Individual Differences in Text Predictive Inferences

by

Jeffrey Caley Doering

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF ARTS

Department of Psychology

University of Manitoba

Winnipeg, Manitoba, Canada

Table of Contents

Abstract

The activation of a predictive inference relies largely on the amount of contextual support for the inference in the text (Cook, Limber, & O'Brien, 2001). However, few sources of individual differences in predictive inference activation have been isolated. An experiment designed to identify possible sources of individual differences was conducted. One hundred and one participants completed an inference task using a long-passage *correct rejection* paradigm that included passages with varying levels of contextual support occurring mid-passage. Participants also completed a reading span measure (Daneman & Carpenter, 1980) and a knowledge access measure (Potts & Peterson, 1985). Results from ANOVA and regression analyses suggest that readers with better knowledge access abilities are better able to correctly reject inference concepts and are less affected by a change in inference-facilitating contextual support. It is suggested that higher knowledge access readers are able to construct and maintain a more specific representation of the text.

Keywords: reading, text comprehension, individual differences, predictive inference, situation model, knowledge, access, reading span, working memory

Acknowledgments

I would like to offer my most sincere thanks for Dr. Murray Singer of the Department of Psychology of the University of Manitoba for his invaluable assistant and guidance during the preparation of this manuscript and throughout the course of my degree as a whole. I would also like to thank Dr. Randy Jamieson of the Department of Psychology of the University of Manitoba as well as Dr. Terry Janzen of the Department of Linguistics of the University of Manitoba for this assistance in the preparation of this manuscript as members of my thesis committee.

List of Tables

List of Figures

List of Appendices

Individual Differences in Text Predictive Inferences

Though it pervades almost every modern human activity, reading is a skill that most people largely take for granted. Despite its commonness, the process of text comprehension is quite complicated. The reader is required to deal with multiple dimensions of a text at once, and various components of the text must be integrated into a holistic understanding. Readers make inferences – logical inductions or deductions about the text – as they read in order to comprehend the text

Many of these inferences are necessary. Without them, discourses would seem disjointed. These inferences are incorporated into a reader's understanding of the text as a whole. However, there are processes that a reader engages in that are not necessary for a basic understanding of a text. These *elaborative inferences* are made all the time. Elaborative inferences are based on general world knowledge, aid in building a more complete understanding of a discourse, and are comparatively distinct from many other kinds of inferences in other ways as well; namely, in that a reader will only generate an elaborative inference when there is enough context or background in the story or passage to facilitate it (Cook, Limber, & O'Brien, 2001; Klin, Guzman, & Levine, 1999; Lassonde & O'Brien, 2009). Thus, readers who generate many elaborative inferences are often going above and beyond what is required of them. This presents a unique opportunity to investigate the way that readers build an understanding of a discourse.

Researchers have long been interested in what cognitive processes govern various aspects of reading comprehension. Many experimenters have taken an *individual differences* approach in their study of language comprehension. The individual differences approach attempts to account for reader differences in comprehension, and allows the researcher to determine what cognitive processes are most crucial for a given task.

With this in mind, one might then ask what cognitive processes determine how elaborative inferences are made, particularly under differing amounts of supportive context in the passage, and what this tells us about the way that readers build a coherent understanding of a discourse. This paper presents possible answers to these questions.

Inferences and the Situation Model

Much of the information that we extract and understand from a text is not stated directly. For example, if a person is asked what a novel is about, it would be rather odd and unexpected for that person to attempt to describe his or her understanding of the novel via a number of direct quotes. This person may instead choose to quickly describe the setting, basic plot, characters, and themes of the novel. This reliance on higher order information is founded in our ability to construct a situation model (van Dijk & Kintsch, 1983) or "mental model" (Johnson-Laird, 1983). The situation model refers to a reader's mental representation of what the text is about. The situation model is useful because it is the ultimate goal of the reader to understand the meaning of the text (Graesser, Singer, & Trabasso, 1994).

The situation model includes many inferences. Inferences require the reader to understand more than what is explicitly stated in the text and typically occur within local domains in a text. Local domains of text comprise no more than a few sentences. For example, a reader may encounter sentences such as the following (Singer, Halldorson, Lear, & Andrusiak, 1992b):

Mary poured the water on the bonfire.

The fire went out.

The reader must make several inferences in order for these sentences to make sense. First, the reader must infer that they are connected in some way. Because they occur together, meaning can be attributed to the situation as a whole. The reader may infer that because Mary poured water on the fire, the fire went out. This inference connects the two sentences and is thus referred to as a bridging inference. The reader may also infer that the water caused the fire to go out. This latter type of inference is known as a causal inference.

The notion of inference generation being tied to situation model construction derives from the *constructionist* theory of text comprehension, which proposes a "search after meaning" principle (Graesser et al., 1994). This principle makes three assumptions regarding the generation of inferences. These assumptions are that (a) the reader's mental representation of the text is based on the readers' goals (information extraction, for example; the *reader goal* assumption), (b) readers attempt to construct a mental representation that is organized at local and global levels (the *coherence* assumption), and (c) readers attempt to explain events described by the text (the *explanation* assumption).

Inferences come in many forms. A comprehensive taxonomy of inferences was presented by Graesser et al. (1994). They listed 13 kinds of inferences: *referential, case structure role argument, antecedent causal, superordinate goal, thematic, character emotion, causal consequence, instantiation noun category, instrument, subordinate goal action, state, reader's emotion*, and *author's intent.* However, inference categories can also be organized into smaller groups. Graesser et al. (1994) similarly suggested that three types of inferences are constructed most of the time. These are: superordinate goals (a goal motivating an action), causal antecedents (explains why something is mentioned), and global thematic inferences (integrates large chunks of text). The taxonomy demonstrates the degree to which inference processing can aid in text comprehension. It is important to note that the taxonomy does not include elaborative inferences. This is likely due to the fact that elaborative inferences, like bridging or causal inferences, can occur in several different forms and encompass more than just one class of inference. Most importantly, Kintsch (1998) suggested that inferences can be broken into two broad categories: necessary and elaborative. A bridging inference is an example of a necessary inference. The reader must link the text ideas in order to make sense of the discourse as a whole. However, necessary inferences are not the focus of the study.

Elaborative Inferences

Elaborative inferences draw on general world knowledge and are used to build a more detailed situation model of the discourse. However, elaborative inferences are not necessary for the preservation of message coherence. McKoon and Ratcliff (1989) proposed a theoretical framework that describes the way that inferences vary in the degree to which they are constructed and encoded. In this framework, a well-encoded inference will incorporate and utilize relevant semantic information. This theory, though somewhat robust, can encounter problems. Singer (1979) provides the following example:

(1a) The boy cleared the snow with a shovel.

(1b) The boy cleared the snow from the stairs.

(2) The shovel was heavy.

Participants were presented with sequence (1a) – (2) or (1b) – (2), and had to signal as soon as they understood (2). Response times were faster in sequence (1a) – (2), which indicates that readers had not made the inference that the snow was cleared using a shovel, even though it was implied by sentence (1b). Singer suggested that readers had either only partially drawn the inference or not computed it at all. This is an elaborative inference about instruments. Instrumental inferences are drawn based on previously held world knowledge.

O’Brien, Shank, Myers, and Rayner (1988) subsequently asked whether elaborative inferences are generated on-line (during reading) or as part of a later reconstructive process during retrieval. They recorded eye-movements and determined that, with sufficient context and facilitating conditions (the target concept involved an implicit antecedent and no specific memory category), a simple elaborative inference is typically made on-line. However, further investigation determined that without a demand sentence, readers might not generate the elaborative inference. A demand sentence is a sentence requiring the reader to make the inference like in sentence (2) from Singer (1979), above.

Several theories addressed the reason for this omission. In general, it seems reasonable to suggest that elaborative inferences are generated pragmatically; when the number of available elaborations is too high, cognitive resource limitations prohibit the drawing of the inference. This requires something of a balancing act on the part of the reader because elaborative information does aid in later retrieval of the target concepts and generally facilitate better understanding of the text, as was demonstrated by O'Brien, Plewes, and Albrecht. (1990). A follow-up study to O'Brien et al. (1988) determined that elaborative inferences might occur in several ways. The inference generation process is *passive* when the context of the discourse merely encourages the reader to make an inference, but *active* when a demand sentence is present (Garrod, O'Brien, Morris, & Rayner, 1990). Garrod et al.'s distinction between passive and active elaborative inference generation is based on top-down (active, and a deductive process) and bottom-up (passive, and an inductive process) processing. In passive generation, the context of the text aids in defining a referent. Low-level priming activates related concepts in memory and leads to an inference. This activation is unguided and nonstrategic (Cook, Limber, & O'Brien, 2001). This leads to a lower activation of the elaborative inference, and only occurs because the context is so facilitating of such an inference. Active inferences, in contrast, will occur because, for example, the demand sentence increases the predictability of a future event. These inferences are specific and goal-oriented (Graesser et al., 1994) and draw on long-term memory resources (see Ericsson & Kintsch, 1995). The fact that elaborative inferences are not necessary and therefore are not always made is the most important distinction between those and bridging inferences.

Similarly, elaborative inferences do not appear to follow the same time-course as other types of inferences. Garnham, Oakhill, and Reynolds (2002) conducted an experiment in which the gender of a character in the experimental text was implied but never stated explicitly. They presented participants with sentences like those below, one component at a time:

(1a) The *housekeeper/soldier* was rushed to the hospital,

(1b) taken to a private ward,

(1c) and gave birth a half hour later.

The critical word of the sentence either stereotypically matched (housekeeper) or mismatched (soldier) a later event in the third component (1c) of the sentence. Reading times were recorded for component (1c) as a measure of comprehension. When the latter component of sentence contradicted the previously implied gender, comprehension was slower. The authors argued that this indicates that the elaborative inference regarding gender must have been made immediately.

Predictive Inferences

Predictive inferences are elaborative inferences that activate one or several possible future outcomes in a discourse. Consider an example from McKoon and Ratcliff (1986):

Predicting sentence: The director and the cameraman were ready to shoot close-ups when suddenly the actress fell from the 14th story.

