Hill
2007

Approved by:

Advisor: Professor Peter C. Gordon
Reader: Professor Jennifer Arnold
Reader: Professor Kelly Giovanello
Reader: Professor Neil Mulligan
Reader: Professor Abigail Panter
ABSTRACT
MARY ANN MICHAEL
Inference and Context: Individual Differences in Integration and Interpretation
Under the direction of Peter C. Gordon

Individual differences in interpretation of ambiguous fronted adjunct sentences were correlated with measures of cognitive ability and reading experience in two experiments. The initial ambiguity was either consistent with a plausible inference or not. Correct interpretation of the inconsistent sentences was found to be correlated with a number of different cognitive components. Responses following sentences for which the inference was consistent showed few correlations with any of the cognitive components measured. In the second experiment, prior context was either supportive of the inference or neutral; passages that required bridging inferences were also included for comparison. Again there were few correlations with response. However, there were correlations of reading times with RTs from the lateralized lexical decision task and with a measure of reading experience. Differences by visual field on the lexical task were found to correlate with RTs and accuracy on the task. Relationships between the measures and reading and response were explored using structural equation modeling.
ACKNOWLEDGEMENTS

Many people deserve to be acknowledged for the help that they have extended to make completion of this dissertation possible. However, I mainly wish to acknowledge the help and support of my husband, Martin Michael.
TABLE OF CONTENTS

| LIST OF TABLES | xiii |
| LIST OF FIGURES | xiv |

CHAPTER

1. INTRODUCTION ...................................................... 1

   BACKGROUND .......................................................... 3

   A. Studies of integration ............................................. 4
      1. Context ............................................................ 5
      2. Inference .......................................................... 6
      3. Priming ............................................................ 8
      4. Integration ......................................................... 10

   B. Individual differences ........................................... 12
      1. Lexical Access .................................................... 12
      2. Vocabulary ....................................................... 15
      3. Working Memory ................................................ 16

   C. Methodology .......................................................... 18
      1. Reading time measures ......................................... 19
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Responses and reading times to garden-path sentences</td>
<td>41</td>
</tr>
<tr>
<td>2. Responses and reading times to sentences with prior context</td>
<td>42</td>
</tr>
<tr>
<td>3. Responses and reading times to sentences in Experiment 1</td>
<td>56</td>
</tr>
<tr>
<td>4. Descriptive statistics on component measures in Experiment 1</td>
<td>57</td>
</tr>
<tr>
<td>5. Differences in accuracy and response time by side of presentation</td>
<td>58</td>
</tr>
<tr>
<td>6. Correlations with response variables for Experiment 1</td>
<td>60</td>
</tr>
<tr>
<td>7. Correlations with reading times for Experiment 1</td>
<td>62</td>
</tr>
<tr>
<td>8. Response accuracy and reading times in Experiment 2</td>
<td>79</td>
</tr>
<tr>
<td>9. Descriptive statistics for component measures in Experiment 2</td>
<td>80</td>
</tr>
<tr>
<td>10. Effects of difference in error rates and RTs in the lateral LDT</td>
<td>81</td>
</tr>
<tr>
<td>11. Correlations with reading times for Experiment 2</td>
<td>83</td>
</tr>
<tr>
<td>12. Correlations between tasks in Experiment 2</td>
<td>85</td>
</tr>
<tr>
<td>13. Response accuracy for different passage types by cluster</td>
<td>87</td>
</tr>
<tr>
<td>14. Fit of Quadratic and Linear Models for Lateralized lexical decision task</td>
<td>136</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Proposed model showing how measures of individual difference represent the processes involved in reading comprehension</td>
<td>101</td>
</tr>
<tr>
<td>2. Plot of difference in reaction time and reaction time to nonwords in the lateralized lexical decision task for Experiment 1 showing quadratic model</td>
<td>102</td>
</tr>
<tr>
<td>3. Measurement model for Experiment 1 showing factor loadings and correlations between latent variables</td>
<td>103</td>
</tr>
<tr>
<td>4. Structural equation model for Experiment 1 responses to GP1 and filler sentences showing standardized regression weights on latent variables</td>
<td>104</td>
</tr>
<tr>
<td>5. Structural equation model for Experiment 1 sentence reading times showing standardized regression weights on latent variables</td>
<td>105</td>
</tr>
<tr>
<td>6. Plots of word RT difference and word RT in the LLDT for Experiment 1 and 2</td>
<td>106</td>
</tr>
<tr>
<td>7. Plot of pooled RTs and RT difference to words in the lateralized lexical decision task, for Experiment 1 and 2</td>
<td>107</td>
</tr>
<tr>
<td>8. Measurement model for Experiment 2 showing factor loadings and correlations between latent variables</td>
<td>108</td>
</tr>
<tr>
<td>9. Structural equation model for Experiment 2 responses showing factor loadings on latent variables</td>
<td>109</td>
</tr>
<tr>
<td>10. Structural equation model for Experiment 2 reading times to target sentences showing factor loadings between latent variables</td>
<td>110</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Research into the processes involved in sentence reading generally assumes that comprehension and interpretation vary little amongst skilled adult readers. While this is generally true for early reading processes such as word recognition, which become increasingly automatic as readers become more experienced, later interpretive processes show greater variability even among “normal” adult readers (Michael, 2004). These differences of interpretation increase when comprehension of the sentence requires the use of inference or integration of information, either from earlier in the text or from external sources. Readers are known to be inconsistent in the way they use contextual information and also in how they apply inference: they may focus either on what the words mean or on what is meant (Benjamin, 1968).

Interpretation of a text relies on integrating ideas, initially at the superficial text-based and proposition levels to produce a locally coherent representation, then at the discourse level to create a globally coherent discourse representation (Kintsch, 1998). This later level may include higher-level interpretive processes that make use of contextual information and inference. While most adult readers can perform the first stage of integration adequately,
although not always completely (Hanna and Daneman, 2004), there are several points at which the later processes of integration and interpretation may break down: initial encoding of relevant information may be inadequate, retrieval or reactivation of that information may fail, or, despite completing the first two processes, readers may fail to integrate that earlier information with the recently read text into a fully comprehensive discourse representation. Recent research (Long and Chong, 2001, and Noordman and Vonk, 1997) has tended to suggest that problems, when they occur, arise specifically at this later stage of integration, rather than at encoding or retrieval, and the idea of an integration bottleneck has emerged (Oakhill, 1993). Yuill and Oakhill (1991) have suggested that problems faced by children who are poor comprehenders relate to problems in integration. However, research with children by Spooner, Gathercole, and Baddeley (2006) found deficits in memory for text by less-skilled comprehenders but failed to find evidence of differences in the ability to integrate.

When going beyond the literal text to make inferences about the intended interpretation readers of lesser ability may again fail to make full use of all the information available. Deciding which of a range of possible interpretations is the most appropriate requires rapid suppression of those ideas that are irrelevant. Every word we read has associated with it a flood of information that potentially could be activated. Most of this deluge never reaches consciousness. However, problems can arise when sufficiently strong associations cause irrelevant information to interfere with comprehension.
Research suggests that readers of different ability levels differ in the strategies they use to cope with this problem (Stanovich, 1980): able readers may selectively attend to relevant information, while less able readers fail to inhibit irrelevant information (Engle, Conway, Tuholski, & Shisler, 1995).

This dissertation examines how later processes of inference and integration vary between competent adult readers and how these differences may relate to individual differences in cognition. Sentences used contained a fronted adjunct ambiguity providing an opportunity to examine the ways in which understanding varies when alternate interpretations are possible. The emphasis is on exploring how differences in these later processes influence comprehension and thus limit the extent to which less able readers are able to go beyond the literal meaning to understand fully what the writer intends. The possible effects of differences at earlier levels of processing on reading and interpretation are also examined. Although I use the terms early and late to describe the processes involved, no strong claims about the temporal sequence are being made, rather early and late refer to lower and higher levels of language representation.

Background

Research into processes of inference and integration during language comprehension has examined the ways in which the interpretation of a target sentence can be impacted by prior information, focusing on three main areas:
inconsistency effects, priming, and integration. A small number of the studies that have explored these issues in reading comprehension have also considered the effect of individual differences. Clearly there are large individual differences in reading ability in the normal adult population. Inferential processes appear to be particularly susceptible to individual differences in reading ability (Murray and Burke, 2003). Differences have also been found in the ability to integrate prior context and in the use of contextual information to support the use of inference (Calvo, 2001).

I discuss the literature describing research into the areas mentioned: first studies of integration, second, the possible role of individual differences in comprehension, and finally, issues of methodology.

A. Studies of integration

Studies using inconsistencies in text provide evidence of how readers process text and integrate new information with old to create a coherent discourse representation. Inconsistencies may take the form of the “Moses illusion” (Erickson and Mattson, 1981) in which participants answer “two”, when asked “How many animals of each kind did Moses take on the ark?”, without apparently noticing the inappropriate name of the main character. Alternatively earlier information may be inconsistent with information in a target sentence; for example, a protagonist being a vegetarian is inconsistent with a later description of her eating a burger (Albrecht & O’Brien, 1993, Albrecht & Myers, 1995). Reading of a target word or sentence is slower when the latest input is
inconsistent with earlier information that has become part of the discourse representation.

1. Context

Albrecht, O’Brien and colleagues (Albrecht & O’Brien, 1993, O’Brien, Rizzella, Albrecht, & Halleran, 1998) performed a series of experiments that explored the way in which distant information is accessed by the reader and the factors that facilitate or limit that process. The main focus of this research was on the elements that contributed to the discourse representation constructed by the reader and how well these were maintained over time and in the presence of contradictory information. For instance, in the case of the burger-eating vegetarian did information that resolved the inconsistency, such as that she occasionally ate fast food when out with friends, reduce the effect or is activation of prior information automatic? Myers and O’Brien (1998) proposed that earlier information “resonated” with the reader but the effects of this automatic activation could be modulated by a number of different factors that would determine how far slowing occurred. For instance, both intervening information and the nature of the inconsistency can influence the size of the effects on reading times.

Various factors, such as focus and fit, have been found to influence the activation of prior information. Readers are more likely to notice inconsistency if the prior information is in focus (Sanford, Moxey, & Patterson, 1996) and are much less likely to notice the error if the discrepancy is in other respects.
felicitous, as in the use of an appropriately biblical name instead of Noah to complete the “Moses illusion”. Thus, if the vegetarian eats a tuna salad rather than a burger the effect of the inconsistency is less marked, presumably because it seems less inconsistent with being vegetarian. In a series of experiments, Kaup and Foss (2005) explored how readers dealt with different types of inconsistency, relational versus direct, as measured by reading times and detection rates. Contrary to their expectation relational inconsistencies, such as differences of size – smaller rather than larger, caused more slowing and were detected more frequently than direct inconsistencies, for instance of the color of an item mentioned. Only when participants were directly warned of inconsistencies did the detection rate for the direct inconsistencies of color equal those for the relational inconsistencies.

When readers have been pretested for reading ability (Long & Chong, 2001, Murray & Burke, 2003), more able readers showed significant slowing in the inconsistent condition, however less able readers showed no effects of inconsistency, so that reading times for the target sentence did not differ significantly by condition. In the study by Long & Chong (2001), participants read passages with target sentences that were locally coherent, but global consistency with details of character provided earlier in the passage was varied. Ken is either a small man who dislikes contact sports or a large man who loves them. This leads to the apparently contradictory finding that the target sentence *Ken decided to enroll in boxing classes* is read more slowly by able readers when the
target was inconsistent with the character description; this effect was not found for less able readers. Thus it appears that more able readers are aware of the inconsistency and affected by it while less able readers are not.

2. Inference

Research exploring the use of context frequently assumes the use of inference on the part of the reader, about character goals (McKoon and Ratcliff, 1992), causal relations (Rizzella, & O'Brien, 1996), and also on temporal or spatial relations. Studies using inconsistency have examined how inferences about a character’s goals affect reading of a target sentence when information is inconsistent with those inferences. Longer reading times have been found for textual inconsistencies both with implicit and explicit information. As with explicit information in the prior context, these effects of implicit information have been found to be modulated by factors such as focus and fit.

The ability to make inferences is known to be an important part of good reading comprehension. Inferential processes appear to be particularly susceptible to individual differences. Less able readers have more problems with implicit than with explicit information (Oakhill, 1984, Hare, Rabinowitz, & Schieble, 1989). This limits their comprehension, since making inferences from text has been shown to play an important role in understanding (Sanford, 1990; McKoon & Ratcliff, 1992; Graesser, Singer, & Trabasso, 1994). Although these differences are apparent mainly for those inferences that occur late as part of higher-level interpretive processes, it may be that earlier apparently “automatic”
inferences are also vulnerable. Data from my recent research (Michael, 2004) show that plausible inferences, which might be expected to be made equally by all participants, were made unevenly. Although there may be a normative, or majority, response, there was considerable variability in how sentences were interpreted. In recent research, Murray and Burke (2003) found college-level students, pretested for reading ability, used inference unevenly, with a small proportion of less able readers applying inference hardly at all while some of the others employed it indiscriminately. Is it the ability to make inferences or to decide which inferences are appropriate that is limited in poor comprehenders or are these separable?

Studies that explored individual differences in inferential processing have used a number of different tests to separate groups of different ability. (I discuss details of the tests used below.) Results show that more able readers are more likely to show evidence of the use of inference in the form of slowing, as described above; however when the level of inference is comparable between groups (Calvo, 2001), because the inference is clear and necessary for comprehension, more able readers read more quickly and show less need for rereading.

3. Priming

The extent to which prior information effectively primes the reader for new information has generally been tested using a probe task. Priming paradigms use reaction times to probe words in a lexical decision task to explore
the extent to which ideas in the text are activated at different points in the text. Reading is interrupted when a word is flashed on to the screen for recognition. Words that have been effectively primed by the prior context should result in more rapid response rates. Priming shows the extent to which information has been activated or remains active at a given point in the text. Results from both Long & Chong (2001) and Murray & Burke (2003) suggest that disparities between readers of different abilities lie less in memory capacity than in processing. While readers who are poor comprehenders encode information into memory, as demonstrated by equivalent priming, they fail to make appropriate use of that reactivated material by integrating it into their current discourse representation.

Priming has tended to be considered as evidence of activation. By contrast, negative priming may provide evidence of inhibition, the suppression of irrelevant information (Hasher & Zacks, 1988), and might also be valuable in explaining individual differences in integration. Perhaps the reason why some readers fail to integrate earlier information effectively is that they reactivate indiscriminately. However, negative priming is not necessarily considered to be a reliable measure of interference but rather as evidence of selective attention or an active process of inhibition. Controlling proactive interference has been shown to be costly (Kane & Engle, 2002) and thus to rely on active rather than passive processing.
Inhibition has been suggested as an explanatory factor in individual differences in cognition. The ability to inhibit irrelevant information is correlated with measures on the reading span task, designed to measure working memory (WM), (Engle, et al. 1995, Engle, Tuholski, Laughlin & Conway, 1999). Engle and his colleagues found negative priming for information that was task irrelevant by high-span participants, demonstrated by longer response times, suggesting that such inhibition requires active, resource dependent, controlled attention, rather than a passive process of inhibition. The idea that differences lie in selective attention fits with work by Egner and Hirsch (2005), which suggests irrelevant information is encoded into memory, and the mechanism that causes negative priming is one of attention to relevant information following episodic retrieval, rather than inhibition of irrelevant information. In support of this interpretation, negative priming was found to be associated with increased activation of the right dorsolateral prefrontal cortex, a region linked to episodic memory retrieval. Research using ERPs (Perfetti, Wlotko, & Hart (2005) also suggests the role of episodic memory in enhancing semantic memory.

4. Integration¹

Experiments that directly explore later processes of integration and interpretation are few. Vonk and Noordman (1990, Noordman, Vonk, & Kempft, 1992) explored processes of necessary inference, requiring the integration of prior

¹ I here distinguish integration, where all the information is available to the reader, from inference, where the reader is required to go beyond the text. A “necessary inference” as the term is used by Vonk and Noordman requires integration of all the relevant information given in the text.
information. They suggest readers are quite “parsimonious” (p.462) when making such inferences from information in the text, balancing costs and benefits, and only make those necessary inferences that are either easy or which match their reading goals. Efforts at integrating information are, they suggest, related to the interest or knowledge that a reader has in the subject. In a later study using eye-tracking (Noordman and Vonk, 1997), they investigated the time course of late integration processes in sentences requiring complex inferences following the conjunction because. When the required analytical inference is made online, slowing occurs in processing the final words of the sentence, apparently as part of later sentence wrap up, rather than as part of constructing the earlier more superficial text and propositional-based levels of representation.

There is considerable evidence from a number of sources that while most readers perform the early more automatic processes of comprehension equivalently, later interpretive processes show more variation (Murray & Burke, 2003). (However, it is possible that the early processes are also more variable than generally assumed.) It is important to understand at what stage in processing these differences occur and what specific aspect of “reading ability” is causing those differences. This dissertation seeks to clarify how effectively inferential processes and prior information are employed by readers of different abilities, and which of the specific components capable of influencing reading ability relate to these processing difficulties. The next section discusses those aspects of difference that have been implicated by previous research.
B. Individual Differences

Research into individual differences in reading comprehension has focused on several possible sources of disparity. Baddeley, Logie, Nimmo-Smith, & Brereton (1985) found that reading comprehension depends on a number of “separable components”, particularly lexical access, vocabulary and WM. Recently these were evaluated in a sentence reading paradigm using eye-tracking (Calvo, 2001, 2005) and found to correlate with differences in processing at early and late stages of sentence reading. In addition to these three essential functional components, other contributing factors, such as reading experience (Cunningham, Stanovich, & Wilson, 1990) have been considered, and found to be predictive of differences found between normal readers.

1. Lexical access

   Lexical access is the speed and accuracy with which a reader identifies a word. There is a significant relationship between word recognition and reading ability. The lexical quality hypothesis (Perfetti & Hart, 2001) focuses on individual differences in early processing and suggests that the way in which individual readers form lexical representations has a direct effect on early comprehension processes and is predictive of an individual’s comprehension of text. They point to differences in early ERP components that differ between skilled and less skilled readers as they identify words. Although other researchers into children’s comprehension have tended to separate early processing ability from comprehension (Yuill and Oakland, 1991), it is
indisputable that lexical access is required for comprehension and consequently that early processes are critical to later processes. Early phonological awareness is known to be highly predictive of later reading success (Nation and Snowling, 2004). However, as reading skills develop, and word recognition becomes more automatic, skilled readers rely increasingly on the orthographic route to recognition. Also, the semantic route, which relies in part on contextual clues, becomes increasingly important to understanding of less familiar words (Nash & Snowling, 2006). Differences may arise between skilled and less able readers in their ability to recognize words, in particular in their ability to recognize irregular or exception words that are not rule-based, as the semantic route becomes increasingly important to reading success. Studies into the sources of individual difference in reading using the lexical decision task (described below) have found that skilled readers are faster and more accurate in identifying low frequency words (Schilling, Rayner, & Chumbley, 1998; Lewellen, Goldinger, Pisoni, & Greene, 1993) and that students who are infrequent readers are slower and less efficient in all components of the task, (Lewellen, Goldinger, Pisoni, & Greene 1993, Chateau, & Jared, 2000).

One aspect of lexical access that has recently been given renewed attention is lateralization. While it is well known that the role of the left hemisphere is primary for most language processes, it has become increasingly apparent that the role of the right hemisphere is also important. Engaging the right brain in language processing is required for processing speaker/writer intention,
particularly in all those subtle interpretive processes that help in deciphering humor, irony, indirect requests, and emotional subtext. At its most extreme, failure to engage the right hemisphere in language processing may be related to psychoses, specifically those relating to schizophrenic disorders (Mitchell and Crow, 1995). Damage to the right hemisphere (RHD), arising from neurological insult such as stroke, frequently results in deficits at the discourse level in interpreting inference and contextual information, as well as broad themes and concepts (Beeman, 1993, Myers and Brookeshire, 1996). RHD is associated with inappropriate use of inference and the inability to make use of context (Bryan, 1988), as well as with difficulty in recognizing appropriate use of inference (Schneiderman, Murasugi, Saddy, 1992).²

In the non-patient population, readers who comprehend well may be more fully/rapidly engaging the right hemisphere in processing text in order to maximize their understanding and resolve areas of ambiguity (Faust & Chiarello, 1998). The right middle temporal region in particular may be important in establishing/maintaining global coherence in discourse (St George, Kutas, Martinez, & Sereno, 1999). It has been suggested that good slow readers have learned to do more consciously what good fast readers do more or less automatically; thus they require longer reading times, specifically more rereading, to reach a similar interpretation. This implies a more conscious activation of those right brain resources.