Test word: Dead

In order for proper comprehension to be achieved, the reader may need to complete an inference about an event that has an obvious consequence, namely that the actress fell to her *death*. Readers were presented with the test probe *dead* after reading the predicting sentence. In contrast to most other recognition paradigms, the reader must correctly reject the inference concept. Participants responded *yes* or *no* using keys on the keyboard as to whether the probe word appeared in the previous sentence. When presented with the probe, readers must determine that the word did not occur, despite the fact that it may have been activated, in order to respond correctly, *no.* This experimental paradigm will be referred to as the *correct rejection* paradigm. Indeed, readers were slower to reject the probe (respond: "no") in the experimental set with the *predicting sentence,* shown above, as opposed to a control version of the same story with a control sentence instead of the predicting sentence. For example, the control version from the above story would be: *Suddenly the director fell upon the cameraman, demanding close-ups of the actress on the 14th story.*

Because they are elaborative predictive inferences do not seem to be made as consistently as necessary inferences. Keefe and McDaniel (1993) demonstrated that test probes fail to reveal evidence of a predictive inference if their presentation is delayed. Consider the following sentences:

> *Predictive sentence:* After standing through the three-hour debate, the tired speaker walked over to his chair.
>
> *Control sentence:* The tired speaker moved the chair that was in his way and walked over to the podium to continue his three-hour debate.
>
> *Test word:* sat

Readers were exposed to either the predictive or control critical sentence within a brief discourse. Following the critical sentence, two more sentences occurred in the passage, and participants were then given the test probe *sat.* They were instructed to read the probe word aloud as quickly as they could, a measure called naming time. A faster naming time would indicate that the probe word was already activated. Keefe and McDaniel (1993) failed to find evidence for readers' encoding of predictive inferences after reading text with the predictive sentence condition above (also see Potts, Keenan, & Golding, 1988). However, they found evidence of predictive inference activation if the test probe was presented without the two intervening sentences. They conducted the same experiment as above, but with only one sentence occurring between the critical sentence and the test probe. With a shorter gap between the critical sentence and the test probe, the researchers found a faster naming time in the predictive versus control condition. Thus, predictive inferences are briefly activated and then later deactivated. Keefe and McDaniel suggested that this deactivation occurs because predictive inferences may not be required in order for the reader to form a coherent representation of the text. They suggested that these inferences are later pruned from that representation or never make it into a long-term text representation at all.

Further analysis has suggested that although predictive inferences are only briefly activated, some predictive information may be encoded directly in the situation model. Most other information retrieved from text is encoded into lower levels of representation first (Peracchi & O'Brien, 2004; Schmalhofer, Keefe, & McDaniel, 2002). Predictive inferences are encoded directly into higher levels of representation first because they never occurred in the textbase. Schmalhofer et al.'s analysis also suggests that a certain degree of interconnectivity must exist between the inference and other concepts in the representation in order for it to be maintained. Similarly, local coherence is maintained when small sections of a text (a paragraph at most) are structured and sensible. Predictive inferences can even be made when no break in local coherence occurs in the text (Murray, Klin, & Myers, 1993), whereas some inferences prevent coherence breaks (e.g., bridging inferences). The manner in which they are encoded might cause predictive inferences to be represented as hypothetical facts rather than certain facts (Campion, 2004). This would be a deductive process and would rely on general world-knowledge.

More recently, research has investigated the effect of contextual support within a text on predictive inference activation (Casteel, 2007; Cook et al., 2001; Fincher-Kiefer, 1993, 1995; Gueraud, Taperio, & O'Brien, 2008; Klin, Guzman, & Levine, 1999; Lassonde & O'Brien, 2009). Contextual support is the degree to which the text builds supporting context around a concept; it varies in specificity. Initial research suggested that when too many possible inferences exist, none are made (Klin et al., 1999).

With a comparatively higher degree of contextual support, a predictive inference can be facilitated no matter where in the text the facilitating context occurs, so long as the reader still has access to it later (Cook et al., 2001). An example of their experimental passages includes a story about a boy (Jimmy) who is throwing either a rock or Nerf (foam) ball around a new car. Consider the following situations:

Low context: A car door being dented after being hit by a Nerf ball.

High context: A car door being dented after being hit by a rock.

Experimental passages included several sentences of intervening text between the critical sentences describing a ball/rock hitting the car and an inference-invoking sentence at the end of the passage: *He (Jimmy) missed, though, and he accidentally hit the door of a new car.* After the inference-invoking sentence, participants read the probe word *dent* aloud as quickly as possible. Naming times were faster in the high context condition than the low context condition, suggesting that the inference *dent* had already been activated. Table 1 displays results from Cook et al. (2001). Cook et al. included delays of either 250 ms or 500 ms before presenting the test probes after the last line of the experimental passage in order to identify brief activation (250 ms only) as well. No evidence of only brief activation, such as the pattern of activation found in the experiments run by Keefe and McDaniel (1993), was found. That is, naming times were faster for high context probes than low context probes following both 250 ms and 500 ms delays. This showed that enough contextual support can compensate for intervening text, which should make it more difficult for readers to use previously instated context. Previous research had demonstrated predictive inference activation only with contextual support immediately preceding the test probes (Klin et al., 1999).

Other results from Cook et al. (2001) suggested that predictive inferences are not encoded into long-term memory, whereas previous research had suggested that they may be encoded into long-term memory (Klin et al., 1999). Cook et al. claimed that this might be due to the gap between contextual support and inference-invoking sentences. When there is no gap, the predictive inference represents a specific item. With a gap, the predictive inference degrades to a point where it represents only a more general concept (Cook et al., 2001; known as “minimal encoding”, McKoon & Ratcliff, 1986).

More than one lexical item can be activated by an inference-invoking sentence. If the number of these possible inferences is constrained, fewer inferential concepts are activated (Lassonde & O’Brien, 2009). Consider the passages below from Lassonde and O’Brien (Appendix A shows a full experimental set). The experimental passage begins with the same story as in Cook et al. (2001), in which a boy named Jimmy is throwing rocks when a new family car is nearby. However, only one constraint condition of the text existed; Jimmy is always throwing a rock (not a Nerf ball).

> *Baseline Sentence:* A dog came racing across the street and distracted Jimmy from his throw.
>
> *Inference Sentence:* He missed, though, and he accidentally hit the door of a new car.

Participants read either the baseline or inference sentence at the end of the passage and were then instructed to name a target word aloud as quickly as they could. The test probes for this experimental set includes both the target words *dent* and *damage.* Target 1 (*dent*) is a more specific form of Target 2 (*damage*). Lassonde and O'Brien (2009) found that responses were faster in the inference-invoking sentence than the baseline sentence condition for both target words 1 and 2. Thus, a predictive inference could activate more than one related item as long as the context was vague enough.

Lassonde and O'Brien (2009) also demonstrated that contextual support helps to guide and reduce the number of possible predictive inferences. To do this, they added three to five sentences after each passage's introduction, which were designed to further constrain the context of the passage (see Appendix 1). They found that responses were faster for the inference-evoking sentence than the baseline sentence condition. However, this was only true for Target 1, not Target 2. Thus, further contextual support can guide predictive inference activation. The authors also noted that predictive inferences must be quite specific. Likewise, Klin et al. (1999) demonstrated that the possibility for too many inferences leads to none being made. This is crucial given the *search after meaning* principle of the constructionist theory of comprehension (Graesser et al., 1994). Furthermore, some types of elaborative inferences might have many related concepts activated. The activation might not occur in a single item, but rather in a whole category (Harmon-Vukic, Gueraud, Lassonde, & O'Brien, 2009). Similarly, Sanford and Garrod's (2005) granularity hypothesis suggests that some items might be represented more specifically than others. This could explain why readers failed to activate the concept of *shovel* in Singer's (1979) experiment. The concept of *shovel* might be too broad, or might be amongst a variation of possible tools. This constraining content can occur anywhere in the text (Cook et al., 2001; Gueraud et al., 2008).

Suppression

Not only must inferences be successfully activated, irrelevant inferences and concepts must be suppressed. Gernsbacher, Varner, and Faust (1990) investigated the effect of a reader's ability to suppress information on text comprehension. They selected pairs of words, of which one was ambiguous and the other was unambiguous but semantically comparable. Immediately after reading one of each of the words of the test pair in a sentence, readers were presented with a test word that was associated with the ambiguous word but was not related to the meaning of the sentence. Participants had to decide whether or not the test word matched the meaning of the sentence they had just read. For example:

Ambiguous: He dug with the *spade*.

Unambiguous: He dug with the *shovel*.

Test word: *Ace*

Gernsbacher et al. also assessed reading skill with the Multimedia Comprehension Battery, an index of general text comprehension skill. Participants of all skill levels were exposed to both ambiguous and unambiguous sentence conditions. Results indicated that less skilled readers were slower to decide that the test word was not associated with the sentence in the ambiguous sentence condition. Thus, better readers are more effective at suppressing irrelevant information during text comprehension.

Semantic Priming

The results of experiments in which contextual support immediately precedes the test probes could be explained by semantic priming rather than by inference activation. Semantic priming occurs when activation spreads across a semantic network to closely related concepts, thereby leaving residual activity and increasing the ease with which these concepts can later be accessed. Collins and Loftus (1975) presented an in-depth discussion of spreading activation and semantic priming. However, research within the field of text processing has ruled out semantic priming as an explanation for most experimental results. For example, Lassonde and O'Brien (2009, Experiment 2) examined people's response times to their probe words. Participants were shown words in pairs and were asked to name them aloud as quickly as they could. Some word pairs were composed of the target words and other words pairs contained the word *blank* in place of one of the target words. Lassonde and O'Brien failed to find significantly different naming times between the experimental conditions and thus no evidence of semantic priming between concepts. If semantic priming had occurred following the presentation of the first word, responses would have been faster following the presentation of the second word. Similarly, the temporal delay between the reading and testing phases in the experiments by Cook et al. (2001) should also remove spreading activation as an explanation, as spreading activation would not last that long.

Individual Differences in Reading Comprehension

Many researchers have studied individual differences in discourse comprehension. For example, men and women process text differently (Sternadori & Wise, 2010), and individual differences are present in a reader's level of lexical precision (Andrews & Hersch, 2010). The most commonly investigated dimensions in individual differences in reading comprehension are: *working memory resources, text integration ability,* and *prior knowledge.*

Working Memory Resources

Working memory is the system of memory that temporarily holds information for use in verbal and nonverbal tasks (Becker & Morris, 1999). It serves as both the workspace in which computational processes are executed and as a storage mechanism for concepts that are currently active. Cognitive psychologists have created a variety of tasks in an attempt to measure working memory capacity. Working memory capacity affects numerous language-related functions, such as syntactic processing (King & Just, 1991) and a reader's ability to suppress irrelevant information during retrieval (Long, Seely, & Oppy, 1999).

Baddeley (1986) suggested that working memory both transfers information from short-term and sensory memory into long-term memory and visa-versa. Kane and Engle (2000) suggested that working memory capacity is a chief factor in the amount of information a reader can extract from long-term memory. Similarly, Conway et al. (2005) noted that most working memory span tasks reflect the assumptions of the working memory model of Baddeley (Baddeley & Hitch, 1974; Baddeley, 1986). These tasks, therefore, tend to emphasize the processing function of working memory.

Reading span task. The reading span task, however, is said to measure both the processing and storage functions of working memory (Daneman & Carpenter, 1980). The reading span task requires participants to recall the last word in each of a set of sentences for as many sentences as they can manage, much like many simple span tasks. In simple span tasks, participants recall as many items as they can. These items could be numbers, for example. Consider the following sentences:

After the storm had subsided we broke camp and departed.

Many students do not believe that homework is a good idea.

The one great stumbling block was the lack of capital.

The following day the patient asked for a bedside telephone.