² Specifically RHD patients are unable to make use of thematic information.
More pronounced lateralization of the processes required to understand language might be an important factor in the difference between good and poor comprehenders. If this is the case then, for instance, the two stages of integration may be carried out largely in opposite hemispheres i.e. with good separation. Alternatively, it may be that the important issue is the involvement of and cooperation between the hemispheres; Brysbaert (1994, 2004) suggests that, in many instances, interhemispheric transfer is essential to word recognition. Differences in how language input is processed between the hemispheres could, potentially, be of interest.

2. Vocabulary

Readers who learn to read early and well continue to improve, particularly in the rate at which they acquire new vocabulary (both understood and produced), at least in part because reading a variety of text gives access to a wider range of unfamiliar words than that obtained through speech (Cunningham & Stanovich, 1997). Thus scores on vocabulary tests suggest the level of reading experience. Given this relationship, it is unsurprising to find that measures of vocabulary should be strongly correlated with measures of reading comprehension. Calvo (2005) found vocabulary to be correlated with reading time measures that reflect later interpretive processes. Some correlation between vocabulary and lexical access might also be expected, but in practice this relationship is not strong, since ideally such tests tap different types of processing; Nation and Snowling (2004) considered a measure of expressive
vocabulary at age 8 as a predictor of both word recognition and reading comprehension at age 13 and found that, while vocabulary accounted for a significant portion of the unique variance, scores on phonological processing proved a better predictor of later reading success.

3. Working memory

Tests intended to provide a measure of WM capacity and processing also correlate with measures of reading comprehension. Researchers have found that individual differences in working memory demonstrated by these tests are correlated with the ability to selectively attend to relevant information (Engle et al, 1995, 1999), suggesting that attention may be important component of WM measures. As described above, Engle et al (1995, 1999) found that individuals with a high span score on tests of WM showed negative priming of task irrelevant information, suggesting the use of controlled attention rather than a more or less automatic process of inhibition by which irrelevant information is suppressed. Thus, low span individuals may not have the resources to inhibit irrelevant information, or rather to selectively activate information that is relevant, making the task of integration more difficult. If all associated information were to be activated indiscriminately, the task of integrating prior knowledge would become unmanageable. Miyake, Friedman, Emerson, Witzki, and Howarter (2000) found inhibition was not correlated with IQ, although a test of updating was highly correlated with measures of intelligence. This is one of a handful of papers that try to differentiate components involved in central
executive functioning and their relative importance in driving higher cognitive processes such as reading.

Studies of reading (Murray & Burke, 2003, Calvo, 2001, 2005) have shown that later interpretive processes seem to be influenced by limitations of WM, although it is possible that the relationship is not direct. Whether there may be effects on early processing is unknown, but no clear evidence has been found (to date) of WM effects on early reading processes (Waters and Caplan, 1999).

According to Yuill and Oakhill (1991), it is a deficit in WM that causes children with poor comprehension to produce discourse representations that are less complete and accurate than those of good comprehenders. In particular, poor comprehenders failed to repair inconsistencies once information became available to resolve the difficulty. Such failures may be due less to limits of memory capacity than of processing ability, but this view is not supported by the work of Spooner et al (2006), which suggests that the problem is primarily one of retention, with poor readers showing less accurate recall. Perfetti (1985) describes less able readers as “losing” information from a recently read discourse more quickly than more able readers.

Following Daneman and Carpenter’s (1980) development of the reading span task (discussed below) numerous experiments have used the task to differentiate groups of high and low span participants. Readers distinguished by their span scores have shown differences in late interpretative processes, in particular inference and effective use of contextual information, as mentioned
above. Vos and Freiderici (2003) have shown differences in ERPs between high and low span participants when reading syntactically difficult or ambiguous sentences. Combining behavioral and ERP data suggests that high-span readers are better able to integrate contextual information online. High span readers also showed separation of these processes from those required to resolve problems of syntax.

Although scores on the reading span task have been found to be correlated with scores on a reading comprehension task (Daneman & Carpenter, 1980, Whitney, Ritchie, & Clark, 1991), it is not clear exactly what the span task is measuring. Rather than measuring WM, the span task may be considered a measure of attention (Engle et al, 1995, 1999), and of central executive function (Miyake et al, 2000). Other difficulties with the use of tasks designed to test WM, specifically with the reading span task, are discussed in the section on individual difference measures below.

C. Methodology

Study of the processes involved in reading comprehension have used a number of different approaches ranging from testing with comprehension questions, free recall, recognition, judgments of coherence and grammaticality, and think-aloud protocols, to measures of reading time including online measures using eye-tracking, ERPs and most recently fMRI. In this paper I concentrate on reading time measures. I also discuss the measures of individual differences that have
been used, with an emphasis on those measures that are of interest to the proposed research.

1. Reading time measures

   Much of the earlier research into reading comprehension relied on simple measures of reading time such as total times to read a passage or target sentence. These measures have shown significant effects of manipulations of consistency and of different types of inference, but do not allow the locus of difficulty to be isolated. This can be done when using single word presentation such as rapid single-word visual presentation (RSVP), which allows reading times for specific words to be recorded, but in this case normal reading is disrupted. Self-paced reading allows readers to control the rate at which words are presented and thus also enables more natural reading and measures of reading times on specific words. Most recently, the use of eye-tracking allows participants to read normally and provides evidence for the locus of any difficulties and the strategies readers use to recover from them. Using eye-tracking to examine processing during reading (Rayner, 1998) provides a way of comparing reading of very similar texts.

   Slowing at any stage in reading is taken to be a measure of difficulty and consequent additional effort such as that needed to make an inference. For instance, elaborative inference was examined by O’Brien and colleagues, (1988):

   All the mugger wanted to steal was the woman’s money. But when she screamed, he (stabbed/assaulted) her with the (weapon/knife) in an attempt
to quiet her. He looked to see if anyone had seen him. He threw the **knife** into the bushes, took her money, and ran away.

This required the reader to infer a general concept from a specific one, or the specific from the general. Measures of interest were fixation times on *knife*. When the verb used was *assaulted*, rather than *stabbed*, participants took longer to read the word *knife* and therefore to make the elaborative inference that a knife had been used.

Calvo (2001) used eye-tracking in experiments to examine whether readers of different ability made equal use of predictive inferences. Participants read passages in which the first sentence provided either a predictive context (1) or a control context (2):

1. *Three days before the examination the pupil went to the library looked for a separate table and opened his notebook.*

2. *The pupil, who was a little tired after finishing his examination, forgot his notebook and left it on a table in the library.*

This was followed by the continuation sentence that contained the target word, either *studied*, which could be predicted from the predictive context, or *slept* (3):

3. *The pupil studied/ slept for an hour approximately.*

The different conditions had no effect on first-pass reading, suggesting that predictive inferences are not made at this early stage of processing. The main effects found between the two conditions were in late measures of processing. The effects took the form of facilitation when the context was predictive, shown
by shorter reading times in the final phrase of the target sentence, rather than the
target word itself. Calvo had divided participants into high and low-span
groups based on a WM test. Differences were found in reading times relative to
condition and to group, even though there were no differences between groups
in the accuracy of their responses to comprehension questions. High-span
readers required less rereading time when an inference was required.

Ashby, Rayner, and Clifton (2004) explored the effect of reading skill on
reading measures and found that skilled readers generally read more quickly
and were less subject to the effects of frequency than less able readers. Effects of
word characteristics such as frequency, predictability, and coherence have been
extensively reported (Rayner, 1998), although seldom in the context of individual
differences.

Problems with using eye-tracking to examine higher level processes such
as inference are demonstrated by the Calvo (2001) study. Even when materials
have been designed with the specific goals of the study in mind, localizing the
point of difficulty in a process that is likely to influence late reading measures is
not always successful.

Noordman and Vonk (1997) have demonstrated that, using eye-tracking,
it is possible to determine the temporal sequence and locus of different types of
processing, in particular, to distinguish early integration from late high-level
processes requiring inference and logical deduction. A connective such as
because signals both the need for integration and invites the use of inference.
Both roles were shown to have effects on reading times. Reading times were shorter for words immediately following the connector, but sentence wrap-up times were longer, showing more complete online integration. Longer sentence wrap-up times were associated with greater accuracy in answers to comprehension questions.

2) Individual difference measures:

The value of examining component measures to studying individual differences in reading comprehension, as suggested by Baddeley et al (1985), has been demonstrated by their use in numerous studies of reading development and disability. Cunningham, Stanovich, and Wilson (1990) used measures of lexical access, vocabulary, WM, reading experience and reading comprehension to explore individual differences in reading ability in a group of introductory psychology students, and recommended this approach to studying differences in how well individuals understand text. Stanovich and West (1998) have also considered individual differences in reasoning ability using a syllogistic reasoning task. Although not directly related to reading comprehension, reasoning ability could be a component in differences between participants who are more or less likely to make appropriate use of inference.

i. Lexical access

Lexical access in reading can be measured using a visually presented lexical decision task (LDT). In its simplest form, the participant must decide whether stimuli are words or not. Response times, and to a lesser extent error
rates, are of interest, and these are found to vary with different experimental
manipulations; effects of frequency are notably strong (Chumbley & Balota,
1984), but effects have also been found for rule-based and exception words
(Weems & Zaidel, 2004), as well as for neighborhood size (Balota, Cortese, &
Sergent-Marshall, 2004) — the number of words that can be created by changing a
single letter in a word. The task has been used in a variety of experiments into
cognitive processes; in studies of reading using dual-task methodology, LD has
frequently been combined with reading a text, in order to give a measure of
priming and provide evidence of activation (Long and Chong, 2001) and of
inferential processing (McKoon & Ratcliff, 1992; O’Brien, Rizzella, Albrecht, &
Halleran, 1998).

In experiments exploring individual differences in competent adult
readers, speed and accuracy on the LDT were found to correlate with both
reading comprehension, $r = .37$, and also with reading span, $r = .26$, (Dixon,
LeFevre, and Twilley, 1988), and with dual task performance, when the task
involved naming and probe-detection (Herdman, & LeFevre, 1992); decrements
in performance on the dual task, compared with the tasks performed separately,
were predicted by scores on both the LDT and a measure of WM.

Schilling, Rayner and Chumbley (1998) compared performance by college
students on an LDT with measures of reading obtained through eye-tracking.
The effects of word frequency on the LDT were larger than those found during
reading, especially for low frequency words. Response times were correlated
with gaze duration times during reading, suggesting that effects on the LDT are consistent with effects on early processing in reading. Lewellen, Goldinger, Pisoni, and Greene (1993) also found that students reporting more reading experience and scoring higher in a vocabulary test responded faster and were more accurate in an LDT task, in particular in rejecting pseudohomophones, words that sound like, but do not look like, actual words (e.g. brane and joak).

Chateau and Jared (2000) found that print exposure was related to performance on an LDT in college students, beyond that explained by comprehension ability. Students who reported higher print exposure responded more quickly and accurately overall and especially to pseudohomophones, while low exposure participants showed larger effects of frequency. Thus, even students who were successful readers showed differences in efficiency of lexical access that were related to the amount they read.

Debate about the effectiveness of the LDT as a measure of lexical access has focused on a number of issues that have little relevance to the proposed study. However, effects on task performance of meaning related variables such as frequency (Chumbley, & Balota, 1984), congruence, ambiguity, and neighborhood size are relevant. Balota et al (2004) found neighborhood effects, especially for non-words, that were more pronounced in older adults and slow young adult readers. This suggests not only that the LDT is influenced by meaning-level variables, but also that meaning is accessed very early in the process of word recognition and, in the LDT, facilitates the decision process.
Differences in meaning activation between the LDT and a naming task have been demonstrated by comparing the effect of reading pseudohomophones. Such non-words produced faster naming but slower LDT response times than those for pronounceable non-words that are not homophones (Seidenberg, Peterson, MacDonald, & Plaut, 1996).

The use of the Divided Visual Field (DVF) lexical decision task makes a range of additional analyses possible. In this version of the LDT, lexical access is measured by randomly presenting words and non-words to either the right or left visual field. While accuracy and response time may provide evidence for different types of processing, especially for non-words, differences in accuracy and response time between the hemispheres may also be informative; as well as the right hemisphere having an important role in discourse processing (Beeman, 1993), it may maintain alternate meanings, especially in situations where there is ambiguity (Faust and Chiarello, 1998), and perhaps also has better error monitoring (Iacoboni, Rayman, and Zaidel, 1997) than the left hemisphere. The right hemisphere is thought to be involved in more top-down processing and to make greater use of orthographic and semantic information; the left-brain makes greater use of phonology (Pugh et al, 1996) and is generally quicker and more accurate in word recognition, particularly for rule-based words.

Hemispheric differences may explain the problems that poor comprehenders have with accessing recent information and in suppressing irrelevant information (Gernsbacher, Varner, and Faust, 1990). Recent work
(Long and Chong, 2001) suggests problems are less of access than of integration. However, an alternative explanation is that good comprehenders are more effective in recruiting the right hemisphere and may have less pronounced lateralization of the processes involved in language processing.

Using a DVF lexical decision task (Weems and Zaidel, 2004, Shears and Chiarello, 2003) could potentially provide additional sources of information while still measuring overall response times and accuracy. While it is possible for words to be visible in both visual fields, in practice there is a lack of behavioral evidence for the use of any overlap. That is, in practice, even when a word falls foveally, there is no evidence that information is bilaterally represented (Brysbaert, 1994), and thus interhemispheric transfer is required. In order to reduce the possibility of saccades from fixation stimuli, presentation time should be kept short—no longer than 165 ms (Pirrozolo & Rayner, 1980).

To sum up, it may be hypothesized that readers who comprehend well are more fully engaging their right hemisphere in language processing, particularly in later processes of integration and interpretation. Rapid and accurate responses by readers on LVF presentation of a lateralized LDT are correlated with good (late) comprehension. Such readers cope well with tasks that require either inference or ambiguity resolution or require integration of contextual information. Rapid and accurate responses by readers following RVF presentation of the same task correspond with good early processing by readers. Responses in both fields are likely to be strongly correlated. A specific
prediction can be made that those readers who respond quickly and accurately to stimuli presented in LVF and thus show less effect of side of presentation will reread less, and thus have shorter total reading times, when presented with complex sentences requiring inference, ambiguity resolution, or monitoring for global coherence, than readers who show larger effects of lateralization.

Lexical access can be tested in a number of different ways apart from LD, including the use reading time measures or an oral word-naming task. A reading time measure has the advantage that it can be incorporated into other tasks, for instance by measuring reading over part of a comprehension task (Calvo, 2001). Calvo found that reading speed recorded this way was correlated with early reading measures, but not with those late measures that are thought to relate to interpretive processes. Others have found the relationship of reading speed and reading comprehension to be weak (Everett and Underwood, 1994). A measure of lexical access obtained this way is likely to be confounded with other tests and probably measures more than lexical access.

ii. Vocabulary and text exposure

The strong relationship of vocabulary knowledge with reading comprehension has been frequently reported and is intuitively transparent. Cunningham, Stanovich and colleagues have explored the relationship in several papers (1990, 1994, 1996), making the connection between reading development, environment, print exposure, and the Matthew effect (Stanovich, 1986). The latter results when children who learn to read early benefit, since they enjoy
reading, read independently for pleasure, and continue to improve, compared to children who have trouble learning to read and thus tend to fall behind.

Vocabulary tests have been a regular component of individual difference testing, mainly in developmental studies of reading in children (Yuill & Oakhill, 1991) and in studies of cognitive aging (Salthouse, 1993) and are considered to relate to crystallized intelligence. There are few studies correlating vocabulary with reading time measures, such as those obtained from eye-tracking; Calvo (2005) showed a measure of vocabulary to be correlated with times required for rereading in sentences requiring inferential processing.

Tests are either of expressive or receptive vocabulary, and frequently ask for definitions, synonyms and antonyms, or analogies, often using sentence completion or cloze tasks, and multiple-choice questions. The Peabody picture test is frequently, but not exclusively, used for testing receptive vocabulary in children and is generally paired with a test of expressive vocabulary. Salthouse and colleagues (Salthouse, Siedlecki, & Krueger, 2006), in an experiment exploring individual differences in memory control, used four different vocabulary tests: the WAIS vocabulary test (Wechsler, 1997) for which participants provide definitions of test items, the WJ picture vocabulary test (Woodcock and Johnson, 1990) which requires naming of pictured objects, and tests which required selecting either the best synonym or antonym of a target word (Salthouse, 1993). Correlations between and reliability of all four tests was high.
In addition to tests of vocabulary, Stanovich et al have developed an author recognition test (1990) to provide a measure of print exposure. Participants are asked to mark the names of 40 popular authors from a list of 80 names. The measure is the number of authors correctly identified, less any identified as authors incorrectly, to discourage guessing. Although a measure of print exposure does not provide a measure of any cognitive component, it does provide a measure of reading experience that may mediate variability in cognitive processes associated with reading.

iii. Working Memory

The need for a reliable test of WM has produced a number of alternatives. The reading span task of Daneman and Carpenter (1980) has been widely used as a way of discriminating between readers of different ability levels in reading and in recall, since it also appears to be a good predictor of readers’ ability to integrate text (Whitney, Ritchie, & Clark, 1991). This task, for which participants read sentences in sets that increase in number from two to seven and then asks them to recall the final word of each sentence at the end of the set, can be scored and run in a number of different ways. As originally conceived, the level an individual completes successfully determines their score. The fact that the test is widely used and recognized, as well as easily administered, is advantageous. However, with its widespread use have come problems and questions: the questions relate to whether the test is really measuring WM and, if so, is it a
measure of WM capacity or is it also a measure of processing. The problems relate to the best way to effectively administer, analyze and use the task.

A number of studies, some already mentioned above, have used the task effectively as a means of determining groups. High and low span groups have been found to respond differently in tasks requiring higher cognitive processes such as inferential processes in reading (Calvo, 2001, Linderholm, 2002). However, other researchers have found problems with applying the task in this way. A serious concern is the limited range in the results that can make it hard to distinguish groups. After studying ways of using the reading span task, Friedman and Miyake (2004a) recommend against using the test to determine groups, in particular to avoid use of a median split. They found that dividing the participants between groups reduced both reliability and predictive power. In addition, they make the case that, where an interaction is expected, any measurement error would have a disproportionate effect on the analysis of variance, and consequently that methods such as regression and correlation are to be preferred over post hoc division into groups. That said, researchers have had some success with dividing participants into groups, in some cases employing the upper and lower quartiles of the range tested (Unsworth, Heitz, Schrock, and Engle et al, 2005). Vos and Freiderici (2004) approached this problem by defining a low span group with scores of 2.7 or below, and a high span group whose scores were at least 4.
Reliability of the task has also been found to be uneven as participants frequently perform more poorly at retest, possibly because of interference (Friedman & Miyake, 2004b). Alternative methods of scoring the task such as counting total words or calculating the proportion of words recalled have produced greater retest reliability (Friedman & Miyake, 2005) and also have the advantage of providing more fine-grained continuous measures. In the latter method forgetting a final word from a set of two sentences has a larger effect on the final score than forgetting a word from a set of five. Both of these scoring measures are best achieved when the complete test is run. Having to take the complete test can be demoralizing for poor testees, so Friedman and Miyake (2005) suggest ordering trials so that short and long blocks are intermixed, allowing for the possibility of success at any stage of the task. The number of sets run at each level of difficulty also produces variability in retest reliability. When three sets were tested at each level reliability was good for scores calculated using total words and proportion of words ($r = .72$ for each) but poorer for correct sets ($r = .59$) and for truncated sets in which participants were allowed to stop once a level had been failed ($r = .52$). Reliability improves for all methods if more sets are tested at each level, but this substantially increases the time needed and presumably unwanted negative effects on participant goodwill. Finally, the question whether the test is self-administered or experimenter-administered has been raised. Despite concerns that testees may make greater
use of strategies to improve recall when the task is self-administered, results were not found to differ significantly (Friedman and Miyake, 2004b).

The interpretation of results and correlations is also complicated by consideration of how well the task may tap central executive function, specifically selective attention, updating and inhibition (Miyake et al, 2000, Engle, et al, 1995, 1999). Redick and Engle (2006) compared tasks designed to test WM with the Attention Network Test and suggest that variation in WM capacity reflects differences in how well individuals are able to direct attention. The importance of efficient updating to maximizing effective reading comprehension (Miyake et al, 2000) is mirrored in the visual domain. Vogel, McCollough, and Machizawa (2005) stress the importance of individual differences in how efficiently space is allocated within WM, so that moment by moment only the most relevant information is represented.

Thus, poor readers may use WM resources less efficiently. Specifically, they may have limited control over how those resources are allocated, leading in turn to poorer scores on the reading span task. Although careful administration and scoring can effectively improve the test’s reliability this does not address problems with validity. The fact that scores on the reading span task are correlated with those for reading comprehension suggests that cognitive processes beyond those that relate strictly to WM are involved. This raises the possibility of overlapping measurement and the need to clearly discriminate WM from components measured by other tasks, in particular those testing vocabulary.
and lexical access. A task with less reliance on reading but which still taxes both storage and processing should provide a more specific test of WM.