After reading, the participant is then asked to recall the end words, with the correct response to the items above being (in any order): *departed, idea, capital, telephone*. This requires the reader to retain previous end words while processing new sentences. After recalling the end words, participants are presented with one of the sentences with two words missing and have to recall them. This is called a *cloze* task. The processing of new sentences while retaining previous end words requires an interaction between the storage and processing functions of working memory. In their initial analysis of the task, Daneman and Carpenter (1980) administered digit span tasks, word span tasks, tests of pronominal reference, fact retrieval, and verbal Standard Aptitude Tests (SATs). The simple span tasks (digit span and word span) were not significantly correlated to reading span. However, the other measures (pronominal reference, fact retrieval, and verbal SAT tests) were correlated with reading span. This suggests that the reading span task is a measure of complex reading ability.

The reading span task has been used many times in individual differences analyses in inference processing (e.g., Singer & Ritchot, 1996) and has guided progress in understanding reading comprehension processes. For example, a larger working memory capacity is required to process more difficult tasks (Anderson, Reder, & Lebiere, 1996).

Singer, Andrusiak, Reisdorf, and Black (1992a) investigated individual difference in bridging inference processing. Participants completed tests of reading span, text integration ability, vocabulary knowledge, general reading comprehension, as well as a bridging inference task. In the bridging inference task, participants had to connect text ideas that were either close together or far apart in the passage. For example, participants read: *"Then she found that the milk* *was three weeks old. The smell turned her stomach."* Participants must infer that the two sentences are related in order to comprehend the passage. In one experimental condition, three related intervening sentences occurred between the two sentences. Probe questions were later presented to the participant. For example, *"Was the milk sour?"* Most readers answered this question faster without the intervening text.

Singer et al. (1992a) used hierarchical regression analyses to identify which individual differences measures best predicted performance on the bridging inference task. Results indicated that reading span significantly predicted performance on the bridging inference task when the text ideas were far apart. Thus, a greater working memory can be helpful in processing more difficult text conditions.

Daneman and Carpenter's (1980) claim that the reading span task is a domain specific measure of working memory has been extensively discussed. For example, Turner and Engle (1989) demonstrated that a non-reading working memory task could accurately predict reading comprehension. Further research has also suggested that working memory capacity includes domain-specific and domain-general components (Hambrick & Ferreira, 2007). Furthermore, Gernsbacher, Varner, and Faust (1990) found high correlations between comprehension of verbal (written, auditory) and non-verbal (picture) narratives, possibly indicating that comprehension has a general component. Similarly, Perfetti and Goldman (1976) determined that memory for raw material, such as a digit span task, could not predict comprehension skill as well as a measure of "structured language".

Working memory and language comprehension. Just and Carpenter (1992) extended the work of Daneman and Carpenter (1980) and created a working memory-based model of text comprehension, called CC Reader. They examined six different domains of text comprehension in their analysis of the contribution of working memory to text comprehension: accessing working memory capacity, modularity of syntactic processing, processing complex embeddings, age-related differences, syntactic ambiguity, extrinsic memory load, and distance effects.

In an experiment of extrinsic memory load, participants were presented with syntactically complex sentences. For example: “The reporter that the senator attacked admitted the error.” King and Just (1991) reported that lower span readers were slower and less accurate in judging the meaning of such a sentence. Just and Carpenter (1992) modified the experiment conducted by King and Just by requiring participants to retain one or two unrelated words while reading the sentences. The researchers used the unrelated words as an *extrinsic memory load* that could be placed on the participant. An extrinsic memory load requires memory resources from the participant for a task unrelated to the primary task that they are completing. They found that readers were able to comprehend the material more accurately in the absence of an extrinsic memory load. Furthermore, comprehension accuracy was particularly reduced in low span readers. Just and Carpenter suggested that because the extrinsic memory load impairs sentence comprehension, both processes must be competing for the same cognitive resources.

Recent research. Recently, Long and Prat (2008) demonstrated that high span readers used information about plausible interpretations when resolving a semantic ambiguity when low span readers did not, even though they had such information. Higher exposure increased low span readers’ use of this information, and high span readers processed slower in some situations (also see Swets, Desmet, Hambrick, & Ferreira, 2007).

Furthermore, attentional control has been to shown to be a factor in one’s reading span task performance (Kane, Beckley, Conway, & Engle, 2001). Researchers have shown that low span readers were slower and less accurate in an anti-saccade task. An anti-saccade task is used to measure eye movements during reading. Readers making fewer saccades (quick eye movements) have better attention than those making more saccades. This is consistent with recent neurobiological findings indicating that working memory and attentional processes, specifically eye fixations, have quite similar neuroanatomical correlates (see Postle, Idizowski, Sala, Logie, & Baddeley, 2011, and Feredoes & Postle, 2009)

On the other end of the memory spectrum, there has also been recent research that suggests that long-term memory is very important when attempting to account for individual differences (Cook & Gueraud, 2005; Ericsson & Kintsch, 1995; Was, 2010). However, this may be heavily linked to working memory because working memory is thought to serve to send information both to and from short-term memory (and sensory memory) and long-term memory (Baddeley, 1986). Further research has confirmed that long-term memory is activated during working memory tasks to aid in that system's functioning (Lewis-Peacock & Postle, 2008).

Text Integration Ability

Measures of a reader's ability to construct and maintain a situation model have multiple forms. Radvansky and Copeland (2001) provide an interesting example of situation model updating. They used the following materials (from Glenberg, Meyer, & Lindem, 1987) to access situation model updating in readers:

> Warren spent the afternoon shopping at the store.
>
> He *picked up/set down* his bag and went over to look at some scarves.
>
> He had been shopping all day.
>
> He thought it was getting too heavy to carry.

The bag is either associated or disassociated with Warren. Readers should therefore have access to the concept of *bag* if they are effectively updating the situation model for the story. Indeed, readers responded to probes related to the critical item (bag) faster when it was physically associated with the protagonist (Warren).

Further research has investigated the dimensions of the situation model (Zwaan, Magliano, & Graesser, 1995). Participants read short stories and were instructed either to read as they normally would, or read for best possible memory of the story. The researchers labeled the sentences as being either temporally, spatially, or causally continuous or discontinuous with the preceding text. The continuity categorizations were based on whether or not a shift in that dimension was introduced to the story in that sentence. The researchers recorded reading times for each sentence, and conducted a regression analysis of the reading times to determine if a shift in any one dimension in a particular condition could produce a change in reading time that was significantly different from 0 (beta weights). The authors found that readers were most sensitive to a shift in temporal and causal dimensions in the situation under normal reading conditions, and that sensitivity to spatial shifts appeared only when the reader had the goal of maximizing memory. Thus, temporality and causality may be the situational dimensions most crucial to an understanding of the text.

Integration task. The integration task (Potts & Peterson, 1985) assesses people's situation model construction ability. It quantifies the degree to which a reader can successfully integrate newly presented with previously held knowledge. This task is most likely reflective of situation model construction ability (Singer & Ritchot, 1996).

The integration task is broken into four categories of probe questions. These are (a) memory probes (measuring text memory), (b) real probes (measuring knowledge access), (c) inference probes (measuring text inferencing), and (d) integration probes (measuring knowledge integration). The integration type questions are also referred to as *knowledge access* or *access* questions. Each type of question probes a different reader trait. The integration questions have been of most theoretical interest in the past (see Singer & Ritchot, 1996). Potts and Peterson noted that the participants' answer accuracies tended to fall into two groups: real and memory questions versus inference and integration questions. This may reflect the fact that the latter two types of questions require the reader to integrate new and old information and the former two types of questions do not.

The integration task was later modified (Hannon & Daneman, 2001) with the aim of increasing its complexity and decreasing the ease with which a participant could memorize the passages. The modified task adds two to four extra pieces of information in each sentence presented and is designed to be more difficult than the original (see Appendix B for an example). This task is a better predictor of reading comprehension than the original integration task (from Potts & Peterson, 1985).

Hannon and Daneman (2001) also compared these results to a modified reading span task and found high correlations between that task and most of their measures with the exception of knowledge access. Hannon and Daneman have also shown, using an individual differences paradigm with the use of an integration task, that older readers decline on all measures (memory, access, inference, integration), but show the most decline on measures related to the use of new information (memory, inference) - more-so than on measures related to the use of prior knowledge (Hannon & Daneman, 2009). This also mirrors the results of Potts and Peterson (1985), as the four measures of the integration task tended to be dichotomized via their correlations into memory - real, and inference - integration questions. However, the modified version of the integration task has not been used as extensively as the original (Singer & Ritchot, 1996, for example). Thus far, research in individual differences in situation model construction is limited in quantity.

Text integration and working memory are relatively unrelated cognitive processes. Radvansky and Copeland (2001) gave participants four measures of working memory, including a word span task (participants remember as many single words as possible) and the reading span task. The results indicated that performance on the working memory test was unrelated to that on the situation model-updating test. Other experimental results also suggest that working memory and text integration are relatively unrelated concepts (Radvansky & Copeland, 2001, 2004; Singer & Ritchot, 1996), but also suggest that better access to recently comprehended information improves general text comprehension (Gernsbacher et al., 1990). Similarly, Singer and Ritchot (1996) suggested that the integration task might be a measure of situation model construction ability. In their experiment, the reading span task and integration task were negligibly correlated.

Prior Knowledge

Predictive inferences necessarily require relevant prior knowledge in order to be constructed. A well-constructed situation model includes information from the reader's general world knowledge (Kintsch, 1988). The amount of prior knowledge about a given text differs greatly among readers. Research has demonstrated that prior knowledge reduces the cognitive load (amount of mental work required for a given task) placed on the reader during on-line comprehension (Fincher-Kiefer, Post, Greene, & Faust, 1988), especially if the reader's prior knowledge is central to the concept in the text (Rizzella & O'Brien, 2002).

Fincher-Kiefer et al. (1988) asked participants to read sets of sentences and then recall the last word of each sentence. Another sentence was added to the set of sentences in each experimental set until the maximum number of recallable end words was reached for that participant. However, some of their participants were baseball “experts” (i.e., they had scored higher on a baseball knowledge test than others), and half of the test sentences were regarding baseball trivia. The results suggested that those readers who were baseball experts had an easier time processing the sentences than those who were not. That is, baseball experts could recall more end words on average than non-experts. Thus, the content of that prior knowledge can also affect comprehension.

Similarly, high-vocabulary readers make more predictive inferences than low-vocabulary readers, and, in some conditions, are the only readers who do (Calvo, Estevez, & Dowens, 2003). Likewise, if no prior knowledge is present, readers might omit predictive inferences all together. Millis and Graesser (1994) investigated readers’ construction of inferences when reading scientific text and concluded that there was no evidence that readers were building predictive inferences when reading this type of text. This omission on the part of the reader of predictive inferences might be due to a higher cognitive load placed on the reader due to the content of the text.

Prior knowledge can make it easier for readers to learn as well. McNamara and Kintsch (1996) investigated how effectively readers with different amounts of prior knowledge learned material from either low coherence or high coherence expository texts. The level of coherence in the text was manipulated by the researchers through several strategies: namely, disambiguating potentially ambiguous phrases, adding descriptive elaborations, adding connectives to indicate relationships between sentences, inserting words to increase conceptual overlap between sentences, adding paragraph headers, adding summarizing sentences to link paragraphs to the rest of the text, and moving sentences to the appropriate section of the text. Texts contained approximately 900 words each. Recall of the texts was measured via both free recall and multiple-choice question about the texts one week after reading the material.