A number of other tasks have been developed to test WM. Most are versions of the span task (e.g. spatial span tasks, Priti & Miyake, 1996). I do not discuss here all of the alternative tests developed to measure WM but concentrate on the operation span or Ospan task (Turner and Engle, 1989) and a related automated version of the same task – the Aospan (Unsworth et al, 2005).

In the Ospan task participants are required to solve a series of mathematical operations and at the end of each set recall in serial order the words presented after each operation. The task takes approximately 20 minutes if experimenter administered. Internal validity and test-retest reliability are high ($r = .83$). The Ospan task has been used by a number of researchers; Yuill, and Oakhill (1991) used a version to test skills related to reading comprehension in children.

Turner and Engle (1989) have shown that the operation span task was as strongly correlated with reading comprehension measures as the reading span task, while Miyake et al (2000) have shown it to be strongly correlated with processes of updating. Efficiency in updating the information held in WM may explain why performance on these measures is also predictive of reading ability (Daneman and Merikle, 1996).

The Aospan task is conducted at each individual’s pace, is mouse driven, self-scoring, and also records response times. Participants are given feedback at
the end of each set and are required to be at least 85% accurate on the mathematical operations to ensure this part of the task is being given appropriate attention. Once the answer has been given to the mathematical operation, a letter is presented which must be recalled later by inputting letters from a matrix in the order presented. Although the test is self-paced, limits are set on time allowed for the mathematical operations based on the participant’s own response time. This restricts the time available for rehearsal of the to-be-remembered items. As with the Ospan task, internal validity for the Aospan task is high, as is test-retest reliability ($r = .83$). Compared to the Ospan task, the automated Aospan task has advantages for administration and, since testees are asked to recall letters and not words, is less likely to be confounded with a lexical access task or with tests of vocabulary.

Scores on the automated span task have been found to correlate with the experimenter administered Ospan and reading span, and to show, by both confirmatory and exploratory factor analysis, moderate correlations with tasks measuring spatial reasoning (Unsworth et al, 2005). This last result provides support for the view (Kane, et al, 2004) that span tasks measure a domain-general component in WM, which may also be predictive of general fluid intelligence. Bunting, 2006, suggests that this component relates to attention, specifically the ability to resist interference, but does not accept its domain generality.

The use of two tests, as recommended by Conway et al (2005), has the advantage of allowing greater confidence in the combined result as a measure of
WM, and would also allow the results to be compared with measures of lexical access to check for confounding. Further, if the group approach is preferred, assigning participants to quartiles is more efficient when two tasks are employed. However, this approach risks over-taxing the participants.

iv. Reasoning

How far reasoning ability might be related to reading comprehension is less clear. Reasoning, particularly on a syllogistic reasoning task has been shown to be correlated with scores on the SAT, $r = .410$ (Stanovich and West, 1998). However, Siddiqui, West and Stanovich (1998) compared a reasoning task with print exposure measures and a test of mental-state verbs, and found that print exposure was a good predictor of verb discrimination but a less effective predictor of reasoning ability. Tests of reasoning ability require responses, normally “True” or “False”, to statements requiring logical reasoning. Stanovich et al (Stanovich & West, 1998, Siddiqui, et al, 1998) used sentences in which the statements were manipulated on believability as well as logical validity (see examples a and b, below).

(a) All mammals walk. Whales are mammals. Therefore whales can walk. (Valid, unbelievable)

(b) All flowers have petals. Roses have petals. Roses are flowers. (Invalid, believable)

Thus correct answers require participants to concentrate on the logic of the statements while ignoring their own world knowledge. Findings showed that high span individuals are more accurate in this task, especially when there is a
conflict with believability. De Neys (2006) found that high span readers showed less effect of a load condition when believability conflicted with the correct response, and were more accurate overall than low span participants. Reading times also tended to be longer for high span readers in the unbelievable conditions. While this finding suggests a probable correlation of scores on this task with those on a WM task it also would point to a failure of inhibition and thus a possible correlation with False alarms on the LD task. When considering the use of inference, it is possible that participants who score high on the reasoning task may be less prone to make use of inference in conditions in which its use is not well supported by the text.

v. Speed of Processing

Individual differences in speed of processing have been suggested as a factor in memory control in aging (Salthouse, Siedlecki, and Krueger, 2006), and thus might mediate the cognitive components under consideration in this experiment. Of the tasks used by Salthouse et al to test this hypothesis the pattern comparison task seemed most appropriate for this experiment, in particular because it shows the least overlap with other tasks.

The role the factors, discussed above, have in influencing comprehension are likely to be interrelated. Thus it is important that tests employed in the experiments overlapped as little as possible in the methods and materials used. If the tasks are distinct then relationships between the measures of cognitive
components that are also correlated with the dependent reading measures will suggest a joint role in comprehension for those components. Thus it is likely that vocabulary measures may correlate with the measure of print exposure, while each is also correlated with reading comprehension. Reading experience, as measured by the author recognition test, may also mediate the role of other cognitive components in improving readers’ understanding. In addition, correlations between measures suggest that related constructs are being tested by different tasks.
CHAPTER 2

EXPERIMENTS

For this dissertation I completed two experiments that explored how individual differences in component cognitive processes are related to differences in how inference and contextual information are used in the interpretation of text. Participants were asked to complete several tasks designed to measure individual differences in cognitive processes as well as a reading task in which the meaning of target sentences containing a fronted adjunct clause will be manipulated to make the use of inference more or less plausible. Experiment 1 was expected to show significant relationships between patterns of inference, interpretation and individual difference measures. Experiment 2 examined more complex inference patterns by embedding target sentences in passages that supplied context that was either supportive or neutral to the inference.

The use of sentences containing a fronted adjunct allowed an opportunity to examine the ways in which understanding varies when alternate interpretations are possible. The initial adjunct clause allows for two possible interpretations of the meaning. Previous experiments (Christianson, Hollingworth, Halliwell and Ferreira, 2001, and Ferreira, Christianson, and Hollingworth, 2001) have shown
that readers do not consistently arrive at a single interpretation of these “garden path” sentences. Readers’ interpretations were influenced by both syntax and semantics, but also by their use of inference. Christianson et al suggest that differences in interpretation arose because some readers relied on an incomplete process of reanalysis, which they called “good enough processing”: this explanation was not supported by my research using similar materials.

In a series of experiments (Michael, 2004), I have examined reading of sentences (see (1) and (2) below) containing a fronted adjunct clause. Sentences were syntactically ambiguous “garden-path” (GP) sentences or were disambiguated by punctuation.

(1) As Tom tattooed(,) the girl paced up and down.
(2) As Tom tattooed(,) the girl tried not to move.

These sentences tend to cause readers to make an initial incorrect analysis that the girl is the object of the first verb (V1), tattooed. Thus, in the main clause, at the second verb (V2), (paced or tried in the examples) it becomes clear there is a problem with the initial analysis that needs to be corrected. The syntactic difficulty can be resolved by reanalysis from V2, but the problem of interpretation is more easily resolved using semantic information from the final part of the main clause. This later information helps disambiguate meaning but requires accepting a plausible inference to create a coherent representation, such that in the dual-meaning sentences (2), the girl is being tattooed by Tom. The manipulation of meaning had a large effect of responses to statement (3):
(3) True or False? Tom tattooed the girl.

Participants were more likely to respond “True” following sentences such as (2). However the effect of the disambiguating punctuation was not significant. There were also overall differences in reading times with longer times for single-meaning sentences such as (1).

<table>
<thead>
<tr>
<th></th>
<th>Garden-path 1</th>
<th>Garden-Path 2</th>
<th>Punctuated 1</th>
<th>Punctuated 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of responses “True”</td>
<td>19%</td>
<td>66%</td>
<td>16%</td>
<td>72%</td>
</tr>
<tr>
<td>Total sentence reading time</td>
<td>4824ms</td>
<td>4860ms</td>
<td>4300ms</td>
<td>3828ms</td>
</tr>
</tbody>
</table>

Table 1: Responses and reading times to garden-path and punctuated sentences

This could be interpreted as being the result of a longer process of reanalysis for sentence (1) but an analysis contingent on response showed this was not the case. Reading times for participants making the more common “False” response for (1) and “True” for (2) had similar reading times for sentences of both types. Differences in overall reading times were accounted for by longer reading times when readers made the less usual responses.

In a later experiment the target GP sentences followed a context sentence that was either consistent (4) and (6) or inconsistent (5) with the most common interpretation of the target sentence.

Supportive of single meaning; single target
(4) Tom started work on the boy's tattoo as his girlfriend watched.
    As Tom tattooed the girl paced up and down.

Supportive of single meaning; dual target
(5) Tom started work on the boy's tattoo as his girlfriend watched.
As Tom tattooed the girl tried not to move.

Supportive of dual meaning; dual target
(6) Tom started work on the girl's tattoo as her boyfriend watched.
As Tom tattooed the girl tried not to move.

Responses to statement (3) and reading times were influenced by the prior context.

<table>
<thead>
<tr>
<th>Proportion of responses “True”</th>
<th>(4) consistent</th>
<th>(5) inconsistent</th>
<th>(6) consistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sentence reading time for both sentences</td>
<td>15%</td>
<td>29%</td>
<td>88%</td>
</tr>
<tr>
<td>Total time for target sentences</td>
<td>9838ms</td>
<td>9963ms</td>
<td>8896ms</td>
</tr>
</tbody>
</table>

Table 2: Responses and reading times to sentences with prior context

As expected, the manipulation of context produced more responses of “True” when the context confirmed the inference in the target sentence (6) and many fewer when it did not. There was also an effect of the prior discourse context on the way in which identical target sentences were read. Reading was particularly slowed in the inconsistent condition, a result expected following the finding of Murray and Burke (2003) that readers are slower when reading a target sentence for which a plausible inference was inconsistent with information from the prior context.

The earliest processing differences were found in First-Pass reading in the final postverb area, but more marked differences were found in later measures. Of particular interest was the interaction between sentence type and response in regression-path reading and rereading. This showed that readers were selective in where they concentrated their effort in reanalysis, and that longer rereading
did not result in greater accuracy but was evidence of greater processing difficulty. This difficulty is especially apparent when reading times for expected and unexpected responses are compared. Rereading and reanalysis appear to be more strongly influenced by how a sentence is understood than by any initial misanalysis or subsequent problems with analyzing syntax; further, making a plausible inference can reduce the need for the reader to reanalyze in this type of ambiguous sentences. Thus, reanalysis is not perfunctory in these cases, and although it may be “good enough” as suggested by Christianson et al (2001), that is because most readers take only as long as is necessary to confirm the interpretation that they ultimately settle on. There is no evidence, from these results, to suggest that readers who take less time rereading are failing to process completely. Instead, employing inference when it matches input is relatively easy and consequently results in less rereading.

Although the inclusion of a context sentence allowed the examination of reading in situations when the use of plausible inference was inappropriate, the manipulation does not replicate the situation, seen in the earlier experiment, in which a reader fails to make appropriate use of plausible inference. Nor does it explain differences seen in how effectively prior contextual information was employed. The post hoc analysis contingent on response is also problematic.

The two experiments described in this paper were designed to explore the ways in which individual differences in interpretation by competent adult
readers are differentially related to the component processes that contribute to effective reading comprehension. Previous research suggested that these differences were more likely to relate to later processes of inference and integration; thus differences in interpretation were most likely to correlate with measures of WM and vocabulary. However, there is also some evidence to suggest that the effects of earlier levels of processing should be considered. Participants were asked to complete the reading task followed by a series of tests that have been widely used to measure important components of cognitive function. The tests specifically relating to aspects of reading comprehension are the DVF lexical decision task (LDT) as a measure of lexical access, a test of vocabulary in which participants were asked to identify synonyms, and the Aospan task which tests WM capacity and processing. These tests of individual difference were chosen to supply clear discriminability of each component process so that any correlation found was unlikely to be due to overlap of the tasks. Of these, only the LDT was expected to be correlated with earlier stages of processing. Although lexical access can be tested in a number of different ways, including the use of a reading time measure or word naming, the LDT has the potential to provide more extensive information, in the form of error rates and reaction times (RTs) especially when the lateralized DVF version of the task is employed. The synonym detection task has advantages in terms of ease of administration and in the contrast it provides with the other tasks. The Aospan task was chosen to measure WM as this task relies less on verbal decoding skills
than either the reading span or traditional operation span tasks. Letter recognition, which is required for the recall portion of the task, is highly automated even in quite young children, so is unlikely to be confounded with other word related measures.

In addition participants were asked to perform an author recognition task to provide a measure of reading experience, a pattern comparison task, and a test of syllogistic reasoning. Scores on the author recognition task were expected to be related to expected responses, while the test of reasoning may be related to a more conservative style of response, leading to less use of inference and fewer answers of “True”. The pattern comparison task provides a speed of processing measure using simple figures consisting of line segments, and therefore relies on materials that are unlike those used in any of the other tasks.

In Experiment 1, responses to the reading comprehension questions following the target sentences with fronted adjuncts were compared with scores on the tasks measuring individual differences. This allowed correlation of individual differences of interpretation with core cognitive components. Those components found to be of interest in Experiment 1 were further examined in Experiment 2 in which the target sentences were shown with two prior sentences that provided context. The information in the prior context biased readers either to use inference or was neutral in its effect on interpretation of the target sentence. This allowed investigation of the way prior contextual information is integrated and how it interacts with interpretive processes. In addition, in
Experiment 2, passages requiring bridging inference provided comparison with the ambiguous fronted adjunct sentences. Together the two experiments provide important information on the cognitive processes that support effective reading comprehension. In doing so, they also cast light on the problems that some adult readers experience.

**Experiment 1**

The first experiment examined the relationship between selected measures of individual difference and how ambiguous sentences are understood. Participants were asked to complete a series of tasks, including a reading task that contained the target garden-path sentences and an equal number of unambiguous filler sentences. Responses to comprehension questions following the garden-path sentences indicated the extent to which participants were willing to use inferential processes. Responses following the filler sentences gave a measure of reading accuracy. To test individual differences, participants also completed the divided visual field LDT, a simple centrally presented LDT (LDT2), synonym selection, the Aospan task, an author recognition test, the pattern comparison task and a test of syllogistic reasoning. With the exception of the two LDT tasks, each task was selected to test separable components of cognition—lexical access, vocabulary, WM, and reasoning ability, as well as the experiential factor of print exposure —that each potentially contributes to reading comprehension. The way in which each measure is expected to relate to
reading comprehension is shown in Figure 1. In addition, participants were asked to complete the Edinburgh Handedness Inventory to confirm that all participants are right handed.

In this first experiment the reading task tested how each participant understood sentences of different levels of difficulty and ambiguity. Sentences (1) and (2) (shown again below, without a disambiguating comma) were presented on the computer. After reading each sentence, participants pressed the space bar to continue to a comprehension question (see (3), below) on the next screen to which participants were asked to respond “True” or “False”.

Participants were expected to show a consistent bias towards answering either “True” or “False” depending on the condition: more answers of “False” are expected following sentences such as (1), and more answers of “True” following sentences such as (2).

(1) As Tom tattooed the girl paced up and down.
(2) As Tom tattooed the girl tried not to move.
(3) True or False: Tom tattooed the girl.

Readers who fail to respond as expected, either answering “False” following sentences such as (2), or answering “True” to sentence such as (1), are either failing to make use of inference, or are making inappropriate use of inference when information in the main clause makes the inference implausible. It was expected that there would be a relationship between responses to the comprehension questions and scores on the tests of individual difference,
particularly the measure of WM. Previous research has shown that high span
readers are more likely to make use of inference in situations where an inference
is warranted by prior context and need for coherence (Calvo, 2001, 2005). I
hypothesized that the difference in responses that we have seen in previous
experiments (Michael, 2004) would be similarly related and that high span
readers would make greater use of the relevant of information in the final clause
than would low span readers. Using fronted adjunct sentences had a number of
advantages over the manipulation used in other experiments that have examined
individual differences. The most important is the variability in interpretation,
clearly shown by the variability in response found in both my research and in
that of Christianson et al (2001), which makes response of particular interest.
Previous research has mainly relied on reading times (Calvo, 2001), and accuracy
(Long and Chong, 2001) predicated on the assumption that there was a “correct”
answer and therefore tended to be subject to ceiling effects.

In addition responses to comprehension questions following the filler
sentences, which varied in complexity but were not ambiguous, provided a more
general measure of reading accuracy.

Method

Participants

One hundred and twenty four students of UNC Chapel Hill received credit in an
introductory psychology course in return for their participation. All were right-
handed and native speakers of English.
Materials and Design

Participants took a series of tests after first completing the reading task, which involved self-paced reading of the experimental sentences and responding to the subsequent comprehension questions. All tests were completed on a computer in a single one-hour session. All parts of the experiment except the Aospan task were linked so that students could work their way through each in their own time. Tasks took from one to twenty minutes. On completion each participant was then asked to complete the Aospan task. Most participants finished the sequence of tests without evident problems.

Materials

There were 36 target sentences identical to those used in prior experiments (see (1) and (2) above). Sentences had been pretested to provide an estimate of mean response using an online web experiment. The sentences were matched for length by number of syllables and, as far as possible, by number of words. The critical word was expected to be the second verb and this was matched for length (mean length 6.5 letters), and for frequency using the Thorndike-Lorge written frequencies, found in the Medical Research Council database, maintained by UWA. There was a start block of 12 sentences and equal numbers of fillers and experimental sentences during which feedback was provided.

Procedure

Sentences were presented on a computer and reading was self-paced. A normal reading speed was encouraged. Once the participants have read the sentence they
pressed the space bar to continue to the next screen, which will show a statement such as (3), to which they responded either “True” or “False.” This response was followed by the next sentence.

Tests of Cognitive Processing

1) Divided visual field lexical decision task

For the lateralized LDT words and non-words were shown in black text on a white background on either side of a central fixation point and were on the computer screen for no longer than 165 ms to limit the opportunity for saccades from fixation (Pirrozolo & Rayner, 1980). The screen refresh rate was set at 85hz. After presentation of stimuli to either the right or left visual field, participants will be required to make a bimanual yes/no response to indicate whether this is a word or not. Both accuracy and response time were measured. The yes/no response was preferred to a go/no go response because of the possible importance of the latency of response to non-words, in particular from the left visual field (LVF). The bimanual response negates concerns relating to interhemispheric communication in response selection (Weems and Zaidel, 2005). A right visual field (RVF) advantage for response times and accuracy is expected. Response accuracy and latency to non-words may also be informative as these have been shown to be correlated with reading comprehension (Weems and Zaidel, 2004) and with reading times (Schilling, Rayner and Chumbley, 1998).
Participants were asked to sit close to the monitor, and to maintain their
gaze direction at a central fixation point. 120 words and orthographically
plausible non-words were presented on one side or the other of fixation. Since
the angle of display could be an issue, words were presented at a distance no
closer than one degree of visual angle from fixation and extending no more than
three degrees to either side. This was calculated based on a viewing distance of
between 50 and 65cm, a comfortable viewing distance. A screen showing only
the fixation point then replaced the stimulus screen until a response had been
made, at which point a new stimulus was displayed.

Words were frequent or infrequent, orthographically regular or irregular,
and from 3 to 6 letters long. Frequent words occurred between 200 and 800 times
in a million according to Kucera-Francis written frequencies, while infrequent
words occurred 1-2 times, determined using the University of Western
Australia’s MRC Psycholinguistic Database. Nonwords were within a single
letter of an actual word. (Stimuli used are listed in Appendix 2)

In addition to the lateralized task a briefer version, hereafter called the
LDT2, with central presentation of words and non-words was also administered
as a check on the effectiveness of the lateralized task. In this version, words
varied in number of near neighbors but were matched for frequency.

2) Synonym detection

This task provides a highly reliable test of receptive vocabulary. Participants
were presented with a target word and were then asked to pick a synonym from
three alternate words, indicating their choice with a key press. There were 60 target words (see Appendix 3). This task differed in style from the other tasks.

3) Aospan task

The Aospan is programmed in E-prime and is available online from the Engle Laboratory website, http://psychology.gatech.edu/renglelab/. The task takes approximately 20 minutes to run and is self-paced. Theoretical advantages of the task have been discussed above.

The task generates a score for the Operation span task, the Ospan score, an Ospan total correct, as well as error measures for different parts of the task, mathematics, speed, and overall error rate. While the Ospan scores and Ospan total are the main measures of interest, the error scores could also be tested for correlation with other components.

4) Author recognition test

A revised version of the Stanovich (1989) list was used as a measure of test experience. The list is available online and has been updated, as suggested by the authors, to reflect current popular writers, such as J.K. Rowling (Appendix 4). The names of 40 popular authors were included in a list of 80 names. As each name appeared on the screen the participant had to respond whether it was that of a popular author or not. Scores reflected those authors correctly identified, less any false alarms.