The authors found that readers with more prior knowledge actually learned material from the low coherence texts better than the high coherence texts. They suggested that this is because the low coherence texts required the readers to make more inferences than the high coherence texts did. This is largely because bridging inferences were required to connect disjointed text ideas in the low coherence text. These inferences might provide a deeper level of encoding and greater elaboration, which leads to better recall later. Low knowledge readers learned the materials from the text better when it was highly coherent. Further research supports these results as well (McNamara, 2001). Similarly, when readers are asked to explain the contents of an expository text after reading it, comprehension improves, particularly when the text is less coherent (Ozuru, Briner, Best, & McNamara, 2010).

Individual Differences in Predictive Inference Generation

Relatively little research has examined individual differences in predictive inference generation. Research has suggested that working memory capacity might be an important factor in predictive inference generation (Linderholm, 2002). Linderholm compared readers' working memory capacities with their predictive inference generation ability using 250 ms and 500 ms time delays between the reading and test phases in low, moderate, and high causality (constraint) conditions in an inference task. Working memory capacity was measured using a modified version of a working memory span task used by Singer (Singer, Andrusiak, Reisdorf, & Black, 1992; Singer & Ritchot, 1996). In this modified task, participants gave naming (spoken aloud) responses to test probes rather than written responses. Results suggest that low working memory readers make very few predictive inferences. Readers with higher working memory capacities make more inferences and make inferences in more difficult conditions.

Murray and Burke (2003) suggested that readers of all skill levels were able to encode predictive inferences. They used the Nelson-Denny Reading Comprehension test to measure reading skill and a test probe at 500 ms after the experimental passage as well as a comprehension test after the passage was complete. However, they claimed that only those with higher reading skill are activating these inferences *passively* (Murray & Burke, 2003).

Overview of the Study

Researchers have demonstrated that a high degree of contextual support increases the likelihood of predictive inference being made (Casteel, 2007; Cook et al., 2001; Fincher-Kiefer, 1993, 1995; Klin et al., 1999; Lassonde & O'Brien, 2009; Sanford & Garrod, 2005). However, very little individual differences research has been conducted regarding predictive inferences, and even less has thus far examined the propensity of readers with different cognitive traits to build predictive inferences under varying conditions of contextual support. Linderholm's (2002) article provided a good analysis of working memory capacity but failed to include any other reader traits, and was conducted using a different experimental paradigm. Knowledge access (Potts & Peterson, 1985) has yet to be investigated in relation to predictive inference generation in any of these experiments.

Experiment Design

An individual differences analysis using multiple reader traits and performance on a predictive inference task with varying conditions of contextual support was conducted. In the experiment, participants completed the reading span task (Daneman & Carpenter, 1980; Singer & Ritchot, 1996), the integration task (Potts & Peterson, 1985), and a predictive inference task

(Cook et al., 2001). The integration task measured knowledge access. The predictive inference task compared low versus high context passages. Participants made recognition judgments; they were presented with experimental texts and then asked if they had seen an inference word before. The inference words never actually occurred in the passage, and the correct answer to these probes was always "no."

Hypotheses

First, it was hypothesized that knowledge access (Potts & Peterson, 1985) would interact with the context variable. Better knowledge access on the part of the reader may aid in situation model construction and thus make predictive inference generation easier. More specifically, it was hypothesized that low-access readers will take longer to reject an inference probe word in the low-context condition compared to the high-context condition. This difference in time required to successfully generate an inference will either not occur in readers with a better text integration ability, or will be less pronounced. It was also hypothesized that reading span would not affect predictive inference activation. No significant pattern regarding accuracy of responses to probes in the predictive inference task was initially anticipated.

Secondly, it was predicted that a set of results analogous to those of previous researchers (Cook et al., 2001; McKoon & Ratcliff, 1986) would be found using a correct rejection paradigm. It was predicted that response times for high context passages would be slower than those for low context passages. This would indicate that readers did activate a predictive inference and thus had more difficulty rejected the inference word.

Higher error rates and response times than those previously observed are anticipated here because (a) participants must correctly reject inference concepts that they may think they have seen, and (b) the experimental passages herein are much longer than those used by McKoon &

Ratcliff (1986), who provide the only recognition-based predictive inference task administered to college-aged adults from which we can predict an outcome, and (c) recognition response times tend to be slower than naming times. However, there is no analogous experiment in the literature to which this one can be accurately compared, and the number of predictive inference investigations involving either a correct rejection paradigm or another recognition paradigm is quite limited. For example, researchers have used the correct rejection paradigm, which has tended to yield a higher average response time (1200-1500 ms, Zipin, Tompkins, & Kasper, 2000; 700-1150 ms McKoon & Ratcliff, 2013), but these investigations were with older adults. McKoon & Ratcliff (1986) used a correct rejection paradigm, but with passages of only two sentences in length, which yields the potential for only a significantly smaller situation model for the reader and thus a quicker recall (see the *fan effect,* Anderson, 1974). Fincher-Kiefer (1995) investigated the effect of breaks in the coherence of a discourse on predictive inference generation, and recorded average response times of roughly 1250-1475 ms. This may be the most analogous in that the task was confusing for readers, though Fincher-Kiefer did not use a correct rejection-type paradigm. Thus, we could speculate based on the above comparisons that readers might find the correct rejection paradigm following long passages unusually difficult due to the differences noted.

Method

Participants

One hundred and eight participants were recruited from the online Subject Pool website through the Department of Psychology at the University of Manitoba. All participants were enrolled in the class *Introduction to Psychology* (PSYC 1200) at the University of Manitoba. All participants' first language was English. This study complied with both CPA and APA ethical guidelines.

Materials

The participants completed three tasks, namely, the reading span task (Daneman & Carpenter, 1980; Singer & Ritchot, 1996), the integration task (Potts & Peterson, 1985), and a predictive inference task (Cook et al., 2001). The aim of the experiment was to identify cognitive factors that promote the generation of predictive inferences.

Reading span task. The original version of the reading span task was used, with few modifications (Singer & Richot, 1996; Daneman & Carpenter, 1980). The reading span task required participants to recall the last word in each of a set of sentences. Sets began at a size of two sentences at a time, up to a size of five sentences at a time. Three sets of each size (two, three, four, and five sentences) were included in the task. Two *cloze items* within a single sentence were always presented following each set of sentences, regardless of the set size. Cloze questions require the reader to fill in missing words from one of the previously viewed sentences. The experimental sets were preceded by two practice sets of size two. Reading span scores were the total number of words that participant could recall (Singer & Ritchot, 1996; Friedman & Miyake, 2005).

Integration task. The integration task (Potts & Peterson, 1985) aims to quantify a reader's ability to integrate previously known world knowledge with newly learned information. Participants were presented with a set of statements, such as those below:

A *JAL* is larger than a *TOC*.

A *beaver* is larger than a *CAZ*.

A *TOC* is larger than a *pony*.

These concepts are correctly ordered in size: JAL > TOC > pony > beaver > CAZ. The task consisted of five experimental passages, preceded by one practice passage. Each passage was followed by 18 yes/no probe statements, which contained a combination of two of the above concepts. The participant was asked to respond YES or NO in regard to the statement, via keys labeled on the keyboard. For example: Is a TOC larger than a CAZ?

The integration task includes four categories of probes. These are (a) memory, (b) real, (c) inference, and (d) integration. The memory questions require the participant to recall a single fact, verbatim that was previously presented in the material set. The real questions require the participant to respond to a probe question containing only real life concepts that should already be known; for example: *Is a beaver larger than a pony?* The inference questions require the participant to use information from two statements at once; for example: *Is a JAL larger than a pony?* If the participant is to answer this question correctly, they must identify that a) *A JAL is larger than a TOC* and b) *A TOC is larger than a pony*. The integration or "access" questions are of the most theoretical interest here. These questions require the reader to use all of the information presented to them if they are to answer the probe question correctly. For example: *Is a JAL larger than a CAZ?* To answer correctly, the reader must identify that a) *A JAL is larger than a TOC*, b) *A TOC is larger than a pony,* c) *A pony is larger than a beaver,* and d) *A beaver is larger than a CAZ.*

The experimental passages and questions were presented in the same previously-generated random order to all participants. Each experimental set included 8 integration questions, 6 memory questions, 2 inference questions, and 2 real questions. The correct answer to half of the questions was no, and the other half, yes. Two of the 18 questions from each set contained the same combination of two concepts, to a total of nine combinations of concepts, but with their order in the probe statement reversed so as to reverse the truth of that statement.

Inference task. The predictive inference task of Cook et al. (2001, Experiment 1; similar versions in Klin et al., 1999 and Lassonde & O'Brien, 2009) was used, with few modifications. Each experimental passage occurred in one of two conditions: low context and high context. Experimental passages include an introduction section, of four or five sentences in length. For example:

(1) Jimmy was the new kid on the block.

(2) Although his parents urged him to go meet the other kids in the neighborhood, he was shy and hadn't made any new friends.

(3) One Saturday morning, his mom asked him to go to the store for her.

(4) While he was walking back home, Jimmy ran into some of the kids from the neighborhood.

(5) They asked him if he wanted to play with them.

The introduction section was, on average, 62.0 words in length (range: 42 – 72 words). Either a low-context or high-context section of two sentences in length followed. For example:

Low-Context:

(6a) Jimmy was delighted and ran across the street to play with them.

(7a) They taught him a fun game that involved throwing Nerf balls at a target to get points.

High-context:

(6b) Jimmy was delighted and ran across the street to play with them.

(7b) They taught him a fun game that involved throwing rocks at a target to get points.

This section reduces or increases the likelihood of a certain outcome in the story. The final section of each passage was an elaborative section of three or four sentences in length. For example:

(8) Jimmy and his friends were having a great time.

(9) Jimmy even won the game once or twice.

(10) He stepped up to take his turn and aimed at the target.

(11) He missed, though, and he accidentally hit the door of a new car.

The mean length of the elaborative section was 62.6 words in length (range = 50-76 words). The final sentence of the inference-evoking section facilitates the activation of a predictive inference. The mean length of the inference-evoking sentence was 41.7 characters (range = 38-46 characters). This inference is much more likely in the high-context condition.

Participants are shown the probe word following their reading of the passage and are required to respond as to whether the word occurred in the story. Only one-syllable words were used as probe words. Participants also answered a general comprehension question about each passage, following the test probe. Comprehension questions unrelated to the inference but were regarding a relatively central concept from the story; these were used to ensure general comprehension of the passage.

The recognition task required lure probes (a "no" probe) and target probes (a "yes" probe) as well. For example, the recognition probe regarding the inference in the above example is *dent*. The correct answer to the probe is *no,* as the reader did not actually see the word *dent* in the passage. Target probes were words that did occur in the passage. Finally, a lure probe was added. Not only did the participant not see that word, but it is altogether unrelated to the passage. Each passage used only 1 probe.