5) Syllogistic reasoning

A syllogistic reasoning task was included to test logical reasoning. The
task required participants to correctly identify whether the third statement followed logically from the first two. Statements were either valid or not, and factually either believable or not. Materials were based on Stanovich and West (1998):

- All mammals walk. Whales are mammals. Therefore whales can walk. (Valid, unbelievable)

- All flowers have petals. Roses have petals. Roses are flowers. (Invalid, believable)

- All fish can swim. Tuna are fish. Tuna can swim. (Valid, believable)

- All fish can swim. People can swim. People are fish. (Invalid, unbelievable)

All of the valid unbelievable and invalid believable passages used in this task had been pilot tested. Participants were scored on accuracy and susceptibility to content in the invalid believable condition. Reading times were also recorded.

6) Pattern Comparison

The pattern comparison task involved a same/different choice with RTs and accuracy recorded to provide a measure of speed of processing. Stimuli consisted of pairs of line-segment patterns that the participant was asked to classify as “same” or “different” as rapidly as possible. Half of the pairs were identical while the other half differed. Pairs requiring a different response were constructed by altering a single line segment in one member of the pair. The line patterns were connected lines in an invisible $3 \times 3$ matrix, with three, six, or nine line segments in each member of the pair, following Salthouse and Babcock.
(1991). 32 pairs were created for each set size (i.e., with three, six, or nine segments); see Appendix 5.

Edinburgh Handedness Inventory

Prior to taking the test participants were screened for handedness using the Edinburgh Handedness Inventory, which asks the preferred hand used for various everyday tasks. Although only right-handed participants were recruited, the inventory provided a further check of handedness. Excluding left-handed subjects was necessary since language processing is more frequently dominant in the right hemisphere than is usual for the rest of the population.

Accuracy was the main measures of interest for all the tests of component processes, except the pattern comparison and lexical decision tasks. Response times were available for all the tests and were of primary interest for the LDTs, the pattern comparison task, but also for the test of reasoning. Previous research (De Neys, 2006) had shown that high span readers were likely to have longer reading times when the correct answer is unbelievable. False alarms were considered relevant for the LDT, especially those to non-words so scores were calculated in terms of error rate. Responses to the comprehension question supplied information on interpretation and were expected to correlate with measures on the tests of cognition. Total reading times were expected to vary by condition and response.
Results

The data presented here are for 120 participants who completed all parts of the experiment. Four participants were omitted as outliers: they had scores that were at the bottom of the range on reading accuracy and were extreme on one or more of the component tasks. A further participant had scores on a single component, the LDT2, omitted because of extreme scores resulting from having accidentally skipped the instructions.

Analysis of this complicated data set took place in several stages. After considering descriptive statistics and confirming that scores and main effects were as expected, I looked at the correlations between the various tasks. Following examination of the correlations and the completion of principle component analyses, I went on to use factor analysis and structural equation modeling to explore the relationships of the measures of cognitive and experiential components with the outcome measures from the reading task. Lastly I considered the results of cluster analysis. I will first give descriptive statistics and any main effects found for the different tasks in Experiment 1, then detail correlations of theoretical interest, and finally discuss the modeling and cluster analysis.

1. Descriptive Statistics

Sentence reading and Comprehension

The pattern of responses to the comprehension questions following garden-path sentences was similar to that seen in prior experiments, \( F(1,119) = \)
21.34, \( p < .001 \), responses are shown as a proportion of “expected” responses.

<table>
<thead>
<tr>
<th></th>
<th>Gp1 sentences “False”</th>
<th>Gp2 sentences “True”</th>
<th>Fillers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>82%</td>
<td>69%</td>
<td>88%</td>
</tr>
<tr>
<td>Reading time</td>
<td>4990ms</td>
<td>4715ms</td>
<td>5269ms</td>
</tr>
</tbody>
</table>

Table 3: Responses and reading times to sentences in Experiment 1

Reading times for the garden path sentences varied by condition being shorter for GP2 sentences, \( F(1,119) = 5.87, \ p = .017 \), but this was reversed for RTs to the comprehension questions which were longer for the GP2 sentences. Longer reading times had no significant effect on response in either condition.

NB: 16 readers who failed to press the spacebar after reading the sentences and thus were allowed all of the available 10 seconds to read every sentence did not have their reading times recorded as this was not considered to represent normal reading.

Measures of component processes

Mean scores and reaction times (RTs) are given in Table 4, below. I provide and discuss descriptive statistics for each measure of component processes, including analyses of main effects before discussing the correlations of measures of component processes with the reading task. Both the synonym detection task and the author recognition task produced a wide range of scores with no evidence of floor or ceiling effects. Ospan scores and totals spanned the range of possible scores. Scores on the pattern comparison task were high as expected for this relatively simple task.
### Table 4: Descriptive statistics for component measures in Experiment 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>SD</th>
<th>Out of</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synonym detection</td>
<td>66</td>
<td>10.5</td>
<td>100</td>
<td>3.8</td>
<td>90</td>
</tr>
<tr>
<td>Author recognition</td>
<td>12</td>
<td>5.3</td>
<td>40</td>
<td>-1</td>
<td>30</td>
</tr>
<tr>
<td>Lateralized LDT – frequent word error rate</td>
<td>.18</td>
<td>.10</td>
<td>1</td>
<td>0</td>
<td>.53</td>
</tr>
<tr>
<td>frequent word RT (ms)</td>
<td>627</td>
<td>120</td>
<td>-</td>
<td>397</td>
<td>1046</td>
</tr>
<tr>
<td>infrequent word error rate</td>
<td>.42</td>
<td>.14</td>
<td>1</td>
<td>.09</td>
<td>.81</td>
</tr>
<tr>
<td>infrequent word RT (ms)</td>
<td>742</td>
<td>165</td>
<td>-</td>
<td>410</td>
<td>1339</td>
</tr>
<tr>
<td>nonword error rate</td>
<td>.24</td>
<td>.11</td>
<td>1</td>
<td>.15</td>
<td>.62</td>
</tr>
<tr>
<td>nonword RT (ms)</td>
<td>769</td>
<td>165</td>
<td>-</td>
<td>429</td>
<td>1300</td>
</tr>
<tr>
<td>LDT2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word error rate</td>
<td>.09</td>
<td>.07</td>
<td>1</td>
<td>0</td>
<td>.30</td>
</tr>
<tr>
<td>word RT (ms)</td>
<td>699</td>
<td>95.4</td>
<td>-</td>
<td>526</td>
<td>1056</td>
</tr>
<tr>
<td>nonword error rate</td>
<td>.06</td>
<td>.09</td>
<td>1</td>
<td>0</td>
<td>.40</td>
</tr>
<tr>
<td>nonword RT (ms)</td>
<td>808</td>
<td>129.5</td>
<td>-</td>
<td>390</td>
<td>1100</td>
</tr>
<tr>
<td>Syllogistic reasoning</td>
<td>17.5</td>
<td>3.8</td>
<td>24</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td>Syllogistic reasoning RT (ms)</td>
<td>1707</td>
<td>819</td>
<td>-</td>
<td>502</td>
<td>5500</td>
</tr>
<tr>
<td>Pattern comparison score</td>
<td>55</td>
<td>2.9</td>
<td>60</td>
<td>43</td>
<td>59</td>
</tr>
<tr>
<td>Pattern comparison RT</td>
<td>1346</td>
<td>229.9</td>
<td>-</td>
<td>951</td>
<td>2156</td>
</tr>
<tr>
<td>Ospan Score</td>
<td>46.9</td>
<td>17.6</td>
<td>75</td>
<td>0</td>
<td>75</td>
</tr>
<tr>
<td>Ospan Total</td>
<td>60.6</td>
<td>12.5</td>
<td>75</td>
<td>7</td>
<td>75</td>
</tr>
</tbody>
</table>

**Lexical decision**

In the lateralized lexical decision task there were significant effects on accuracy for words compared with nonwords, \(F(1, 119) = 32.87, p < .001\), and by visual field \(F(1, 119) = 10.90, p = .001\), frequency \(F(1, 119) = 189.06, p < .001\), regularity, \(F(1, 119) = 8.63, p = .004\), and length, \(F(1, 119) = 79.99, p < .001\). The same held for RTs, although the effect of regularity was marginal, \(F(1, 119) = \)
3.18, \( p = .077 \). There were no significant interactions of these variables.

**Laterlization**

Effects of the lateralization manipulation were small but consistent and significant, see Table 5.

<table>
<thead>
<tr>
<th>Stimulus type</th>
<th>Visual field</th>
<th>Mean Correct</th>
<th>SD</th>
<th>Mean RT(ms)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequent word</td>
<td>Right</td>
<td>.817</td>
<td>.140</td>
<td>613</td>
<td>138</td>
</tr>
<tr>
<td>Frequent word</td>
<td>Left</td>
<td>.807</td>
<td>.135</td>
<td>629</td>
<td>135</td>
</tr>
<tr>
<td>Infrequent word</td>
<td>Right</td>
<td>.591</td>
<td>.180</td>
<td>722</td>
<td>193</td>
</tr>
<tr>
<td>Infrequent word</td>
<td>Left</td>
<td>.561</td>
<td>.164</td>
<td>741</td>
<td>178</td>
</tr>
<tr>
<td>Nonword</td>
<td>Right</td>
<td>.787</td>
<td>.126</td>
<td>752</td>
<td>179</td>
</tr>
<tr>
<td>Nonword</td>
<td>Left</td>
<td>.727</td>
<td>.147</td>
<td>765</td>
<td>184</td>
</tr>
</tbody>
</table>

Table 5: Differences in accuracy and response time by side of presentation.

When words were presented in the RVF accuracy was greater, \( F(1, 119) = 10.90, p = .001 \), and response times were shorter, \( F(1, 119) = 32.87, p = .004 \), than for the LVF (see Table 5). The effects on accuracy were slight for frequent words and increasingly pronounced for less frequent words and non-words. T-tests showed that differences between means were significant for nonwords and infrequent words, and differences in RT were significant for words but not for nonwords. However, there was no significant interaction of RTs and frequency.

Difference scores between the left and right visual fields were calculated and then tested for their predictive effects on the errors and RTs for the lateralized task. Testing with a linear model showed no statistically significant effects. However the possibility that difference from zero was critical suggested the use of a quadratic model (for details see Appendix 12). In the analyses using the quadratic model the effects on non-words were statistically significant with
the error rate $F(1,119) = 5.29$, $p = 0.023$, and with RTs, $F = 10.58$, $p = 0.001$. The effect on RTs to words of difference in RTs was also significant, $F(1,119) = 3.81$, $p = 0.05$, but the effect on error rates was not, $F(1,119) = 2.72$, $p = .10$.

As the difference between error rates and RTs became closer to zero the fewer errors or faster RTs the participant tended to make overall (see Figure 2, for an example).

*Reasoning and Believability*

As predicted, correct judgments of the validity of the syllogism were strongly influenced by believability when the syllogism was invalid, thus although there was not an overall effect of believability there was an interaction of believability and validity, $F(1,119) = 123.02$, $p < .001$. Valid but unbelievable syllogisms were judged to be valid less frequently ($M = .63$) than valid believable ($M = .97$). However, invalid unbelievable syllogisms were correctly found to be invalid more frequently ($M = .80$) than those that were invalid but believable ($M = .48$).

2. *Correlations*

Tables 6 and 7 give correlations discussed.

Correct comprehension of explicit information in the filler sentences and correct answers of “False” to the GP1 sentences, although not significantly correlated with each other, $r = .171$, $p = .062$, were correlated with a number of other variables, notably the vocabulary measure, and the Ospan score. Responses of “True” to GP2 sentences were positively correlated with filler accuracy, $r = .183$, $p = .046$, suggesting these answers reflected good
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Filler Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>GP1 “False”</td>
<td>.171</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>GP2 “True”</td>
<td>.183*</td>
<td>-.213**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>LLDT word error</td>
<td>-.236**</td>
<td>-.075</td>
<td>.006</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>LLDT word RT</td>
<td>.128</td>
<td>.254**</td>
<td>-.045</td>
<td>-.128</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>LLDT nonword error</td>
<td>.048</td>
<td>-.066</td>
<td>.203**</td>
<td>-.181*</td>
<td>-.204*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>LLDT nonword RT</td>
<td>.154</td>
<td>.279**</td>
<td>-.085</td>
<td>-.283**</td>
<td>.923**</td>
<td>-.134</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Synonym detection</td>
<td>.211*</td>
<td>.272**</td>
<td>-.089</td>
<td>-.277**</td>
<td>.194**</td>
<td>-.225*</td>
<td>.190*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Author recognition</td>
<td>.129</td>
<td>.202**</td>
<td>-.101</td>
<td>-.274**</td>
<td>.228*</td>
<td>-.174</td>
<td>.224**</td>
<td>.619**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Pattern score</td>
<td>.229**</td>
<td>.123</td>
<td>-.019</td>
<td>-.252**</td>
<td>.165</td>
<td>-.045</td>
<td>.233**</td>
<td>.214*</td>
<td>.215*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Pattern RT</td>
<td>.010</td>
<td>.006</td>
<td>-.078</td>
<td>.110</td>
<td>.245**</td>
<td>-.116</td>
<td>.161</td>
<td>-.146*</td>
<td>-.113</td>
<td>-.006</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Syllogism scores</td>
<td>.056</td>
<td>.219**</td>
<td>-.241**</td>
<td>-.083</td>
<td>.178</td>
<td>-.121</td>
<td>.149</td>
<td>.333**</td>
<td>.181**</td>
<td>.204*</td>
<td>-.026</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Invalid Believable Syllogisms</td>
<td>.004</td>
<td>.234**</td>
<td>-.299**</td>
<td>-.080</td>
<td>.228*</td>
<td>-.199**</td>
<td>.183*</td>
<td>.323**</td>
<td>.200*</td>
<td>.231</td>
<td>.056</td>
<td>.872**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Ospan Score</td>
<td>.202*</td>
<td>.361**</td>
<td>-.094</td>
<td>-.092</td>
<td>.277**</td>
<td>-.096</td>
<td>.275**</td>
<td>.328**</td>
<td>.278**</td>
<td>.165</td>
<td>-.017</td>
<td>.172*</td>
<td>.234**</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>Ospan Task</td>
<td>.179*</td>
<td>.407**</td>
<td>-.088</td>
<td>-.169</td>
<td>.279**</td>
<td>-.114</td>
<td>.276**</td>
<td>.369**</td>
<td>.335**</td>
<td>.187*</td>
<td>-.077</td>
<td>.191*</td>
<td>.280**</td>
<td>.920**</td>
</tr>
</tbody>
</table>

Table 6: Correlations with response variables for Experiment 1

*significant at $\alpha = .05$,  **significant at $\alpha = .01$
comprehension, and negatively, although not significantly, with answers of “False” to GP1 sentences, suggesting that readers who were more likely to accept the inference in the GP2 sentences and answer “True” were also somewhat less likely to answer “False” in the GP1 condition. GP1 accuracy was also correlated with syllogistic reasoning and nonword recognition on the lateralized LDT task.

GP2 responses of “True” were negatively correlated with nonword recognition and accuracy in the syllogistic reasoning task, particularly in the invalid believable condition. By contrast response accuracy to the filler sentences was correlated with word and not with nonword recognition on both the lateralized LDT and the simpler LDT2, and not with accuracy on the syllogistic reasoning task.

Reading times were correlated across conditions, see Table 7, including times for the syllogistic reasoning task, and with RTs on the pattern comparison task. Although reading times for the filler sentences were correlated with RTs on the LDT, there was less evidence for correlations with the garden-path sentences (GP1 reading time and LDT2 nonword RTs was an exception), however differences in RT between visual fields were correlated. (This last finding is discussed in more detail below.) It is also notable that reading times were not related to accuracy of response.

*Lexical decision*

Unsurprisingly accuracy in word recognition was related to the vocabulary measure and also to scores on the author recognition task, as was
<table>
<thead>
<tr>
<th></th>
<th>Filler reading time</th>
<th>1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>GP1 reading time</td>
<td>.737**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>GP2 reading time</td>
<td>.724**</td>
<td>.362**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Syllogism RT</td>
<td>.315**</td>
<td>.190*</td>
<td>.142</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>LLDT Frequent RT</td>
<td>.240**</td>
<td>.997</td>
<td>.095</td>
<td>.210*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>LLDT Infrequent RT</td>
<td>.258**</td>
<td>.181*</td>
<td>.184</td>
<td>.123</td>
<td>.811**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>LLDT nonword RT</td>
<td>.226**</td>
<td>.133</td>
<td>.157</td>
<td>.160</td>
<td>.834**</td>
<td>.929**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>LLDT word RT diff</td>
<td>.187</td>
<td>.139</td>
<td>.235**</td>
<td>.172</td>
<td>.122</td>
<td>.182</td>
<td>.154</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>LLDT nonword RT diff</td>
<td>.193*</td>
<td>.209**</td>
<td>.236**</td>
<td>.056</td>
<td>.031</td>
<td>.077</td>
<td>.037</td>
<td>.202*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LLDT2 word RT</td>
<td>.185</td>
<td>.172</td>
<td>.131</td>
<td>-.025</td>
<td>.236**</td>
<td>.170</td>
<td>.141</td>
<td>.083</td>
<td>.075</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>LLDT2 nonword RT</td>
<td>.132</td>
<td>.137</td>
<td>.064</td>
<td>.113</td>
<td>.153</td>
<td>.301*</td>
<td>.174</td>
<td>.172</td>
<td>-.050</td>
<td>.505**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Pattern RT</td>
<td>.238**</td>
<td>.149</td>
<td>.240**</td>
<td>.178</td>
<td>.215*</td>
<td>.227*</td>
<td>.147</td>
<td>.064</td>
<td>-.069</td>
<td>.362**</td>
<td>.270**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Synonym detection</td>
<td>.011</td>
<td>-.161</td>
<td>-.033</td>
<td>.137</td>
<td>.202*</td>
<td>.130</td>
<td>.179</td>
<td>.184</td>
<td>.064</td>
<td>-.011</td>
<td>-.088</td>
<td>.104</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>Author recognition</td>
<td>-.153</td>
<td>-.272**</td>
<td>-.165</td>
<td>-.001</td>
<td>.191*</td>
<td>.166</td>
<td>.191</td>
<td>.203*</td>
<td>-.129</td>
<td>-.031</td>
<td>-.139</td>
<td>.120</td>
<td>.641</td>
</tr>
<tr>
<td>15</td>
<td>Syllogism score</td>
<td>.045</td>
<td>.016</td>
<td>-.071</td>
<td>.253**</td>
<td>.149</td>
<td>.158</td>
<td>.158</td>
<td>.167</td>
<td>.023</td>
<td>.041</td>
<td>.102</td>
<td>.023</td>
<td>.349**</td>
</tr>
<tr>
<td>16</td>
<td>Copan Score</td>
<td>.194*</td>
<td>.072</td>
<td>.186*</td>
<td>.069</td>
<td>.296**</td>
<td>.363**</td>
<td>.272**</td>
<td>.068</td>
<td>.077</td>
<td>.054</td>
<td>.017</td>
<td>.083</td>
<td>.388**</td>
</tr>
</tbody>
</table>

Table 7: Correlations of reading and response times Experiment 1

*significant at \( \alpha = .05 \), ** significant at \( \alpha = .01 \)
accuracy to nonwords. Errors in nonword recognition correlated negatively with answers of “True” to GP2 sentences, $r = -.260$, $p = .004$, but not to response accuracy for GP1 or the unambiguous filler sentences. This was reversed for errors to words, which were correlated with filler accuracy, $r = -.239$, $p = .009$, but not with GP responses. LDT2 word scores were correlated with GP1 responses but other measures on LDT2 were not significantly correlated with reading responses or times.

RTs to the lateralized LDT were consistently correlated with the Ospan, $r = .272$ for infrequent words with the Ospan score, $r = .275$, $p = .002$, for nonwords, however error rates were not significantly related and nor were measures on LDT2. For other correlations see Table 6, and for correlations with reading times see Table 7.

The two LD tasks were expected to be strongly correlated and correlations between the tasks were significant; RTs to words for the two tasks, $r = .232$, error rates to words, $r = .210$; non-word errors were most strongly correlated, $r = .318$. That the relationships were not stronger was perhaps not surprising since the demands of the lateralization manipulation makes the tasks somewhat different; in the simpler task the letter string is on the screen until a response is made at which point the next word appears, whereas in the lateralized task the word is only on the screen for 165ms followed by a central fixation point which remains on the screen until a response is made. In addition, the frequency manipulation
in the lateralized task had a greater effect than the neighborhood manipulation in
the simpler task.

**Differences in Lateralization**

Differences in RT proved to be significantly correlated with reading times; differences in RTs to words were positively correlated with reading times to GP2
sentences, \( r = .236, p = .016 \), and with between VF differences in nonword RT, \( r = .211, p = .021 \); differences in nonword RT were positively correlated with reading
times to GP1 sentences, \( r = .229, p = .019 \), to GP2 sentences, \( r = .239, p = .014 \), and
also to filler sentences, \( r = .195, p = .048 \). These results suggest that even in
normal readers smaller differences in processing time between visual fields may
be related to improved efficiency in the form of faster reading times and, in some
situations, greater accuracy.

**Vocabulary and Reading Experience**

The synonym detection task and the author recognition task proved to be
strongly correlated both with each other, as predicted by previous research
(Cunningham, Stanovich, and Wilson, 1990), and with a number of the other
measures (see Table 5), in particular to accuracy in responses and reading times
for the GP1 sentences and with the Ospan task. This supports the notion that
these two tasks are tapping related aspects of semantic knowledge and reading
experience that are, in turn, important components in reading comprehension.
The central role of these components was further confirmed by factor analysis as
discussed below.
**Reasoning and Believability**

Accuracy on the syllogistic reasoning task was negatively related to responses of “True” to the GP2 sentences, \( r = -.211 \), and positively to answers of “False” to GP1, \( r = .220 \). The relationships were only slightly stronger for believable but invalid syllogisms.

Reasoning scores were also correlated with those for vocabulary and to a lesser extent author recognition. Overall scores were not correlated with scores on the Ospan task but there was a correlation of the Ospan with accuracy to believable but invalid syllogisms, \( r = .235 \), suggesting that susceptibility to content was negatively related to WM.

**Aospan task**

The automated span task produces an Ospan score and an Ospan total as well as three error measures. Both measures show strong correlations with other cognitive measures (see table 6). Many of these have been addressed in the sections above.

**Speed of processing**

RTs to the pattern comparison task were found to correlate to other measures of time, in particular reading times and RTs in the LDTs. There were effects of RT on reading times for GP2 and filler sentences, but not for GP1 sentences: \( F(1, 119) = 6.46, p = .012 \), and \( F(1, 119) = 5.99, p = .016 \) respectively. However there was no correlation between this measure and measures of comprehension, even though some of the measures with which there was
correlation, such as LDT RTs, were themselves correlated with comprehension measures. Accuracy on this task might be considered to reflect attention and was correlated with accuracy to the filler sentences, \( r = .229 \) but not with responses to either of the garden path sentences.

3. Modeling

*Principle components*

Principle components analysis was used as a way to explore how each of the different components contributed to overall variability. In particular examination of the first principle component extracted suggested the important role that components such as vocabulary and author recognition had in explaining total variance with a smaller role for WM and RTs for infrequent words and non words in the lateralized LDT. This analysis helped inform choices used in modeling the relationship of cognitive components and reading comprehension measures.

*Factor analysis and structural equation models*

The relationship between the measured variables was further explored with factor analysis and structural equation modeling using Amos 7.0 (Analysis of Moment Structures; Arbuckle, 2006). Factor analysis allowed a more integrated examination of relationships observed in the measured variables by combining them in a single model so that the contribution of each of the components can be assessed. Model fit is considered to be good when the value for \( \chi^2 \) is low, and the associated probability that the null hypothesis is correct is
not significant. In addition I include two other widely accepted measures of fit, the comparative fit index (CFI) which shows good fit as it approaches one, and the root mean square error of approximation (RMSEA) which show improved goodness of fit as it approaches zero, and probability is not significant.

Structural equation modeling uses maximum likelihood estimates to explain the relationships of a set of variables within a single model. The same fit indices will be reported for the initial factor analyses. It should be emphasized that latent variables have been named in line with theoretical predictions about their role in reading but other descriptors could be used. Thus the “early processing” variable described below could perhaps equally well be named “word recognition speed”.

Measures for vocabulary, author recognition, and scores for the invalid believable syllogisms, variables expected to be related to higher level processes and thus to processes of inference and integration, were expected to load on to a latent variable for “Late Processing”. This was then included in a measurement model with a working memory factor and the fit was found to be good: $\chi^2 (5, N = 119) = 5.0, p = .285, CFI = .997, RMSEA = .048$. The two factors in the model were significantly correlated, $r = .454$. Although the contribution of the invalid believable syllogisms to the late processing component was small it was retained for theoretical reasons. Without this measure the fit of the model improved slightly, $\chi^2 (4, N = 119) = .5, p = .494, CFI = 1, RMSEA = 0$, and the correlation was hardly altered, $r = .446$. 
A further model, which included an early processing factor, using RTs for frequent and infrequent words and also nonwords on the lateralized LDT, was then tested and the fit was also encouraging: $\chi^2 (11, N = 119) = 8.3, p = .682, CFI = 1, RMSEA = 0$. All three factors were found to be correlated: WM with early processing, $r = .279$, early processing with late processing, $r = .266$, and WM with late processing, $r = .456$. Including the invalid believable syllogisms in the model (Figure 3) increased the value of $\chi^2 (17, N = 119) = 17.9, p = .394$, but the fit remained good, $CFI = .999, RMSEA = .021, p = .620$.

The measurement model was then tested as part of a structural equation model that included the means for sentence comprehension responses as outcome measures. Using structural equation modeling allowed the exogenous variables to be included in a model with the endogenous variables related to reading comprehension, so an additional latent variable for comprehension was included (see Figure 4). The initial model included measures of response accuracy for all three sentence types and the fit of this model was reasonable: $\chi^2 (38, 119) = 45.3, CFI = .990, RMSEA = .040, p = .620$. However answers of “True” to the GP2 sentences were negatively correlated with other components. When the GP2 sentences were excluded from the model, the fit was improved: $\chi^2 (29, 119) = 26.3, p = .908; CFI = 1, RMSEA = 0, p = .890$. The model shows all three factors loading on to the latent variable for comprehension. This variable in turn predicts GP1 responses well and significantly predicts accuracy for the filler sentences.
In this model working memory and early processing variables were correlated, $r = .262$, in contrast with a similar model with slightly improved fit using mean RTs from the simple LDT in which they were not. When the path from WM was held to 0, the fit for the original model was significantly worse, $\chi^2(30, 119) = 34.9$, $p = .246$, CFI = .993, RMSEA = .037, $p = .634$.

The close relationship between the Ospan score and Ospan total produced a Heywood case for the latter, in the form of a correlation greater than one. However when the model implied covariance matrix was checked it was found to be proper; all Eigenvalues were greater than zero. Since the matrix was proper, the likelihood solution was also proper, and the fit indices were found to be well defined (personal communication, Daniel Serrano). Rather than adopting a theoretically less satisfactory model the Heywood case was retained.

The same model was then tested with reading times as the endogenous variable. In this version of the model relative regression weights calculated showed that the latent variable for early processing contributes more to the reading times, as might be expected (see Figure 5). The fit of this model was reasonable, $\chi^2(38, 119) = 49.43$, $p = .101$, CFI = .987, RMSEA = .050, $p = .467$. It is notable that, although their contribution to the model is slight, late processing measures contribute negatively so that high scores are related to faster reading times, but this is not true for the working memory scores on the span task, which are instead positively related to reading times in this experiment.
Cluster analysis

The value of cluster analysis for this dataset was explored using both hierarchical and K-means cluster analysis. The clusters of interest were based on sentence comprehension measures. (Reading time measures were also explored but were ultimately of less interest.) Clusters based on three or four groups were examined and using all three sentence types or garden-path sentences only. Examining the results of the K-means analysis using three groups suggested that very similar groupings occurred whether or not the filler sentences were included but since was accuracy to the fillers sentences was also a variable of interest I concentrate on the analysis based on the garden-path sentences.

The K-means analysis produced three unevenly sized clusters (12 of 13 participants in Group 3 were female, while the gender distribution in the other groups approximated that in the database as whole): Group 1, N = 76, had mean responses in the direction expected that were equivalent on both GP1 and GP2 sentences; Group 2, N = 31, answered as expected for GP1 but not for GP2; and Group 3, N = 13, answered as expected for GP2 but not for GP1 sentences. Filler accuracy was poorest for Group 2 (M = .84), and very similar for the other two groups, $F(1, 119) = 3.97$, $p = .048$.

On most of the cognitive components Groups 1 and 2 had very similar scores and RTs, however Group 3 performed worse on a number of measures, significantly for the Ospan score, $F(1, 119) = 5.13$, $p = .025$, and also for believable invalid syllogisms (M = .28). Scores on these difficult syllogisms were
numerically different between the groups: Group1 (M = .49), Group 2 (M = .60), however the difference was not significant. Although reading times did not differ across groups, reaction times for the laterialized LDT words and nonwords were significantly shorter for Group 3; however the number of errors also increased, significantly for infrequent words. (Groups were not significantly different in performance on the simpler LDT).

Discussion

The hypothesis that patterns of response to questions about the garden-path sentences would correlate with other measures of cognition was only partly supported. While correct responses to the GP1 sentences were correlated with a number of the cognitive components measured, this was not the case for GP2 sentences, suggesting that for the latter the preference for relying on inference and therefore responding “True” is not easily accounted by individual differences in those cognitive components generally associated with reading comprehension.

The correlations of correct answers of “False” to the GP1 sentences with scores on Ospan score and total is compatible with the Turner and Engle (1989) finding that the operation span task was correlated with reading comprehension, and also with that of a strong correlation with processes of updating (Miyake et al, 2000). Daneman and Merikle (1996) suggest that it is specifically efficiency in updating information in WM that explains why span measures are predictive of reading ability. Responses to GP1 sentences were also correlated with a number
of measures that collectively relate to reading comprehension, seen in the structural equation model (Figure 4), specifically the measures of vocabulary, reading experience, and scores on the LD, as well as the syllogistic reasoning tasks.

While the correlations found for GP1 responses were in contrast with the lack of such correlations for GP2 responses, the two correlations found may be revealing; answers of “True” to comprehension questions relating to the GP2 sentences were negatively related the non-word accuracy on the LDT and to accuracy in response to syllogisms that were believable but invalid. Accuracy on both of these measures potentially relates to error monitoring, although possibly at different stages of processing since factor analysis did not show the two components to load on to the same factor. Answers of “False” could result from either failure to make use of inference when it is warranted by semantic information, which in the case of GP2 sentences seems unlikely since the inference is effectively given by the fronted adjunct construction, or a preference for a more literal, and syntactically preferred, interpretation of the sentences. Following Stanovich and West (1998), this type of literal interpretation is considered likely to be associated with high scores on the measure of WM provided by the Aospan task. It is also possible that the error signal caused by the garden-path ambiguity actually makes some readers more likely to reject the plausible inference. Thus in responding to GP2 sentences, readers may end up
questioning their initial interpretation, especially at the question answering stage, leading to the longer question RTs.

Accuracy in response to questions relating to the filler sentences showed some correlation with other measures but these correlations were generally not as strong or as numerous as those with responses to GP1 sentences. This suggests that correct interpretation of the GP1 sentences places specific demands on participants that may exceed those generally required for reading comprehension. The correlations of responses on GP1 sentences and accuracy on the syllogistic reasoning task would support this interpretation, which is also supported by the finding of Siddiqui, West and Stanovich (1998) that reasoning ability and reading experience were not strongly correlated.

Vocabulary and reading experience, as measured by the synonym detection task and author recognition test, were both found to be good predictors of late processing required for correct comprehension of GP1 sentences and the more difficult filler sentences. Both measures are also positively correlated with scores on the Aospan test and the syllogistic reasoning task. The relationship between these measures is possibly best understood as a carry-over of the Matthew effect (Stanovich, 1986) in which the positive effect of high WM on early reading acquisition continues to enhance reading performance into later life. It is also possible that the clear analytic thinking required to perform well on the syllogistic reasoning task is also helpful in the synonym detection task. By contrast, a more heuristic approach could hurt scores on both tasks.
Although the effect of the early processing component in the model of sentence processing was only marginally significant, correlations of RTs from the lateraled LDT with reading comprehension measures suggests that even in normal readers efficiency in early word level processing has some influence on later sentence level interpretive processes. Efficiency in accessing and integrating information from different levels of processing appears to be critical to effective processing (Lewellen et al, 1993, Daneman & Merikle, 1996, Vogel, McCollough, & Machizawa, 2005). That this should be found at the level of differences in processing between visual field was unexpected.

The finding of an effect on the difference in response times and accuracy by visual field of presentation on both error rates and response times in the lateraled task is interesting, and provides some support for the idea that there are advantages to a reader in having both hemispheres actively involved in language processing.

The model for reading times, Figure 5, shows support for the view that reading times, even for some complex sentences, are related more to individual differences in early processes such as lexical access, than to cognitive components that influence later processing or to working memory.

In Experiment 2, the relationship of complex sentences, which require readers to use inference, was further explored with the component processes considered above. Sentence comprehension was examined with and without an
informative prior context, therefore allowing the effects of demands for integration as well as inference to be examined concurrently.

Experiment 2

Reading of target sentences in Experiment 2 were again correlated with scores on measures of component processes of reading. The GP2 sentences used in Experiment 1 were compared with sentences requiring a bridging inference (see examples 9 and 10 below) using an established paradigm (Haviland and Clark, 1974). Target sentences followed two context sentences (see examples (7), and (8), below) were intended to provide a neutral non-biasing condition (7), or to bias the reader to make a plausible inference (8). The effect of the neutral prior context should be to make readers more likely to respond “True” when the context confirms the inference, as in (8), compared to responses made when the sentences are read with a neutral context.

Although there is no inconsistency involved in the garden-path passages, alternative interpretations are possible depending on whether or not readers make a pragmatic inference that improves coherence and how the prior contextual information is integrated. Individual differences in comprehension ability are known to relate to differences in whether readers make use of inference in reading. Previous research has suggested a variety of possible causes. Inability to make inferences may cause poor comprehension or vice
versa. Problems may relate to limits to WM, specifically those relating to processing and central executive function rather than to capacity.

Problems with the processing component of WM are also thought to be at the root of difficulties that less-able readers have with integration. Possibly they are less efficient at selectively activating the most relevant information. However it is possible that the problem is simpler: perhaps poor comprehenders tend to have poorer recall than good comprehenders. Research into reading comprehension in children (Goff, Pratt, & Ong, 2005) suggests that problems may be due to difficulties with activation and integration of information from long-term memory. Thus readers with poor WM are less likely to be influenced by the supportive context than readers with high WM.

Method

Participants

One hundred and nineteen students of UNC Chapel Hill were included in the analysis. All received credit in an introductory psychology course in return for their participation and were right-handed and were native English speakers. None of these participants had taken part in Experiment 1. In addition, six participants who were not fully right-handed were excluded, as well as two outliers, and four participants who did not follow directions correctly and failed to press the space bar after reading each sentence in the passage study.
Materials and Design

Materials

There were 36 experimental passages, each consisting of three sentences, the final target sentences being identical to the GP2 sentence used in Experiment 1. The first two sentences provide a prior context that either disambiguated or was neutral to the interpretation of the target sentence. When the context was neutral (7) Tom is tattooing the customer who it could be inferred is the girl in the target sentence. However, when the context is supportive (8) he is tattooing the girl.

Neutral context
(7) Tom started work on the customer’s tattoo as a friend watched. + The design was of a Chinese dragon would be vividly colored when finished. + While Tom tattooed the girl tried not to move.

Supportive of inference
(8) Tom started work on the girl’s tattoo as her friend watched. + The design was of a Chinese dragon would be vividly colored when finished. + While Tom tattooed the girl tried not to move.

T or F: Tom tattooed the girl.

All target sentences and passages were globally coherent.

In addition to passages including the GP2 sentences used in Experiment 1 24 passages were used that in one version (9) required a bridging inference to be understood, while in the other (10) the information was given. The example shown is based on the classic work by Haviland and Clark (1974). The comprehension question is the same in both conditions.
9) Stephanie packed up a cooler for their picnic lunch. + It looked as if they should have a glorious day at the beach. + At the beach, Stephanie found that the beer was warm.

10) Stephanie packed up a cooler for their picnic lunch. + She made sure to put in some beer and sodas from the fridge. + At the beach, Stephanie found that the beer was warm.

T or F: Stephanie packed beer for the picnic.

This material introduces a comparison with the other experimental sentences, as in this case there is no problem of local coherence, since the target sentences are unambiguous. However, the target sentence is expected to be read more slowly when the inference that beer is included in the picnic supplies must be inferred, as in (9), rather than is given, as in (10).

Comprehension questions following these passages were more frequently correctly answered “False” to balance the response contingencies, and did not in most cases relate to the target sentence.

Procedure

Participants read the passages at their own pace. Each of the three sentences was shown separately on the screen. After reading each sentence the participant pressed the space bar to move on either to the next sentence or, after the third sentence, to the comprehension question. Once a key had been pressed to respond either “True” or “False” to the question the first sentence of the next passage appeared.

78
The tests for the individual reading measures were completed as in the first experiment, although the simpler LDT2 was omitted for brevity. The whole experiment took participants just under an hour to complete.

**Results**

Analyses for Experiment 2 duplicated in most respects those used for Experiment 1.

1. **Descriptive Statistics**

_Sentence reading and Comprehension_

Responses and reading times were again the main variables of interest. Responses to the passages containing garden-path sentences were more frequently “True” when the context was supportive than neutral, $F(1, 118) = 100.87, p < .001$. However, it should be noted that there were generally more answers of “True” in the garden-path condition following the passages than for the sentences in Experiment 1.

<table>
<thead>
<tr>
<th></th>
<th>GP neutral context “True”</th>
<th>GP supportive context “True”</th>
<th>Bridging - inference</th>
<th>Bridging - no inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>88%</td>
<td>93%</td>
<td>91%</td>
<td>94%</td>
</tr>
<tr>
<td>Sentence 3</td>
<td>2493ms</td>
<td>2576ms</td>
<td>2214ms</td>
<td>2138ms</td>
</tr>
</tbody>
</table>

Table 8: Response accuracy and reading times to target sentences in Experiment 2
As expected reading times for the target sentence were shorter in the supportive condition, $F(1, 118) = 17.62, p < .001$. As predicted, reading times for the target sentences were slower when the bridging inference was required, $F(1, 118) = 47.0, p < .001$. Accuracy following the bridging passages was high, with more than 25% of participants getting all questions correct.

**Measures of component processes**

Mean scores in response to the synonym detection, author recognition, pattern comparison, and Aospan tasks were almost identical to those found in Experiment 1.

<table>
<thead>
<tr>
<th>Task</th>
<th>M</th>
<th>SD</th>
<th>Out of</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synonym detection</td>
<td>.67</td>
<td>.10</td>
<td>1</td>
<td>.30</td>
<td>.90</td>
</tr>
<tr>
<td>Author recognition</td>
<td>12</td>
<td>5.3</td>
<td>40</td>
<td>-4</td>
<td>26</td>
</tr>
<tr>
<td>Lateralized LDT – frequent word error rate</td>
<td>.20</td>
<td>.11</td>
<td>1</td>
<td>0</td>
<td>.62</td>
</tr>
<tr>
<td>frequent word RT (ms)</td>
<td>665</td>
<td>185</td>
<td>-</td>
<td>117</td>
<td>1513</td>
</tr>
<tr>
<td>infrequent word error rate</td>
<td>.42</td>
<td>.13</td>
<td>1</td>
<td>.09</td>
<td>.78</td>
</tr>
<tr>
<td>infrequent word RT (ms)</td>
<td>767</td>
<td>208</td>
<td>-</td>
<td>152</td>
<td>1459</td>
</tr>
<tr>
<td>nonword error rate</td>
<td>.26</td>
<td>.13</td>
<td>1</td>
<td>.03</td>
<td>.59</td>
</tr>
<tr>
<td>nonword RT (ms)</td>
<td>807</td>
<td>215</td>
<td>-</td>
<td>174</td>
<td>1613</td>
</tr>
<tr>
<td>Syllogistic reasoning</td>
<td>18</td>
<td>3.8</td>
<td>24</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>Syllogistic reasoning RT (ms)</td>
<td>1747</td>
<td>947</td>
<td>-</td>
<td>498</td>
<td>5392</td>
</tr>
<tr>
<td>Pattern comparison score</td>
<td>55.6</td>
<td>2.6</td>
<td>60</td>
<td>44</td>
<td>59</td>
</tr>
<tr>
<td>Pattern comparison RT</td>
<td>1315</td>
<td>211</td>
<td>-</td>
<td>979</td>
<td>2143</td>
</tr>
<tr>
<td>Ospan Score</td>
<td>45.9</td>
<td>16.8</td>
<td>75</td>
<td>3</td>
<td>75</td>
</tr>
<tr>
<td>Ospan Total</td>
<td>60.1</td>
<td>11.6</td>
<td>75</td>
<td>22</td>
<td>75</td>
</tr>
</tbody>
</table>

Table 9: Descriptive statistics for component measures in Experiment 2
Overall accuracy in response to syllogisms was marginally better than in Experiment 1. In particular, invalid believable syllogisms were correctly found to be invalid slightly more often (M = .52). There was also one subject who answered every one of the 24 syllogisms correctly. Stanovich and West (1998) also found a low rate of participants answering all syllogisms correctly.