There were 18 experimental passages and 24 filler passages – 12 target passages and 12 lure passages - for a total of 42 passages. The filler passages were modified from O'Brien, Albrecht, Rizella, and Halleran (1998). Two experimental lists were created as to vary the experimental condition for each experimental passage in task. The lists were also counterbalanced so that half of each of the experimental, target, and lure passages occurred in each half of the experiment. Six practice passages preceded all other passages. Example passages can be seen in Appendix A.

Procedure

The aforementioned tasks were administered in Singerlab in Duff Roblin Building at the University of Manitoba. Up to four participants were tested at a time, each at their own computer workstation and in their own room. Participants received partial course credit in compensation for their participation.

The experiment was broken into two experimental sessions, and the order in which the tasks were administered was counterbalanced. Participants completed either the inference task or the individual differences measures – the reading span and integration tasks – in the first session, and then the remaining task(s) in the latter session. The order in which the individual differences were administered was also counterbalanced, so that some participants completed the reading span task first and some the integration task first. This was done as to control for order effects, to be discussed shortly. Each experimental session lasted up to one hour. The sessions were conducted two days apart, each at exactly the same time of day.

All tasks were administered through MEL (Micro Experimental Laboratory) software. Predictive inference task and integration task data was gathered via computer responses. Reading span task responses were gathered via a pencil and paper answer sheet. For all tasks, participants received a brief set of verbal instructions as well as a more complete set of instructions via paper.

Reading span. The reading span task was presented to participants via a computer screen, and participants responded on a pencil and paper answer form. At the start of each trial, participants saw the words “NEXT TRIAL” displayed for 1000 ms. Next, a fixation point (an X) was displayed for 500 ms. The sentences then appeared one-at-a-time for 8 seconds each. After, the words “WRITE END WORDS” appeared for 12 seconds. The participant then recorded their answer. A tone sounded, and the cloze item for the set appeared on the screen for 12 seconds. The participant then recorded his or her answer on the paper for the cloze items. Another tone sounded to indicate the end of the trial, and the next trial began. This continued over 14 trials.

Integration task. Participants were instructed to press the space bar to initiate each trial. Each trial began with a fixation point, which was displayed for 500 ms. Next, the three sentences of the experimental passages were displayed simultaneously, each on a new line. No time limit was imposed on study. The participant next pressed the space bar when they were ready to continue to the test questions. A 2500 ms interval then occurred, followed by the message “TEST ITEMS”, which was displayed for 1000 ms. Another fixation point was then presented for 500 ms prior to each probe question. Participants answered with the “.” and “x” keys, which were labeled “YES” and “NO”, respectively, on the keyboard. After responding, the probe disappeared and a new trial began once the participant indicated readiness.

Predictive inference task. The words, <PRESS THE SPACE BAR FOR MORE TEXT>, remained at the top of the screen for the duration of each passage. After the first space-bar press, the first sentence of the passage appeared. Sentences appeared at the vertical center, left aligned, on the screen. At each bar press, the previous sentence disappeared, and the next sentence appeared on the screen in the same location. Each sentence in the passage remained on the screen for a maximum of ten seconds before the next sentence replaced it. After the last sentence of the passage, the message at the top of the screen disappeared, and the participant was shown a fixation point for 500 ms. Next, the probe word appeared where the fixation point had been. Participants responded YES or NO via the corresponding keys on the keyboard as to whether or not they had seen that exact word in the preceding passage. The “x” and “.” keys were labeled NO and YES, respectively. A 2500 ms break then occurred, followed by another fixation point for 500 ms, and a comprehension question then appeared. Participants then responded to the comprehension question using the same YES and NO keys. After the comprehension question, another 2500 ms break occurred, and the message, <PRESS THE SPACE-BAR FOR MORE TEXT> again appeared and remained at the top of the screen. At the halfway point of the experiment, the word “REST” was displayed on the screen for 40 seconds in order for the participants to rest briefly before continuing. Both accuracy and response time data were recorded for the probe and comprehension questions in the predictive inference task.

Results

One hundred and eight 108 participants attended the experiment. Of these, seven participants failed to complete the inference task in the allotted time. Of the 101 participants that completed all three tasks, three yielded illegible reading span response forms.

One participant scored 0 on the cloze score. This suggested that the participant attended only to the end words, so his/her data were disregarded. A criterion of three standard deviations was used for trimming scores for all tasks, with the exception of accuracy scores on the inference task. This criterion has been used frequently in similar experiments in the past (Albrecht & O’Brien, 1993; Cook, Limber & O’Brien, 2001; O’Brien & Albrecht, 1992). The data of three participants were removed from the data set on the basis of their inference task probe response times. All three had average response times of over 4500 ms for at least one of either low- or high-context type probes. No other participants yielded any average scores over three standard deviations from the mean on any other measure.

The data of 94 participants were analyzed. Forty-four and 50 participants completed each of list 1 and 2, respectively, of the inference task. Twenty-five, 24, 25, and 20 participants completed the inference task (INF), reading span task (RS), and integration (INT) task; in the following orders, respectively: INT, RS, INT; INF, INT, RS; RS, INT, INF; INT, RS, INF.

Individual Differences

For the purpose of these analyses, reading span and access score were subjects to a median split. Cut points of 33 (*max* = 42) and 25 (*max* = 40) were applied for the reading span and access scales, respectively. Participants who scored below the cut points were classified as low span or low access. The resulting assignment of number of subjects to the four reading span x access conditions appear in Table 2. Descriptive statistics for the individual differences measures appear in Table 3.

Preliminary analysis. Participant responses were first analyzed using an ANOVA (analysis of variance) that included all of the variables in the experimental design. Context (high or low) was a within-subjects variable, and reading span, access, list, and order were between-subjects variables. In the recognition time analysis, there was a main effect of context, $F(1,63) = 3.85$, $MSE = 210{,}692.07$, $p = .05$. No other statistically significant main effects or interactions were present, all *p*s > .05. In an analysis of accuracy scores, there was a significant Context x List interaction: Participants responded more accurately to some list items than others, which varied across conditions of context, $F(1,63) = 57.58$, $MSE = 54.58$, $p < .01$. No other significant effects were present, all *p*s > .05.

Main analyses. Because the design of the preliminary analysis yielded 64 groups over 94 subjects, order and list were disregarded in the main analysis. The data were analyzed using a context (within-subjects variable) by reading span by access (between-subjects variables) ANOVA. Recognition time to probes was the dependent variable. The inference task recognition times as a function of context, reading span, and access appear in Table 4. There were no new significant main effects or interactions above those of the preliminary analysis present, all *p*s > .05. Figure 1 shows low and high access readers' recognition times in low and high context conditions. This interaction was hypothesized, though was not significant.

ANOVA was also applied to the inference task accuracy rates, with the same context by reading span by access design as above. The mean accuracies as a function of context, reading span, and access appear in Table 5. Participants responded marginally more accurately to probes following low context passages (mean = 87.0%) than high context passages (mean = 83.4%) $F(1, 90) = 3.30$, $MSE = 642.46$, $p = .07$. The access x context effect was significant, $F(1, 90) = 5.61$ $MSE = 1{,}092.85$, $p = .02$. Tests of simple main effects showed that low access readers were less accurate in responding to high context probes than low context probes, $F(1, 92) = 8.81$, $p < .01$, but high access readers were not, $F(1, 92) < 1.00$. Figure 2 shows the low and high access readers' accuracy rates in the low and high context conditions.

Regression analyses. The data were also submitted to hierarchical regression analysis. This type of analysis has been used in past investigation of individual differences in inference processing (for example, Singer, Andrusiak, et al., 1992). The differences between the average low context and high context scores were used as the dependent variable. This was done using each response times and accuracy scores, and yields a measure for the degree to which readers are affected by the context variable. However, the regression analyses yielded no results that were informative above those of the ANOVAs: Only access scores served to predict the difference between low and high context scores, and only in regard to response accuracy, $F(1, 92) = 12.51$, $R^2 = .12$, $MSE = 36.19$, $p < .01$. See Figure 3 for a regression plot of accuracy difference scores against access scores.

Predictive Inference Task

An analysis of the data, disregarding the individual differences variables, was conducted in order to evaluate the effects of context in more detail because this predictive inference task was newly created. The average rate of correct responses to comprehension questions was 87.0%. Further analyses of low context versus high context responses were conducted alternately treating participants (*F*1) and experimental items (*F*2) as the random effect (Clark, 1973). The data were analyzed first using response times as the random variable. List and order were not included in the design because the addition of further variables would reduce power. This analysis involves only 9 means per condition.

There was a significant main effect of context: Responses were slower to inference probes following high context passages (mean recognition time = 1690 ms) than low context passages (mean recognition time = 1621 ms), though *F*2 was not statistically significant, $F1(1, 93) = 4.13$, $MSE = 223{,}629.02$, $p = .045$, $F2(1, 16) = 1.79$, $MSE = 33{,}795.70$, $p = .20$. There was a main effect of context when analyzing accuracy scores following low and high context passages as well; responses to probes were less accurate following high context passages (mean accuracy = 83.4%) than low context passages (mean accuracy = 87.0%), though *F2* was only marginally statistically significant, *F1*(1, 93) = 2.94, *MSE* = 591.02, *p* = .09, *F2*(1, 16) = 2.78, *MSE* = 162.14, *p* = .12.

Probe types comparison. A comparison of responses to target, lure, and inference probes was conducted to evaluate the recognition paradigm because no previous experimentation provides a guide to expectations regarding these results. Thus, this analysis was conducted for exploratory purposes, though the results of the analysis had no bearing on hypotheses. ANOVA was applied to both recognition times and accuracies. The design was the same for both dependent variables: list and order were between-subjects variables and probe type was the within subjects variable. In an analysis of response times, the probe type effect was significant. Participants took longer to respond to inference probes (mean recognition time = 1656 ms) than to target probes (mean recognition time = 1567 ms) or lure probes (mean recognition time = 1523 ms), *F*(2, 172) =12.56, *MSE* = 4,578.25, *p* < .01. The data for average recognition times across probe types and task orders is displayed in Table 6. Pairwise contrasts were conducted to evaluate differences between probe types. There was a significant difference between inference probe and target probe response times, *t(93)* = 2.42, *p* = .02, and between inference and lure probe response times, *t(93)* = 4.86, *p* < .01. There were no other significant main effects or interactions.

In the analysis of recognition accuracies, there was a main effect of probe type. Participants were more accurate in rejecting inference probes (mean accuracy = 85.2%) than in accepting target probes (accuracy = 76.2%) but less accurate than in rejecting lure probes (mean accuracy = 90.5%), *F*(2, 86) = 9.11, *MSE* = 451,234.99, *p* < .01. There was also a significant probe type by order interaction, *F*(2, 172) = 3.00, *MSE* = 148,491.28, *p* < .01. The data for that interaction are shown in Table 7. Pairwise contrasts were conducted to evaluate differences between probe types. There was a significant difference between inference probe and target probe accuracy rates, *t(93)* = 5.22, *p* < .01, and between target and lure probe accuracy rates, *t(93)* = 4.67, *p* < .01. There were no other significant effects.

Discussion

Several findings were of theoretical interest, namely, those related to individual differences and, more specifically, the access by context interaction. Those effects specific to the predictive inference task and effects of the counterbalancing variables of the experiment are also of interest. These findings are discussed, as well as future directions for research in light of these findings.