In the lateralized LDT, RTs tended to be longer and more variable than in Experiment 1, while accuracy was marginally poorer. However there were no significant differences between the experiments. There were large main effects of frequency and word/nonword on RTs, $F(1, 118) = 631.89, p < .001$ and $F(1, 118) = 756.82, p < .001$ respectively, and of accuracy on frequency, $F(1, 118) = 18.02, p < .001$ but not on word/nonword recognition.

<table>
<thead>
<tr>
<th>Response</th>
<th>Experiment</th>
<th>N</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDT Word Error Score</td>
<td>1</td>
<td>120</td>
<td>2.72</td>
<td>.102</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>119</td>
<td>20.59</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Pooled 1 and 2</td>
<td>239</td>
<td>21.08</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LDT Non-Word Error Score</td>
<td>1</td>
<td>120</td>
<td>5.29</td>
<td>.023</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>119</td>
<td>19.30</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Pooled 1 and 2</td>
<td>239</td>
<td>28.07</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LDT Word Mean Time (ms)</td>
<td>1</td>
<td>120</td>
<td>3.81</td>
<td>.053</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>119</td>
<td>19.72</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Pooled 1 and 2</td>
<td>239</td>
<td>24.98</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LDT Non-Word Mean Time (ms)</td>
<td>1</td>
<td>120</td>
<td>10.58</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>119</td>
<td>18.70</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Pooled 1 and 2</td>
<td>239</td>
<td>32.70</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Table 10: Effects of difference in error rates and RTs in the lateral LDT.
The effects of lateralization between the two experiments were remarkably similar.

Differences in word error rates and RTs between visual field were tested with a quadratic model, as in Experiment 1, and were found to have significant effects on accuracy and RTs in the LDT; effects of difference in RTs to words on RTs to words, \( F(1, 118) = 19.72, p < .001 \), of difference in error rates for words on error rates to words, \( F(1, 118) = 20.59, p < .001 \), of nonword RT differences on nonword RTs, \( F(1, 118) = 18.70, p < .002 \), and difference in error rates to nonwords on error rates in recognizing non-words, \( F(1, 118) = 19.30, p < .001 \) (See Figures 6 and 7).

When the results of the two experiments were pooled (see Table 10) effects remained consistent. In each case the pooled results provided stronger evidence for the effects of difference. Differences closer to zero have the effect of reducing response times and error rates.

2. Correlations

Reponses to the garden-path sentences across the two passage manipulations were strongly correlated, \( r = .680, p < .001 \). Participants who answered “True” in the supportive condition were more likely to answer “True” in the neutral condition. There were no other correlations of response, possibly in part because the accuracy in response to sentences following the bridging passages was high.

Reading times were correlated across sentences and by condition (see
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sentence 1 BR supported</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Sentence 1 BR inference</td>
<td>.663</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Sentence 1 GP supported</td>
<td>.669</td>
<td>.713</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Sentence 1 GP inference</td>
<td>.745</td>
<td>.718</td>
<td>.772</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Sentence 2 BR supported</td>
<td>.447</td>
<td>.566</td>
<td>.574</td>
<td>.514</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Sentence 2 BR inference</td>
<td>.649</td>
<td>.452</td>
<td>.440</td>
<td>.609</td>
<td>.523</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Sentence 2 GP supported</td>
<td>.564</td>
<td>.529</td>
<td>.533</td>
<td>.619</td>
<td>.676</td>
<td>.671</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Sentence 2 GP inference</td>
<td>.503</td>
<td>.560</td>
<td>.587</td>
<td>.581</td>
<td>.724</td>
<td>.603</td>
<td>.672</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Sentence 3 BR supported</td>
<td>.523</td>
<td>.452</td>
<td>.499</td>
<td>.451</td>
<td>.735</td>
<td>.502</td>
<td>.639</td>
<td>.540</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Sentence 3 BR inference</td>
<td>.516</td>
<td>.392</td>
<td>.406</td>
<td>.337</td>
<td>.437</td>
<td>.651</td>
<td>.506</td>
<td>.533</td>
<td>.535</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Sentence 3 GP supported</td>
<td>.352</td>
<td>.341</td>
<td>.322</td>
<td>.270</td>
<td>.406</td>
<td>.252</td>
<td>.379</td>
<td>.448</td>
<td>.401</td>
<td>.398</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Sentence 3 GP inference</td>
<td>.376</td>
<td>.364</td>
<td>.358</td>
<td>.431</td>
<td>.419</td>
<td>.502</td>
<td>.541</td>
<td>.553</td>
<td>.427</td>
<td>.467</td>
<td>.363</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>LLD T Frequent RT</td>
<td>4.00</td>
<td>3.76</td>
<td>4.02</td>
<td>3.25</td>
<td>3.25</td>
<td>3.26</td>
<td>3.49</td>
<td>2.59</td>
<td>2.69</td>
<td>1.90</td>
<td>1.54</td>
<td>1.38</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>LLD T nonword RT</td>
<td>.356</td>
<td>.321</td>
<td>.318</td>
<td>.305</td>
<td>.268</td>
<td>.250</td>
<td>.178</td>
<td>.224</td>
<td>.199</td>
<td>.196</td>
<td>.107</td>
<td>.078</td>
<td>.851</td>
<td>.932</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>Pattern RT</td>
<td>.077</td>
<td>.035</td>
<td>.165</td>
<td>.115</td>
<td>.034</td>
<td>.059</td>
<td>.117</td>
<td>.019</td>
<td>.019</td>
<td>.024</td>
<td>.138</td>
<td>.299</td>
<td>.255</td>
<td>.240</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>Invalid Believable RT</td>
<td>-.031</td>
<td>-.023</td>
<td>.039</td>
<td>.044</td>
<td>.122</td>
<td>.036</td>
<td>.118</td>
<td>.096</td>
<td>.079</td>
<td>.004</td>
<td>.289</td>
<td>.132</td>
<td>-.009</td>
<td>-.072</td>
<td>-.113</td>
</tr>
<tr>
<td>18</td>
<td>Author recognition</td>
<td>-.073</td>
<td>-.228</td>
<td>-.180</td>
<td>-.169</td>
<td>-.315</td>
<td>-.214</td>
<td>-.366</td>
<td>-.231</td>
<td>-.101</td>
<td>-.147</td>
<td>-.127</td>
<td>-.040</td>
<td>-.068</td>
<td>-.013</td>
<td>-.240</td>
</tr>
<tr>
<td>19</td>
<td>Synonym detection</td>
<td>-.025</td>
<td>-.164</td>
<td>-.098</td>
<td>-.100</td>
<td>-.197</td>
<td>-.237</td>
<td>-.242</td>
<td>-.171</td>
<td>-.242</td>
<td>-.017</td>
<td>-.017</td>
<td>-.136</td>
<td>-.114</td>
<td>.022</td>
<td>.052</td>
</tr>
</tbody>
</table>

Table 11: Correlations with reading times for Experiment 2
Table 11). Reading times for the target sentence in the bridging passages with and without the need for inference are correlated $r = .535$, and reading times for the sentences requiring the bridging inference are also correlated with those for GP sentences requiring inference, $r = .467$. This correlation is stronger than that for the GP sentences in the two context conditions, $r = .363$.

The main other correlations of interest in Experiment 2 were those of reading times and RTs to the LDT. Correlations are strongest with sentence 1 in each manipulation, are slightly reduced for sentence 2, and are not present for sentence 3 in the garden-path condition. They are also stronger for words than for nonwords. Faster reading times also tended to be related to higher scores on the author recognition test and to a lesser extent with the synonym detection task.

Expected responses of “True” to the garden-path sentences requiring inference were negatively correlated with reading times for garden-path sentences where no inference was required, $r = -.270, p = .003$, and also with target sentence reading times in the bridging condition, $r = -.206, p = .024$, suggesting that failure to answer “True” was not related to difficulty in making the inference.

In contrast to Experiment 1, there were few correlations of responses following the passage reading task with other measures. However, as most of those correlations seen in Experiment 1 were with GP1 responses of “false” and with accuracy to the fillers, not with the GP2 sentences used in Experiment 2, this
is not surprising. Responses following the GP sentences requiring inference were again negatively correlated with scores for the invalid believable syllogisms, $r = -.208$, $p = .023$, but were not significantly correlated with nonword accuracy on the LDT.

Correlations across the tests of the cognitive components were similar to those seen in Experiment 1. Scores on the synonyms, author recognition, pattern comparison, syllogisms, and Ospan were all positively correlated. However, RTs and error rates on the LDT showed little correlation. Although there were no significant correlations of the difference in scores by visual field with responses to the passage questions, word accuracy differences were correlated with scores on the syllogistic reasoning task, $r = .195$, and with the invalid believable syllogisms, $r = .184$, and with other measures on the LDT.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synonym detection</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author recognition</td>
<td>.563**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern score</td>
<td>.303**</td>
<td>.360**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern mean RT</td>
<td>-.083</td>
<td>-.240**</td>
<td>-.095</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllogistic reasoning</td>
<td>.422**</td>
<td>.242**</td>
<td>.274**</td>
<td>.153</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invalid Believable syllogisms</td>
<td>.412**</td>
<td>.189*</td>
<td>.255**</td>
<td>.103</td>
<td>.903**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ospan score</td>
<td>.375**</td>
<td>.263**</td>
<td>.327**</td>
<td>.069</td>
<td>.357**</td>
<td>.346**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Ospan Total</td>
<td>.384**</td>
<td>.249**</td>
<td>.358**</td>
<td>.041</td>
<td>.376**</td>
<td>.363**</td>
<td>.927**</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 12: Correlations between tasks in Experiment 2
3. **Modeling**

The measurement model tested for Experiment 1 (Figure 2) was tested with the data from Experiment 2 (see Figure 8). The resulting $\chi^2 (17, N = 118) = 28.7, p = .038$, is problematic. Even though the fit was reasonable, CFI = .985, RMSEA = .074, $p = .188$, and the difference in $\chi^2$ is not significantly different from that found for Experiment 1, there is a significant probability that the null hypothesis should not be rejected. Correlations between the latent variables were similar for Late processing and WM, $r = .450$, and only slightly less than that found in Experiment 1 for Early processing and WM, $r = .194$. However, the correlation between Early and Late processing was not significant, $r = .053$.

Further, when the model tested in Experiment 1 was tested on the data from Experiment 2 with the responses relating to the garden-path sentences the model was inadmissible due to the error associated with the GP sentences that required inference. Omitting these responses and testing on responses to the bridging passages also produced an inadmissible outcome. Although, when the nested model was tested, which included only the responses for those passages that did not require inference, it produced reasonable fit, $\chi^2 (29, 118) = 39.7, p = .090$, CFI = .986, RMSEA = .054, $p = .407$, the standardized regression weights support the conclusion that the measures tested are not strongly related to comprehension of these sentences (see Figure 9).

Given the pattern of correlations observed, a model for reading times was also examined as in Experiment 1. A model with target sentence reading times
as the endogenous variables produced $\chi^2(41, 118) = 54.8, p = .074, CFI = .979, RMSEA = .052, p = .444$. (A more complicated model which included lateralization differences and error rates on the lateralized LDT was tested but had poor fit and is not included.) In the model shown here (see Figure 10) the relationship of reading times with the early processing variable is again clear, while WM and late processing make a smaller, non-significant contribution; weights for these two latent variables are both negative since higher scores are related to shorter reading times. This is in contrast to Experiment 1 for which the relationship with the span measures was positive.

**Cluster Analysis**

K means cluster analysis was performed as for Experiment 1. The four groups produced consisted of a large group of 51 who achieved overall good scores, a smaller group of 39 participants whose accuracy was reduced in the bridging where inference was required. 15 participants had low scores in the bridging passages, and 14 had low scores compared to those expected following the garden-path passages.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Number</th>
<th>Bridging Supported</th>
<th>Bridging Inference</th>
<th>Garden path supported</th>
<th>Garden path Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51</td>
<td>.913</td>
<td>.968</td>
<td>.978</td>
<td>.923</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>.989</td>
<td>.955</td>
<td>.871</td>
<td>.927</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>.861</td>
<td>.766</td>
<td>.903</td>
<td>.851</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>.952</td>
<td>.922</td>
<td>.738</td>
<td>.642</td>
</tr>
</tbody>
</table>

Table 13: Response accuracy for different passage types by cluster.
Group 3 had generally higher scores on the syllogistic reasoning task, especially when the syllogism was believable but invalid.

Discussion

Although responses and reading times for the passage reading task produced results that were as predicted, there was very little evidence of correlations with scores and RTs on the other tests completed with the exception of the lateralized LDT. When the context supported the inference in the garden-path sentences, there was a greater percentage of responses that were consistent with the inference than when the context was neutral. However neither of these effects appeared to show much relationship with the other cognitive components measured. High scores on the Aospan task and other component measures had been expected to correlate with more accurate responding, as found for the fillers and GP1 sentences in Experiment 1, and by extension with shorter reading times but this was not found.

The effects and correlations of most interest in Experiment 2 related to the lexical decision task. These findings support the suggestion that, when reading is not hampered by difficulties of interpretation, reading time is most strongly correlated with measures that relate to early levels of processing such as lexical decision. Thus, in Experiment 1 there were correlations with filler sentence reading times that were not found for the garden-path sentences. Further, although each of these sentences was presented separately on the screen, the existence of context creates a more natural reading situation. The almost
complete lack of correlation of reading times with scores on the Aospan task would also suggest that reading of this level of difficulty makes few demands on higher level cognitive processes. The correlations of the author recognition task with reading times but not with scores on the LDT (see Table 8) further illustrate the complexity of these relationships.

The effects of difference between visual field on scores on the LDT show the indirect effect that a difference in processing at one level may have on a higher level. Although correlations were found with the LDT they were not found with reading times, and did not fit easily into a model of reading.

The higher proportion of answers of “True” following the garden-path sentences in Experiment 2 may result from the lack of strong counter examples to the acceptability of employing plausible inference provided by the GP 1 sentences. However, it is also possible that reading these sentences with a prior context, even when that context is neutral to the plausible inference, makes accepting the inference easier.

The fact that the negative correlation of scores on the invalid believable syllogisms was found in this experiment as well as in Experiment 1 suggests that this is a consistent relationship. Readers who are better able to ignore the effects of believability may also be less likely to accept a plausible inference, especially when it arises from an ambiguous text.

The presence of context made comprehension of the garden-path sentences easier, as reflected by shorter reading times. Since this ease of
integration reduces the need for higher level cognition the lack of correlation with measures of higher level cognitive processes is not surprising, especially since responses for these garden-path sentences (GP2) were not correlated with those measures in Experiment 1. The correlation of test of syllogistic reasoning with a tendency to make the nonstandard response and with the Aospan task suggests that a proportion of participants who perform well on the reasoning task choose a more literal interpretation over the interpretation that requires inference. Correlation of reading experience as measured by the author recognition task but not with the span task suggest that responses and reading may have more to do with readers’ expectations of text and their ability to keep relevant information active in long term memory.
CHAPTER 3
GENERAL DISCUSSION

This study explored the way in which individual differences in interpretation and of reading ability are related to other measures of cognitive ability. Using sentences which contained a fronted adjunct ambiguity provided an opportunity to explore a wide range of interpretation, since previous research (Christianson et al, 2001, Michael, 2004) had found that there are large individual differences in how these sentences are understood. Since differences in interpretation had also been shown to influence reading times these provided an additional avenue of enquiry.

The expectation that differences in interpretation would be correlated with other measures of cognition was supported in the case of the GP1 sentences for which the syntax and correct interpretation were consistent and the initial ambiguity required reanalysis to reach a correct interpretation, and the correct answer is “False”. Correct answers by readers following these sentences were correlated with higher scores on a number of other cognitive tasks, in particular the vocabulary and span task but also with the syllogistic reasoning task, but slower reaction times to the lateralized lexical decision task. However, individual differences in response to the GP2 sentences, which required
accepting a plausible inference to arrive at the expected answer, showed very little correlation with other cognitive measures. The negative correlations with the syllogistic reasoning task, particularly to the invalid but believable syllogisms, suggest that participants who are less susceptible to the believability manipulation are also less likely to make use of the plausible inference unless it is well supported by the text. Thus answers of “False” for the GP1 sentences are positively correlated with answers of “False” for the GP2 sentences. However, the positive correlation of responses of “True” following the GP2 sentences with accuracy of response to the filler sentences suggest that these responses do not imply a lack of accuracy in response but instead a difference in how the sentences are being understood. The results of the cluster analysis confirmed that most readers, 63% in Experiment 1, answer both sentence types as expected. Of the remainder, 13 participants answered the GP2 sentences as expected, preferring to accept the plausible inference, while the remaining 26% generally did not. In Experiment 2, 76% answered the GP2 sentences “True” as expected even when the context was neutral.

The results of Experiment 2 further support the finding from Experiment 1 that responses to GP2 sentences are not correlated with other cognitive processes, with the exception of the syllogistic reasoning task. Thus the choice to accept a plausible inference when that inference arises out of ambiguity is idiosyncratic, varying between individuals in a way that is proving difficult to predict.
The role of higher level cognitive processes

An important finding of Experiment 1 is the confirmation of the role of higher level processes in reading comprehension in situations where there is inconsistency or difficulty. The correlations of reading accuracy for both the filler sentences and correct responses of “False” to the GP1 sentences with a number of other measures, in particular the measures of WM, vocabulary, and reading experience suggest that for sentences of this type, for which more conscious or top-down processing is required, more able readers have an advantage. Further, given the way in which these measures are also correlated with each other, that readers who have, for instance, good WM and good vocabulary skills are also likely to have considerable print exposure and are more capable readers. The role of these skills is likely to be interactive in an extension of the Matthew effect described by Stanovich (1986). Having abilities that make reading easy fosters an interest in reading which in turn further enhances those abilities. Vocabulary knowledge is strongly associated with reading rather than exposure to speech or other forms of media, so the strong relationship between the synonym detection task and the author recognition task seen in both experiments is not surprising. However, the synonym detection task also makes demands on a participant’s analytic skills and attention, as well as their semantic memory, to select the correct response; for instance, to determine that the correct synonym for generic is commonplace and not hereditary or customary.
In this context, the syllogistic reasoning task could be considered as a special case of reading difficulty, as the task makes demands on the reader that exceed those needed for more “normal” reading. As expected, scores on this task correlate with scores for WM, synonym detection, and the author recognition. Correlations are still more pronounced for the invalid believable syllogisms for which the demands are greater, since the reader must ignore real world knowledge that suggests that the syllogism is valid. De Neys (2006) showed that in a dual task situation low span readers were more impaired by this condition than high span readers.

*The role of lower level cognitive processes*

The correlation of sentence reading times with the RTs and error rates from the LDT provides evidence for the role that early processes such as word recognition play in reading, not just of children but of “normal” adult readers, at least in situations where the text is easily understood. However, it is also clear that as complexity increases, as in the case of the garden-path sentences in Experiment 2, that relationship of reading time with these relatively automatic processes are submerged as more conscious higher levels of processing are recruited. Calvo (2005) found effects of reading speed on early level processes observed during eye-tracking of reading sentences requiring inference but no effects on later reading measures.

Given the findings of correlations of the lateralized LDT with a number of other measures, the relative absence of similar correlations for the simpler LDT2
task is striking. The short presentation time required for the lateralized task put demands on memory, albeit briefly, that is not required in the simpler task. This effectively reduces the quality of the representation for the reader. For frequent words and/or skilled readers this is not a problem, however for low frequency words, nonwords, or less skilled readers the limited viewing time is likely to be problematic. Perfetti and Hart (2001) suggest that when word recognition is efficient the reader has more resources available for comprehension. Poor readers, by contrast, have fewer resources available to cope with text complexity. However, the evidence of the experiments presented here suggests that the advantage provided by efficient lexical processing may be valuable when text is demanding, but is less relevant when text is ambiguous or inconsistent.

Perfetti and Hart (2001) presented lexically ambiguous stimuli at variable intervals of 150, 450, and 2000ms. Less skilled readers were slower to respond in all conditions; however the delay caused by the homophones appeared at the longer presentation intervals while for the more skilled readers it appeared only at the shortest interval. The apparent discrepancy is similar to that seen in the Long and Chong (2001) study, in which less able readers read inconsistent text more quickly than more able readers.

By limiting presentation time, it seems that the lateralized task effectively recreates the pressures put on less effective readers by more demanding text. Readers pressed to make a response after brief presentation of a letter string are
reasonably fast and accurate if the stimulus is a familiar word but much less accurate and slower if the word is less familiar or not a word at all.

Using information from both visual fields

Perhaps the most surprising result found from these two experiments is that the ability to use information from both visual fields as equally as possible appears to be valuable in improving overall speed and accuracy on the LDT. The intriguing underlying possibility that this suggests is that readers who activate information in both hemispheres equally may have an advantage in language processing. Further research using techniques that could test this hypothesis more directly would be needed before such a claim could be advanced.

Speed of processing

The pattern comparison was included to check whether speed of processing had an effect on individual differences in reading. RTs on the pattern comparison task was correlated with both LDT tasks and with some sentence reading times but not interpretation in Experiment 1, and not for any sentence reading times in Experiment 2 (see Table 9?). Further, including pattern RTs in the model for reading reduced the model fit. The combined evidence seems to suggest that individuals who are fast and accurate on one task are likely to be fast and accurate on another but that reading speed and accuracy is little influenced by more generalized speed of processing.
Single or Dual resource for language processing

An issue that this research did not set out to address, but which it has become apparent is relevant, is the question whether language processing, in particular reading, relies on the same cognitive components as other verbally mediated cognitive processes. The debate has divided over whether there is a single resource (SR) as hypothesized by Just and Carpenter (1992), which supports all verbally mediated processes including those for language, or a resource specific to language processing, the Separate Language Interpretation Resource (SLIR) hypothesis proposed by Caplan and Waters (1999).

Caplan and Waters (1999) use the term post-interpretive processes, which might better be called late interpretive processes, for that stage of processing in reading which exceeds average reading times, and involves reasoning, problem solving or resolution of other problems such as inconsistency or ambiguity. In eye-tracking these late processes are represented mainly by rereading and spillover effects, and may relate to some level of comprehension difficulty. Although there is considerable evidence that more complex structures make greater processing demands, Caplan and Waters make the point that there has been little success in finding effects of WM even in reading quite difficult syntactic constructions.

The research described here would support the view that when reading is difficult due to inconsistency, or demands on memory then, as described above, differences in higher cognitive processes have an effect on processing. However
there has been a lack of evidence for such effects when the text being read makes no more than moderate demands on processing resources, although reading is influenced by earlier processes such as those measured by the lateralized LDT.

**The role of context and integration**

As expected, readers were able to make use of the supportive context to facilitate processing in both the bridging and garden-path conditions. However there was little evidence that the ability to make use of contextual information was related to the other cognitive abilities measured.

The role of contextual information in minimizing problems caused by ambiguity are well-known and had been thought to relate to WM (Gernsbacher, 1997). However the results of Experiment 2 provide no support for this. There was no evidence of a significant relationship between scores on the Aospan task and reading times or responses in Experiment 2 and, in particular, no evidence that effective use of the supportive context was related to individual differences in WM. Thus, there was no evidence for an “integration bottleneck”, although there was probably insufficient difficulty or inconsistency in the passages used here to test that idea.

**Effectiveness of Measures**

In attempting to address the role of individual differences choices had to be made amongst the many tests that could have been used. The use of the automated Aospan task rather than a version of the reading span task was justified by results; a test of verbally mediated WM, which does not require
reading or word recognition, still was seen to be correlated with certain reading processes. This was particularly the case for sentences where demands on processing were high, as in cases of inconsistency.

Both the synonym detection task and author recognition test were dependable across the two experiments and in the way they correlated with each other and with other measures. This consistency made interpreting differences in correlation particularly valuable, for instance for Experiment 2, in which correlations for the author recognition test provided evidence of the important role of print exposure in passage reading.

The syllogistic reasoning task also produced results that helped clarify why some readers were less likely to make use of a plausible inference even when most other readers had no such problem.

Overall, the differences between the tasks was advantageous at a practical level in keeping participants interested, but more importantly it allowed correlation that were found to be interpreted without concerns about overlapping task demands.

Conclusion

Reading is a highly overlearned, largely automatic process. The massive amount of experience that most of us have with text allows us to make use of a depth of knowledge much of which we may not consciously be aware of, such as preferred usage or the probabilities of word pairings. However that knowledge can serve to disrupt the reading process when we come across an unusual word
use or an uncommon structure, a fact that psycholinguists have been exploring fruitfully for many years.

Most normal reading makes minimal demands on resources or processing (i.e. on WM) for experienced adult readers. As we read a representation of the discourse is assembled, then held in LTM to be reactivated as necessary. It is only when there are severe demands on the reader either due to the difficulty of the text, or the need to recall content accurately, or because of distraction, that individual differences in processing ability or capacity become apparent. Thus research has shown that low span readers have problems with making use of inference and of integrating text in the face of inconsistency (Long and Chong, 2001). However most “normal” readers’ everyday experience of text may not be particularly difficult or demanding, nor do readers often experience inconsistency or ambiguity as they read.

Not every reader interprets text the same way and the interesting and important question that this research sought to answer is why or, more properly, why not. This research showed that the answer is complex but that certain patterns are predictable. Individual differences in processes like lexical access generally considered to occur early in comprehension affect reading of simple text. As text becomes more difficult these differences continue but other factors such as experience, semantic knowledge and WM become increasingly relevant.

Interpretation of text happens in the head not on the page. Understanding does not happen in a vacuum and certain interpretations are expected, but not
inevitable. Prior experience may constrain how we interpret a sentence and makes certain interpretations predictable, but not for everyone, not even for every “normal” reader. Real world knowledge also acts to guide our interpretation of what we read. How processes of logical thinking are applied can also influence an individual’s interpretation.

In a great deal of research into language comprehension, normal adult readers are treated as equal, yet large differences on wide range of measures of language processing can be found within the undergraduate community of a selective university such as the University of North Carolina at Chapel Hill. Students may not be getting everything from their reading that might be expected and yet are generally unaware that their understanding is imperfect. A better understanding of the individual differences that drive these differences in comprehension would be a major contribution to cognitive research into reading. The research described here suggests some future directions in which this research might go. In particular, better understanding of some of the underlying processes and how they interact with processes at higher levels of cognition could be both valuable and interesting.
Figure 1: Proposed model showing how measures of individual difference represent the processes involved in reading comprehension
Figure 2: Plot of difference in reaction time and reaction time to nonwords in the lateralized lexical decision task for Experiment 1 showing quadratic model.
Figure 3: Measurement model for Experiment 1 showing factor loadings and correlations between latent variables.
Figure 4: Structural equation model for Experiment 1 responses to GPI and filler sentences showing standardized regression weights on latent variables.
Figure 5: Structural equation model for Experiment 1 sentence reading times showing standardized regression weights on latent variables.
Figure 6: Plots of word RT difference and word RT in the LLDT for Experiment 1 and 2.
Figure 7: Plot of pooled RTs and RT difference to words in the lateralized lexical decision task, for Experiment 1 and 2.
Figure 8: Measurement model for Experiment 2 showing factor loadings and correlations between latent variables.
Figure 9: Structural equation model for Experiment 2 responses showing factor loadings on latent variables.
Figure 10: Structural equation model for Experiment 2 reading times to target sentences showing factor loadings between latent variables.
Appendix 1: Garden-path sentences in Experiment 1  
(shown in Courier New)

/ indicates where words differ between conditions. Words between are for GP1, words after for GP2.

1. While Dave drank the brandy /arrived from the cellar/soothed his sore throat.  
True or False: Dave drank the brandy.

2. As Sarah painted the picture /dried in the next studio/ improved beyond recognition.  
True or False: Sarah painted the picture.

3. As the team climbed the mountain /dominated the horizon/ loomed above them.  
True or False: The team climbed the mountain.

4. As Laurel wrote the new novel /sold well at the bookstore/ grew more and more complicated.  
True or False: Laurel wrote her new novel.

5. While Martin programmed his old computer /landed in the recycling/ persisted in causing problems.  
True or False: Martin programmed his old computer.

6. While the chef cooked the cake /stood ready on the side table/ filled the room with a pleasant smell.  
True or False: The chef cooked the cake.

7. As Rowan read the letter /dropped into her mailbox/ confused her increasingly.  
True or False: Rowan read the letter.

While the team sailed their new yacht /lay in dry dock/ developed a leak.  
True or False: The team sailed their new yacht.

8. As Jeff parked the van /collided with him /scratched the curb.  
True or False: Jeff parked the van.

10. While Mandy studied the textbook /tumbled off the shelf/ provided the answer.  
True or False: Mandy studied the textbook.

11. As the magician juggled the plates /stood ready for the next trick/ flew up higher and higher.  
True or False: The magician juggled the balls.

12. As Patrick walked his dog /stayed in its pen/ tugged on the leash.  
True or False: Patrick walked his dog.

13. As Rebecca polished the vase /fired in the kiln/ acquired a nice shine.
True or False: Rebecca polished the vase.

14. As the secretary typed the urgent letter /left express mail/ the urgent letter was long overdue.
True or False: The secretary typed the urgent letter.

15. While Nick filmed the next scene /required a rewrite/ ran quite perfectly.
True or False: Nick filmed the next scene.

16. As Janet edited the article /returned from the layout artist/ became easier to understand.
True or False: Janet edited the article.

17. While the children played the piano /was quite neglected/ sounded out of tune.
True or False: The children played the piano.

18. As Kevin grilled the sausages /cooled on the rack/ burnt to a crisp.
True or False: Kevin grilled the sausages.

19. As Gloria sculpted the statue /appeared at the gallery/ drew praise from her students.
True or False: Gloria sculpted the statue.

20. As the boys wrestled Jim /watched in excitement/ struggled to break free.
True or False: The boys wrestled Jim.

21. While Matt drew the actress /admired his paintings/ struck a dramatic pose.
True or False: Matt drew the actress.

22. While Marilyn lectured her student /researched the next topic/ listened very carefully.
True or False: Marilyn lectured her student.

23. As the lawyer cross-examined the defendant /whispered to his attorney/ collapsed in the witness box.
True or False: The lawyer cross-examined the defendant.

24. While Wallace explored the island /disappeared beyond the horizon/ lay spread out before him.
True or False: Wallace explored the island.

25. While the cat groomed her kitten /scurried under the sofa/ protested vigorously.
True or False: The cat groomed her kitten.

26. As Harry drove the car /emerged from a small side road/ began to make a strange noise.
True or False: Harry drove the car.

27. While Glenn tattooed the girl /paced anxiously outside/ attempted to stay still.
True or False: Glenn tattooed the girl.
28. As the friends cleaned the living room /needed doing next/ looked much better.
True or False: The friends cleaned their house.

29. While Tom washed up the dirty dishes /soaked in the sink/ continued to accumulate.
True or False: Tom washed the dirty dishes.

30. While Sandra counted the children /hurried to hide/ formed a line.
True or False: Sandra counted the children.

31. As the marine attacked the general /observed from the dug out/ tried to defend himself.
True or False: The marine attacked the general.

32. While Linda photographed the model /returned from lunch/ posed languidly.
True or False: Linda photographed the model.

33) As Eric carved the table /wanted some polish/ started to take shape.
True or False: Eric was carving the table.

34) While Sally rode her pony /rested in its stall/ broke into a canter.
True or False: Sally rode her pony.

35) As Fred steered the boat/headed straight towards him/ bumped the riverbank.
True or False: Fred steered the boat.

36) While Laura ate the ice cream /remained in the freezer/ tasted really great.
True or False: Laura ate the ice cream.
Appendix 2: Words used in Lateralized LDT

<table>
<thead>
<tr>
<th>Frequent</th>
<th>Regular</th>
</tr>
</thead>
<tbody>
<tr>
<td>art</td>
<td>away</td>
</tr>
<tr>
<td>car</td>
<td>fact</td>
</tr>
<tr>
<td>bed</td>
<td>long</td>
</tr>
<tr>
<td>sun</td>
<td>free</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Infrequent</th>
<th>Regular</th>
</tr>
</thead>
<tbody>
<tr>
<td>wig</td>
<td>bred</td>
</tr>
<tr>
<td>bog</td>
<td>cask</td>
</tr>
<tr>
<td>zip</td>
<td>jest</td>
</tr>
<tr>
<td>gut</td>
<td>rash</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonwords</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>hox</td>
<td>cesk</td>
</tr>
<tr>
<td>ert</td>
<td>tash</td>
</tr>
<tr>
<td>pud</td>
<td>warl</td>
</tr>
<tr>
<td>rin</td>
<td>lerk</td>
</tr>
<tr>
<td>fut</td>
<td>mest</td>
</tr>
<tr>
<td>ket</td>
<td>samp</td>
</tr>
<tr>
<td>dap</td>
<td>cust</td>
</tr>
<tr>
<td>lig</td>
<td>mant</td>
</tr>
<tr>
<td>zay</td>
<td>soof</td>
</tr>
<tr>
<td>mut</td>
<td>loat</td>
</tr>
<tr>
<td>tal</td>
<td>skik</td>
</tr>
<tr>
<td>ret</td>
<td>ince</td>
</tr>
<tr>
<td>faw</td>
<td>anch</td>
</tr>
<tr>
<td>ped</td>
<td>mest</td>
</tr>
<tr>
<td>zad</td>
<td>fach</td>
</tr>
<tr>
<td>ret</td>
<td>fown</td>
</tr>
</tbody>
</table>
# Appendix 3: Words used in Synonym Detection

**Stimuli**

<table>
<thead>
<tr>
<th>lesion</th>
<th>wound</th>
<th>diminish</th>
<th>seminar</th>
</tr>
</thead>
<tbody>
<tr>
<td>officious</td>
<td>overbearing</td>
<td>organized</td>
<td>authorized</td>
</tr>
<tr>
<td>zealot</td>
<td>religious</td>
<td>fanatic</td>
<td>agnostic</td>
</tr>
<tr>
<td>yoke</td>
<td>hitch</td>
<td>egregious</td>
<td>hick</td>
</tr>
<tr>
<td>prodigious</td>
<td>enormous</td>
<td>puzzle</td>
<td>exorbitant</td>
</tr>
<tr>
<td>conundrum</td>
<td>confusion</td>
<td>perfidy</td>
<td>dispute</td>
</tr>
<tr>
<td>fealty</td>
<td>duplicity</td>
<td>celebration</td>
<td>fidelity</td>
</tr>
<tr>
<td>carnage</td>
<td>slaughter</td>
<td>forbid</td>
<td>scarlet</td>
</tr>
<tr>
<td>prohibit</td>
<td>scheme</td>
<td>skillful</td>
<td>advance</td>
</tr>
<tr>
<td>detritus</td>
<td>harmful</td>
<td>fault</td>
<td>rubbish</td>
</tr>
<tr>
<td>fleck</td>
<td>spot</td>
<td>save</td>
<td>weave</td>
</tr>
<tr>
<td>credentials</td>
<td>beliefs</td>
<td>qualifications</td>
<td>savings</td>
</tr>
<tr>
<td>devise</td>
<td>cunning</td>
<td>appliance</td>
<td>invent</td>
</tr>
<tr>
<td>predict</td>
<td>forecast</td>
<td>preach</td>
<td>progress</td>
</tr>
<tr>
<td>hue</td>
<td>disease</td>
<td>color</td>
<td>slash</td>
</tr>
<tr>
<td>excise</td>
<td>use</td>
<td>train</td>
<td>remove</td>
</tr>
<tr>
<td>contrary</td>
<td>opposing</td>
<td>joining</td>
<td>agreeing</td>
</tr>
<tr>
<td>daunting</td>
<td>horrible</td>
<td>frightening</td>
<td>mysterious</td>
</tr>
<tr>
<td>mimic</td>
<td>repeat</td>
<td>reply</td>
<td>copy</td>
</tr>
<tr>
<td>coherent</td>
<td>consistent</td>
<td>illogical</td>
<td>traditional</td>
</tr>
<tr>
<td>fueled</td>
<td>creased</td>
<td>rolled up</td>
<td>uncurled</td>
</tr>
<tr>
<td>opt</td>
<td>lose</td>
<td>fine</td>
<td>choose</td>
</tr>
<tr>
<td>euphoric</td>
<td>ecstatic</td>
<td>surprised</td>
<td>indignant</td>
</tr>
<tr>
<td>dour</td>
<td>dark</td>
<td>stern</td>
<td>erratic</td>
</tr>
<tr>
<td>fickle</td>
<td>choosy</td>
<td>careful</td>
<td>resistant</td>
</tr>
<tr>
<td>optimistic</td>
<td>hopeful</td>
<td>reliable</td>
<td>fabricated</td>
</tr>
<tr>
<td>authentic</td>
<td>honest</td>
<td>genuine</td>
<td>avoid</td>
</tr>
<tr>
<td>evade</td>
<td>explore</td>
<td>conquer</td>
<td>costly</td>
</tr>
<tr>
<td>plausible</td>
<td>believable</td>
<td>problematic</td>
<td>confound</td>
</tr>
<tr>
<td>chagrin</td>
<td>perplexity</td>
<td>disappointment</td>
<td>scarcity</td>
</tr>
<tr>
<td>death</td>
<td>surfeit</td>
<td>demise</td>
<td>diverse</td>
</tr>
<tr>
<td>nebulous</td>
<td>vague</td>
<td>subtle</td>
<td>heritability</td>
</tr>
<tr>
<td>geniality</td>
<td>legacy</td>
<td>hospitality</td>
<td>commonplace</td>
</tr>
<tr>
<td>generic</td>
<td>hereditary</td>
<td>customary</td>
<td>fallible</td>
</tr>
<tr>
<td>banal</td>
<td>trite</td>
<td>concrete</td>
<td>harshness</td>
</tr>
<tr>
<td>aplomb</td>
<td>confusion</td>
<td>panache</td>
<td>cheerful</td>
</tr>
<tr>
<td>jovial</td>
<td>plump</td>
<td>quick</td>
<td>entice</td>
</tr>
<tr>
<td>jettison</td>
<td>discard</td>
<td>replace</td>
<td>immovable</td>
</tr>
<tr>
<td>intractable</td>
<td>irregular</td>
<td>obstinate</td>
<td>negligent</td>
</tr>
<tr>
<td>remiss</td>
<td>inaccurate</td>
<td>hurried</td>
<td>dormant</td>
</tr>
<tr>
<td>transient</td>
<td>ephemeral</td>
<td>moveable</td>
<td></td>
</tr>
<tr>
<td>temperate</td>
<td>slight</td>
<td>moderate</td>
<td></td>
</tr>
<tr>
<td>tenuous</td>
<td>flagrant</td>
<td>fragrance</td>
<td></td>
</tr>
<tr>
<td>potable</td>
<td>drinkable</td>
<td>container</td>
<td></td>
</tr>
<tr>
<td>parlous</td>
<td>intensive</td>
<td>perilous</td>
<td></td>
</tr>
<tr>
<td>venal</td>
<td>proud</td>
<td>introspective</td>
<td></td>
</tr>
<tr>
<td>otiose</td>
<td>redundant</td>
<td>clear</td>
<td>specious</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>lugubrious</td>
<td>gentle</td>
<td>mournful</td>
<td>humiliated</td>
</tr>
<tr>
<td>nuance</td>
<td>deviance</td>
<td>demeanor</td>
<td>subtlety</td>
</tr>
<tr>
<td>bucolic</td>
<td>pastoral</td>
<td>platonic</td>
<td>annual</td>
</tr>
</tbody>
</table>
Appendix 4: Names used in Author recognition task

Authors
J K Rowling
Mitch Albom
Maya Angelou
Judy Blume
Isaac Asimov
Michael Chabon
Tom Clancy
Stephen Coonts
Richard Dawkins
Dave Eggers
Helen Fielding
Ian Fleming
Charles Frazier
John Grisham
S. E. Hinton
Alan Furst
David Halberstam
Stephen King
Dean Koontz
Judith Krantz
Robert Ludlum
Dan Brown
Cormac McCarthy
Orhan Pamuk
Chinua Achebe
Annie Proulx
Kathy Reichs
David Sedaris
Danielle Steele
Alvin Toffler
JRR Tolkein
John Updike
Janet Evanovich
Toni Morrison
John Gray
Jonathon Franzen
Gabriel Garcia Marquez
Leon Uris
Tom Wolfe
Bob Woodward
Appendix 5: Figures used for Pattern Comparison task
Appendix 6: Syllogisms

Valid unbelievable

All fruit are animals.
Bananas are not animals.
Bananas are not fruit.

All geometrical forms are squares.
Circles are geometrical forms.
Circles are squares.

All metal objects are alive.
Automobiles are made of metal.
Automobiles are alive.

All weapons are dangerous.
Machine guns are not dangerous.
Machine guns are not weapons.

All mammals walk.
Whales are mammals.
Whales can walk.

All buildings are round.
Pyramids are buildings.
Pyramids are buildings.

Invalid believable

All fruit has peel.
Oranges have peel.
Oranges are fruit.

All birds have wings.
Dogs are not birds.
Dogs do not have wings.

All flowers have petals.
Roses have petals.
Roses are flowers.

All houses have doors.
Books are not houses.
Books do not have doors.

All doctors examine patients.
Surgeons examine patients.
Surgeons are doctors.
All seals have flippers.  
Dolphins are not seals.  
Dolphins have flippers.

Valid Believable

All fish can swim.  
Tuna are fish.  
Tuna can swim.