Individual Differences

The access measure. Perhaps most noteworthy of the results is that low access readers were more affected by context than high access readers. In the paradigm used in the predictive inference task, the reader had to reject an inference probe that did not actually occur. This was made more difficult by additional facilitating context because the facilitating context should have made it more likely that the inference probe be previously activated. Drawing the inference would cause a reader to either respond incorrectly or more slowly to the inference probe because they would need more time to reject an activated concept. Indeed, low access readers had much more difficulty in rejecting the inference probe in the high context condition as opposed to the low context condition and were able to do this more accurately. See Figure 1 for an overview of the access by context interaction.

Access is said to be reflective of the ease with which a reader can retrieve concepts from prior world knowledge and integrate them into the current situation model (Singer & Ritchot, 1996). Each of these steps (access and integration) could theoretically aid in yielding a lower accuracy rate given enough facilitating context because the inference might be better encoded or drawn more easily with the addition of more facilitating context. The results indicate that even high access readers must have failed to realize some portion of the time that the probe had not actually occurred and merely represented an inferred concept. Interestingly, higher access readers had an easier time rejecting the inference probes in the high context condition. This, in conjunction with the fact that low access readers were more effected by the shift in context, suggests that higher access readers are more aware of the original state of the text, a result not supported by previously experimentation (Schmalhofer, McDaniel, & Keefe, 2002), which has suggested that predictive inferences are likely to be encoded directly into the situation model. Were it the case that predictive inferences were encoded directly into the situation model, higher access readers would have made more errors given that they would have better encoded their predictive inferences, and thus been more likely to recall them.

Given this contradictory set of results, two possibilities emerge. First, it is possible that knowledge access is reflective of an ability to retain a more specific representation of the text than originally thought; this is different than retaining a verbatim representation of the text. Were it identical to a verbatim representation of the text, reading span would have predicted probe responses in some way. A better situation model might include a more detailed and more structured model of facts regarding the current textbase. If higher access readers simply had more immediate and ready access to their situation models, they would have been able to respond faster but not necessarily more accurately. Thus, higher access readers must be able to maintain a more specific representation of the text in their situation models that allows them, on demand, to make more accurate judgments regarding inferred concepts.

Alternatively, it is possible that the access by context interaction in the accuracy data reflects that predictive inferences might be represented as hypotheticals rather than absolute facts

(Campion, 2004). This would mean that, when incorporating the information from the inference-evoking sentence into the situation model, higher access readers might be aware of the restrictions of the inferences that they are making, or of the time frame in which these events would occur. Indeed, temporal information is thought to be an important dimension included in situation models (Zwaan et al., 1995). Furthermore, higher access readers might be able to retain the information necessary to know that the predictive event was inferred but did not explicitly happen. Previous research has already appeared in support of the hypothesis that predictive inferences might be represented as hypothetical facts (Campion, 2004), and others have made similar suggestions (McKoon & Ratcliff, 2013, pp. 241). This possible explanation of the results is not mutually exclusive to that of higher access readers building more detailed situation models. However, the hypothesis regarding predictive inferences' encoding as hypothetical facts has many limitations, and lacks proper supporting evidence that is necessary prior to its positing. For example, Campion (2004) used a lexical decision paradigm in his experiments, the limitations of which have already been discussed. Any direct, further evidence to that of Campion regarding the representation of predictive inferences as hypothetical facts is, to the knowledge of this researcher, currently non-existent.

Reading span. The lack of reading span by context effects in this case is not surprising; the inferential concept is initially instated mid-passage. Roughly five sentences occur between the instantiation of the inference concept and its related probe – far out of the range of what is usually considered working memory (1-2 sentences). Were the inference concept initially instated within the range of working memory or the passages shorter, then higher span readers might have more readily been able to recall the exact form of the textbase and thus whether the inference word actually occurred. Given the gap between the initial instantiation of the inference concept and the probe, only information in the situation model should remain to inform the reader as to whether an event occurred (the inference concept). More surprising, though, are the results here in contrast to those of Linderholm (2002), who found that readers with better working memory made more predictive inferences. The reason for this result might be in differences in task design. Linderholm used a naming paradigm for the inference task, but a name-aloud procedure for the reading span task as well, which is a highly unusual paradigm. It is possible that participants in Linderholm's experiments performed similarly on both tasks due to the similarity in the way responses are collected between the two tasks.

Predictive Inference Task

The predictive inference task was also scrutinized apart from individual differences effects in order to better examine its success in creating context effects and differences across probe types, particularly because it had not been used before in a recognition paradigm. This material set (from Cook et al., 2001) did succeed in producing context effects. However, the items-random analysis inference probe responses failed to produce significant effects, though this is most likely due to the low power of the statistic analysis – only 9 items occurred in each group. There would have been no theoretical reason to anticipate that this material set would produce context effects in one paradigm (naming paradigm) but not yield significant effects in a new paradigm. However, its failure to do so would have created concerns regarding the strength previous findings (for example, Cook et al., 2001; Klin et al., 1999; Lassonde & O'Brien, 2009).

Effects of the Counterbalancing Variables.

The list effects observed would suggest that some experimental items are more difficult in one condition – either low context or high context – than another. This is to be expected, and is of limited concern given the extent to which this set of materials has been used prior (Cook et al., 2001; Klin et al., 1999; Lassonde & O'Brien, 2009). Given an approximately equal number of participants encountering each experimental list, these effects should wash out over the whole population. However, given that list interactions were present, it is possible for their effects to skew group means, which should be kept in mind when interpreting results.

The pattern of results regarding task order yields similar conclusions. Upon visual inspection, it appears that participants tended to respond to probes faster when the inference task was completed in the first experimental session, but there was no main effect of order nor interaction between task order, list, context, reading span, or access. However, there was a significant order by probe type interaction. The nature of the interaction was such that participants responded to different probe types faster than others in different task orders, which limits conclusions that can be made regarding participant responses across probe types. Like list effects, effects of task order will skew the results.

Most importantly, task order effects found are informative as to best practices for executing experiments with multiple tasks. Future researchers might choose to either counterbalance task order because it has been found that task order effects are present, or select the task order with the most favourable outcome. For example, it was found that participants tended to respond faster to probes when the inference task was completed in the first session. The latter has more commonly been the case: no other work exists in the individual differences literature where task order has been mixed, which makes the effects found here more significant.

Limitations

Primary concerns regarding the limitations of this experiment fall into two categories. First, average response times are higher than is usually the case, and average accuracy scores are lower than those usually observed. For example, use of a similar inference task yielded average response times of roughly 1250-1450 ms, and average accuracy rates of 85-90% depending on the experimental condition (Fincher-Kiefer, 1995). The initially predicted pattern of results regarding responses to low and high context probes was partially correct. It was anticipated that there would be differences between response times, but not between accuracy rates. Indeed, it has been suggested that response time and accuracy rates are largely interdependent (Ratcliff, 2002), and no evidence supporting a speed-accuracy trade-off was present in this dataset. Participants were both faster and more accurate in rejecting inference probes in the low context condition than the high context condition, regardless of the effects of other variables (individual differences variables, counterbalancing variables). However, as already discussed, no previous work yields an analogous mode of comparison that might indicate what should be expected in these regards. The higher than usual response times might be reflective of the longer than usual passage length – approximately 9 sentences instead of approximately 2 sentences - and the lower than average accuracy scores light be reflective of the difficulty of the task. Theoretically, there would have been a bias to answer "no" to probes as well, because more than half of the correct answers to probes were "no."

Second, conclusions based on the accuracy scores might be limited due to concerns regarding the certainty with which the results confirm that participant responses are reflective of inference activation. As already noted, the length of the average response time does not permit conclusions regarding whether the inference was made initially on-line or not until the probe demanded it, though this may be somewhat immaterial as the process is none-the-less inferential. However, it could be posited that even if a reader correctly rejects the inference concept, they might still have made the inference. This could be the case, and might render some of the response time data less meaningful. However, this process would presumably take longer than correct rejection of the inference probe wherein the participant never made the inference, and with these aforementioned timeframes one can still infer that comparatively slower rejection of and more difficulty in rejecting an inference probe indicates that the inference was more likely to be made following low context probes than high context probes. Indeed, slower response times for high context probes were coupled with lower accuracy scores.

Future Directions

Results obtained raise questions regarding both predictive inferences and knowledge access. Another similar individual differences analysis might shed light on the time course in which predictive inferences are made access reader types, particularly in regard to how robust their encoding is (see Keefe & McDaniel, 1993) and what information better readers build into a situation model. Access has thus far been shown to play a role in bridging text concepts in expository text (Doering & Singer, 2011), text validation (Singer & Doering, in press), and inference processing (Singer & Ritchot 1996), often in stark contrast to working memory, which frequently appears to play a role in processing opposite to knowledge access and text integration. Indeed, given the extensive range of roles that working memory has been shown to play (see Just & Carpenter, 1992, for example), it is interesting to find an individual differences measure so opposing. Beneficial individual differences projects might include an investigation of the role of knowledge access in tasks such as that administered by Zwaan et al. (1995). Such a project might indicate which dimensions of the situation model higher access readers are more likely to make.

In some ways, the results obtained here are in no way surprising, knowledge access and integration appears to play a role in predictive inference generation. Generating a predictive inference requires the reader to extract relevant prior world knowledge from long-term memory stores and use that information to generate a prediction regarding what happens next. On the contrary, it is in some ways more surprising that higher access readers make fewer errors when prompted with inference probes; they are aware that the probes did not actually occur, and it is this result that prompts questions regarding the nature of efficacious situation model construction, and what traits make for a better reader.

References

Albrecht, J. E., & O'Brien E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1061-1070.

Anderson, J. R. (1974). Retrieval of prepositional information from long-term memory. *Cognitive Psychology*, *6*, 451-474.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30*, 221-256.

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masking neighbor priming. *Journal of Experimental Psychology: General, 139,* 299-318.

Baddeley, A. D. (1986). Working Memory. Oxford, England: Oxford University Press.

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

Becker, J. T., & Morris, R. T. (1999). *Working memory*. *Brain and Cognition, 41,* 1-8.

Calvo, M. G., Estevez, A., & Dowens, M. G. (2003). Time course of elaborative inferences as a function of prior vocabulary knowledge. *Learning and Instruction, 13,* 611-631.

Campion, N. (2004). Predictive inferences are represented as hypothetical facts. *Journal of Memory and Language, 50,* 149–164.

Casteel, M. A. (2007). Contextual support and predictive inferences: What do readers generate and keep available for use? *Discourse Processes, 44,* 51–72.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335-359.

Collins, A. M., & Loftus, A F. (1975). A spreading-activation theory of semantic priming.

Psychological Review, 82, 407-428.

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

Cook, A. E., & Gueraud, S. (2005). What have we been missing? The role of general world knowledge in discourse processing. *Discourse Processes, 39,* 265-278.

Cook, A. E., Limber, J. E., & O'Brien, E. J. (2001). Situation-based context and the availability of predictive inferences. *Journal of Memory and Language, 44,* 220-234.

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

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Doering, J. C. & Singer, M. (2011, June). Inferences from expository text: An individual differences analysis. Poster presented at the meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science, Winnipeg, MB., Canada.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211-245.

Feredoes, E., & Postle, B. R. (2009). Prefrontal control of familiarity and recollection in working memory. *Journal of Cognitive Neuroscience, 22,* 323-330.

Fincher-Kiefer, R. (1993). The role of predictive inferences in situation model construction. *Discourse Processes, 16,* 99-124.

Fincher-Kiefer, R. (1995). Relative inhibition following the encoding of bridging and predictive inferences. *Journal of Experimental Psychology, 21,* 981-995.

Fincher-Kiefer, R., Post, T. A., Green, T. R., & Faust, J. F. (1988). On the role of prior knowledge and task demands in the processing of text. *Journal of Memory and Language, 27,* 416-428.

Friedman, N. P., & Miyake, A. (2005) Comparison of four scoring methods for the reading span test. *Behaviour Researching Methods, 37,* 581-590.

Garnham, A., Oakhill, J., & Reynolds, D. (2002) Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory and Cognition, 30,* 439-446.

Garrod, S., O'Brien, E. J., Morris, R. K., & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16,* 250-157.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 430-445.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26,* 69-83.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371-395.

Guéraud, S., Tapiero, I., & O'Brien, E. J. (2008). Context and the activation of predictive inference. *Psychonomic Bulletin & Review, 15,* 351–356.

Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory capacity in syntactic ambiguity resolution. *Journal of Experimental Psychology, 136,* 64-81.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component process of reading comprehension. *Journal of Educational Psychology, 93,* 103-128.

Hannon, B., & Daneman, M. (2009). Age related changes in reading comprehension: An individual differences perspective. *Experimental Age Research, 35,* 432-456.

Harmon-Vukic, M., Guéraud, S., Lassonde, K. A., & O'Brien, E. J. (2009). The activation and instantiation of instrumental inferences. *Discourses Processes, 46,* 467–490.

Johnson-Laird, P. N. (1983). *Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness*. Cambridge: Harvard University Press.

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

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of experimental psychology: General, 130,* 169-183.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26,* 336-358.

Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language, 32*, 446–463.

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

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163-182.

Klin, C. M., Guzman, A. E., & Levine, W. H. (1999). Prevalence and persistence of predictive inferences. *Journal of Memory and Language, 40,* 593–604.

Lassonde, K. A., & O'Brien, E. J. (2009). Contexual specificity in the activation of predictive inferences. *Discourse Processes, 46,* 426-438.

Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary activation of long-term memory supports working memory. *The Journal of Neuroscience, 28,* 8765-8771.

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes, 34,* 259-280.

Long, D. L., & Prat, C. S. (2008). Individual differences in syntactic ambiguity resolution: Readers vary in their use of plausible information. *Memory and Cognition, 36,* 375-391.

Long, D. L., Seely, M. R., & Oppy, B. J. (1999). The strategic nature of less skilled readers' suppression problems. *Discourse Processes*, *27*, 281-302.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 82-91.

McKoon, G., & Ratcliff, R. (1989). Semantic associates and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 326-338.

McKoon, G., and Ratcliff, R. (2013). Aging and predictive inference: a diffusion model analysis. *Journal of Memory and Language, 68,* 240-254.

McNamara, D. S. (2001). Reading both high-coherence and low coherence texts: Effects of text sequencing and prior knowledge. *Canadian Journal of Experimental Psychology, 55,* 51-62.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22,* 247-288.

Millis, K. K., & Graesser, A. C. (1994). The time-course of constructing knowledge-based inferences for scientific texts. *Journal of Memory and Language, 33,* 583-599.

Murray, J. D., & Burke, K. A. (2003). Activation and encoding of predictive inferences: The role of reading skill. *Discourse Processes, 35,* 81-102.

Murray, J. D., Klin, C. M., & Myers, J. L. (1993). Forward inferences in narrative text. *Journal of Memory and Language, 32,* 464-473.

O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 777-784.

O'Brien, E. J., Albrecht, J. E., Rizzella, M. L., & Halleran, J. G. (1998). Updating a situation model a memory-based text processing view. *Journal of Experimental Psychology, 24*, 1200–1210.

O'Brien, E. J., Plewes, P. S., & Albrecht, J. E. (1990). Antecedent retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 241-249.

O'Brien, E. J., Shank, D. M, Myers, J. L., & Rayner, K. (1988). Elaborative inference during reading: Do they occur on-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 410-420.

Ozuru, Y., Briner, S., Best, R., & McNamara, D. S. (2010). Contributions to self-explanation of high- and low-cohesion texts. *Discourse Processes, 47,* 641-667.

Peracchi, K. A., & O'Brien, E. J. (2004). Character profiles and the activation of predictive inferences. *Memory and Cognition, 32,* 1044-1052.

Perfetti, C. A., & Goldman, S. R. (1976). Discourse memory and reading comprehension skill. *Journal of Verbal Learning and Verbal Behavior, 15,* 33-42.

Postle, B. R., Idzikowski, C., Sala, S. D., Logie, R. H., & Baddeley, A. D. (2011). The selective disruption of spatial working memory by eye movements. *The Quarterly Journal of Experimental Psychology, 59,* 100-120.

Potts, G. R., Kennan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language, 27,* 399-415.

Potts, G. R., & Peterson, S. B. (1985). Incorporation versus compartmentalization in memory for discourse. *Journal of Memory and Language, 24,* 107-118.

Radvansky, G. A., & Copeland, D. E. (2001). Working memory and the situation model. *Memory and Cognition, 29,* 1073-1080.

Radvansky, G. A., & Copeland, D. E. (2004). Working memory span and situation model processing. *The American Journal of Psychology, 117,* 191-213.

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin and Review, 9,* 278-291.

Rizzella, M. L., & O'Brian, E. J. (2002). Retrieval of concepts in script-based texts and narratives: The influence of general world knowledge. *Journal of Experimental Psychology, 28,* 780-790.

Sanford, A. J., & Garrod, S. C. (2005). Memory-based approaches and beyond. *Discourse Processes, 39,* 205–224.

Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes, 33,* 105-132.

Singer, M. (1979). Processes of inference during sentence reading. *Memory and Cognition, 7,* 192-200.

Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. (1992a). Individual differences in bridging inference processes. *Memory & Cognition*, *20*, 538-548.

Singer, M., & Doering, J. C. (in press). Exploring individual differences in validation processing. *Discourse Processes.*

Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992b). Validation of causal bridging inferences. *Journal of Memory and Language, 31,* 507-524.

Singer, M., & Ritchot, K. (1996). Individual differences in inference validation. *Memory & Cognition*, *24*, 733-743.

Sternadori, M. M., & Wise, K. (2010). Men and women read news differently. *Journal of Media Psychology: Theories, Methods, and Applications, 22,* 14-25.

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

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 386-397.

Zipin, L. M., Tompkins, C. A., & Kasper, S. C. (2000). Effects of foregrounding on predictive inference generation by normally aging adults. *Aphasiology, 14,* 115–131.

Table 1

Results of Cook et al. (2001); mean naming times following a test probe

	Context	
Delay	High	Low
500 ms	506	521
250 ms	531	543

Note. Response times are presented in milliseconds (ms).

Table 2
Assignment of subjects to low and high reading span and access groups

	Low Reading Span	High Reading Span
Low Access	20	25
High Access	25	24

Table 3
Descriptive statistics regarding individual differences measures

Measure	Mean	Standard deviation	Standard error of the mean
Reading Span	32.83	5.83	0.59
Cloze	9.53	4.19	0.42
Integration	27.19	7.82	0.79
Memory	22.29	4.83	0.49
Inference	7.53	2.15	0.22
Real	7.68	2.11	0.21

Note. Reading span maximum score = 42, cloze maximum score = 24, integration maximum score = 40, Memory maximum score = 30, Inference maximum score = 10, real maximum score = 10.

Table 4

Participants' mean correct recognition times by context, reading span and access

	Low Access		High Access	
	Low Span	High Span	Low Span	High Span
Low Context	1712 (116)	1574 (95)	1634 (95)	1550 (103)
High Context	1769 (124)	1674 (102)	1685 (102)	1634 (110)

Note. Vales are recognition times in milliseconds. Standard error is shown in brackets.

Table 5

Participant's mean response accuracies by context, reading span, and access

	Low Access		High Access	
	Low Span	High Span	Low Span	High Span
Low Context	.87 (.03)	.88 (.02)	.83 (.02)	.90 (.03)
High Context	.84 (.04)	.81 (.03)	.85 (.03)	.88 (.04)

Note. Values are accuracy scores as portions correct /1. Standard error is shown in brackets.

Table 6

Participants' mean correct recognition times by probe type and task order

	Probe Type		
Task order	Inference	Target	Lure
Inf, Rdsp Int	1590 (92)	1550 (103)	1507 (84)
Inf, Int, Rdsp	1633 (95)	1735 (107)	1545 (87)
Rdsp, Int, Inf	1597 (92)	1468 (103)	1518 (84)
Int, Rdsp, Inf	1811 (104)	1535 (116)	1544 (95)

Note. Values shown are recognition times in milliseconds. Inf = inference task; Int = integration task; Rdsp = reading span task. Standard error of the mean is in brackets.

Table 7

Participants' mean accuracy rates by probe type and task order

	Probe Type		
Task order	Inference	Target	Lure
Inf, Rdsp Int	.86 (2.5)	78.9 (1.8)	1.00 (5.9)
Inf, Int, Rdsp	.87 (2.6)	71.2 (1.8)	.86 (6.1)
Rdsp, Int, Inf	.86 (2.5)	74.7 (1.8)	.86 (5.9)
Int, Rdsp, Inf	.81 (2.8)	80.3 (2.0)	.89 (6.6)

Note. Values shown are accuracy rates as portions correct. Standard error of the mean is shown in brackets. Inf = inference task; Int = integration task; Rdsp = reading span task