All cats have whiskers.  
A lion is a cat.  
Lions have whiskers.

All boats can float.  
Canoes are boats.  
Canoes can float.

All boats can sink.  
Submarines are boats.  
Submarines can sink.

All birds can fly.  
Robins are birds.  
Robins can fly.

All jackets have sleeves.  
Windbreakers are jackets.  
Windbreakers have sleeves.

Invalid unbelievable

All fish can swim.  
People can swim.  
People are fish.

All bicycles have wheels.  
Cars have wheels.  
Cars are bicycles.

All clothes are made of cotton.  
Pants are not made of cotton.  
Pants are not clothes.

All monkeys can climb.  
Children can climb.  
Children are monkeys.

All cars can float.  
Lily pads can float.
Lily pads are cars.

All mice drink soda.
Cats drink soda.
Cats are mice.
Appendix 7: Words for LDT2

<table>
<thead>
<tr>
<th>Many neighbors</th>
<th>Few neighbors</th>
<th>Nonwords</th>
</tr>
</thead>
<tbody>
<tr>
<td>band</td>
<td>coax</td>
<td>poce</td>
</tr>
<tr>
<td>cast</td>
<td>snip</td>
<td>voun</td>
</tr>
<tr>
<td>hill</td>
<td>fact</td>
<td>joat</td>
</tr>
<tr>
<td>last</td>
<td>film</td>
<td>rube</td>
</tr>
<tr>
<td>rice</td>
<td>jump</td>
<td>wefe</td>
</tr>
<tr>
<td>slow</td>
<td>next</td>
<td>tymb</td>
</tr>
<tr>
<td>mine</td>
<td>open</td>
<td>gaik</td>
</tr>
<tr>
<td>tent</td>
<td>taut</td>
<td>bist</td>
</tr>
<tr>
<td>wart</td>
<td>writ</td>
<td>fost</td>
</tr>
<tr>
<td>pest</td>
<td>trap</td>
<td>tuss</td>
</tr>
</tbody>
</table>
Appendix 8: Form for **Edinburgh Handedness Inventory**

Please indicate your preferences in the use of hands in the following activities *by putting a check in the appropriate column*. Where the preference is so strong that you would never try to use the other hand, unless absolutely forced to, *put 2 checks*. If in any case you are really indifferent, *put a check in both columns*.

Some of the activities listed below require the use of both hands. In these cases, the part of the task, or object, for which hand preference is wanted is indicated in parentheses.

Please try and answer all of the questions, and only leave a blank if you have no experience at all with the object or task.

<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Writing</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>2. Drawing</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>3. Throwing</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>4. Scissors</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>5. Toothbrush</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>6. Knife (without fork)</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>7. Spoon</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>8. Broom (upper hand)</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>9. Striking Match (match)</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>10. Opening box (lid)</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td><strong>TOTAL (count checks in both columns)</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 9: Passages with bridging inference in Experiment 2

1) Stephanie packed up a cooler for their picnic lunch. / It looked as if they should have a glorious day at the coast. / She made sure to put in some beer and sodas from the fridge. / At the beach, Stephanie found that the beer was warm.
+ True or False: Stephanie packed beer for the picnic.

2) Steve got the / automobile repair kit / wrench and screwdriver / out of the trunk of the car. A loose fan belt can be very dangerous. He needed the screwdriver to tighten the fan belt.
+ True or False: Steve keeps his tools in the trunk.

3) The ship’s captain watched the instruments / radar screen / closely. Thanks to technology they had some warning of the danger. The radar was picking up a storm due south of their position.
+ True or False: The radar showed a storm to the north.

4) The hiker checked the food / trail mix / and spare clothes in her pack before setting out. She liked to have some high calorie food like chocolate when she walked. The trail mix had all the M&Ms picked out.
+ True or False: The hiker liked candy.

5) Beth worked on the orchards all summer picking fruit / peaches. It was hard work and the pay was poor. By the end she could not stand the sight of peaches.
+ True or False: Beth liked peaches.

6) Zack was looking forward to the championship / basketball / match. It would be worth the long drive even if the team lost. He loved watching basketball.
+ True or False: Zack liked football.

7) Bill was desperately searching for his fishing gear / rod. He was going to the lake with his parents. The rod had been a present from his father.
+ True or False: Bill wanted to go fishing.

8) Jim loves to have the radio / pop music / playing all day long. He also likes to have the windows open. The whole neighborhood is annoyed by the loud music.
+ True or False: Jim likes to play music.

9) Jessica went shopping for salad greens / a fresh lettuce / at the nearby store. She planned on making lunch for her friends. None of the lettuce looked particularly fresh.
+ True or False: Jessica liked fresh vegetables.
10) Carol was trying to repair the clock’s delicate mechanism / mainspring. It required very precise work. The mainspring was hopelessly twisted.
+ True or False: Carol could not repair the clock.

11) Amy’s little boy liked to play with her cooking pans / frying pan. Almost anything makes a good toy for a toddler. He simply loves banging on the frying pan.
+ True or False: Amy won’t let her son play with the frying pan.

12) Jerry’s desk is covered with piles of files as well as notes and messages / telephone numbers on scraps of paper. He really should get better organized. Recently he lost an important telephone number.
+ True or False: Jerry always keeps his desk tidy.

13) The most valuable picture / Monet / in a downtown gallery was stolen yesterday. The curator was very upset because it was also his personal favorite. It was a mystery how the thieves had got away with the huge Monet.
+ True or False: The picture was not particularly valuable.

14) Joan lost all her personal possessions in the fire including important papers / her new passport. She is very upset. Particularly as she now has to replace her passport.
+ True or False: Joan was lucky to have lost nothing of importance in the fire.

15) The flight engineer carefully examined the airplane wings for dangerous stress fractures / hairline cracks. Even metal will fatigue with time. Hairline cracks were visible on the underside of the wing.
+ True or False: The flight engineer found no problems.

16) As an infantryman, Michael was issued with his own weapon / rifle. He was careful to follow the maintenance schedule. His rifle was carefully cleaned and oiled daily.
+ True or False: Michael’s rifle was dirty.

17) Crystal likes to listen to classical / Beethoven’s / music. Her favorite pieces are the great symphonies. Her all time favorite piece is Beethoven’s Ninth.
+ True or False: Crystal prefers to listen to pop music.

18) Justin hated taking his prescription medicine / antibiotic. Sometimes the cure seemed worse than the disease. The antibiotic made him sick to his stomach.
+ True or False: The medicine made Justin nauseous.
19) The friends could talk of nothing but camping and boating and swimming. They had been planning their trip to the lake all semester. They hoped the water would be warm enough for swimming.

True or False: The friends hoped to swim at the lake.

20) Nico works as a researcher for a local radio / news channel. He was enjoying his new job now he had earned more responsibility. He thinks providing local news is important.

True or False: Nico works for a local news channel.

21) Bob decided to clean the engine / carburetor of his truck. A blocked fuel system can ruin a vehicle. He was pleased he had bothered when he found an oily residue was blocking his carburetor.

True or False: Bob’s engine did not need cleaning.

22) Dan went to the supermarket to get drinks and snacks / soda and lemonade / for the frisbee tournament. He was wondering how much variety to get. He knew not everyone likes soda.

True or False: Dan was planning on getting soda.

23) When Frank emptied the dishwasher he found that the silverware was / knives and forks were still spotty. He would have to run them through the washer again. He wanted the knives and forks spotless for the reception.

True or False: Everything in the dishwasher was clean.

24) John realized that someone else in the apartment had been using his toiletries / toothpaste. He felt annoyed because he was particular about his things. He would never squeeze his toothpaste in the middle.

True or False: The toothpaste had been squeezed in the middle.
Appendix 10: Passages with garden path sentences in Experiment 2

1) Laura finished her salad and then had ice cream for dessert. She wondered how the chef had come up with such a great combinations of flavors. While Laura ate the ice cream tasted really great. True or False: Laura ate the ice cream.

2) Sarah asked her student to varnish his picture next door, while she worked on her new project. She had decided to lighten the background color. As Sarah painted the picture improved dramatically. True or False: Sarah painted the picture.

3) Joe's cat took great care while washing her kitten even though it was now almost weaned. She liked to sit in a patch of sun from the window. While the cat groomed her kitten protested vigorously. True or False: The cat groomed her kitten.

4) Fred was enjoying his afternoon out helping on his friend’s boat. He was not yet used to the river. As Fred steered the boat bumped the riverbank. True or False: Fred steered the boat.

5) Despite its age, Martin liked his old computer and continued trying to program it. Even though he found the small screen rather irritating the response time was much better. While Martin programmed his old computer persisted in having problems. True or False: Martin programmed his old computer.

6) The chef carefully checked the temperature of the main oven. Although he still had a great deal to do for the wedding reception, the most important item would be ready. While the chef cooked the cake filled the room with a pleasant smell. True or False: The chef cooked the cake.

7) Rowan spent a long time trying to understand the surveyor’s letter. She was worried about what her mother had said on the phone, and had hoped the letter would explain. As Rowan read the letter worried her increasingly. True or False: Rowan read the letter.

8) The crew was pleased that the newly finished yacht was ready to sail. They hoped to win their next race in her. While the crew sailed the new yacht developed a leak. True or False: The crew sailed their new yacht.

9) Jeff's struggled to park his van into the tight space. The vehicle was new and he had almost finished making all his payments on his loan. As Jeff parked the van scraped the curb. True or False: Jeff parked the van.

10) Mandy checked the notes she had made in class and wondered whether she should call her friend. Normally, she had no trouble with biology. While Mandy studied the textbook provided the answer. True or False: Mandy studied the textbook.
11) The magician spun the plates high into the air, as the audience watched in amazement. This was last portion of the act. As the magician juggled the plates flew up higher and higher. True or False: The magician juggled the plates.

12) Patrick was busy planning his day as he walked. He was looking forward to finishing an important sale later. As Patrick walked his dog tugged on the leash. True or False: Patrick walked his dog.

13) Rebecca enjoyed cleaning her mother's vase, so she always left it until after the brass candlesticks that had belonged to her grandmother. The silver vase had special memories. As Rebecca polished the vase acquired a nice shine. True or False: Rebecca polished the vase.

14) The chairman’s secretary hurried to finish the typing that her colleague had asked her to get in the mail. She knew her friend had been worried about it being late. As the secretary typed the urgent letter was long overdue. True or False: The secretary typed the urgent letter.

15) Eric put in some work on the table that his most important client had commissioned. The cherry wood he had chosen was a particularly rich red. As Eric carved the table started to take shape. True or False: Eric carved the table.

16) Sally was glad to have a chance to visit her friend. It was a nice afternoon for a ride and not far to her friend's house. While Sally rode her pony broke into a canter. True or False: Sally rode her pony.

17) The children played the new piano duet, as they waited for the piano tuner. The summer holidays were almost over and they needed to get in some practice. While the children played the piano sounded out of tune. True or False: The children played the piano.

18) Linda set up her camera as she waited for the famous fashion model. The setting was beautiful and she hoped the light would hold. While Linda photographed the model posed languidly. True or False: Linda photographed the model.

19) Although her own work would be very different, Gloria began work on the statue after studying the work of her predecessor. She loved the way the marble caught the light. As Gloria sculpted the statue drew praise from her students. True or False: Gloria sculpted the statue.

20) Jim was in his best clothes when he went out to play with his friends. He knew he would get into trouble. As the boys wrestled Jim struggled to break free. True or False: The boys wrestled Jim.

21) Matt did a sketch of the actress as she sat on the bridge. The sun was setting over the river. While Matt drew the actress looked down at the water.
True or False: Matt drew the actress.

22) The marine charged up the hill to attack the command post. The ground was steep and muddy and clouds were hiding the moon. As the marine attacked the general tried to defend himself.
True or False: The marine attacked the general.

23) The prosecution lawyer continued to question the defendant. The case was not going well. As the lawyer cross-examined the defendant collapsed in the witness box.
True or False: The lawyer cross-examined the defendant.

24) Wallace followed his guide into the heart of the jungle. The trees were thick even close to the shore and the calls of birds filled the air. While Wallace explored the island experienced the monsoon.
True or False: Wallace explored the island.

25) The team reached the mountain and started the long ascent to the summit. The sky was clear but the forecast was not good, so they needed to keep moving. As the team climbed the mountain loomed above them.
True or False: The team climbed the mountain.

26) Harry left to get the children from school and worried he was late. He was tired and it was raining heavily. As Harry drove the car began to make a strange noise.
True or False: Harry drove the car.

27) Glenn started work on the girl's tattoo as her boyfriend watched. The design was of a Chinese dragon and would be vividly colored when finished. While Glenn tattooed the girl attempted to stay as still as she could.
True or False: Glenn tattooed the girl.

28) The friends planned on spring-cleaning the whole house instead of leaving it for the maid. They wanted the whole place to look great for their party to celebrate the end of the semester. As the friends cleaned the living room looked much better.
True or False: The friends cleaned the living room.

29) Tom went to clean up and started on the dirty dishes. His wished his roommates would help. While Tom washed up the dirty dishes continued to accumulate.
True or False: Tom washed up the dirty dishes.

30) Sandra checked that all the children were there for the school trip. Even though they were excited the children tried to stay quiet. While Sandra counted the children waited in line.
True or False: Sandra counted the children.

31) Kevin watched the sausages on the grill while his son kept the campfire blazing. Cooking outside was always fun and it was a beautiful clear night. As Kevin grilled the sausages burnt to a crisp.
True or False: Kevin grilled the sausages.
32) While Janet edited the next day's front page the intern worked on the layout. There was a great picture to accompany the main story. As Janet edited the article became easier to understand.
True or False: Janet edited the article

33) As Dave was finishing his meal the waiter asked if he would like a brandy. It sounded like a great way to complete the dinner. While Dave drank the brandy soothed his sore throat.
True or False: Dave drank the brandy.

34) Laurel had high hopes for her new book. It was the second in a trilogy and the first part had been well reviewed. As Laurel wrote her new novel grew more and more complicated.
True or False: Laurel wrote her new novel.

35) Nick was worried about the dialogue in the next scene. He wondered if it would sound too rehearsed. While Nick filmed the next scene ran quite perfectly.
True or False: Nick filmed the next scene.

36) Marilyn was angry with the student and felt she should explain her annoyance. They had always worked well together until now. While Marilyn lectured her student listened very carefully.
True or False: Marilyn lectured her student.
Appendix 11: Filler sentences from Experiment 2

1) The objections that Philip raised were disregarded by the rest of the group. True or False: Everyone agreed with Philip.

2) The movie that Krista saw was criticized by most of the press. True or False: The movie was praised by the press.

3) The politician who Vic admired ended up resigning. True or False: The politician quit.

4) Carla and Ethan couldn't count the number of games they had watched with their father. True or False: They watched many games with their father.

5) Even though it was corny Craig wouldn't leave Venice without going on a gondola ride. True or False: Craig didn't have any plans for the Venice trip.

6) Early that morning Erin drove to the dock to see where the ships were anchored. True or False: Erin watched the ships from the cliff top.

7) Over lunch Skip worked on the application that was due later in the week. True or False: The application needed to be submitted that day.

8) For her birthday Ike and Maude tried to make their mother pancakes but they failed miserably. True or False: They weren't good at making pancakes.

9) While leaving class Tanya complained about the homework that the teacher assigned. True or False: Tanya didn't like the homework assignment.

10) Before work Elise sewed the button on her jacket and mended the sleeve. True or False: She was able to fix the jacket before work.

11) The businesses that Homer called had contracts with the bankrupt firm. True or False: The firm had escaped bankruptcy.

12) Cher left the cookies out on Christmas Eve in hopes that Santa would come. True or False: Cher no longer believed in Santa Claus.

13) By accident Yvonne set off the alarm that automatically alerted the police. True or False: The police were not alerted.

14) Finally Fiona got rid of her fever but her father still kept her home from school. True or False: Her fever was gone.
15) The next day Edward sighed with relief when he printed out the finished report.
True or False: Edward was pleased that the report was done.

16) On Tuesday Melanie took the entire office out to lunch.
True or False: The employees never have lunch together.

17) Finally, Tina told the customer the secret while nobody was listening.
True or False: Tina told the secret to the customer.

18) Late Sunday night the executive prepared Graham a report to present Monday morning.
True or False: Graham prepared a report for the executive.

19) As a favor, Emily lent Julia an umbrella because it was raining.
True or False: Julia lent Emily an umbrella.

20) This afternoon the plumber asked the electrician the time when they met at the construction site.
True or False: The plumber asked the time of the electrician.

21) Yesterday the housewife showed Kent a photograph of her grandchild.
True or False: Kent showed a photograph to the housewife.

22) Thoughtfully, Rhonda grabbed Debbie a sandwich at lunchtime.
True or False: Rhonda grabbed Debbie a sandwich.

23) Last winter the dancer sold Mitch a stereo at a yard sale.
True or False: Mitch sold a stereo to the dancer.

24) At the stoplight Teresa presented the secretary with a package.
True or False: Teresa presented a package to the secretary.

25) For security, Denise opened the milkman an account at the local bank.
True or False: Denise opened a bank account for the milkman.

26) Early one morning Pamela slid Elaine the preserves at the breakfast table.
True or False: Elaine slid the preserves to Pamela.

27) From home the editor faxed Frank the documents with a number of comments.
True or False: Frank faxed the documents to the editor.

28) One cold day the drummer fixed Arnold a lunch with chicken noodle soup and crackers.
True or False: Arnold fixed a lunch for the drummer.

29) Hoping to make amends, Ramona left her roommate some candy on the kitchen table.
True or False: Ramona left her roommate some candy.

30) While camping, the guide dug the climber a pit for the fire.
True or False: The climber dug a fire pit for the guide.
Appendix 12: Analysis of the effects of the laterality manipulation

Testing with a quadratic model in Experiment 1 showed significant effects of differences between visual fields on the LDT measures. Nonword error rates were found to be significantly effected by differences in error rate between visual field, $F(1, 119) = 3.21, p = .044$, as were nonword RT differences on nonword RTs $F(1, 119) = 6.05, p < .003$; smaller difference were associated with fewer errors and shorter RTs. The effects on words were not significant.

In each of the quadratic analyses, the quadratic term was statistically significant and positive, and the linear term was non-significant. This suggests that the minimum difference in response was at or close to zero and therefore it is appropriate to drop the linear term from the quadratic model and fit the quadratic term only (personal communication, Martin Michael). This analysis forces the minimum response to be at zero difference. Table 14 (following page) shows the analysis including the linear term.
<table>
<thead>
<tr>
<th>Response</th>
<th>Experiment</th>
<th>N</th>
<th>F</th>
<th>p</th>
<th>R2</th>
<th>Difference at Trough</th>
<th>p&gt;\text{t} (linear)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDT Word Error Score</td>
<td>1</td>
<td>120</td>
<td>1.44</td>
<td>0.24</td>
<td>2.4%</td>
<td>0.04</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>119</td>
<td>10.33</td>
<td>&lt;0.001</td>
<td>15.1%</td>
<td>-0.06</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>1 and 2</td>
<td>239</td>
<td>10.54</td>
<td>&lt;0.001</td>
<td>8.2%</td>
<td>-0.03</td>
<td>0.76</td>
</tr>
<tr>
<td>LDT Non-Word Error Score</td>
<td>1</td>
<td>120</td>
<td>3.21</td>
<td>0.04</td>
<td>5.2%</td>
<td>0.05</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>119</td>
<td>9.85</td>
<td>&lt;0.001</td>
<td>14.5%</td>
<td>-0.04</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>1 and 2</td>
<td>239</td>
<td>13.99</td>
<td>&lt;0.001</td>
<td>10.6%</td>
<td>0.004</td>
<td>0.90</td>
</tr>
<tr>
<td>LDT Word Mean Time (ms)</td>
<td>1</td>
<td>120</td>
<td>2.63</td>
<td>0.076</td>
<td>4.3%</td>
<td>-106.66</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>119</td>
<td>10.59</td>
<td>&lt;0.001</td>
<td>15.4%</td>
<td>37.50</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>1 and 2</td>
<td>239</td>
<td>13.15</td>
<td>&lt;0.001</td>
<td>10.0%</td>
<td>27.76</td>
<td>0.37</td>
</tr>
<tr>
<td>LDT Non-Word Mean Time (ms)</td>
<td>1</td>
<td>120</td>
<td>6.05</td>
<td>0.003</td>
<td>9.4%</td>
<td>-22.91</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>119</td>
<td>9.41</td>
<td>&lt;0.001</td>
<td>14.0%</td>
<td>-8.36</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>1 and 2</td>
<td>239</td>
<td>17.13</td>
<td>&lt;0.001</td>
<td>12.7%</td>
<td>-14.86</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Table 14: Fit of Quadratic and Linear Models for Lateralized lexical decision task
REFERENCES


Goff, D. A., Pratt, C., & Ong, B., (2005) The relations between children's reading comprehension, working memory, language skills and components of