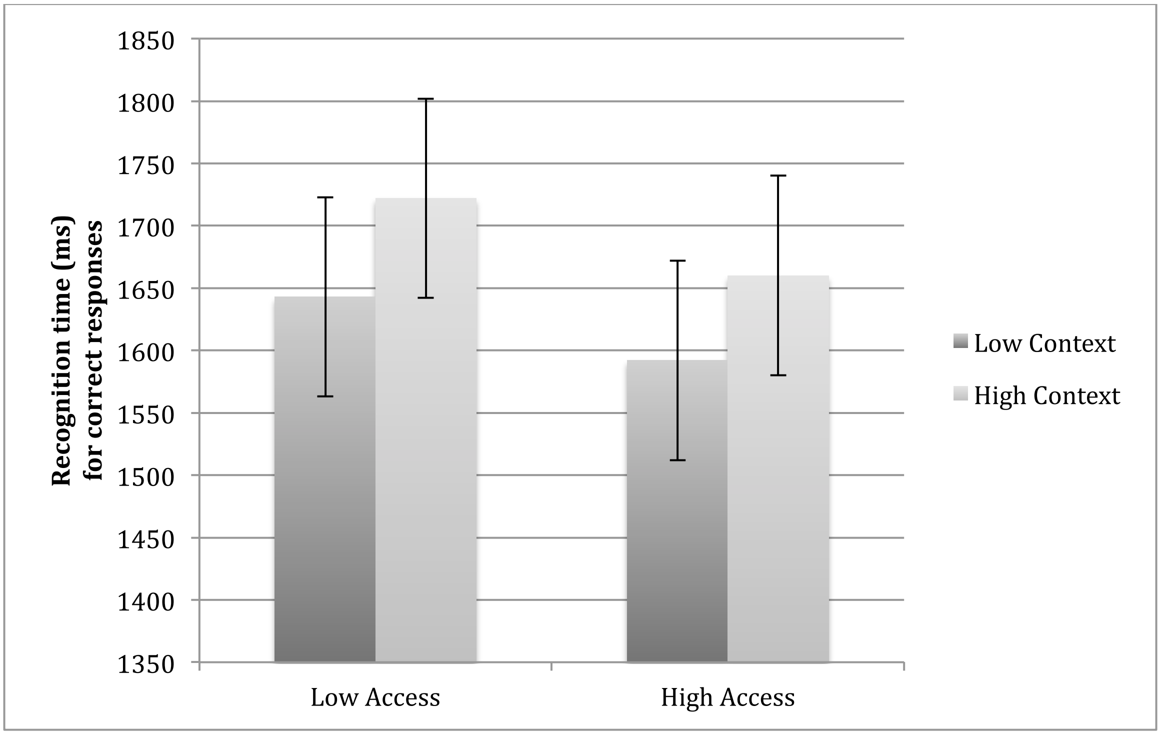

Figure 1. Participants' mean recognition times by context and access

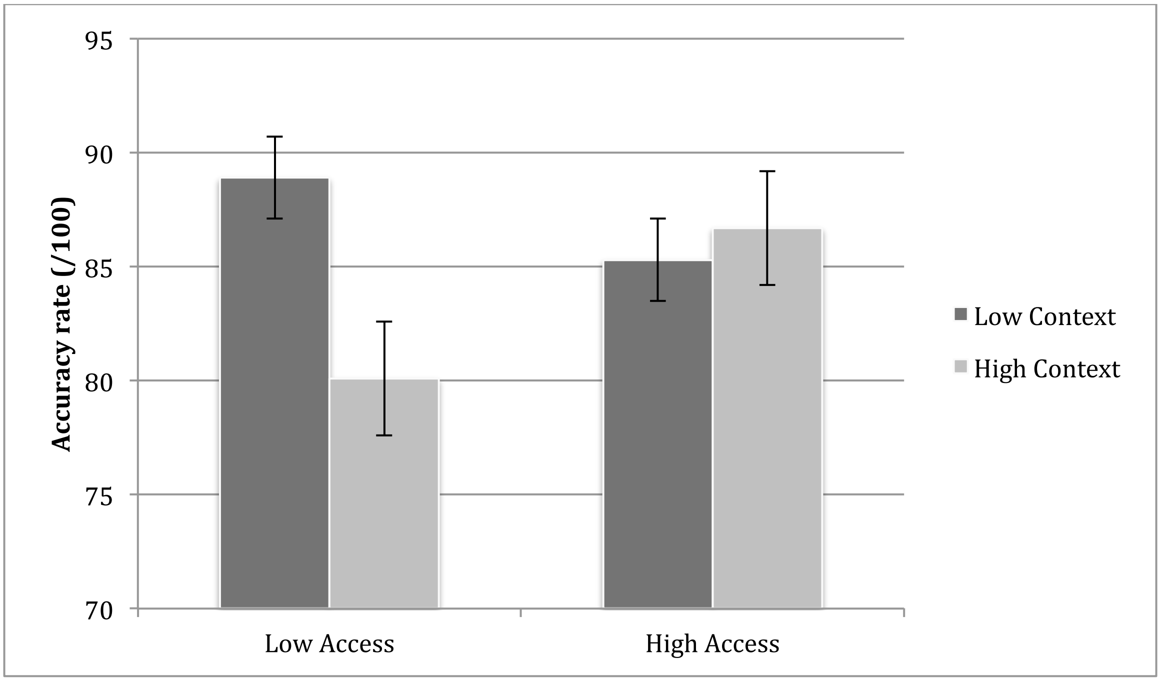

Figure 2. Participants' mean response accuracies by context and access

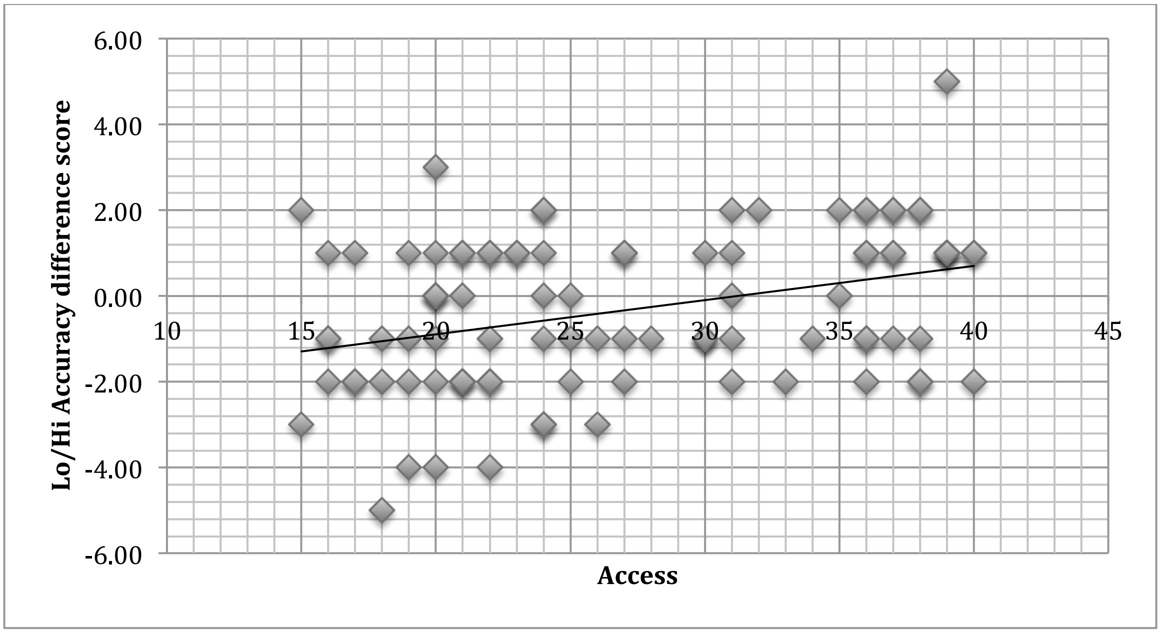

Figure 3. Access raw scores regressed onto accuracy difference scores R^2 = .120, *p* = .001

Appendix A: Experimental passages from the predictive inference task

---

1.

Jimmy was the new kid on the block.
Although his parents urged him to go meet the other kids in the neighborhood, he was shy and hadn't made any new friends.
One Saturday morning, his mom asked him to go to the store for her.
While he was walking back home, Jimmy ran into some of the kids from the neighborhood.
They asked him if he wanted to play with them.
Jimmy was delighted and ran across the street to play with them.

High context (HC):
They taught him a fun game that involved throwing rocks at a target to get points.

Low Context (LC):
They taught him a fun game that involved throwing Nerf balls at a target to get points.

Jimmy and his friends were having a great time.
Jimmy even won the game once or twice.
He stepped up to take his turn and aimed at the target.
He missed, though, and he accidentally hit the door of a new car.
Probe: dent
Comprehension Question: Was Jimmy watching TV with his friends?

2.

Elliot really liked working from home rather than being cooped up in an office in a noisy building downtown.
He got to spend more time with his kids and wife as well as have a more peaceful atmosphere in which to work.
The only problem was he could never seem to keep track of where he left things, because he carried them all over the house with him.
He was working diligently one day when he heard a ruckus in the kitchen.

High context (HC):
He got up to go investigate, absentmindedly bringing a piece of stapler with him.

Low Context (LC):
He got up to go investigate, absentmindedly bringing a piece of paper with him.

Elliot saw that his children had left the stopper in the kitchen sink and the water running.
Water was getting everywhere.
Elliot reached into the sink to pull out the stopper and accidentally dropped what he had been holding.

Probe: splash
Comprehension Question: Did Elliot like working from home?

3.
Amy had just gotten a new car from her parents for her birthday.
It was a good thing, too, because her old one had just broken down the week before and Amy couldn't afford to have it repaired.
Her new car had a standard transmission, though, and Amy was only used to cars with an automatic transmission.
She had a hard time getting used to all of the differences in driving it.
Today her father was going to take her out driving in it to teach her some things.

High context (HC):
When Amy got home from school she parked the car on the steep mountain road next to her parents' house.

Low Context (LC):
When Amy got home from school she parked the car on the road next to her parents' house.

Amy got into the house and immediately realized that she had forgotten to put the emergency brake on.
She put her things down on a chair and walked back outside.
She was just in time to see her car slowly begin to roll backwards away from where it had been parked.

Probe: crash
Comprehension Question: Did Amy get a new car?

4.
Arnie ran a 24hour convenience store in a bad part of town.
He hated working the night shifts because some pretty dangerous characters sometimes came into the store and gave him trouble.
He was good friends with the policemen who patrolled the area, but Arnie still didn't feel very safe working there at night.
There were always lots of teenagers hanging around outside the store.
Tonight a group of guys were loitering outside the door.

High context (HC):
Arnie couldn't help but notice that one of the guys had a pistol and was shooting it in the air.

Low Context (LC):
Arnie couldn't help but notice that one of the guys had a water gun and was shooting it in the air.

Arnie was stocking shelves in the back of the store when he heard some of the guys come in.
He walked around the corner to the front of the store to see what they wanted when all of a sudden the guy he had seen shot him with the gun.

Probe: kill
Comprehension Question: Was Arnie's convenience store in a safe part of town?

5.
Jeffrey was a science teacher at a local middle school.
He liked to teach by demonstration.
He always had something interesting and educational for his students.
He had been lecturing for about half an hour and thought he might be losing the students' interest, so he decided to have them do an experiment to wake them up.
He asked one of the students to come up to the front of the room and choose one of three objects he had put on the desk.

High context (HC):
Susie walked up to the desk and picked up a pin.

Low Context (LC):
Susie walked up to the desk and picked up a feather.

Jeffrey asked the student to poke a variety of things with the chosen object to see what would happen.
Finally, he had the student poke a balloon full of air.

Probe: pop
Comprehension Question: Was Jeffrey a teacher?

Appendix B – Sample material set from the modified integration task

---

A MIRT resembles an OSTRICH but is larger and has a longer neck.

A COFT resembles a ROBIN but is smaller and has a longer neck.

A FLIP resembles a COFT but is smaller, has a longer neck, and nests on land.

The correct size ordering for the items above is: MIRT > OSTRICH > ROBIN > COFT > FLIP.